

QUANTUM ARITHMETIC AND NUMERICAL ANALYSIS USING REPEAT-UNTIL-SUCCESS CIRCUITS

NATHAN WIEBE

*Quantum Architectures and Computation Group, Microsoft Research, One Microsoft Way
Redmond, WA 98052, USA*

MARTIN ROETTELER

*Quantum Architectures and Computation Group, Microsoft Research, One Microsoft Way
Redmond, WA 98052, USA*

Received November 17, 2014

Revised October 23, 2015

We develop a method for approximate synthesis of single-qubit rotations of the form $e^{-if(\phi_1, \dots, \phi_k)X}$ that is based on the Repeat-Until-Success (RUS) framework for quantum circuit synthesis. We demonstrate how smooth computable functions f can be synthesized from two basic primitives. This synthesis approach constitutes a manifestly quantum form of arithmetic that differs greatly from the approaches commonly used in quantum algorithms. The key advantage of our approach is that it requires far fewer qubits than existing approaches: as a case in point, we show that using as few as 3 ancilla qubits, one can obtain RUS circuits for approximate multiplication and reciprocals. We also analyze the costs of performing multiplication and inversion on a quantum computer using conventional approaches and find that they can require too many qubits to execute on a small quantum computer, unlike our approach.

Keywords: Quantum computing, quantum algorithms, quantum circuit synthesis

Communicated by: R. Jozsa & M Mosca

1 Introduction

Classical arithmetic has a long history in quantum computation. Shor’s algorithm [1], linear systems algorithms [2] and general purpose quantum simulation algorithms [3] all deeply rely on arithmetic functions, which are traditionally implemented using reversible circuits (see, e.g., [4, 5, 6, 7]). Beyond direct applications, there is a substantial body of literature that focuses specifically on implementing arithmetic functions in a reversible fashion on a quantum computer, see e.g., [8, 9, 10, 11, 12, 13, 14]. Unfortunately, the inability of reversible circuits to forget previous parts of the calculation carries with it a heavy price: the number of qubits required can be large. For example, the number of qubits required to use Newton’s method to compute reciprocals in linear systems algorithms can easily stretch to several hundred qubits if ten or more bits of precision are required [15]. This result is interesting because it suggests that the number of qubits required to implement certain inexpensive classical algorithms (such as Newton’s method) may be far greater than the number of qubits needed for the remainder of the quantum algorithm. Thus new methods for performing arithmetic

and function synthesis may be needed to enable computationally useful examples of linear systems, or related algorithms, to be run on a small scale quantum computer.

We address this problem by introducing an alternative method for computing functions on quantum computers. The idea behind our approach is to encode numbers in the amplitudes of a qubit, or more properly as polar angles on the Bloch sphere. For example, we represent the number ϕ as the quantum state $e^{-i\phi X} |0\rangle$ where X is the Pauli-X operator. Our approach, in effect, consumes copies of resource states drawn from a set of inputs $\{e^{-i\phi_1 X} |0\rangle, \dots, e^{-i\phi_k X} |0\rangle\}$ to approximate a unitary $e^{-if(\phi_1, \dots, \phi_k)X}$ for some smooth computable function $f : \mathbb{R}^k \mapsto \mathbb{R}$. Here f can be an elementary function such as multiplication or it can be a more complicated function like a trigonometric function or a reciprocal. In this sense, existing circuit synthesis results [16, 17, 18, 19, 20] reduce to solving this problem for cases where f is the constant function. If we think of this task as arithmetic, rather than a unitary synthesis task, then outputting the result as a rotation may at first glance seem unnatural; however, several important algorithms including quantum linear systems and fitting algorithms [2, 15, 21, 22] require the result to be output in exactly this fashion.

Computing arithmetic in amplitudes circumvents the use of a qubit representation for the output and removes many of the ancillas needed in the computation. This results in a substantial reduction in the number of qubits used relative to conventional approaches. A further advantage of our approach is that ϕ need not be input as a qubit string: a quantum circuit that is promised to output $e^{-i\phi X}$ suffices. A drawback is that it requires the use of amplitude estimation [23] or phase estimation when a qubit representation of $f(\phi_1, \dots, \phi_k)$ is needed.

We leverage the ‘‘Repeat-Until-Success’’ (RUS) paradigm of circuit design. This paradigm broadly has two main features: (a) it allows probabilistic execution of unitaries, in particular the conditional application of unitary operations to parts of the quantum data depending on the outcomes of earlier measurements, and (b) all failures must be detectable and lead to an error that is correctable by a Clifford circuit. In this sense, our approach is reminiscent of the KLM proposal for performing a CNOT gate in linear optical quantum computing [24]. See also Figure 1 for a visualization of an RUS protocol. The repeat until success moniker is thus earned because such algorithms do not fail but rather can be corrected and repeated until a successful outcome is obtained.

Most of the work on RUS circuits focuses on the Clifford+ T gate set which is the universal gate set [25] given by $\{H, S, T, \text{CNOT}\}$, where H denotes the Hadamard gate, $S = \text{diag}(1, i)$ is the phase gate, $T = \sqrt{S}$, CNOT is the controlled-NOT gate, and we are allowed to apply these generators to any pair of qubits. The most costly gate in the gate set is assumed to be the T gate because it is by far the most expensive operation to perform fault tolerantly in error correcting codes such as the surface code. For these reasons, we also use the number of T -gates used, or T -count, as a metric to gauge the time-efficiency of our methods.

The RUS paradigm has gained interest after it was realized that it offers advantages for synthesizing single qubit unitaries [17, 18, 19]. In particular: [17] shows that this paradigm provides very low depth circuits for approximating small rotations and also elementary circuits that can approximately implement the square of the product of the input rotation angles. The circuits used there are RUS in the sense of Figure 1, i.e., in the failure case they implement a Clifford gate that can be inexpensively corrected. In [18, 19] RUS circuits are used for

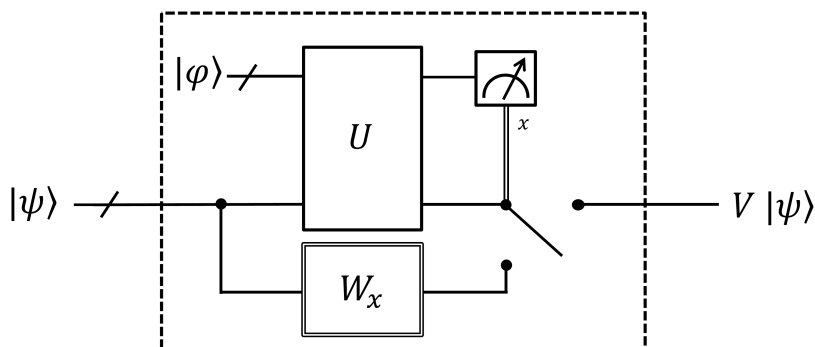


Fig. 1. Implementation of a unitary transformation V via a Repeat-Until-Success (RUS) circuit protocol. The unitary transformation U acts on the input state $|\psi\rangle$ and an ancilla state $|\varphi\rangle$ which has been prepared independent of the input. After application of U the register holding the ancilla state is measured and in case the desired result x is obtained—which, without loss of generality, can be chosen to be the outcome “ $(0, \dots, 0)$ ”—the state $V|\psi\rangle$ is propagated to the output. Otherwise, a state is obtained that differs from the input $|\psi\rangle$ only by an application of a Clifford operation W_x which itself might depend on the measurement outcome x . Hence we can “route” the output back to the input of U , together with a fresh copy of $|\varphi\rangle$. The routing is indicated by the switch that is controlled by x being different from “ $(0, \dots, 0)$ ”. After W_x has been applied, another attempt to compute $V|\psi\rangle$ can be started. The classical control of the Clifford gate W_x is also indicated by double lines. The large dashed box indicates that this procedure is repeated until for the first time the desired outcome is measured, in which case the classically controlled “switch” is set in such a way that the output state $V|\psi\rangle$ can exit the RUS circuit.

single qubit synthesis and the probabilistic nature of these circuits can be used to achieve an expected T -count of $1.15 \log_2(1/\varepsilon)$ for approximating a single qubit axial-rotation using the Clifford+ T gate set, up to an error of ε . This beats a lower bound on the average T -count of roughly $3 \log_2(1/\varepsilon)$ that is known for the ancilla-free synthesis using Clifford + T gates.

The repeat until success paradigm has several features that are especially valuable for quantum arithmetic. First, they allow non-linear functions of the input angles to be computed. This is very useful for arithmetic because multiplication is itself a non-linear function of its inputs. Second, the correctability of the circuits allow them to be applied deterministically to an unknown quantum state (although the run time required to apply the transformation will vary). This is significant because their success probability would shrink exponentially if such failures were not easily correctable. Third, the transformations are irreversible. This means that intermediate computations do not have to be retained for the entire calculation. Finally, the qubit overheads of this form of arithmetic are extremely low since (in principle) the input, output and intermediate results can be stored in amplitudes rather than qubit strings.

The paper is organized as follows. We discuss the two circuits that form the core of our approximations in Section 2. We show that they can be used to provide an arbitrarily accurate approximation to $e^{-if(x)X}$ for any continuous function f in Section 3 before discussing how to apply these methods to problems that have entangled inputs in Section 4. We then apply the ideas of Section 3 to implement multiplication of rotation angles in Section 5 and then use both of these ideas to show how to implement reciprocals in Section 6 before concluding.

2 RUS circuit elements

RUS circuits are a class of non-deterministic circuits that have the following property:

Definition 1 *The unitary U is an RUS circuit if for any input $|0^n\rangle |\psi\rangle \in \mathbb{C}^{2^{n+m}}$ there exists a set of unitary operations $\{W_x\}$ that can be exactly implemented using a finite-size deterministic circuit on the quantum computer and a unitary operation V such that*

$$U |0^n\rangle |\psi\rangle = a_0 |0^n\rangle V |\psi\rangle + \sum_{x=1}^{2^n-1} a_x |x\rangle W_x^\dagger |\psi\rangle. \quad (1)$$

These circuits have been extensively used in circuit synthesis because of their ability to expediently synthesize operations that cannot be exactly synthesized without the use of measurement [25, 17, 18, 19]. Assuming $a_0 \neq 0$, Definition 1 shows that an RUS circuit can deterministically apply V to $|\psi\rangle$ by applying U and measuring the first register. If the register reads $|j\rangle$ for $j > 0$ then the state $|\psi\rangle$ can be re-prepared by applying W_x^\dagger and the protocol can be repeated until a success is observed.

The core idea behind RUS arithmetic (or function synthesis) is to find repeat until success circuits, that upon success, perform a rotation that is a non-linear function of input rotation angles. More specifically, given a set of rotations $\{e^{-i\phi_j X}\}$ we wish to find a circuit implement upon success $e^{-if(\phi_1, \dots, \phi_k)X} |\psi\rangle$ for a known function f . For example, this function could implement multiplication of ϕ_1 and ϕ_2 , or it could be the reciprocal of ϕ_1 . In order to implement a broad class of functions in this sense, we need to have the ability to perform non-linear transformations on x amongst other requirements. This suggests that measurement in general, and RUS circuits in particular, will be of great use in implementing such rotations. We specifically seek a class of circuits that obeys the following four properties:

1. The circuit performs a non-linear mapping that takes single qubit rotations whose angles are in \mathbb{R}^k and maps these to a single qubit rotation whose angle is in \mathbb{R} upon success.
2. The circuit will succeed with non-zero probability for all input rotations that lie in a compact interval centered about $\phi = 0$.
3. The circuit is an RUS circuit in the sense of Definition 1.
4. For any non-negative integers p_1, \dots, p_k , the family of circuits considered must be able to exactly implement functions that scale as $\Theta(\phi_1^{p_1} \dots \phi_k^{p_k})$ in the limit as $\max_j \{|\phi_j|\} \rightarrow 0$.

We use two different classes of circuits that, when combined, allow us to satisfy all four of these objectives. We refer to these circuits as the *gearbox circuit* and the *generalized PAR circuit*. The gearbox circuit satisfies the first three criteria but cannot achieve the fourth without incurring prohibitive costs because $\tan^2(x)$ has only even terms in its Taylor series expansion and shifting the argument to produce odd terms leads to rapidly diverging coefficients in the Taylor series expansion. These coefficients can be controlled, but the number of additional operations required to do so can be prohibitive. Similarly, the PAR circuit satisfies all these criteria but fails the third unless combined with an expensive amplitude amplification algorithm or constrained to act on specific input states. We therefore use both types of circuits to implement elementary functions in order to reduce the costs of the resultant circuits.

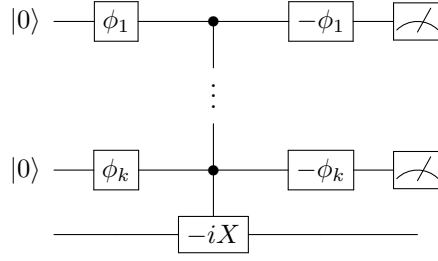


Fig. 2. Gearbox circuit with k inputs, where the gate ϕ_j denotes $e^{-i\phi_j X}$. Success is achieved if *every* measurement reads 0.

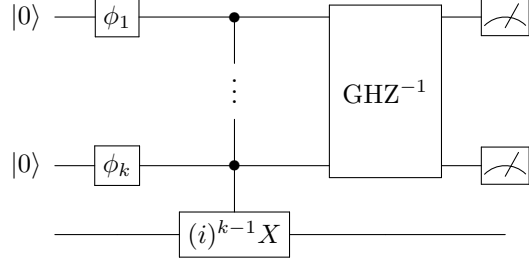


Fig. 3. Generalized PAR circuit with k inputs, where GHZ^{-1} is the inverse GHZ measurement. Success is achieved if *every* measurement reads 0.

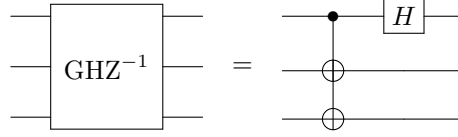


Fig. 4. Circuit for GHZ^{-1} demonstrated for 3 qubits.

Circuit	Function	Success probability	Correction circuit
$e^{-i\phi_1 X} e^{-i\phi_2 X}$	$\phi_1 + \phi_2$	100%	—
$\text{PAR}(\phi_1, \dots, \phi_k)$	$\prod_{i=1}^k \phi_i + O(\max_i \phi_i ^{k+2})$	$\frac{1}{2} (\prod_{i=1}^k \cos(\phi_i)^2 + \prod_{i=1}^k \sin(\phi_i)^2)$	1 or $2\text{PAR}(\phi_1, \dots, \phi_k)$
$\text{RUSPAR}(\phi_1, \dots, \phi_k)$	$\prod_{i=1}^k \phi_i + O(\max_i \phi_i ^{k+2})$	$(\prod_{i=1}^k \cos(\phi_i)^2 + \prod_{i=1}^k \sin(\phi_i)^2)$	1
$\text{GB}(\phi_1, \dots, \phi_k)$	$\prod_{i=1}^k \phi_i^2 + O(\max_i \phi_i ^{2k+2})$	$(1 - \prod_{i=1}^k \sin^2(\phi_i))^2 + \prod_{i=1}^k \sin^4(\phi_i)$	$e^{-i\pi/4 X}$ (Clifford)

Fig. 5. Circuits elements that we use to synthesize arbitrary functions: shown are the circuit names, the corresponding functions, the success probabilities, and the corresponding correction circuits. The correction circuit for PAR is either the identity operation or two repetitions of the PAR circuit (depending on the measurement outcome). RUSPAR, as introduced in Lemma 2, is an RUS analogue of PAR that can be implemented using amplitude amplification.

2.1 Properties of Gearbox circuits

The gearbox circuit is introduced in [17] as a means for expediently generating small rotation angles and rescaling the rotation angles output by circuit synthesis methods. They earned their name because in the circuits can transform coarse input rotations into fine output rotations in analogy to a gear box. The following result, proven in [17], provides justification for this claim as well as those made in Figure 5.

In the following, we will use GB to denote the transformation that the gearbox circuit has on a set of input rotation angles as per

$$\text{GB} : (\phi_1, \dots, \phi_k) \mapsto \arctan(\tan^2(\arcsin(\sin(\phi_1) \cdots \sin(\phi_k))))). \quad (2)$$

A diagram of the corresponding circuit is given in Figure 2 and the success probability of the gearbox circuits and their correction operations are given in Figure 5. In most applications, the

$\text{GB}(\phi_1, \phi_2) = \text{GB}(\phi_2, \phi_1).$	(Commutativity)
$\text{GB}^{\circ k}(\phi) = \arctan(\tan^{2^k} \phi).$	(Composition)
$\text{GB}(\phi_1 + m\pi, \phi_2 + n\pi) = \text{GB}(\phi_1, \phi_2).$	(Periodicity)
$\text{GB}((-1)^p \phi_1, (-1)^q \phi_2) = \text{GB}(\phi_1, \phi_2).$	(Evenness)
$\text{GB}(\phi_1, \dots, \phi_k) = \phi_1^2 \cdots \phi_k^2 + O(\max_k \phi_k)^{2k+2}.$	(Non-linearity)
$\forall \phi_1 \in (-\frac{\pi}{4}, \frac{3\pi}{4}) \lim_{k \rightarrow \infty} \text{GB}^{\circ k}(\phi_1) = \frac{\pi}{2}(H_{\text{Heaviside}}(\phi_1/\pi - 1/4)).$	(Square Wave Limit)

Fig. 6. Table of properties of Gearbox circuits. $H_{\text{Heaviside}}$ is the Heaviside function. We use the notation $\text{GB}^{\circ k}$ to represent composition to emphasize the difference between the composed gearbox circuit and GB^k which we use to denote k repetitions of the gearbox circuit on an input state.

input rotation angles will be quantum states that control the rotations given in the diagram. $\text{GB}^{\circ k}(x) = \text{GB}(\text{GB}(\dots \text{GB}(x) \dots))$ is used to denote composition of the gearbox circuit with itself k times and is not typically used in cases where there are more than one input. The following theorem states the success probability and resultant transformation yielded by GB.

Theorem 1 [Wiebe, Kliuchnikov] *Given that each measurement in $\text{GB}(\phi_1, \dots, \phi_k)$ yields 0, the circuit enacts the transformation $|\psi\rangle \mapsto e^{-iX \arctan(\tan^2(\phi))} |\psi\rangle$, where*

$$\sin^2(\phi) = \sin(\phi_1)^2 \cdots \sin(\phi_k)^2.$$

This outcome occurs with probability $\cos^4(\phi) + \sin^4(\phi)$ and all other measurement outcomes result in the transformation $|\psi\rangle \rightarrow e^{i\pi X/4} |\psi\rangle$, regardless of the choice of ϕ_1, \dots, ϕ_k .

The gearbox circuit is therefore a repeat until success circuit, which means that the user can correct the result and try again upon failure just like a Las-Vegas algorithm. This provides a huge benefit here because it means that these circuit elements can be relied upon to produce the desired transformation. The only downside, apart from low success probability near $\phi = \pi/4$, is that the rotations cannot be pre-cached into qubits and teleported into the system as per [26]. Non-RUS variants of GB that have this property are given in Appendix C. These circuits are Clifford circuits and have the further advantage of lower online costs than GB circuits but cannot be reliably applied to an unknown quantum state. Such circuits can also be used to simplify the state factory used in applications of floating point synthesis [17] that are optimized for low online T -counts.

Further properties of the Gearbox circuits are provided for reference in Figure 6. Properties 1–6 are trivial consequences of Theorem 1. Property (7) follows from the following argument.

Lemma 1 *For all $\phi_1 \in (-\frac{\pi}{4}, \frac{3\pi}{4}) \lim_{k \rightarrow \infty} \text{GB}^{\circ k}(\phi_1) = \frac{\pi}{2}(H_{\text{Heaviside}}(\phi_1/\pi - 1/4))$*

Proof. Let $\phi_1 = \pi/4$ then $\text{GB}^{\circ k}(\phi_1) = \pi/4$ since $\tan(\pi/4) = 1$. Thus for all k , $\text{GB}^{\circ k}(\phi_1) = \pi/4 = \frac{\pi}{2}(H_{\text{Heaviside}}(\phi_1/\pi - 1/4))$ if $\phi_1 = \pi/4$ since Heaviside function is 1/2 at 0.

Next if $\phi_1 \in (\pi/4, 3\pi/4)$ then $|\tan(\phi_1)| > 1$ and hence $\lim_{k \rightarrow \infty} \arctan(\tan^{2^k}(\phi_1)) = \pi/2$ as required. Now if $\phi_1 \in (-\pi/4, \pi/4)$ then $|\tan(\phi_1)| < 1$. It then similarly follows that

$\lim_{k \rightarrow \infty} \arctan(\tan^{2^k}(\phi_1)) = 0$. Thus GB^{ok} converges to a shifted and rescaled Heaviside function on this domain as claimed. \square

This property will be used extensively later when we discuss how gearbox circuits can be used to synthesize functions as a sum of square waves in the appendix.

2.2 Properties of Generalized PAR

The generalized PAR circuit takes a very similar form to the programmable ancilla rotation circuit (or PAR circuit) proposed by Jones, Whitfield et al [26] for teleporting a pre-cached axial rotation that is stored in an ancilla qubit into a target qubit. The principal differences between our circuits and those of [26] are: (a) our circuits are adapted to X -rotations and (b) they map multiple inputs to a single output rotation. Since these circuits are clearly a generalization of the PAR concept, we refer to them collectively as PAR. A diagram for these circuits is given in Figure 3 and they perform the following rotation as a function of input angles:

$$\text{PAR} : (\phi_1, \dots, \phi_k) \mapsto \pm \arctan(\tan(\phi_1) \cdots \tan(\phi_k)). \quad (3)$$

We only consider the positive branch to be a success. It is instructive to note that if $k = 1$ then $\text{PAR} : \phi_1 \mapsto \pm \phi_1$ and thus it reduces to a X -rotation version of the PAR concept introduced in [26]. This claim is demonstrated in the following theorem.

Theorem 2 $\text{PAR}(\phi_1, \dots, \phi_k)$ performs either $\exp(i \arctan(\tan(\phi_1) \cdots \tan(\phi_k))X)$ or the operation $\exp(-i \arctan(\tan(\phi_1) \cdots \tan(\phi_k))X)$ with probability

$$\cos^2(\phi_1) \cdots \cos^2(\phi_k) + \sin^2(\phi_1) \cdots \sin^2(\phi_k)$$

for $\phi_j \in \mathbb{R}$, and the positive and negative branches of the rotation occur with equal probability. All other outcomes result in the identity operation being performed on $|\psi\rangle$.

Proof. The initial rotations and the multiply controlled $(i)^{k-1}X$ gate in the PAR circuit perform the following mapping:

$$\begin{aligned} |0\rangle^k |\psi\rangle &\rightarrow (\exp(-i\phi_1 X) \otimes \cdots \otimes \exp(-i\theta_k X) |0\rangle^{\otimes k} \\ &\quad - \cos(\phi_1) \cdots \cos(\phi_k) |0\rangle^{\otimes k} - (-i)^k \sin(\phi_1) \cdots \sin(\phi_k) |1\rangle^{\otimes k}) |\psi\rangle \\ &\quad + |0\rangle^{\otimes k} (\cos(\phi_1) \cdots \cos(\phi_k)) |\psi\rangle + |1\rangle^{\otimes k} (\sin(\phi_1) \cdots \sin(\phi_k)) (-iX) |\psi\rangle. \end{aligned} \quad (4)$$

Now let us define the following states:

$$\begin{aligned} \text{GHZ} |0\rangle |0\rangle^{k-1} &:= |\text{GHZ}^+\rangle = (|0\rangle^k + |1\rangle^k) / \sqrt{2} \\ \text{GHZ} |1\rangle |0\rangle^{k-1} &:= |\text{GHZ}^-\rangle = (|0\rangle^k - |1\rangle^k) / \sqrt{2} \\ |\phi\rangle &:= (\exp(-i\phi_1 X) \otimes \cdots \otimes \exp(-i\theta_k X) |0\rangle^k \\ &\quad - \cos(\phi_1) \cdots \cos(\phi_k) |0\rangle^k - (-i)^k \sin(\phi_1) \cdots \sin(\phi_k) |1\rangle^k) |\psi\rangle. \end{aligned} \quad (5)$$

where $|\phi\rangle$ is subnormalized and orthogonal to $|\text{GHZ}^+\rangle$ and $|\text{GHZ}^-\rangle$. The right hand side of (4) can be written as

$$\begin{aligned} |\phi\rangle + \frac{1}{\sqrt{2}} |\text{GHZ}^+\rangle (\cos(\phi_1) \cdots \cos(\phi_k) \mathbf{1} - i \sin(\phi_1) \cdots \sin(\phi_k) X) |\psi\rangle \\ + \frac{1}{\sqrt{2}} |\text{GHZ}^-\rangle (\cos(\phi_1) \cdots \cos(\phi_k) \mathbf{1} + i \sin(\phi_1) \cdots \sin(\phi_k) X) |\psi\rangle. \end{aligned} \quad (6)$$

The first register is then measured in the circuit and if either $|\text{GHZ}^+\rangle$ or $|\text{GHZ}^-\rangle$ is observed then the outcome is

$$\exp\left(\mp i \arctan\left(\frac{\sin(\phi_1)\cdots\sin(\phi_k)}{\cos(\phi_1)\cdots\cos(\phi_k)}\right)X\right) = \exp(\mp i \arctan(\tan(\phi_1)\cdots\tan(\phi_k))X), \quad (7)$$

and the probability of either measurement outcome occurring is

$$P_{\pm} = \frac{\cos^2(\phi_1)\cdots\cos^2(\phi_k) + \sin^2(\phi_1)\cdots\sin^2(\phi_k)}{2}. \quad (8)$$

Finally, since the control register is disentangled with the target register in $|\phi\rangle$ we have that

$$\text{Tr}_1(|\phi\rangle\langle\phi|) = |\psi\rangle\langle\psi|, \quad (9)$$

and hence it is clear that if a measurement outcome other than $|\text{GHZ}^+\rangle$ or $|\text{GHZ}^-\rangle$ is observed then the output state is $|\psi\rangle$ and hence no correction operation needs to be applied before repeating the circuit. \square

The generalized PAR circuit satisfies three of the four required properties but fails the third criterion. Theorem 2 shows that PAR cannot satisfy the third criteria because it either implements $e^{i \arctan(\tan(\phi_1)\cdots\tan(\phi_k))X}$ or $\mathbf{1}$ upon failure and the former error cannot be exactly corrected with a deterministic circuit for arbitrary input states. One approach for correcting such errors involves using the doubling down strategy of [26]. This strategy attempts to correct backwards-rotation errors by performing a rotation with twice the original rotation angle. Should this succeed, the circuit will implement the desired target rotation. Should it fail, then the correct rotation can be implemented by succeeding on a rotation with four times the initial rotation angle. This process can be repeated until a successful measurement is observed.

The success probability for PAR also shrinks as $\Theta(2^{-k})$ for cases where the ϕ are each approximately $\pi/4$. This does not directly affect the viability of our approach because we focus on cases where the input rotation angle is small in order to ensure that the rotations are sufficiently accurate. This means that the number of repetitions needed before a success is observed will not be prohibitive in practice, given that we have the ability to run PAR as an RUS circuit.

The PAR circuits do act as RUS circuits on particular inputs. For example, if $|\psi\rangle = |0\rangle$ and in such cases the direction of the rotation can be switched by applying a Z -gate if necessary:

$$e^{-i(\arctan(\tan(\phi_1)\cdots\tan(\phi_k)))X} |0\rangle = Z e^{i \arctan(\tan(\phi_1)\cdots\tan(\phi_k))X} |0\rangle. \quad (10)$$

Thus PAR can be repeated until success, at twice the success probability quoted above, if the generalized PAR gates are applied to a fresh ancilla. We also show below that generalized PAR circuits (which also includes the PAR circuit [26]) can be converted into a repeat until success circuit for arbitrary inputs using oblivious amplitude amplification.

Lemma 2 *The operation $\text{PAR}(\phi_1, \dots, \phi_k)$ can be converted to a repeat until success circuit, $\text{RUSPAR}(\phi_1, \dots, \phi_k)$, that has success probability*

$$\cos^2(\phi_1)\cdots\cos^2(\phi_k) + \sin^2(\phi_1)\cdots\sin^2(\phi_k)$$

and uses three $\text{PAR}(\phi_1, \dots, \phi_k)$ and a constant sized Clifford circuit. Upon failure, the correction operation is the identity gate.

$$\begin{aligned}
 \text{PAR}(\phi_1, \phi_1) &= \text{GB}(\phi_1). && \text{(Equivalence)} \\
 \text{PAR}(\text{PAR}(\phi_1, \dots, \phi_k), \phi_{k+1}) &= \text{PAR}(\phi_1, \dots, \phi_{k+1}). && \text{(Associativity)} \\
 \text{PAR}(\phi_1, \phi_2) &= \text{PAR}(\phi_2, \phi_1). && \text{(Commutativity)} \\
 \text{PAR}(\phi_1 + m\pi, \phi_2 + n\pi) &= \text{PAR}(\phi_1, \phi_2). && \text{(Periodicity)} \\
 \text{PAR}(\phi_1(-1)^{p_1}, \dots, \phi_k(-1)^{p_k}) &= (-1)^{\sum p_k} \text{PAR}(\phi_1, \dots, \phi_k) && \text{(Oddness)} \\
 \text{PAR}(\phi_1, \dots, \phi_k) &= \phi_1 \cdots \phi_k + O(\max_k |\phi_k|)^{k+2}. && \text{(Non-linearity)}
 \end{aligned}$$

Fig. 7. Table of properties of output rotation angles generalized PAR circuits upon success.

A proof of this lemma is given in Appendix A. The idea of the proof is to use oblivious amplitude amplification to eliminate the backwards rotation failure branch. This is possible to do deterministically because the probability of the PAR circuit falling into that failure branch is known and because the only measurements used inside U are implemented using RUS subcircuits that can be trivially inverted by switching the sign of the controlled X gate in Figure 2 or Figure 3. Note that the basic case $\text{RUSPAR}(\phi_1)$, as introduced in Lemma 2 is entirely de-randomized by this procedure and all other cases are reduced to repeat until success circuits.

These primitives can be composed by replacing the rotation used in the above circuits with an RUS sub-circuit. For example, we can implement $\arctan(\tan^2(\phi_1) \tan(\phi_2))$ using the following circuit:

$$\text{PAR}(\text{GB}(\phi_1), \phi_2) \equiv$$

(11)

The circuit is intended to be implemented by repeating the top-most measurement until a “successful outcome” of 0 is measured. Unsuccessful attempts can be corrected with a Clifford circuit. Note that the output of the entire circuit is equivalent to $\text{PAR}(\phi_1, \phi_1, \phi_2)$ upon success; however, the above method typically requires fewer T gates and has a lower online cost. Equation (11) can also be promoted to an RUS circuit using oblivious amplitude amplification since GB is an RUS circuit.

3 Completeness of Gearbox, PAR and addition

We saw in Section 2 that the PAR and GB circuits have many properties in common with multiplication. Since multiplication and addition can be used to approximately implement any continuous function on a compact domain, it is natural to expect that compositions of these circuit elements should also. We refer to arithmetic performed in the fashion as RUS arithmetic.

Input: A smooth function $f \in C^M$, number of function variables k and order of approximation $m \leq M$.

Output: Vector of functions such that $|f(\phi_1, \dots, \phi_k) - \sum_{\kappa=0}^m v[\kappa]|$ is bounded by $O(\max\{|\phi_1|, \dots, |\phi_k|\}^{m+1})$ and each $v[\kappa]$ can be implemented using Gearbox and PAR.

function GENERATEAPPROXIMANT(f, k, m)

$F[0] \leftarrow$ Zeroth order Taylor series expansion of $f(\phi_1, \dots, \phi_k)$ about $\phi_1 = \dots = \phi_k = 0$.

$v[0] \leftarrow F[0]$ ▷ Store constant offset in first entry of output vector.

$f \leftarrow f - v[0]$

for $i \in 1 \rightarrow m$ **do**

$F[i] \leftarrow$ Lowest-order Taylor series expansion of $f(\phi_1, \dots, \phi_k)$.

$v[i] \leftarrow$ Approximate $F[i]$ using Gearbox and PAR circuits to lowest order.

▷ Approximation always exists but is not unique.

$f \leftarrow f - v[i]$

end for

return v

end function

Fig. 8. Taylor series based approximation algorithm.

Our goal in this section is not to provide efficient circuits for performing RUS arithmetic. Instead we seek to show a proof of principle demonstration that a broad class of functions can be implemented using these techniques. We give more efficient methods for implementing particular functions in subsequent sections.

In order to see the validity of this claim formally, we need to introduce a lemma that shows that monomials can be approximated to within arbitrarily small error using GB and PAR via time-slicing ideas reminiscent of those used in Trotter-Suzuki formulas.

Lemma 3 (“time slicing lemma”) *Let $|a| \leq 1$, b be a positive integer and $f(x) = ax^b$ be defined on a compact domain U . Then for any $\epsilon > 0$ the function f can be implemented as an RUS circuit using GB and PAR such that the maximum difference between the output of the RUS circuit and $f(x)$ is at most ϵ for all $x \in U$. The number of GB and PAR circuits needed to achieve error ϵ scales as $r^b \in O((\max\{x : x \in U\})^{b+b^2/2}/\epsilon^{b/2})$.*

Proof. Assume b is even and $0 < a \leq 1$. Then there exists k such that $b = 2k$. From the non-linearity property of GB there exists a gearbox circuit with $k + 1$ inputs such that

$$\text{GB}(\arcsin(\sqrt{a}), x, \dots, x) = ax^b + O(x^{b+2}). \tag{12}$$

Now assume than $-1 \leq a < 0$. Since $ZXZ = -X$, we can apply a Z -conjugated gearbox circuit to implement the appropriate negative rotation. Thus we can approximate this function to $O(x^{b+2})$ if $|a| \leq 1$.

The monomial ax^b for odd b can be approximated using GB, PAR and addition in a similar manner. Since b is odd, we can always express $b = 2k + 1$ and hence the linearity property of PAR and (12) then imply that for any $|a| \leq 1$

$$\text{PAR}(x, \text{GB}(\arcsin(\sqrt{a}), x, \dots, x)) = ax^b + O(x^{b+2}). \tag{13}$$

Thus odd Taylor series can be formed by nesting PAR and GB circuits (such terms can also be generated directly using PAR circuits). Note that here and in the following we use PAR to refer to both the PAR circuit and RUSPAR as appropriate.

In order to use time slicing ideas to make the error less than ϵ for any $\epsilon > 0$ we will need to convert (13) into an RUS circuit. This can be achieved by using oblivious amplitude amplification to convert PAR into a repeat until success circuit as per Lemma 2. Once the circuits have been converted to RUS circuits we can write a large rotation as a sum of small rotations because repeat until success circuits can be reliably corrected when errors occur and addition can be performed deterministically using serial composition (i.e. $e^{-iaX}e^{-ibX} = e^{-i(a+b)X}$). Since the error in small rotations shrinks faster than the size of the rotation, this process allows arbitrarily small errors to be achieved in exact analogy to the use of Trotter-Suzuki algorithms in quantum simulation [3]. To see that slicing can make the error arbitrarily small, let $r > 0$ be a parameter that is the b^{th} root of the number of slices used and note that (for odd b)

$$\begin{aligned} r^b \text{PAR}(x/r, \text{GB}(\arcsin(\sqrt{a}), x/r, \dots, x/r)) &= r^b (a(x/r)^b + O(x/r)^{b+2}) \\ &= ax^b + O(x^{b+2}/r^2). \end{aligned} \quad (14)$$

Thus for any x , the error can be made less than ϵ for any $\epsilon > 0$ by choosing $r \in \Theta(x^{1+b/2}/\sqrt{\epsilon})$. Because x is taken from a compact domain, it is possible to maximize over the values of r to find a value that uniformly makes the error at most ϵ . The case for even b is similar except since GB gives a repeat until success circuit oblivious amplitude amplification is not needed in such cases. This implies that the L_2 distance between ax^b and its approximant can be made less than ϵ for all $\epsilon > 0$ under these assumptions. \square

If $|a| > 1$ then the proof of Lemma 3 does not directly follow because $\arcsin(\sqrt{a})$ is not well defined. We can circumvent this problem by noting that $|a|/2^{\lceil \log_2 a \rceil} \leq 1$ and $ax^b = 2^{\lceil \log_2 a \rceil} \left(\frac{a}{2^{\lceil \log_2 a \rceil}}\right) x^b$. Thus it follows from (12) that

$$2^{\lceil \log_2 a \rceil} \text{GB} \left(\arcsin \left(\sqrt{\frac{a}{2^{\lceil \log_2 a \rceil}}} \right), x, \dots, x \right) = ax^b + O(x^{b+2}). \quad (15)$$

Thus we can assume without loss of generality that $|a| \leq 1$ since all other cases can be found by adding the results of several gearbox circuits with $|a| \leq 1$. This addition step can be performed by serial composition (i.e. running one circuit after another using the same output qubit).

It is important to note that this algorithm scales efficiently with the relative error only in the case where b is fixed. Although the above method is not efficient in many cases, we present it because it is a simple easy to evaluate result that can be used to show that in principle the circuit elements that we present can implement any continuous function without ever needing to resort to using a reversible classical circuit to compute a function value. More efficient formulas for special cases will be given later. However, for the moment our goal is to show that these tools can be used to implement a broad class of functions. We illustrate this with the following theorem.

Theorem 3 *For any $\epsilon > 0$ any function that is analytic on a compact domain $U \subset \mathbb{R}$ can be implemented within maximum error ϵ using PAR and GB and addition as an*

RUS circuit. Furthermore, if the function is approximated with a polynomial of degree $p(\epsilon)$ then the number of GB and PAR circuits needed scale at most as $O(p(\epsilon)(\max\{x : x \in U\})^{p(\epsilon)+p(\epsilon)^2/2}/(\epsilon/2p(\epsilon))^{p(\epsilon)/2})$ for fixed U .

Proof. Let f be analytic on \mathbb{R} then by definition there exists, for any $\epsilon > 0$ and x , $\{a_j\}$ and $p(\epsilon) \in \mathbb{Z}_+$ such that

$$\left| f(x) - \sum_{j=0}^{p(\epsilon)} a_j x^j \right| \leq \epsilon/2. \tag{16}$$

Let $A = \max\{a_j\}$, it then follows that f/A satisfies (16) but with each $a_j \leq 1$. Thus by rescaling the problem we can assume without loss of generality that $a_j \leq 1$. Lemma 3 shows that for any $\epsilon > 0$ an RUS circuit can be constructed that approximates $\tilde{f}_j(x)$ using GB, PAR and addition such that

$$|\tilde{f}_j(x) - a_j x^j| < \epsilon/(2p(\epsilon)). \tag{17}$$

Then two uses of the triangle inequality and (16) and (17) gives us

$$\left| \sum_{j=0}^{p(\epsilon)} \tilde{f}_j(x) - f(x) \right| \leq \sum_{j=0}^{p(\epsilon)} |\tilde{f}_j(x) - a_j x^j| + \left| f(x) - \sum_{j=0}^{p(\epsilon)} a_j x^j \right| \leq \epsilon. \tag{18}$$

Finally, each monomial requires $O((\max\{x : x \in U\})^{p(\epsilon)+p(\epsilon)^2/2}/(\epsilon/2p(\epsilon))^{p(\epsilon)/2})$ RUS circuit elements. There are $p(\epsilon) + 1$ such monomials in the approximation and so the number of GB and PAR circuits required is it most $p(\epsilon)$ multiplied by the number of circuits required to implement the most costly operation in the polynomial. The result then follows. \square

A natural generalization of this theorem is given below.

Corollary 1 *For any $\epsilon > 0$ and continuous function f on a compact domain $U \subset \mathbb{R}$ can be approximated to within maximum error ϵ using an RUS circuit formed using PAR, GB and addition.*

Proof. Proof trivially follows from Theorem 3 and the Weierstrass approximation theorem [27]. \square

Any function that is continuous and computable can therefore be synthesized using repeat until success circuits. This process is concretely summarized in Figure 8. It is further easy to see that this generalizes to piecewise continuous functions that have a finite number of isolated discontinuities at known locations by using elementary control logic.

We have not discussed the success probability of the RUS circuits in the above discussion. As r increases, the size of the largest rotation considered decreases linearly. The success probability for each of the RUS circuit elements used tends to a constant in that limit. The mean number of repetitions needed for each of these RUS elements follows a geometric distribution with mean equal to the reciprocal of the probability of success. As such the complexity of these simple proof of principle methods is not infinite, although they may be expensive in cases where the Taylor series expansion of the function is slowly converging or a large input angle is required.

4 Entangled inputs

We focused in the previous examples on cases where the input is stored as a product state. More generally, the qubits fed into the gearbox circuit could be entangled with another regis-

ter. For example, rather than taking the qubit state $|\phi\rangle$ as input to a gearbox could consider input $\sum_a \alpha_a |a\rangle |\phi_a\rangle$ for some sets of states $\{\phi_a\}$ and amplitudes $\{\alpha_a\}$. Such cases are not pathological: they are representative of typical use cases in linear systems algorithms or quantum simulation algorithms. We will see that the analysis of such cases is much more complicated than the case of unentangled inputs considered in Section 3.

It is tempting to imagine that our methods will run without modification given entangled inputs. After all, if a failure occurs then the correction operation is a fixed Clifford gate regardless of the value of ϕ_a . Thus regardless of whether the input is entangled or not, the operation $e^{-if(\phi_a)X}$ is applied to the target qubit $|\psi\rangle$. The problem with this reasoning is that the probability of success or failure is intimately linked with the value of ϕ_a . This means that if success is measured then amplitudes of all states that lead to low success probability will be suppressed; whereas if failure is measured then terms with large α_a will drop in amplitude. This is because of the information disturbance property of quantum mechanics. Thus if we want to apply our RUS circuits to such states without disturbing the distribution then we must guarantee that success or failure is insufficiently informative for the coefficients to be significantly altered by the measurements used in RUS circuits.

As an example, let us focus on the case of implementing a gearbox circuit with entangled inputs. The gearbox circuit has success probability $\cos(\phi)^4 + \sin(\phi)^4$. This means that, upon success, a state proportional to

$$\sum_a \alpha_a \text{GB}(\phi_a) |\psi\rangle = \sum_a \alpha_a \sqrt{\cos(\phi_a)^4 + \sin(\phi_a)^4} \alpha_a e^{-i \arctan(\tan^2(\phi_a))X} |\psi\rangle, \quad (19)$$

is prepared.

In cases where almost all of these probabilities are equal, the multiplicative factor due to the success probability drops out in normalization. Although this may seem like a trivial case, there are important examples where this actually happens. For example, the circuit $\text{PAR}(\phi_a)$, which is the original PAR circuit of [26], satisfies this and hence works for entangled inputs without modification. Thus balancing all probabilities to be nearly equivalent will ensure that the RUS circuits will not have a meaningful impact on the result. In other cases the probability distribution may be naturally be flat enough such that these errors can be neglected.

The problem of converting GB and PAR to circuits that leak negligible information about the ϕ_a can be solved by using product formulas, oblivious amplitude amplification and time-slicing. We will focus on GB first. For small input angles, the success probability scales as $\cos^4(x) + \sin^4(x) = 1 - 2x^2 + (8/3)x^4 + O(x^6)$. These probabilities approach 1 as x approaches zero. Unfortunately, we cannot implement a squaring circuit using time-slicing because if we use r^2 slices then the deviation of the probability from 1 scales as $O(r^2(x/r)^2)$ which does not shrink as r increases. In order to make this approach work we will have to construct a new function that has a flatter distribution of success probabilities. Consider the following function

$$\text{GB}\left(\frac{x}{2} + \pi/4\right) \text{GB}^*\left(\frac{x}{2} + \pi/4\right) \text{GB}(x) \equiv \text{GB}(x), \quad (20)$$

where $\text{GB}^*(x)$ denotes the gearbox circuit with a controlled iX gate rather than a controlled $-iX$ gate. This circuit is, upon success, equivalent to the inverse of the gearbox circuit. Since it is only equivalent to the gearbox circuit upon success it is not generally true that

$\text{GB}^* = \text{GB}^\dagger$. Eq. (20) is equivalent to the gearbox circuit if all three operations succeed because the two left most operations invoke equal and opposite rotations upon success. The probability of all three of these operations succeeding on the first attempt is

$$(\cos(x/2 + \pi/4)^4 + \sin(x/2 + \pi/4)^4)^2(\cos(x)^4 + \sin(x)^4) = \frac{1}{4} - \frac{1}{4}x^4 + O(x^6). \quad (21)$$

For $x \leq 0.1$, the deviation from success probability $1/4$ is roughly 10^{-5} . Since the deviation from uniformity scales like $O(x^4)$ here, we can use r^2 time slices to reduce the scaling of the skewness of the distribution to $O(x^4/r^2)$ while retaining the initial rotation angle (and at the same time improving its fidelity with an ideal squaring operation) if all such measurements succeed.

A major problem is that failures do leak $O(x^2)$ information about the value of x . If we forgo measurement and mark all states that would correspond to three successful measurements as a “good” state then it is straight forward to see from [28] that a single iteration of oblivious amplitude amplification will make the probability of success $1 - O(x^4)$. Thus by slicing we can make this probability of success arbitrarily high while minimizing the information leaked.

A similar expansion also exists for PAR:

$$\text{GB}\left(\frac{a}{2} + \frac{\pi}{4}\right) \text{GB}\left(\frac{b}{2} + \frac{\pi}{4}\right) \text{PAR}(a, b),$$

which leads to success probability $1/4 + (a^2b^2 - b^4 - a^4)/4 + O(x^6)$ where $x \geq \max\{|a|, |b|\}$. Here it is not possible to set up the two gearbox circuits to create rotations that cancel each other out. Instead, by choosing an even number of time-steps, these residual rotations can be canceled in subsequent steps; thereby removing any information about the values of a and b . In particular, the subsequent sequence of operations needed to cancel these rotations is

$$\text{GB}^*\left(\frac{a}{2} + \frac{\pi}{4}\right) \text{GB}^*\left(\frac{b}{2} + \frac{\pi}{4}\right) \text{PAR}(a, b)$$

Thus GB and PAR can be applied in cases where entangled inputs are used. This raises the possibility that these constructions might be able to substantially reduce the space requirements for linear systems algorithms. Formally analyzing the time complexity of this application would, however, require carefully trading off both errors in the amplitudes as well as the errors in implementing $|a\rangle e^{-iX/a} |0\rangle$ for each $|a\rangle$. We leave analyzing such applications for subsequent work.

5 Multiplication

Multiplication is perhaps the most important application for RUS function synthesis because of its ubiquity in both quantum algorithms and numerical approximations to other functions such as reciprocals. The key result that we will show in this section is that repeat until success circuits can be used to implement a form of multiplication that requires a constant number of ancilla qubits. In contrast, most methods that have been proposed thus far for implementing multiplication require a number of qubits that scales at least logarithmically in the number of bits of precision needed. As per the ideas of the previous section, we do not calculate the

value of the product into a qubit string but instead we provide methods for approximately implementing

$$|\phi_1\rangle |\phi_2\rangle |0\rangle \mapsto |\phi_1\rangle |\phi_2\rangle e^{-i\phi_1\phi_2 X} |0\rangle. \quad (22)$$

For simplicity, we will first assume that the user can perform a set of rotations, provided as black boxes, that can implement the single-qubit rotations $\{e^{-i\phi_j X} : j = 1, \dots, k\}$ and also has access to a circuit based quantum computer. Later, we consider specific applications we estimate the total number of T -gates required by the algorithm when these oracles are replaced by Clifford + T circuits.

The PAR circuit provides simplest approximation to multiplication within our framework but is only accurate to $O(\phi^4)$. The relative error will be minuscule for cases where the input angles are small; however, if $\phi \approx 0.5$ then the relative error in the circuit is 16%. This means that more accurate multiplication formulas will be needed for multiplying modestly large numbers if the error tolerance is small. The following lemma shows that a high-order multiplication formula can be constructed that utilizes a small number of qubits. It can be viewed as providing an estimate of the space complexity a specific application of the algorithm in Figure 8.

Lemma 4 *Let $\max\{|\phi_1|, |\phi_2|\} = x$, then $e^{-i\phi_1\phi_2 X}$ can be approximated to within error $O(x^{q+2})$ for any even integer $q \geq 2$ on a compact domain $U \subset \mathbb{R}$. At most $O(\log q)$ qubits are needed to perform the multiplication.*

Proof. Our proof proceeds inductively. The non-linearity property of PAR shows that an approximate multiplication circuit $M_4(\phi_1, \phi_2) = \text{PAR}(\phi_1, \phi_2)$ implements multiplication to within error $O(x^4)$. This demonstrates the case for $q = 2$. Furthermore, Taylor's theorem shows that

$$\phi_1\phi_2 - M_4(\phi_1, \phi_2) = \phi_1\phi_2 \sum_{j'+j=2} c_{j,j'} \phi_1^j \phi_2^{j'} + O(x^6), \quad (23)$$

where $c_{j,j'} \in \mathbb{R}$ are some constants. In (23) we use the property that the error is of order $O(x^6)$ rather than $O(x^5)$ as $M_4(x)$ is an even power series in ϕ_1 and ϕ_2 .

Now, let us assume that for some $q \geq 4$ and some $c_{j,j'}$

$$\phi_1\phi_2 - M_q(\phi_1, \phi_2) = \phi_1\phi_2 \sum_{j+j'=q-2} c_{j,j'} \phi_1^j \phi_2^{j'} + O(x^{q+2}). \quad (24)$$

Lemma 3 shows that each term in the above power series can be implemented within error $O(x^{q+2})$. In particular,

$$c_{j,j'} \phi_1^{j+1} \phi_2^{j'+1} = c_{j,j'} \times \text{PAR}(\phi_1, \phi_2, \text{GB}(\phi_1^{j/2}, \phi_2^{j'/2})) + O(x^{q+2}). \quad (25)$$

It then follows that

$$\phi_1\phi_2 - M_q(\phi_1, \phi_2) - \sum_{j+j'=q-2} c_{j,j'} \times \text{PAR}(\phi_1, \phi_2, \text{GB}(\phi_1^{j/2}, \phi_2^{j'/2})) = O(x^{q+2}). \quad (26)$$

This process can be iterated in order to implement each ϕ_1^j and $\phi_2^{j'}$ using GB and PAR. Now let us define $M_{q+2}(\phi_1, \phi_2)$ to be the approximant formed in this manner. Note that because GB is an even function and PAR is an odd function, we have that there exist $c'_{j,j'}$ such that

$$\phi_1\phi_2 - M_{q+2}(\phi_1, \phi_2) = \phi_1\phi_2 \sum_{j+j'=q} c'_{j,j'} \phi_1^j \phi_2^{j'} + O(x^{q+4}). \quad (27)$$

This demonstrates the induction step of our proof. Now using (23) as our induction hypothesis, we arrive at the conclusion that we can approximate multiplication to within error $O(x^{q+2})$ for any even $q \geq 2$.

The composed gearbox circuit $\text{GB}^{2^k}(\phi)$ yields a rotation angle of $\phi^{2^k} + O(\phi^{2^k+2})$ using k recursive applications of the circuit. Each application of the circuit requires 1 additional qubit. Therefore it is trivial to see by induction that ϕ^{2^k} can be approximated using at most $k + 1$ qubits since ϕ^2 requires 2 qubits. If we consider a binary expansion of j then we see that at most $\lceil \log_2(q/2 - 1) \rceil$ bits are needed to encode $j/2$. Since the qubits used each of the ϕ^{2^k} in the decomposition $\phi^{j/2} = \phi^2 \phi^4 \dots$ can be recycled and used to implement the other $\phi^{2^{k'}}$ terms, the number of qubits required to perform each of these terms is at most the number of qubits required for the most memory intensive calculation. This means $\lceil \log_2(q/2 - 1) \rceil$ qubits will suffice, if we exclude the output qubits. Since there are at most $\lceil \log_2(q/2 - 1) \rceil$ terms in the decomposition, the total number of output qubits is $O(\log q)$ as well. An additional 5 qubits are needed to store the remaining arguments to the function and implement the multiply controlled Toffoli gate inside the gearbox circuit. Thus the space total space requirements for this circuit scale as $O(\log q)$. \square

Lemma 4 gives a procedure that can be used to construct a multiplication formula that has error that has arbitrarily high order error scaling using a number of qubits that scales logarithmically with the order of the multiplication formula. This allows us to trade off space usage and time-complexity for the multipliers.

We also do not-discuss how the constants in the Taylor series expansion $c_{j,j'}$ are computed. For multiplication, the coefficients can be found by performing Taylor expansions of products of elementary functions. This means that the Taylor series can be explicitly computed using computer algebra packages. These coefficients can also be estimated numerically if symbolic computation is too time consuming.

An important remaining issue is that many angles that appear naturally in problems, such as reciprocal calculation using the binomial method, will be approximately 1. These rotations can be implemented using time slicing as per Lemma 3 but the cost of doing so may be prohibitive. Instead, it makes sense to use a Taylor series expansion centered around $x = 1$ rather than $x = 0$. We formally state this in the following corollary.

Corollary 2 *Assume that $\phi_1 \approx \phi_2 \approx 1$ then $\phi_1\phi_2$ can be implemented within error $O(\max\{|1-\phi_1|, |1-\phi_2|\}^{q+2})$ using at most $O(\log q)$ qubits. Similarly, if $\phi_1 \approx 0$ and $\phi_2 \approx 1$ then $\phi_1\phi_2$ can be implemented within error $O(\max\{|\phi_1|, |1-\phi_2|\}^{q+2})$ also using at most $O(\log q)$ qubits.*

Proof. Assume that $\phi_1 \approx \phi_2 \approx 1$ then from Lemma 4 we have that

$$\begin{aligned} \phi_1\phi_2 &= -1 + \phi_1 + \phi_2 + (1 - \phi_1)(1 - \phi_2), \\ &= -1 + \phi_1 + \phi_2 + M_q(1 - \phi_1, 1 - \phi_2) + O(\max\{|1 - \phi_1|, |1 - \phi_2|\}^{q+2}), \end{aligned} \quad (28)$$

and this operation can clearly be implemented using at most $O(\log q)$ qubits.

Now let us assume that $\phi_1 \approx 0$ and $\phi_2 \approx 1$. We can then use similar reasoning to show that

$$\begin{aligned} \phi_1\phi_2 &= \phi_1 - \phi_1(1 - \phi_2), \\ &= \phi_1 - M_q(\phi_1, 1 - \phi_2) + O(\max\{|\phi_1|, |1 - \phi_2|\}^{q+2}), \end{aligned} \quad (29)$$

Name	Multiplication formula	Error	Qubits (RUS)
M_4	$\text{PAR}(\phi_1, \phi_2)$	$O(x^4)$	4
M_6	$\text{PAR}(\phi_1, \phi_2, \frac{\pi}{4} - \text{GB}(\gamma_2, \phi_1) - \text{GB}(\gamma_2, \phi_2))$	$O(x^6)$	5
M_8	$\text{PAR}(\phi_1, \phi_2, \frac{\pi}{4} - \text{GB}(\gamma_2, \phi_1) - \text{GB}(\gamma_2, \phi_2), \frac{\pi}{4} - \text{GB}(\gamma_3, \phi_1) - \text{GB}(\gamma_3, \phi_2) + \text{GB}(\gamma_2, \phi_1, \phi_2))$	$O(x^8)$	7

Fig. 9. Lowest three orders of multiplication formula designed using the method of Lemma 4 but optimized for execution as an RUS circuit when acting upon $|0\rangle$. We use the constants $\gamma_2 = \arcsin(1/\sqrt{6})$ and $\gamma_3 = \arcsin(1/\sqrt{15})$. Circuits are optimized for width and are meant to be executed from right to left in the PAR to allow the left most qubits to be used as ancillas for the GB operations appearing to their right.

x	$ x^2 $	$ M_4(x, x) - x^2 $	$ M_6(x, x) - x^2 $	$ M_8(x, x) - x^2 $
0.01	1×10^{-4}	6.7×10^{-9}	6.6×10^{-14}	5.5×10^{-17}
0.05	2.5×10^{-3}	4.2×10^{-7}	1.0×10^{-9}	2.2×10^{-11}
0.10	1×10^{-2}	6.7×10^{-5}	6.6×10^{-8}	5.5×10^{-9}
0.5	2.5×10^{-1}	4.0×10^{-2}	9.4×10^{-4}	1.9×10^{-3}
1.0	1.0	0.18	0.054	0.088

Fig. 10. Errors in the first three orders of multiplication formulas as a function of input angles.

and again the resultant rotation can be implemented using $O(\log q)$ qubits. \square

5.1 Sixth-order multiplication formulas

As an example, we will show how to derive a sixth-order multiplication formula, M_6 , from a fourth-order multiplication formula M_4 , which we take to be the output of the PAR circuit. If x is a small parameter then we can evaluate the behavior of the function for two small inputs by examining the Taylor series of $\text{PAR}(ax, bx)$. By Taylor expanding the function in powers of x (i.e. using $\arctan(x) = x - x^3/3 + \dots$ and $\tan(x) = x + x^3/3 + \dots$) we find

$$\text{PAR}(ax, bx) = \arctan(\tan(ax) \tan(bx)) = abx^2 + \frac{1}{3} (ab^3x^4 + ba^3x^4) + O(x^6). \quad (30)$$

(30) shows that M_4 behaves as an ideal multiplication circuit but with $O(x^4)$ error. These error terms could be canceled by using the fact that $\text{GB}(ax) = a^2x^2 + O(x^4)$ and then applying two more PAR circuits in series and Taylor expanding the result:

$$\begin{aligned} & \text{PAR}(ax, bx) - \text{PAR}(ax, bx, \text{GB}(ax), \tan^{-1}(1/3)) - \text{PAR}(ax, bx, \text{GB}(bx), \tan^{-1}(1/3)) \\ & = abx^2 + O(x^6). \end{aligned} \quad (31)$$

This process can then be repeated to make the $O(x^6)$ terms zero and so forth.

We do not use (31) in practice because it uses qubits too greedily and is the sum of three different PAR circuits. Because PAR is only an RUS circuit when it acts on $|0\rangle$, the sum of three outputs from PAR is not an RUS circuit since the three rotations that constitute it are applied in series and hence cannot possibly all act on $|0\rangle$ in the limit of small x (unless

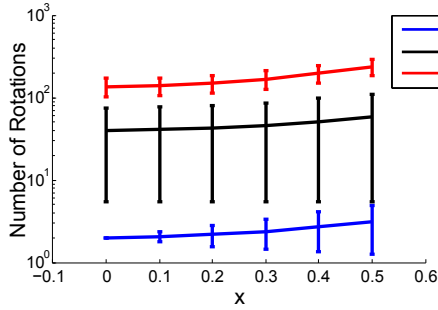


Fig. 11. Distribution for number of applications of ϕ_1 and ϕ_2 needed to approximate $e^{-i\phi_1\phi_2 X}$ using M_4, M_6 and M_8 for different values of $x = \max\{|\phi_1|, |\phi_2|\}$.

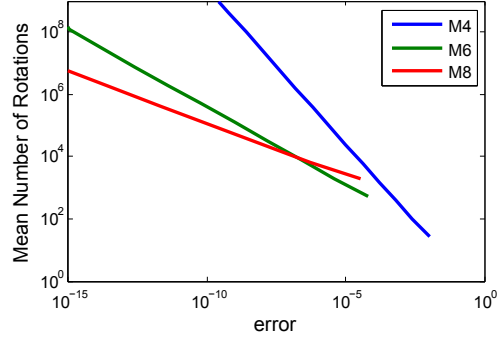


Fig. 12. Mean number of rotations (i.e. number of queries to an oracle that yields $e^{-i\phi X} |0\rangle$) needed to implement $e^{-i\phi^2 X}$ for $\phi = 0.5$ using M_4, M_6 and M_8 versus approximation error for a number of time steps, r^2 , ranging over $2^2, 2^4, \dots, 2^{20}$.

$ab = 0$). Thus oblivious amplitude estimation will be needed to convert two of the three generalized PAR circuits into repeat until success circuits. Lemma 2 shows that this involves roughly tripling the cost of the circuit and so it is desirable to optimize the circuits to avoid this when possible. M_6 uses one particular strategy to address the problem.

The inspiration behind M_6 comes from noting that each term in (30) consists of at least one a and b . This means that we can multiply $(1 - a^2x^2/3 - b^2x^2/3)$ by $\text{PAR}(ax, bx)$ to achieve the desired result. An efficient way to do this is to note that

$$\tan(x + \pi/4) = 1 + 2x + O(x^2). \tag{32}$$

Thus for any analytic function $f(x)$ we have that

$$\begin{aligned} &\arctan(\tan(ax) \tan(bx) \tan(f(x) + \pi/4)) \\ &= \arctan(\tan(ax) \tan(bx) + 2f(x) \tan(ax) \tan(bx) + O(f(x))^2). \end{aligned} \tag{33}$$

The choice of $f(x)$ used in M_6 is $f(x) = -\text{GB}(ax, \arcsin(\sqrt{1/6})) - \text{GB}(bx, \arcsin(\sqrt{1/6})) = -a^2x^2/6 - b^2x^2/6 + O(x^4)$ which then, along with (30), gives us that

$$\begin{aligned} \tan^{-1}(\tan(ax) \tan(bx) \tan(f(x) + \pi/4)) &= \tan^{-1}(\tan(ax) \tan(bx) - a^3bx^4/3 - b^3ax^4/3 + O(x^6)) \\ &= \tan^{-1}(\tan(ax) \tan(bx)) - a^3bx^4/3 - b^3ax^4/3 + O(x^6) \\ &= abx^2 + O(x^6). \end{aligned} \tag{34}$$

M_6 therefore gives a sixth order approximation to the product of two numbers. If only one time slice is needed to achieve the accuracy threshold for the problem then these circuits will be inexpensive; whereas if more than one time slice is needed then oblivious amplitude amplification will be needed to convert the PAR into an RUS circuit. This causes the costs of these circuits to jump substantially during the transition from one to two slices.

There are several ways in which the circuits could be optimized further for execution in cases where multiple time slices are needed. One of the issues that arises stems from the fact

that shifting the argument of these functions by $\pi/4$ (as per (32)) comes at a steep price: it reduces the success probability of the circuit by roughly a factor of two. This means that it should be used sparingly. In cases where one slice is needed, its use results in an increase of a factor of 2 in the expected cost of the circuit, which is superior (for small arguments) to the three-fold increase that would result from using amplitude amplification to allow the PAR circuits to be run in series as RUS circuits. In cases where two or more slices will be needed, this trick becomes unnecessarily costly and the resulting circuits can be further optimized by opting instead for a strategy that is closer to (31).

5.2 *Performance of multiplication circuits*

How well do these multipliers work for concrete inputs and concrete errors? We address this question by providing a table of elementary multiplication formulas for small rotation angles in Figure 9. We focus on formulas that are accurate for $x \approx 0$, but formulas adapted for $x \approx 1$ can be derived from these using the approach of Corollary 2. The formulas M_4 , M_6 and M_8 are highly accurate if $\max\{\phi_1, \phi_2\} \leq 0.1$, but fail to accurately approximate multiplication for large rotations. This is particularly noticeable with M_8 , which is actually less accurate than M_6 for $\max\{\phi_1, \phi_2\} \geq 0.5$. This is because the convergence of the Taylor series for M_8 is slowed due to the presence of large coefficients on the $O(x^8)$ terms that are introduced in this construction.

The cost of implementing the multiplication in terms of the number of times that the angles ϕ_1 and ϕ_2 need to be used is given in Figure 11. There we see that these costs are minimal for small angles but increase with x because of the probability of failure in the gearbox and PAR circuits. These costs are given in Corollary B.1. We ignore the costs of the Toffoli gates needed to perform these rotations because we assume that synthesizing the rotation will be much more expensive than applying a Toffoli gate.

Figure 11 shows that M_4 costs approximately 2 rotations, M_6 costs 40 rotations and M_8 costs roughly 120 rotations for $x \leq 0.1$. If time-slicing is used then all these circuits must be converted to genuine RUS circuits using Lemma 2, which triples the cost of all rotations. Regardless, we can use $r = 5$ for M_4 at roughly the same cost as a single iteration of M_6 and $r = 8$ for the same cost as an iteration of M_8 . The question remaining is, under what circumstances will using M_4 be preferable to its higher order brethren? We see from Figure 12 that each of these methods for multiplying two rotation angles works best in a different regime. M_4 is preferable for low accuracy rotations; whereas M_6 and then M_8 become methods of choice as the error tolerance shrinks.

5.3 *Comparison with classical methods for multiplication*

In this section we compare the resources required for RUS circuits for approximate multiplication with the traditional approaches of implementing multiplication by means of reversible circuits. Comparing classical approaches to implement a function $f(\phi_1, \dots, \phi_m)$ of several inputs ϕ_i , $i = 1 \dots, m$ —which are all assumed to be integers with the same precision, i.e., they are given by bit-strings of length n —with the ones described in this paper is not entirely straightforward: our methods assume that the inputs are given in form of rotations, whereas in classical approaches usually the inputs are given in a bit-string that encodes a basis state. To make the models comparable, we force the input and output types to be the same, i.e., a bit-string for the inputs and rotations for the outputs: for implementations based on classical

Multiplier method	$n = 2$		$n = 4$		$n = 8$		$n = 16$	
	T -count	qubits	T -count	qubits	T -count	qubits	T -count	qubits
Carry-ripple	2.34E+02	4	7.84E+02	8	2.80E+03	16	1.06E+04	32
Table-lookup	3.38E+03	2	3.26E+06	2	3.98E+09	2	1.13E+13	2
M_4	6.11E+01	3	1.97E+03	4	4.64E+04	4	3.00E+07	4
M_6	7.71E+02	4	1.67E+03	4	3.82E+03	4	5.21E+05	5

Fig. 13. A comparison of the resources required for space efficient multiplication on a quantum computer. Shown are circuit size (number of T -gates) and number of required qubits for two $n = 2, 4, 8$ and 16 bit numbers. M_8 is not given because it is strictly more expensive than M_6 for this data set. RUS synthesis was used to implement single-qubit rotations and a Toffoli construction that uses 7 T -gates was used. Extra qubits required for controlled Toffoli gates in M_6 are assumed to be recycled from prior steps. All operations are assumed to be performed sequentially, the mean time required by M_4 and M_6 fall substantially if parallel execution is permitted.

circuits this means to encode the output from a reversible implementation U_f of f into a rotation. For this we use the circuit *Enc* shown in Figure 15. Overall, we get a unitary circuit as shown in Figure 14 (a) to compute the function by way of a classical reversible implementation and an RUS circuit as shown in Figure 14 (b) to compute the function in Repeat-Until-Success style using measurements. In the remainder of this section we give estimates on the (expected) required resources for both cases (a) and (b), where we instantiate the function f to be a multiplier of two n -bit integers ϕ_1 and ϕ_2 . Later in Section 6 we perform a similar analysis for the case where the function f is the reciprocal function applied to n -bit integers. Our cost estimates for circuit size are based on the total number of T gates used. Our cost estimates for required number of qubits do not include the qubits required to encode the inputs but only those required for everything else, i.e., the output qubits and any ancilla qubits that might be needed in the computation.

5.3.1 Comparison with RUS methods

We compare the two methods by choosing a problem for which both the inputs and outputs are well defined. The problem that we use to benchmark these algorithms is one where two input numbers are provided as qubit strings: $|\phi_1\rangle|\phi_2\rangle$ and from these qubit strings we wish to implement the rotation $e^{-i\phi_1\phi_2 X}$ within error at most $2^{-(n+1)}$, meaning that we have n -digits of precision in the output rotation. We further constrain all algorithms to use gates only from the Clifford and T library and take the cost to be the number of T -gates used. The RUS synthesis method of [19] is used to implement the rotations required in the inputs of M_4 and M_6 and the outputs of the carry-ripple and table-lookup multipliers. In all cases we assume that $\phi_1 = \phi_2 = 1/2$ since that maximizes the error incurred for multipliers centered about $\phi = 0$ because for larger ϕ we can take $\phi = 1 - \theta$ and then consider a small expansion in θ . We also assume that non-entangled inputs are used to provide a proof of principle. The costs in entangled settings are potentially greater, however in the limit of small ϕ these costs become negligible and our formulas can be used directly.

A comparison between the resources required for the carry-ripple and the table-lookup

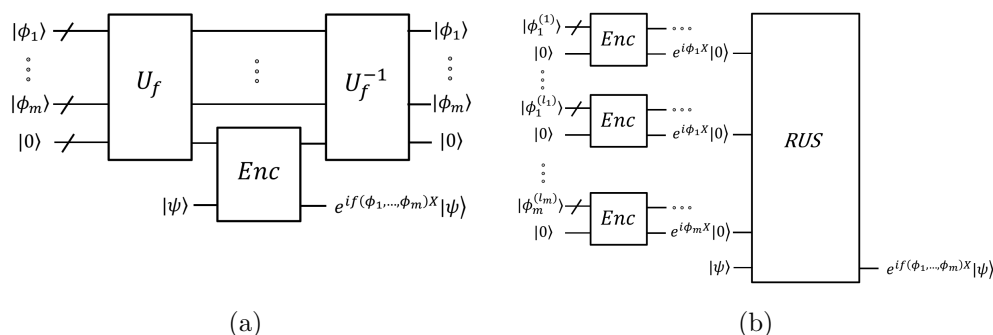


Fig. 14. Comparison of implementing a function f as a rotation of a single target qubit $|\phi\rangle$. Shown are two implementations: (a) via a classical, reversible circuit U_f that first computes f and then encodes the resulting bit-string $f(\phi_1, \dots, \phi_m)$ as a rotation using a phase encoding circuit Enc . The implementation of Enc in turn is shown in Figure 15. And (b) via an RUS circuit as in the methods presented in this paper. Note that in the RUS case, the input bit strings are encoded directly into rotations which are then consumed by the RUS circuit. In contrast to the classical case, several rotations might be required to implement the target rotation: as shown in the figure $l_i \geq 1$ copies of the rotation corresponding to angle ϕ_i , $i = 1, \dots, m$ are used, denoted by $|\phi_i^{(1)}\rangle, \dots, |\phi_i^{(l_i)}\rangle$.

multipliers with the methods M_4 and M_6 (which are given in Section 5) can be found in Figure 13. The data was found by simulating the RUS circuits action on a state and computing the mean and variance of the number of T gates in these simulations. The values were found to agree with theoretical expectations given in Appendix B.

We see that the number of qubits needed is substantially lower than those required to obtain comparable accuracy using the carry-ripple multiplier, although the T -count required for these implementations of RUS arithmetic are several times higher (except for M_4 at two bits of precision). In contrast, lookup tables also require a constant number of qubits since but the T -count required by them is prohibitive for $n > 4$. We see from this data that M_4 and M_6 give viable alternatives to performing multiplication using traditional methods on a fault tolerant quantum computer. Perhaps most significantly, both methods require fewer than 5 qubits to implement which means that they can be performed on existing quantum computers unlike carry-ripple multiplication.

As our RUS implementations are highly space efficient—recall that for instance M_4 requires only 4 additional qubits to approximately compute the product of two n -bit numbers—we focus on classical implementations that optimize circuit width. Specifically, we consider the straightforward way of implementing a multiplication of two n bit numbers using the naïve method of reducing the problem to n additions. The advantage of this method is that it requires only $O(\log n)$ space in addition to the input and output registers. The size of the resulting circuit scales as $O(n^2)$. More advanced algorithms that achieve asymptotically better scaling in terms of total circuit size such as Karatsuba-Ofman or FFT-based methods seem to require a significant higher amount of space [29], so we do not consider them in the comparison.

As our methods require only a constant amount of space, arguably the most meaningful comparison is to compare them with classical multipliers that only use a constant amount of

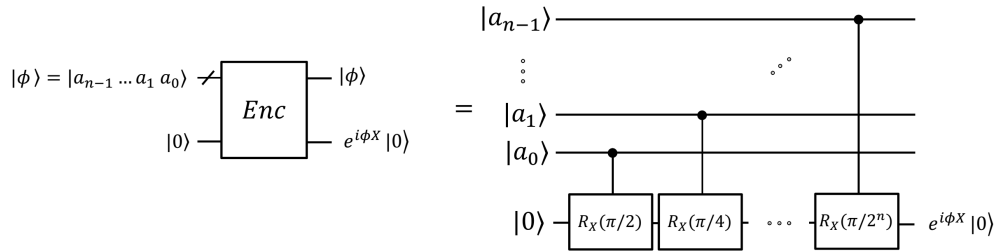


Fig. 15. Implementation of the phase encoding *Enc* of an n -bit string $\phi = a_{n-1} \dots a_1 a_0$ using controlled rotations. The cost for each controlled rotation depends on the overall target error ϵ and scales as $O(\log 1/\epsilon)$ and is assumed to be the same for all rotations shown in the figure.

memory to do so. While there is work on the complexity of computing output bits of the product in the context of space-bounded computation models such as OBDD or branching programs—e.g., on the middle bit [30] or the most significant bit [31]—we are unaware of work that addresses the space-bounded complexity for the *approximate* multiplication problem. In the absence of such results, we instead consider an extremal case of table-lookup computations, i.e., computations that can be implemented with a constant number of additional qubits but which implements the function in a brute force way that does not exploit any features that the function might possess.

We do not consider entangled inputs in this comparison. For M_4 , we can easily see that the errors introduced from measurement can be made comparable to the truncation errors by increasing the gates by a factor of 3 by using the techniques of Section 4. Higher-order methods can also be implemented using these ideas, but are much more involved. For simplicity we have focused on the single input case and leave a full comparison of these methods in cases where there is substantial entanglement for subsequent work.

5.3.2 Carry-ripple multipliers

The naïve method implements multiplication of two n -bit numbers $x = \sum_{i=0}^{n-1} x_i 2^i$ and $y = \sum_{i=0}^{n-1} y_i 2^i$ by performing $n - 1$ additions where the summands are shifted versions of x . Overall, this requires scratch space that is logarithmic in the number of input bits. Moreover, carry-save techniques can be applied to keep circuit depth of the additions small [32], i.e., the first of the $n - 2$ additions can be performed in constant depth using a suitable data structure. As we are interested in optimizing space, we chose a different path and consider a carry ripple adder [10] which we then use $n - 1$ times to produce the desired output. From the quantum circuits for additions that have been studied in the literature [10, 11, 14], we pick one that has the property that a controlled adder can be implemented at relatively low overhead.

While classically the naïve method needs only $\log n$ additional space, in the quantum case we have to store the intermediate results because the overall computation must be reversible. We obtain an upper bound of $3n$ qubits for the total space required, including input and output qubits as we can implement the multiplication using n controlled adders that add shifted copies of the input x to the output register, where the controls depend on the bits of y . The resultant T -count can be upper bounded by n times the cost for an in-place adder that is controlled on a single qubit.

To keep space as well as circuit size small, we choose the in-place adder presented in [10, Figure 5] as this has the particular useful feature that a *controlled* adder can be derived directly from it without using many additional control gates, an observation also used in [33, Section V]. Explicitly, by counting the number of gates we obtain that at most $12n$ Toffoli gates are required. Using the fact that a Toffoli gate can be implemented using 7 T -gates [25, 34] we obtain an upper bound of $84n$ T -gates for a controlled adder. We now use n controlled adders (of input size $n, n+1, \dots, 2n$ bits) conditioned on the bits of y for an overall T -count of $\sum_{k=n}^{2n} 12k = 18n^2 + 18n$ for the carry-ripple multiplier, i.e., for the implementation of U_f as in Figure 14. The cost for implementing Enc can be estimated as follows: to be comparable with the RUS-based multipliers that produce $n+1$ -bit approximations of the rotations, not all $2n-1$ bits of the output of f have to participate in the controlled rotations as in Figure 15. Indeed, it is enough if the highest $n+2$ -bits participate in order to get an $n+1$ -bit approximation. As we then have $n+2$ rotations and we have the target error $\epsilon_{target} = 2^{-(n+2)}$, we obtain that we need at most $1.15 \log_2((n+2)/2^{-(n+2)})$ many T -gates per each rotation in Enc where we choose to distribute errors uniformly, i.e., $\epsilon = \epsilon_{target}/(n+2)$ and we used the upper bound [19] for RUS-based single qubit decompositions. Putting everything together, we get an overall cost of $2 \cdot (18n^2 + 18n) + 1.15(n+2) \log_2((n+2)/2^{-(n+2)})$ many T -gates, where the leading factor of 2 is due to the cleanup of the ancillas. For small values of n , the resulting upper bound estimates on the number of T -gates are shown in Figure 13. Note that the space bound on the number of qubits for the carry-ripple adders are given by $2n-1$ for storing the output of f plus 1 qubit for the finally resulting rotation.

5.3.3 Table-lookup multipliers

The problem of computing the product of two n -bit numbers x and y can be restated as a problem of computing $2n-1$ Boolean functions

$$f_0(x_0, \dots, x_{n-1}, y_0, \dots, y_{n-1}), \dots, f_{2n-2}(x_0, \dots, x_{n-1}, y_0, \dots, y_{n-1}),$$

i.e., one Boolean function $f_i : \{0,1\}^{2n} \rightarrow \{0,1\}$ for each output bit. This function can be stored as a lookup table wherein the individual bits yielded by the Boolean functions are stored in an array. We now consider the complexity of implementing all these functions via lookup tables. As we are interested in uniform families of circuits (as opposed to non-uniform circuit models in which lookup tables can be implemented in $O(n)$ time and $O(n)$ space) we are therefore looking for a quantum circuit that can implement a lookup-table with $2n$ inputs and $n+2$ outputs. Note that we need only the highest order $(n+2)$ of the result to be comparable to the RUS-based implementation, so we do not have to synthesize all $2n-1$ output functions. One simple way to upper bound the cost for implementing such a lookup-table is to assume that each output bit is implemented via a sequence of $2n$ -fold controlled NOT gates $\Lambda_{2n}(\text{NOT})$, where $\Lambda_k(U)$ is defined as the operation $\Lambda_k(|x_0, \dots, x_{k-1}\rangle |\psi\rangle) = |x_0, \dots, x_{k-1}\rangle |\psi\rangle$ if $(x_0, \dots, x_{k-1}) \neq (1, \dots, 1)$ and $\Lambda_k(|1, \dots, 1\rangle |\psi\rangle) = |1, \dots, 1\rangle U |\psi\rangle$.

From [35] follows that we can implement a k -fold controlled NOT (i.e., the case $U = \text{NOT}$) using at most $8k-24$ many Toffoli gates, provided that $k \geq 5$. For small values of k , a case analysis shows that $k=2$ requires 1 Toffoli gate, $k=3$ requires at most 4, and $k=4$ at most 10.

We now break these Toffoli circuits further down over the Clifford+T gate set, specifically, we count the number of T gates. Using known implementations [25, 34] of the Toffoli gate

over Clifford+ T it can be shown that its cost in terms of T -gates is given by 7. Note, however, that often it is useful to consider a Toffoli up to a diagonal phase. It is known that this leads to savings in the T -count, specifically, an implementation of Toffoli up to a phase is known that requires 4 T gates only [35, 36].

An analysis of the Toffoli network for the k -fold NOT given in [35], while using as much as possible the cheaper Toffoli up to phase whenever phase cancellations are possible, reveals that for $k \geq 5$ the T -count can be upper bounded by $32k - 84$. For small values of k , again a case analysis can be done which shows that $k = 2$ requires at most 7 T -gates, $k = 3$ requires at most 22 T -gates, and $k = 4$ at most 52.

Each output bit is a Boolean function of $2n$ inputs and each non-zero line of the truth table is implemented by a $\Lambda_{2n}(\text{NOT})$ gate, up to Clifford gates (NOTs). The space overhead of this implementation is constant, namely we need 1 additional qubit in order to implement the decomposition as in [35]. Note that as above we have to only implement the leading $n + 2$ bits of the product. We make the conservative worst case assumption that the Hamming weight of each output bit could potentially be as high as 2^{2n} and we have $n + 2$ output bits, i.e., we get an upper bound of $(n + 2)(2^{2n}(32(2n) - 84)) = 2^{2n}(64n^2 + 44n - 168)$ T -gates are required for computing the $(n + 2)$ highest order bits of the product of two n -bit numbers using a reversible circuit U_f . Here we used 1 additional ancilla qubit to enable the linear time factorization of the $2n$ -fold controlled NOT gates.

As above we have an overhead of $1.15(n + 2) \log_2((n + 2)/2^{(n+2)})$ T gates to implement the Enc gates. Putting everything together we get an overall cost of $2 \cdot 2^{2n}(64n^2 + 44n - 168) + 1.15(n + 2) \log_2((n + 2)/2^{(n+2)})$ many T -gates, where the leading factor of 2 is due to the cleanup of the ancillas. For small values of n , the resulting upper bound estimates on the number of T -gates are shown in Table 13. Note that for the case $n = 2$ we cannot use the formula as it falls within one of the special cases of small number of controls. In this case we obtain a bound of the T gates arising from the reversible part of the circuit as 6, 656 gates to which the cost for Enc has to be added.

Note that the space bound on the number of qubits for the table lookup implementation is given by 1 qubit for the result of each output function, 1 qubit as an ancilla, and 1 qubit to store the final rotation, i.e., a total of 3 qubits.

It seems possible that this crude upper bound on the number of T gates can be improved by reusing intermediate results and output bits [37, 38] or by applying synthesis techniques based on the Reed-Muller transform [7], however, an exponential lower bound for f_n (the “middle bit”) of $\Omega(2^{n/2})$ is known for branching programs that are allowed to read the inputs a constant number of times [30] and also for the most significant bit an exponential lower bound of $\Omega(2^{n/720})$ is known (for OBDDs which are a special case of branching programs) [31], hence even after optimization, the circuit complexity will be exponential in case there is only a constant amount of memory available.

6 Reciprocals

An important gap in the application of the Harrow, Hassidim and Lloyd quantum algorithm for solving linear systems [2] is the fact that a rotation of the form $e^{-iX/a}$ must be performed for some superposition over the values of a stored in a quantum state. The conventional approach to solving this problem is to provide a classical reversible circuit for the reciprocal

and use it to compute $1/a$ into a qubit string stored in a tensor product with a . With this value in hand, $e^{-iX/a}$ can be performed using a series of controlled rotations. This procedure is discussed in detail in [15] and also in Section 6.3. A major drawback of this approach is that many qubits are required to store $1/a$. A method that goes directly from $|a\rangle$ to $|a\rangle e^{-iX/a} |0\rangle$ without needing to compute $1/a$ in a qubit register would therefore be quite useful.

Newton's method is perhaps the most commonly prescribed method for computing the reciprocal. The reason for its popularity stems from the fact that (i) Newton's method converges quadratically for a good initial guess and (ii) the approach only requires multiplication and addition. In particular, if x_n is an approximation to the value of the reciprocal then Newton's method provides a new approximation

$$x_{n+1} = 2x_n - ax_n^2. \quad (35)$$

This process begins with a reasonable guess for the value of the reciprocal, such as $x_1 = 2^{-\lceil \log_2 a \rceil}$, and (35) is then iterated until the error is sufficiently small. Since the error shrinks quadratically, $n \in \Theta(\log \log 1/\epsilon)$ iterations suffice to reduce the error to at most ϵ . A direct application of Newton's method is not well suited for calculating the reciprocal using RUS arithmetic because each iteration requires four copies of x_n and hence the total number of rotations required scales as $O(4^n)$; making a direct application of this method costly. This approach can be made more viable by unrolling the recurrence relation into a polynomial and then approximating the polynomial using the methods of Section 3, but the cost of implementing the resulting polynomial using RUS arithmetic can be prohibitive because the coefficients in the polynomial diverge exponentially. Caching methods, described in Appendix D, can also be used to reduce the cost of implementing Newton's method using RUS arithmetic at the price of increased circuit width.

We focus on two other methods for computing the reciprocal using RUS circuits. First we discuss directly implementing a Chebyshev approximant to the reciprocal and then consider implementing the binomial method for implementing the reciprocal. Both approaches yield practical methods for approximating $1/a$ using RUS arithmetic. The first step in both of these methods involves rescaling a . This step is important because it circumvents the problem of exponentially diverging coefficients that appears in a direct application of Newton's method. This rescaling can be expressed as

$$\frac{1}{a} = 2^{-\lceil \log_2 a \rceil} \left(\frac{1}{2^{-\lceil \log_2 a \rceil} a} \right), \quad (36)$$

where $2^{-\lceil \log_2 a \rceil} a \in [1/2, 1]$. We can therefore introduce a new variable

$$y = 1 - 2^{-\lceil \log_2 a \rceil} a, \quad (37)$$

where $y \in [0, 1/2]$. In other words, we seek to find a power series approximation in powers of y to

$$2^{-\lceil \log_2 a \rceil} \left(\frac{1}{1 - y} \right). \quad (38)$$

Three natural methods then arise for implementing this: Taylor series, Chebyshev polynomials and the binomial division algorithm. Taylor series tend to provide poor accuracy for this application because of the slow convergence of the series near $y = 1/2$ (as alluded to previously). For this reason, we focus our attention on the remaining two methods.

Gadget name	Formula	Maximum error
R_2	$1.012194 + .608948y + 2.664355y^2$	1.6×10^{-2}
R_4	$1.000359 + .966359y + 1.490195y^2 - 1.362554y^3 + 5.019604y^4$	5.1×10^{-4}
R_6	$1.000012 + .9980208y + 1.059785y^2 + .336629y^3 + 4.386547y^4 - 7.295458y^5 + 9.456853y^6$	1.2×10^{-5}

Fig. 16. Chebyshev approximants to $1/(1 - y)$ on $y = 0 \dots 1/2$ where $y = 1 - 2^{-\lceil \log_2 a \rceil} a$.

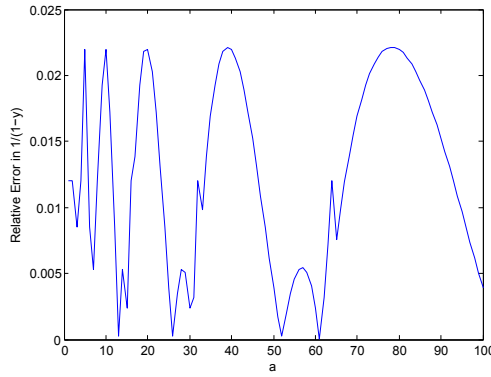


Fig. 17. Error in inversion using R_2 implemented using (40) and (41) as a function of a as discussed below.

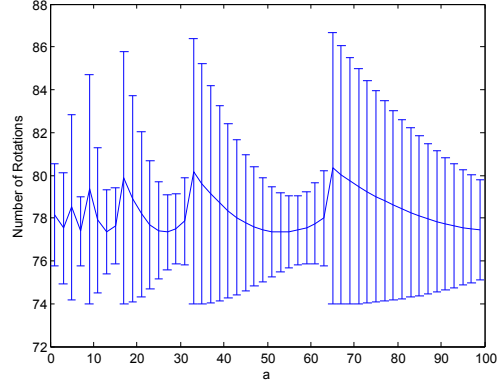


Fig. 18. Cost of implementing R_2 using (40) and (41) as a function of a as discussed below. Error bars give the standard deviation.

6.1 Chebyshev polynomials

Chebyshev polynomials are a complete set of orthogonal polynomials that can be used to represent any piecewise continuous function, such that the infinity norm of the difference between the approximation and the original function is minimized. Taylor series approximations (such as those used in Section 5) provide extremely accurate local approximations to a function but tend not to provide approximations that are accurate throughout the domain of the function. Thus Chebyshev polynomials are often the preferred method for obtaining a polynomial approximation to a function on an interval. The properties of these polynomials are well studied and discussed in detail in [39]. The key point behind this approach is that by doing a Chebyshev polynomial expansion, we can reduce the problem of finding the reciprocal to that of implementing a polynomial. This can be achieved by using the multiplication formulas provided in Section 5. The three lowest-order Chebyshev approximants to the rescaled reciprocal, $1/(1 - y)$, are given in Figure 16.

We could directly implement these formulas for the (rescaled) reciprocal using the multiplication circuits discussed in Section 5 for implementing y^2 but instead using gearbox circuits to implement an approximate squaring circuit. We do this because GB is naturally an RUS circuit so it is much less costly to construct an approximate squaring circuit using these components than it is to use PAR circuits and oblivious amplitude amplification to convert them into RUS circuits.

We also use another optimization that exploits the fact that $y \in [0, 1/2]$. This means that the worst-case errors can be minimized by using a Taylor-series expansion about $y = 1/2$

rather than one centered about $y = 0$. In other words, we express y^2 as The second trick uses the fact that

$$y^2 = (y - 1/4)^2 + \frac{y}{2} - \frac{1}{16}, \quad (39)$$

and use $\text{GB}(y - 1/4) \approx (y - 1/4)^2$.

A further optimization that we consider is using high-order formulas for computing $(y - 1/4)^2$. The methods of Section 3 can be used to show that

$$x^2 = \text{GB}(x) - \text{GB}(x, x, \arcsin(\sqrt{2/3})) - \text{GB}(x, x, x, \arcsin(\sqrt{22/45})) + O(x^8). \quad (40)$$

If we cost all such inputs at one rotation, this approximant requires a minimum of 16 rotations. Similarly, for any constant $\alpha \geq 2/\sqrt{15} \approx 0.52$.

$$\begin{aligned} \alpha x^2 = & \text{GB}(x, \arcsin(\sqrt{\alpha})) - \text{GB}(x, x, \arcsin(\sqrt{\alpha^2 - \alpha/3})) \\ & - \text{GB}(x, x, x, \arcsin(\sqrt{2\alpha^3/3 - 8\alpha/45})) + O(x^8). \end{aligned} \quad (41)$$

Thus $0.664355(y - 1/4)^2$ can be approximated to eighth order using this approach using a minimum of 18 rotations. Since the requisite rotations are nearly half the size of those required for y^2 , we do not need to expand about the midpoint to obtain sufficient accuracy for this rotation. The approximant R_2 can then be realized by combining these ideas and noting that αy can be directly prepared from $|a\rangle$ for any constant α using standard synthesis methods.

Expansion into Chebyshev polynomials provides an excellent way to represent $1/(1-y)$ as a power series in y that minimizes the max-norm of the difference between the approximant and the actual function. Although such expansions can be practical and highly space efficient, we see that the coefficients in R_2 , R_4 and R_6 do not remain small as the order of the polynomial approximation increases. Although the Chebyshev approximation theorem clearly shows that such errors can be made arbitrarily small by increasing the order of the polynomial, the coefficients may increase with the approximation order. This makes the complexity analysis much challenging since the cost of RUS arithmetic depends on these coefficients. We will see below that such problems do not occur when the binomial method is used for division.

6.2 The binomial method

The binomial method is an alternative to Newton's method for computing the reciprocal that also has the property that it converges quadratically. Here quadratic convergence means that the error drops doubly exponentially with the number of iterations used in the method. At its heart, the binomial method is simply a re-grouping of the terms in the Taylor series of $(1 - y)^{-1}$. It reads

$$\frac{1}{1 - y} \approx (1 + y)(1 + y^2) \cdots (1 + y^{2^{n-1}}), \quad (42)$$

and the error in this approximation is at most 2^{-2^n} . The resultant series can be implemented as a series of multiplications of a form that is similar to that in Corollary 2.

Quadratic convergence does not occur for direct implementations of the binomial method using RUS arithmetic. This is because the product $(1 + X)(1 + Y)$ requires $O(1/(\epsilon^{1/(2k)}))$ copies of X and Y if we want to implement the multiplication within error ϵ . By iterating this process, it is then clear that $(1 + y)(1 + y^2) \cdots (1 + y^{2^{n-1}})$ requires $e^{O(1/\epsilon^{1/(2k)})}$ copies of

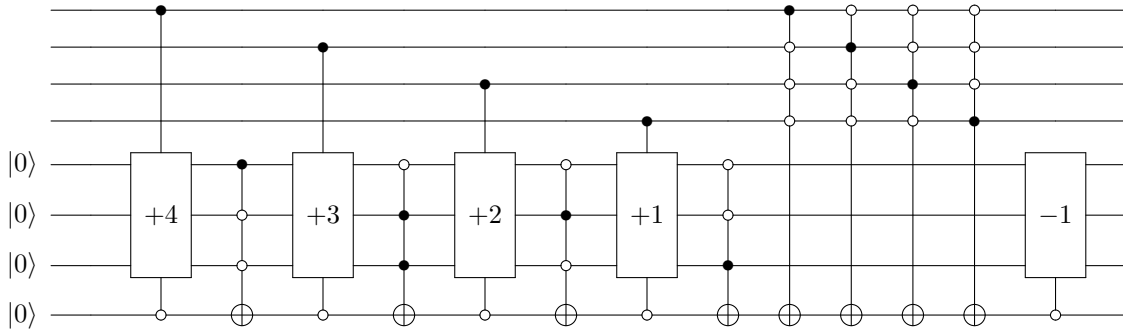
$y, y^2, \dots, y^{2^{n-1}}$. Thus linear, rather than quadratic convergence, is expected if the operations in the product are performed sequentially using RUS arithmetic. Quadratic convergence can be recovered, however, in architectures that can prepare $e^{O(n)} = O(\text{polylog}(1/\epsilon))$ copies of the requisite states in parallel or by using caching strategies (which we discuss in Appendix D). In either case, the price of recovering quadratic convergence is a substantial increase in the width of the resulting circuits.

The following corollary shows that the binomial method can be used to compute the reciprocal in a remarkably space efficient manner.

Corollary 3 *Given a qubit string $|a\rangle \in \mathbb{C}^{2^m}$ the rotation $\exp(-iX/a)$ can be approximated to within distance ϵ as measured by the 2-norm using $\lceil \log_2(m+1) \rceil + O(\log \log(1/\epsilon))$ additional qubits.*

Proof. The error in the binomial approximation to the reciprocal is at most 2^{-2^n} for any n , hence if we wish the error to be ϵ then it suffices to take $n \in \Omega(\log \log(1/\epsilon))$. This also suffices to make the error $O(\epsilon/n)$, which is sufficient to guarantee that the overall error is at most $O(\epsilon)$. All of the monomials present in the binomial expansion can either be constructed using $\text{GB}^{op}(y)$ for $p = 1, \dots, n$ or directly implemented from the qubit representation of y . This construction requires at most n qubits. Since this is an RUS circuit, we can make the error arbitrarily small using time-slicing without resorting to oblivious amplitude amplification, which substantially reduces the cost of method. Corollary 2 can then be used to perform the multiplications using M_2 , which requires only 2 qubits within error at most $O(\epsilon/n)$. Therefore the total error can be made at most $O(\epsilon)$ using the claimed number of qubits.

The remaining issue is that of constructing $y = 1 - 2^{-\lceil \log a \rceil} a$ from $|a\rangle$. This problem is equivalent to bit shifting a to the right until a number in the range $[1/2, 1]$ is attained. This, in effect, becomes the problem of preparing the state $|a\rangle |\lceil \log_2 a \rceil\rangle$. This can be implemented using a simple reversible circuit, as illustrated below for the case of 4 input qubits, where $\boxed{+j}$ represents an adder circuit that increments a register by j .



This requires $\lceil \log_2 m + 1 \rceil$ additional qubits (for technical reasons one additional qubit is also used in the reversible circuit that implements this, but this does not change the scaling). Given this state, the input bit string a can be logically bit shifted so that $\exp(-i2^{-\lceil \log a \rceil} aX)$ can be performed from $|a\rangle$ using a series of controlled rotations. Thus the rotation can be implemented within error ϵ using $\lceil \log_2(m+1) \rceil + O(\log \log(1/\epsilon))$ qubits, as claimed. \square

6.3 Comparison with classical methods for computing reciprocals

Similar to the case of multiplication, we provide a comparison between the resources required for RUS circuits for approximate computation of reciprocals with reversible circuits. We consider three different ways of implementing the computation of the reciprocal of n -bit integers: (i) a reversible implementation of the (extended) Euclidean algorithm following [40], (ii) a reversible implementation of Newton's method following [15], and (iii) a table-lookup implementation.

Similar to our comparison of conventional reversible circuits against RUS arithmetic for multiplication, we will choose a problem for which the inputs and outputs of both algorithms are comparable. We take our problem to be one inspired by the linear systems algorithm [2]. A qubit string encoding a value a is provided consisting of n bits and we wish to use this to perform $e^{-iX/a}$. Again we assume that all operations are decomposed into Clifford and T gates and that the result is accurate to n bits, meaning that the error in the resultant rotation is at most $2^{-(n+1)}$.

6.3.1 Reciprocals via extended Euclidean algorithm

The basic idea is that for an input $x = \sum_{i=0}^{n-1} x_i 2^i$ the first n digits of the reciprocal of x can be computed by running the extended Euclidean algorithm for the computation of the greatest common divisor $\text{GCD}(x, 2^n)$. By performing bit-shifts if necessary, we can assume that x is odd, i.e., we are in the case where the GCD is equal to 1. The extended Euclidean algorithm will then produce two integers r and s such that $rx + s2^n = 1$. If $r = \sum_{i=0}^{n-1} r_i 2^i$, then the bit presentation of x^{-1} is given by $\sum_{i=0}^{n-1} r_i 2^{-n+i}$. The computation of the extended Euclidean algorithm is highly non-trivial as the number of iterations of the basic reduction step depends on the inputs.

A reversible implementation of the extended Euclidean algorithm has been given that synchronizes the computation for any pairs of inputs so that it has the same number of steps as given in [40] and resources estimates are provided in [40, Section 5.4.1]. The synchronization of the computation of the GCD of two n -bit number requires $4.5n$ repetitions of a fundamental cycle which in turn consists of the controlled application—depending on the content of a flag qubits—of 4 adders, one swap, and one comparison circuit. Up to leading order, only the controlled adders matter and as a comparison can be reduced to an adder, we get an overall cost of 5 times the cost for a controlled n -bit adder which we estimated earlier in Section 5.3 to be upper bounded by $84n$ in terms of T -gates. Hence we obtain $4.5n \cdot 5 \cdot 84n = 1,890n^2$ as an upper bound for the number of T -gates. Putting everything together, we obtain the upper bound $2 \cdot 1,890n^2 + 1.15(n+2) \log_2((n+2)/2^{(n+2)})$ on the number of T gates. The space requirements for this method are bounded above by $5n + 4 \log_2(n)$ in [40, Section 5.4.2] plus 1 additional qubit for the output rotation.

6.3.2 Reciprocals via Newton's method

For a given n -bit integer x one can obtain an approximation for $1/x$ by iterating the map $\mu : z \mapsto 2z - xz^2$ which quickly converges toward the fixed point of μ which is given by $\mu(1/x) = 1/x$. In order to achieve an output that has the first n bits equal to those of $1/x$, one needs at most $O(\log n)$ many iterations of μ , when starting from an initial value that can be chosen to be any number between 2^{-n} and 2^{-n+1} . A detailed analysis of the number of

iterations to achieve a target precision of n bits has been provided in [15]: as the number of operations for the computation of μ in each iteration can be bounded by the circuit sizes for two multiplications (one for the square and one for multiplication with x ; the multiplication by 2 can be implemented as a bit shifts), the number of iterations can be bounded by $2 \log_2(n)$ and there is an overhead of a factor of 2 as the computation has to be reversed to disentangle the ancillae used in the algorithm. Overall, we obtain $4 \log_2(n)$ times the cost for one adder and two multipliers. We bound the T -count for an n -bit multiplier in Section 5.3 above by $18n^2 + 18n$ and choose the adder to be the in-place adder from [11, Table 1] for which the cost in terms of Toffoli gates has been bounded by $2n - 1$, i.e., we can bound the number of T -gates for this adder by $14n - 7$. The total T -count is therefore the sum of all these contributions: $4 \log_2(n) \cdot (2(18n^2 + 18n) + (14n - 7)) = (144n^2 + 200n - 28) \log_2(n)$ for the computation of U_f as in Figure 14

Putting everything together, we obtain $2 \cdot (144n^2 + 200n - 28) \log_2(n) + 1.15(n + 2) \log_2((n + 2)/2^{(n+2)})$ many T -gates, where as before the leading factor of 2 is due to the cleanup of the ancillas. For small values of n , the resulting upper bound estimates on the number of T -gates are shown in Table Figure 19.

For the space requirements we obtain that in each iteration $n + 2(2n - 1)$ new ancillas are required, so that in total we get an upper bound of $2 \log_2(n)(5n - 2) = (10n - 4) \log_2(n)$ for the computation of U_f . As in case of the multiplier we now have to compute the result into a rotation which leads to an additional overhead of 1 qubit so that in total we need no more than $\log_2(n)(10n - 4) + 1$ qubits.

6.3.3 Reciprocals via table-lookups

This leads to the same bound as in the case of the table-lookup implementation of the multipliers, with the only difference being that the input and output sizes are n . We get an upper bound of $2 \cdot (n2^n(32n - 84)) + 1.15(n + 2) \log_2((n + 2)/2^{(n+2)}) = 2^{n+1}(32n^2 - 84n) + 1.15(n + 2) \log_2((n + 2)/2^{(n+2)})$ many T -gates that are required to implement the computation of the reciprocal of an n -bit numbers in constant space where $n \geq 5$. For $n = 2$ we obtain the bound $8 \cdot 2 \cdot 7 = 112$ plus the cost for Enc and for $n = 4$ we obtain the bound $2^5 \cdot 5 \cdot 52 = 8320$ plus the cost for Enc . These results are summarized in Figure 19.

As in case of the multiplier we obtain that the circuit can be implemented with an additional of 3 qubits only.

6.3.4 Comparison with RUS methods

A comparison between the resources required for reciprocals based on the Euclidean method, Newton's method, and the table-lookup method and R_2 is given in Figure 19. The comparison was done using RUS synthesis to convert the rotations in R_2 into Clifford and T circuits and using the same Toffoli gate used in previous steps. M_4 and M_6 were used for $n = 2$ and $n = 4$ respectively to rescale the reciprocal after $1/(1 - y)$ was computed using R_2 , and in both cases 99% of the error tolerance was used in the synthesis steps on the rotations in R_2 because these constitute the inner loop of the algorithm. Making these rotations as inexpensive as possible helps reduce the cost of the overall algorithm because they are repeated many more times than the rescaling operation used after $1/(1 - y)$ is computed.

The first feature to note is that the number of ancilla qubits used in a reversible implementation of Euclid's algorithm and Newton's method are quite daunting. Our estimates suggest

Reciprocal method	$n = 2$		$n = 4$		$n = 8$		$n = 16$	
	T -count	qubits	T -count	qubits	T -count	qubits	T -count	qubits
Euclid	1.51E+04	12	6.05E+04	23	2.42E+05	44	9.68E+05	85
Newton	1.92E+03	17	6.21E+03	73	2.17E+04	229	8.05E+04	625
Table-lookup	1.40E+02	3	8.38E+03	3	7.05E+05	3	8.98E+08	3
R_2	3.17E+03	6	1.53E+05	6	NA		NA	

Fig. 19. A comparison of the resources required for space efficient computation of reciprocals on a quantum computer. Shown are mean circuit sizes (number of T -gates) and number of required qubits for $n = 2, 4, 8, 16$ bit numbers. We assume $a \in (1, 2]$ for simplicity in the case of R_2 .

Input: Function f , endpoints for interpolation range x_{\min} and x_{\max} , number of square waves used in the synthesis N , k recursion order for Gearbox circuit.

Output: $\{a_j : j = 1, \dots, N\}$ such that $f(x) \approx \sum_j a_j \frac{4}{\pi} (\text{GB}^{\circ k}(xN\pi/(2j(x_{\max} - x_{\min}))) + \pi/4) - \pi/4$.

```

function SQUAREWAVEAPPROX( $f, x_{\min}, x_{\max}, N$ )
   $x_j \leftarrow x_{\min} + (x_{\max} - x_{\min})(j - 1/2)/N$ 
   $T_j \leftarrow 2(x_{\max} - x_{\min})j/N$ 
   $A_{j,k} = \frac{4}{\pi} (\text{GB}^{\circ k}(x_j\pi/T_k + \pi/4) - \pi/4)$ 
   $\vec{y} \leftarrow f(x_j)$  ▷ Compute function on mesh
  return  $A^{-1}\vec{y}$  ▷ Solve system of equations,  $A\vec{a} = \vec{y}$ , for coefficients  $a_j$ 
end function

```

Fig. 20. Square Wave Synthesis Algorithm.

that dozens to hundreds of qubits will be needed to perform a computationally useful quantum linear systems algorithm. The Chebyshev approximation implemented using RUS synthesis methods, R_2 , requires 6 qubits in contrast to these results and requires a number of T -gates that is comparable to Euclid's method but substantially greater than Newton's method. Unfortunately, because Chebyshev approximants have fixed accuracy, it is impossible to provide more than 4 bits of precision. Perhaps surprisingly, for the circuit sizes considered, table-lookup proved to be a viable approach because the table is only one dimensional. We see that it is clearly a method of choice for small inversion problems, but its poor asymptotic scaling will make higher-order variants of R_2 much less expensive for more inversion problems with more stringent requirements on the error tolerance. These results show that, unless further optimizations are used for division algorithms, the qubit requirements involved in performing an inversion are far beyond the capabilities of existing quantum computers and may exceed those of even near-future quantum computers.

We will now illustrate these ideas by using approximate square waves to implement $1/(1 - y)$ on $y \in [-0.1, 0.6]$. We use $k = 8$ and take $N = 71$ and find from the data in Figure E.1 that the function can be approximated on the subinterval $y \in [0, 0.5]$ to within a maximum

error of 2.1% and a mean error of 0.36%. These spikes in the error occur because if y is in the interval where the function is ramping up from -1 to 1 then the midpoint approximation will fail to be very accurate. The probability of such events can be decreased by increasing k at the price of increasing the circuit size.

The cost of this synthesis process is $O(2^k N \log N)$. However, an advantage of this method is that the rotation-depth of the resulting circuit is substantially smaller: $O(kN \log N)$ [17], meaning that the overwhelming majority of the rotations used can be performed simultaneously in an architecture that allows parallel execution.

As a final note, the square wave method should not just be seen as a method for implementing a piecewise constant function. A broader view is to see each square wave as a logical function that thresholds the rotation angles input into the gearbox circuit. These circuits therefore give a way to approximate logical operations on rotation angles and hence square wave synthesis can be viewed as a special case of logical operations being used to implement a function via a lookup table on N points with piecewise constant interpolation between them.

7 Conclusion

Our main contribution is a fundamentally quantum approach to arithmetic that stores both inputs and outputs not as qubit strings but as amplitudes. More specifically, our method represents the value ϕ as $e^{-i\phi X} |0\rangle$. The approach uses measurement as an active participant in the approximations. This strategy is profitable because of the use of repeat until success circuits, which removes the possibility of a faulty measurement irreparably corrupting an entire multi-step computation. In a deep sense, our work can be thought of as providing a generalization of existing circuit synthesis ideas wherein axial rotations of the form $e^{-if(\phi_1, \dots, \phi_k)X}$ are implemented instead of a constant function of the form e^{-ifX} .

Optimizing the performance of our circuits therefore requires a different flavor of numerical analysis that takes into account the ability of measurement to non-deterministically apply non-linear transformations on the input data. We provide a first foray into this field of quantum numerical analysis by providing Taylor series based methods for finding arbitrarily high-order approximations to smooth functions using repeat until success circuits. Specific examples are given for multiplication and division wherein we find that we can implement both operations using substantially fewer operations than the most popular methods currently proposed. In particular, our methods can provide an arbitrarily accurate multiplier using a constant number of ancilla qubits. Reciprocals can also be implemented in this manner using a number of qubits that scales logarithmically in the number of bits of precision required. This is significant because reversible circuits for division or multiplication can require hundreds of logical qubits, making algorithms like linear systems algorithms outside of the reach of small quantum computers. Our ideas promise to enable such applications of quantum computers.

An interesting point to note, however, is that we find that the costs of implementing reciprocals using a lookup table was found to be surprisingly small for applications where modest precision (at most 16 bits) are required. This suggests that brute force inversion may also be a practical way to implement the inversion step in linear systems algorithms on a small quantum computer.

Looking forward, there are several potential further avenues of inquiry revealed by this work. There may exist other classes of repeat until success circuits that are yet to be found

that could enable new ways of synthesizing functions. This may lead to more efficient, or more natural, ways of implementing arithmetic in our paradigm. In a similar vein, synthesizing functions by way of a Taylor series expansion is perhaps not the most natural way to implement these rotations since multiplication is not natural for large angles in our framework. The gearbox and generalized PAR circuits proposed here may also have applications beyond function synthesis. Also the question of whether sub-polynomial scaling with ϵ^{-1} can be obtained with this method also remains open. Finally, a natural extension of our work on square wave synthesis is to consider a decomposition of f into a sum of approximate Walsh functions (previously considered for circuit synthesis in [41]). This idea would constitute a natural way to approximate arbitrary functions on a quantum computer without using a polynomial expansion. This list is by no means exhaustive and we suspect that the ideas presented here may stimulate the development of new and innovative methods for performing arithmetic, function approximation or state distillation on fault-tolerant quantum computers.

References

1. P. W. Shor (1997), *Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer*, SIAM Journal on Computing, 26, pp. 1484–1509.
2. A. W. Harrow, A. Hassidim and S. Lloyd (2009), *Quantum algorithm for linear systems of equations*, Physical Review Letters, 103, 150502.
3. D. W. Berry, G. Ahokas, R. Cleve and B. C. Sanders (2007), *Efficient quantum algorithms for simulating sparse Hamiltonians*, Communications in Mathematical Physics, 270, pp. 359–371.
4. C. H. Bennett (1973), *Logical reversibility of computation*, IBM J. Res. Develop., 17, pp. 525–532.
5. V. V. Shende, S. S. Bullock and Igor L Markov (2006), *Synthesis of quantum logic circuits*, IEEE Transactions on Computer-Aided Design, 25, pp. 1000–1010.
6. D. Maslov, D. M. Miller and G. W. Dueck (2007), *Techniques for the synthesis of reversible Toffoli networks*, ACM Transactions on Design Automation of Electronic Systems, 12, 42.
7. C. C. Lin and N. K. Jha (2014), *RMDDS: Reed-Muller decision diagram synthesis of reversible logic circuits*, ACM Journal on Emerging Technologies in Computing Systems, 10,14.
8. V. Vedral, A. Barenco and A. Ekert (1996), *Quantum networks for elementary arithmetic operations*, Physical Review A, 54, pp. 147–153.
9. D. Beckman, A. N. Chari, S. Devabhaktuni and J. Preskill (1996), *Efficient networks for quantum factoring*, Phys. Rev. A, 54, pp. 1034–1063.
10. T. G. Draper, S. A. Kutin, E. M. Rains and K. M. Svore (2006), *A logarithmic-depth quantum carry-lookahead adder*, Quantum Information & Computation, 6, pp. 351–369.
11. S. A. Cuccaro, T. G. Draper, S. A. Kutin, and D. P. Moulton (2004), *A new quantum ripple-carry addition circuit*, quant-ph/0410184.
12. Y. Takahashi and N. Kunihiko (2005), *A linear-size quantum circuit for addition with no ancillary qubits*, Quantum Information & Computation, 5, pp. 440–448.
13. S. Beauregard (2002), *Circuit for Shor’s algorithm using $2n + 3$ qubits*, quant-ph/0205095.
14. R. van Meter and K. M. Itoh (2005), *Fast quantum modular exponentiation*, Phys. Rev. A, 71, 052320.
15. Y. Cao, A. Papageorgiou, I. Petras, J. Traub and S. Kais (2013), *Quantum algorithm and circuit design solving the Poisson equation*, New Journal of Physics, 15, 013021.
16. V. Kliuchnikov, D. Maslov and M. Mosca (2013), *Fast and efficient exact synthesis of single-qubit unitaries generated by Clifford and T gates*, Quantum Information & Computation, 13, pp. 607–630.
17. N. Wiebe and V. Kliuchnikov (2013), *Floating point representations in quantum circuit synthesis*, New Journal of Physics, 15, 093041.
18. A. Paetznick and K. M. Svore (2014), *Repeat-Until-Success: Non-deterministic decomposition of*

- single-qubit unitaries*, Quantum Information & Computation, 14, pp. 1277-1301.
19. A. Bocharov, M. Roetteler and K. M. Svore (2015), *Efficient synthesis of probabilistic quantum circuits with fallback*, Physical Review A, 91, 052317.
 20. Neil J Ross and Peter Selinger (2014), *Optimal ancilla-free Clifford+ T approximation of z-rotations*, arXiv:1403.2975.
 21. N. Wiebe, D. Braun and S. Lloyd (2012), *Quantum algorithm for data fitting*, Phys. Rev. Lett., 109, 050505.
 22. Guoming Wang (2014), *Quantum algorithms for curve fitting*, arXiv:1402.0660.
 23. G. Brassard, P. Høyer, M. Mosca, and A. Tapp (2000), *Quantum amplitude amplification and estimation*, quant-ph/0005055.
 24. E. Knill, R. Laflamme, and G. J. Milburn (2001), *A scheme for efficient quantum computation with linear optics*, Nature, 409, pp. 46–52.
 25. M. A. Nielsen and I. L. Chuang (2010), *Quantum computation and quantum information*, Cambridge University Press (New York).
 26. N. C. Jones, J. D. Whitfield, P. L. McMahon, M.-H. Yung, R. Van Meter, A. Aspuru-Guzik, and Y. Yamamoto (2012), *Faster quantum chemistry simulation on fault-tolerant quantum computers*, New Journal of Physics, 14, 115023.
 27. M. H. Stone (1948), *The generalized Weierstrass approximation theorem*, Mathematics Magazine, 21, pp. 237–254.
 28. D. W. Berry, A. M. Childs, R. Cleve, R. Kothari and Somma, R. D. (2014), *Exponential improvement in precision for simulating sparse Hamiltonians*, In Proceedings of the 46th Annual ACM Symposium on Theory of Computing, pp. 283-292.
 29. A. A. Karatsuba (1995), *The complexity of computations*, Proceedings of the Steklov Institute of Mathematics-Interperiodica Translation, 211, pp. 169–183.
 30. I. Wegener and P. Woelfel (2007), *New results on the complexity of the middle bit of multiplication*, Computational Complexity, 16, pp. 298–323.
 31. B. Bollig and J. Klump (2011), *New results on the most significant bit of integer multiplication*, Theory Comput. Syst., 48, pp. 170–188.
 32. P. Pham and K. M. Svore (2013), *A 2D nearest-neighbor quantum architecture for factoring in polylogarithmic depth*, Quantum Information and Computation, 13, pp. 937–962.
 33. M. Saeedi and I. L Markov (2013), *Quantum circuits for GCD computation with $O(n \log n)$ Depth and $O(n)$ Ancillae*, arXiv:1304.7516.
 34. M. Amy, D. Maslov, M. Mosca and M. Rötteler (2013), *A meet-in-the-middle algorithm for fast synthesis of depth-optimal quantum circuits*, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 32, pp. 818–830.
 35. A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. W. Shor, T. Sleator, J. Smolin and H. Weinfurter (1995), *Elementary gates for quantum computation*. Physical Review A, 52, pp. 3457–3467.
 36. P. Selinger (2013), *Quantum circuits of T-depth one*, Phys. Rev. A, 87, 042302.
 37. D. M. Miller, D. Maslov, and G. W. Dueck (2003), *A transformation based algorithm for reversible logic synthesis*, In Proceedings of the 40th Design Automation Conference (DAC'03), pp. 318–323.
 38. A. Shafaei, M. Saeedi and M. Pedram (2013), *Reversible logic synthesis of k-input, m-output lookup tables*, In Proceedings of the Conference on Design, Automation and Test in Europe (DATE'13), pp. 1235–1240.
 39. M. Abramowitz and I. A Stegun (1964), *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, Courier Dover Publications, (Mineola NY).
 40. J. Proos and C. Zalka (2003), *Shor's discrete logarithm quantum algorithm for elliptic curves*, Quantum Information and Computation, 3, pp. 317–344.
 41. J. Welch, D. Greenbaum, S. Mostame and A. Aspuru-Guzik (2014), *Efficient quantum circuits for diagonal unitaries without ancillas*, New Journal of Physics, 16, 033040.
 42. O. Skliar, V. Medina and R. E. Monge (2008), *A new method for the analysis of signals: the square wave method*, Revista de Matemática. Teoría y Aplicaciones, 15, pp. 109–129.

Appendix A. Proof of Lemma 2

The proof of Lemma 2 requires the use of oblivious amplitude amplification [28], which is a technique that allows amplitude amplification to be performed without knowing the initial state that the system is prepared in. We state the result, proven in Lemma 3.6 of [28], in the following lemma.

Lemma A.1 (*Berry, Childs, Cleve, Kothari, Somma*) *Let U and V be unitary matrices on $\mu + n$ qubits and n qubits, respectively, and let $\theta \in (0, \pi/2)$. Suppose that for any n -qubit state $|\psi\rangle$*

$$U |0\rangle |\psi\rangle = \sin(\theta) |0^\mu\rangle V |\psi\rangle + \cos(\theta) |\phi\rangle,$$

where $|\phi\rangle$ satisfies $\langle 0^\mu | \langle 0^\mu \otimes \mathbf{1} | |\phi\rangle = 0$. Let $R := (2 |0^\mu\rangle \langle 0^\mu| \otimes \mathbf{1}) - \mathbf{1}$ and $S := -URU^\dagger R$. Then for any $t \in \mathbb{Z}$,

$$S^t U |0\rangle |\psi\rangle = \sin((2t + 1)\theta) |0^\mu\rangle V |\psi\rangle + \cos((2t + 1)\theta) |\phi\rangle.$$

In our context the R operation in the above lemma can simply be taken to be the Z gate. This result naturally leads to the following corollary:

Corollary A.1 *If $U : |0\rangle |\psi\rangle \mapsto \frac{|0\rangle V |\psi\rangle + |1\rangle |\phi\rangle}{\sqrt{2}}$ and $\theta = \pi/4$ then the state $V |\psi\rangle$ can be prepared deterministically using three applications of U and a constant sized Clifford circuit.*

Proof. Oblivious amplitude estimation cannot be used directly to see this result because $\theta = \pi/4$ in this case. Since the amplitudes of the “good” state vary like $\sin((2t + 1)\theta)$, it is clear that amplitude is not transferred to the good solution by repeatedly using oblivious amplitude amplification. Instead we poison the success probability by defining a new success condition that would happen with probability $1/4$ if amplitude amplification was not used.

$$\tilde{U} |0\rangle^2 |\psi\rangle := H |0\rangle \otimes (U |0\rangle |\psi\rangle) = \frac{1}{2} |0^2\rangle V |\psi\rangle + \sqrt{\frac{3}{4}} |\tilde{\phi}\rangle, \quad (\text{A.1})$$

where $\langle \tilde{\phi} | (|0^2\rangle V |\psi\rangle) = 0$. If the top qubit in the above circuit were to be measured then it would yield the desired transformation on the bottom qubit with probability $1/4$. This means that in essence, these transformations allow us to implement a transformation that logically has the same action upon success but with half the original success probability. This is all that we need in order to make the application of V to $|\psi\rangle$ deterministic.

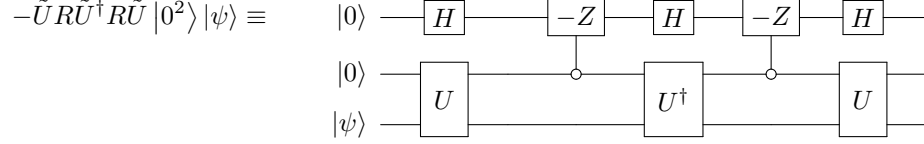
Eq. (A.1) is in exactly in the form required by Lemma A.1 with $\theta = \arcsin(1/2) = \pi/6$ and $\mu = 2$. Thus we see that

$$\begin{aligned} -\tilde{U} R \tilde{U}^\dagger R \tilde{U} |0^2\rangle |\psi\rangle &= \sin((2 + 1)\pi/6) (|0^2\rangle V |\psi\rangle) + \cos((2 + 1)\pi/6) \frac{\sqrt{3}}{2} |\tilde{\phi}\rangle \\ &= |0^2\rangle V |\psi\rangle. \end{aligned} \quad (\text{A.2})$$

Thus the success probability can be boosted to 100% using three calls to \tilde{U} .

The fact that we require a constant sized Clifford circuit can easily be seen from the following circuit, which implements oblivious amplitude amplification for these cases (up to

an irrelevant global phase)



This completes the proof \square

This corollary is significant because it allows us to make any such circuit, such as those used in PAR circuits or non-deterministic circuit synthesis methods [18, 19]. In particular, Corollary A.1 leads directly to a proof of Lemma 2.

Proof of Lemma 2. Consider $\text{PAR}(\phi_1, \dots, \phi_k)$. Eq. (6) shows that the generalized PAR circuits prepare the state

$$\begin{aligned} & \text{PAR}(\phi_1, \dots, \phi_k) : |0\rangle |0\rangle^{k-1} |\psi\rangle \\ & \mapsto \text{GHZ}^{-1} \left(e^{-i\theta_1 X} \otimes \dots \otimes e^{-i\theta_k X} |0\rangle^k |\psi\rangle - \prod_{j=1}^k \cos(\phi_j) |0^k\rangle |\psi\rangle - (-i)^k \prod_{j=1}^k \sin(\phi_j) |1^k\rangle |\psi\rangle \right) \\ & + \frac{1}{\sqrt{2}} |0^k\rangle (\cos(\phi_1) \dots \cos(\phi_k) \mathbf{1} - i \sin(\phi_1) \dots \sin(\phi_k) X) |\psi\rangle \\ & + \frac{1}{\sqrt{2}} |1\rangle |0^{k-1}\rangle (\cos(\phi_1) \dots \cos(\phi_k) \mathbf{1} + i \sin(\phi_1) \dots \sin(\phi_k) X) |\psi\rangle. \end{aligned} \quad (\text{A.3})$$

For notational simplicity in the following we will represent a number v as a bitstring $v_1 \dots v_k$, where products within the ket of the form $|v_1 \dots v_k\rangle$ are interpreted to be the concatenation of the bits so that $|v_1 \dots v_k\rangle \equiv |v\rangle$. Similarly we define powers such that $|v_j^2\rangle \equiv |v_j v_j\rangle$. Then applying the $k-1$ controlled not gates to the state and then the Hadamard on the transform first qubit we find that we can write

$$\begin{aligned} & \text{GHZ}^{-1} \left(e^{-i\theta_1 X} \otimes \dots \otimes e^{-i\theta_k X} |0\rangle^k |\psi\rangle - \prod_{j=1}^k \cos(\phi_j) |0^k\rangle |\psi\rangle - (-i)^k \prod_{j=1}^k \sin(\phi_j) |1^k\rangle |\psi\rangle \right) \\ & = (H \otimes \mathbf{1}) \sum_{0^k \neq v \neq 1^k} \prod_{j=1}^k \cos(\phi_j)^{1-v_j} \sin(\phi_j)^{v_j} (-i)^{\sum_{q=1}^k v_q} |v \oplus 0(v_1)^{k-1}\rangle |\psi\rangle \end{aligned} \quad (\text{A.4})$$

Then by expanding the result we see that it is equivalent to

$$\begin{aligned}
& \sum_{0^k \neq v \neq 1^k} \prod_{j=1}^k \left(\frac{\cos(\phi_j)^{1-v_j} \sin(\phi_j)^{v_j} (-i)^{\sum_{q=1}^k v_q}}{\sqrt{2}} \right. \\
& \quad \left. \times (|v \oplus (v_1)^k\rangle + (-1)^{v_1} |v \oplus (v_1 + 1)v_1^{k-1}\rangle) |\psi\rangle \right) \\
& = \sum_{0^k \neq v \neq 1^k} \prod_{j=1}^k \left(\frac{\cos(\phi_j)^{1-v_j} \sin(\phi_j)^{v_j} (-i)^{\sum_{q=1}^k v_q}}{\sqrt{2}} \right. \\
& \quad \left. \times (|0\rangle |v_2 \cdots v_k \oplus (v_1)^{k-1}\rangle + (-1)^{v_1} |1\rangle |v_2 \cdots v_k \oplus v_1^{k-1}\rangle) |\psi\rangle \right) \\
& := \frac{\sqrt{1-\beta^2}}{\sqrt{2}} (|0\rangle |\chi_0\rangle + |1\rangle |\chi_1\rangle) |\psi\rangle, \tag{A.5}
\end{aligned}$$

for some β that is a function of ϕ_1, \dots, ϕ_k . The quantity β is included in order to ensure that the resulting state has less than unit length. Since $v_2 \cdots v_k \neq (v_1)^{k-1}$ for any of the states in (A.5), $\text{GHZ}^{-1} |\phi\rangle$ is orthogonal to $\text{span}(|0^k\rangle, |1\rangle |0^{k-1}\rangle)$, which means that we can use (A.5) to write for $V_0 = \exp(-i \arctan(\prod_j \tan(\phi_j)) X)$

$$\begin{aligned}
& \text{PAR}(\phi_1, \dots, \phi_k) : |0\rangle^k |\psi\rangle \mapsto \\
& \frac{|0\rangle \left(\beta |0\rangle^{k-1} V_0 |\psi\rangle + \sqrt{1-|\beta|^2} |\chi_0\rangle |\psi\rangle \right) + |1\rangle \left(\beta |0\rangle^{k-1} V_0^\dagger |\psi\rangle + \sqrt{1-\beta^2} |\chi_1\rangle |\psi\rangle \right)}{\sqrt{2}}. \tag{A.6}
\end{aligned}$$

Since the overall state must be normalized to unit length, we can infer from (A.3) that

$$|\beta| = \sqrt{\prod_{j=1}^k \cos^2(\phi_j) + \prod_{j=1}^k \sin^2(\phi_j)}. \tag{A.7}$$

Equation (A.6) is of the form required by Corollary A.1, which implies that we can use oblivious amplitude amplification to deterministically prepare the state

$$\beta |0\rangle^{k-1} V_0 |\psi\rangle + \sqrt{1-\beta^2} |\chi_0\rangle |\psi\rangle,$$

using 3 PAR rotations and a constant sized Clifford circuit. If we then measure the first $k-1$ qubits of this state to be 0^{k-1} then $V_0 |\psi\rangle$ will be performed as required; whereas if we measure any other outcome the identity gate will be applied to $|\psi\rangle$. The probability of implementing V_0 is therefore $\beta^2 = \prod_{j=1}^k \cos^2(\phi_j) + \prod_{j=1}^k \sin^2(\phi_j)$. The circuit is therefore satisfies Definition 1 and is a Repeat-Until-Success circuit with the success probability claimed by Lemma 2. \square

Appendix B. Bounds on mean and variance of cost of RUS circuits

Of course, if we nest several PAR and Gearbox circuits then we have to be able to bound the number of repetitions needed for success to occur with high probability. This can be done using Chebyshev's inequality given the expectation value and the variance of the number of repetitions of the RUS circuits used. These properties are summarized in the following lemma.

Lemma B.1 *Let V be an RUS circuit composed of k RUS circuits whose success probabilities are independent. Let x_1, \dots, x_k be random variables that describe the number of attempts of each of the k RUS circuits that are needed in order to achieve a successful measurement of each of the k RUS circuits and let N be a random variable describing total number of repetitions of V needed for success. The total number of repetitions of x_1, \dots, x_k , R_x , that are needed obeys*

$$\mathbb{E}(R_x) = \mathbb{E}(N) \sum_{j=1}^k \mathbb{E}(x_j) \tag{B.1}$$

$$\mathbb{V}(R_x) = \mathbb{E}(N) \sum_{j=1}^k \mathbb{V}(x_j) + \mathbb{V}(N) \left(\sum_{j=1}^k \mathbb{E}(x_j) \right)^2 \tag{B.2}$$

Proof. Expanding R_x ,

$$R_x := \sum_{i=1}^N \sum_{j=1}^k x_j = \sum_{i=1}^{\infty} \sum_{j=1}^k x_j \mathbf{1}_{i \leq N} = \sum_{i=1}^{\infty} \sum_{j=1}^k (\chi_j + \mathbb{E}(x_j)) \mathbf{1}_{i \leq N}, \tag{B.3}$$

where each χ_j is a zero-mean random variable such that $x_j = \chi_j + \mathbb{E}(x_j)$ and $\mathbf{1}_{i \leq N}$ is the indicator function that is 1 for $i \leq N$ and 0 otherwise.

The random variables N and χ_j are independent and each χ_j has zero-mean, which means that

$$\mathbb{E}(R_x) = \mathbb{E} \left(\sum_{i=1}^{\infty} \sum_{j=1}^k (\chi_j + \mathbb{E}(x_j)) \mathbf{1}_{i \leq N} \right) = \mathbb{E} \left(\sum_{i=1}^{\infty} \mathbf{1}_{i \leq N} \right) \sum_{j=1}^k \mathbb{E}(x_j) = \mathbb{E}(N) \sum_{j=1}^k \mathbb{E}(x_j). \tag{B.4}$$

The variance is a little more difficult to compute

$$\mathbb{V}(R_x) = \mathbb{E}(R_x^2) - \mathbb{E}(R_x)^2. \tag{B.5}$$

Expanding this out we find

$$\mathbb{V}(R_x) = \mathbb{E} \left(\left[\sum_{i=1}^{\infty} \sum_{j=1}^k (\chi_j + \mathbb{E}x_j) \mathbf{1}_{i \leq N} \right] \left[\sum_{i'=1}^{\infty} \sum_{j'=1}^k (\chi_{j'} + \mathbb{E}x_{j'}) \mathbf{1}_{i' \leq N'} \right] \right) - \mathbb{E}(N)^2 \left(\sum_{j=1}^k \mathbb{E}(x_j) \right)^2. \tag{B.6}$$

Since $\mathbb{E}(\chi_j) = 0$, many of the terms in (B.6) are zero. In fact, any term in that is not of the form χ_j^2 or $\mathbb{E}(\chi_j)\mathbb{E}(\chi_k)$ is zero. After this observation we see using the independence of each

random variable that the variance $\mathbb{V}(R_x)$ is

$$\begin{aligned}
& \mathbb{E} \left(\left[\sum_{i=1}^{\infty} \sum_{j=1}^k \chi_j^2 \mathbf{1}_{i \leq N} \right] + \left[\sum_{i=1}^{\infty} \sum_{j=1}^k \mathbb{E} x_j \mathbf{1}_{i \leq N} \right] \left[\sum_{i'=1}^{\infty} \sum_{j'=1}^k \mathbb{E} x_{j'} \mathbf{1}_{i' \leq N'} \right] \right) - \mathbb{E}(N)^2 \left(\sum_{j=1}^k \mathbb{E}(x_j) \right)^2 \\
&= \mathbb{V}(x_j) \mathbb{E}(N) + \mathbb{E}(N^2) \left(\sum_{j=1}^k \mathbb{E}(x_j) \right)^2 - \mathbb{E}(N)^2 \left(\sum_{j=1}^k \mathbb{E}(x_j) \right)^2 \\
&= \mathbb{V}(x_j) \mathbb{E}(N) + \mathbb{V}(N) \left(\sum_{j=1}^k \mathbb{E}(x_j) \right)^2. \tag{B.7}
\end{aligned}$$

This completes the proof. \square

This provides estimates of the mean and the variance of the number of times that the RUS circuits have to be applied in order to achieve the desired rotation. It is easy to go from these results to find estimates for the number of T gates needed to implement either PAR or GB. We focus on the number of T gates required because of their importance in fault tolerant quantum computing wherein T gates are often considered to be the most costly operations because they can require several rounds of magic state distillation.

Corollary B.1 *Let T_1, \dots, T_k be the probability distributions that describe the number of T gates needed to implement the unitaries ϕ_1, \dots, ϕ_k , $T_{\text{PAR}(\phi_1, \dots, \phi_k)}$ be the cost of the PAR circuit (acting on $|0\rangle$) and $T_{\text{GB}(\phi_1, \dots, \phi_k)}$ be the number of T gates needed to implement the garbox circuit then*

$$\mathbb{E}(T_{\text{PAR}(\phi_1, \dots, \phi_k)}) = \frac{4(k-1) + \sum_{j=1}^k \mathbb{E}(T_j)}{P}, \tag{B.8}$$

$$\mathbb{V}(T_{\text{PAR}(\phi_1, \dots, \phi_k)}) = \frac{(4(k-1) + \sum_{j=1}^k \mathbb{E}(T_j))^2 (1-P)}{P^2} + \frac{\sum_{j=1}^k \mathbb{V}(T_j)}{P}, \tag{B.9}$$

where $P = \prod_{j=1}^k \cos(\phi_j)^2 + \prod_{j=1}^k \sin(\phi_j)^2$. Similarly,

$$\mathbb{E}(T_{\text{GB}(\phi_1, \dots, \phi_k)}) = \frac{4(k-1) + 2 \sum_{j=1}^k \mathbb{E}(T_j)}{P}, \tag{B.10}$$

$$\mathbb{V}(T_{\text{GB}(\phi_1, \dots, \phi_k)}) = \frac{(4(k-1) + 2 \sum_{j=1}^k \mathbb{E}(T_j))^2 (1-P)}{P^2} + \frac{2 \sum_{j=1}^k \mathbb{V}(T_j)}{P}, \tag{B.11}$$

where $Q = \prod_{j=1}^k \cos(\phi_j)^4 + \prod_{j=1}^k \sin(\phi_j)^4$.

Proof. Let us first begin with PAR. There are two sources of T gates in PAR: T gates from the k -controlled X gate and T gates from implementing the k rotations:

$$\mathbb{E}(T_{\text{PAR}(\phi_1, \dots, \phi_k)}) = 4(k-1)\mathbb{E}(N) + \mathbb{E} \left(\sum_{i=1}^N \sum_{j=1}^k T_j \right) = \left(4(k-1) + \sum_{j=1}^k \mathbb{E}(T_j) \right) \mathbb{E}(N), \tag{B.12}$$

here we assumed that the Toffoli gate construction of [26] is used to implement the k -controlled X gate. N obeys a geometric distribution with $p = P$, which means that $\mathbb{E}(N) = 1/P$. Hence,

we find by substituting $x_j \rightarrow T_j$ in Lemma B.1 that

$$\mathbb{E}(T_{\text{PAR}(\phi_1, \dots, \phi_k)}) = \frac{4(k-1) + \sum_{j=1}^k \mathbb{E}(T_j)}{P}. \tag{B.13}$$

Similarly, since the variance of the geometric distribution is $(1-P)/P^2$,

$$\begin{aligned} \mathbb{V}(T_{\text{PAR}(\phi_1, \dots, \phi_k)}) &= \mathbb{V}\left(\sum_{i=1}^N \sum_{j=1}^k T_j + 4(k-1)\right) \\ &= \frac{(4(k-1) + \sum_{j=1}^k \mathbb{E}(T_j))^2(1-P)}{P^2} + \frac{\sum_{j=1}^k \mathbb{V}(T_j)}{P}. \end{aligned} \tag{B.14}$$

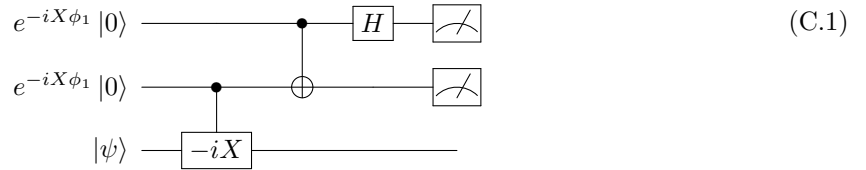
The analogous formulas for GB can be found by noting that the main difference between the T count of GB and PAR is that each ϕ_j is repeated twice in GB, which leads to the costs claimed above. \square

Appendix C. Non-RUS gearbox circuits

GB can be converted into a non-deterministic circuit that requires no online rotations as given below. Such circuits are particularly important in cases where qubits are inexpensive and the architecture allows any parallel operations to be performed simultaneously. This circuit may be preferable to $\text{PAR}(\phi_1, \phi_1)$ for approximately squaring input rotations because the circuit requires only online Clifford operations, which are typically inexpensive in fault tolerant implementations. Such circuits also provide a lower T -depth method for implementing $\text{GB}^{\circ k}(x)$ than that used in Appendix E or [17] because the gearbox circuit is insensitive to the sign of the input rotation angle and also because most of the required operations act upon $|0\rangle$. The latter case is significant because under-rotations can be corrected by applying a Z gate if required.

Lemma C.1 *There exists a non-deterministic version of the gearbox circuit $\text{GB}(\phi_1)$ wherein all of the rotations can be performed offline that has success probability $\frac{1}{2}(\cos^4(\phi_1) + \sin^4(\phi_1))$ and upon failure either performs $e^{i \arctan(\tan^2(\phi_1)X)}$ or applies a Clifford operation.*

Proof. The circuit is



We have from Theorem 2 that the operations affecting the topmost qubit simply act to teleport the rotation to the middle qubit upon success. In this case, we want the rotation direction to flip which happens with probability $1/2$ for $\text{PAR}(\phi_1)$. If this operation is successful then the operations on the bottom two qubits are equivalent to those in $\text{GB}(\phi_1)$, which succeeds with probability $(\cos^4(\phi_1) + \sin^4(\phi_1))$ [17]. The claim regarding the overall success probability then follows from noting that the measurement success probabilities are independent and hence the total success probability is the product of the two probabilities.

If the first measurement yield “1” then the circuit will deviate from the original gearbox circuit. This results in the following effective transformation for the bottom two qubits.

$$e^{-i\phi_1 X} |\psi\rangle \mapsto \cos^2(\phi_1) |0\rangle |\psi\rangle - \sin^2(\phi_1) |0\rangle (-iX) |\psi\rangle - i \cos(\phi_1) \sin(\phi_1) (|1\rangle |\psi\rangle + |1\rangle (-iX) |\psi\rangle). \quad (\text{C.2})$$

If the middle qubit is measured to be 0 then $e^{i \arctan(\tan^2(\phi_1))X}$ is applied on the bottom most qubit, which is a rotation in the opposite direction from what is intended. Alternatively if the middle qubit is measured to be 1 then a Clifford operation, $e^{-i\pi X/4}$, is applied to $|\psi\rangle$. \square

The possibility of failure can also be removed from these circuits if the user has access to ϕ_1 as well as $\{\phi_j : j = 2, \dots, \infty\}$ wherein

$$\phi_{j+1} = \arctan \left(\sqrt{\tan(2 \arctan(\tan^2(\phi_j)))} \right). \quad (\text{C.3})$$

This is because $\text{GB}(\phi_{j+1}) = 2\text{GB}(\phi_j)$ and thus, similar to [26], if the direction of the rotation is opposite to that initially intended then $\text{GB}(\phi_j)$, for any $j \geq 2$, will correct the sum of all failures due to prior attempts:

$$\text{GB}(\phi_j) - \sum_{j=1}^{j-1} \text{GB}(\phi_2) = \phi_1. \quad (\text{C.4})$$

Obviously, the cost of preparing such states using reversible logic or RUS synthesis may be quite high so we do not advocate this approach in general. However, it is interesting to note that many of the costs of RUS synthesis can be reduced if we were to posit the existence of a more complex resource state factory.

Appendix D. Caching strategies

One of the drawbacks of the approaches we outline above is that arithmetic elements designed using the previous methods do not necessarily compose nicely. For example, if function f_1 requires N_1 rotations to implement and f_2 requires N_2 rotations then $f_1 \circ f_2$ requires $N_1 N_2$ rotations. It may be in principle natural to unravel the recursion and approximate the resulting function using RUS arithmetic, but in practice methods such as Newton’s method or the Binomial method that explicitly use recursion and avoiding an exponential slowdown at the price of increased circuit width may be desirable in such cases. Such exponential slowdowns can be avoided by using phase estimation to “cache” the result of the $f_2(x)$ in a register before using that result in $f_1(x)$. These results are also valuable for cases where it is useful to output the result of the RUS arithmetic as a qubit string.

Lemma D.1 *Let f_1 be a differentiable function that requires N_1 rotations to implement and satisfies $|\partial_\phi f_1(\phi)| \leq \kappa$ and let f_2 be a piecewise continuous function that requires N_2 rotations to implement. Then $e^{-i(f_1(f_2(\phi)))X}$ can be implemented within error ϵ with probability at least $1 - \delta$ using*

$$\lceil \log_2(\kappa/\epsilon) \rceil + \lceil \log_2(2 + 1/(2\delta)) \rceil$$

additional qubits and

$$O\left(\frac{N_2 \kappa}{\epsilon \delta} + N_1\right)$$

additional single qubit rotations.

Proof. In order to ensure that the total error is at most ϵ for the computation of $f_1(f_2(\phi))$ it suffices to guarantee that $f_2(\phi)$ is computed within error ϵ because Taylor's remainder theorem implies that

$$|f_1(f_2(\phi)) - f_1(f_2(\phi) + \epsilon/\kappa)| \leq \frac{\epsilon}{\kappa} \max_x |f'(x)| \leq \epsilon, \tag{D.1}$$

by assumption.

Instead of viewing f_2 as faulty, imagine instead that f_2 is given by a qubit string that stores the result of approximating its value using phase estimation. Given a unitary that enacts $e^{-if_2(\phi)X}$ using N_1 fundamental rotations, it then follows that the following unitary can be implemented using N_1 rotations

$$He^{-if_2(\phi)X}H|0\rangle = e^{-if_2(\phi)}|0\rangle. \tag{D.2}$$

Thus phase estimation can be used to estimate $f_2(\phi)$ within error ϵ/κ using t qubits and success probability $1 - \delta$ where t is [25]

$$t = \lceil \log_2(\kappa/\epsilon) \rceil + \lceil \log_2(2 + 1/(2\delta)) \rceil. \tag{D.3}$$

Phase estimation requires 2^t applications of $e^{-if_2(\phi)X}$ in order to learn the state within the required tolerances. This means that the cost of preparing $|f_2(\phi)\rangle$ is given by (D.3) to be $O(N_2\kappa/(\epsilon\delta))$. Now if we cost performing $e^{-if_2(\phi)X}$ using this state as a single rotation then the cost of performing $e^{-if_1(f_2(\phi))X}$ is N_1 . The lemma then follows by summing both terms. \square

Lemma D.1 shows that caching can reduce the cost of using RUS arithmetic to approximate recursive functions from exponential to linear at the price of additional costs incurred by using phase estimation. Such tradeoffs are asymptotically beneficial when $\frac{\epsilon\delta}{\kappa} \in o(N_1)$. Since recursion can be quite expensive in terms of either time or space it may often be of practical interest to find alternatives to these approaches, such as that given below.

Appendix E. Square wave synthesis

An alternative approach to the method described above is to synthesize functions using the square wave property of the GB function. The idea behind our approach builds upon existing classical results that use *exact* square waves to approximate a function [42]. Assume a function $f(x)$ is provided and that the objective is to approximate it on a uniform mesh of points on $[x_{\min}, x_{\max}]$ and define $\{x_i : i = 1, \dots, N\}$ to be the midpoints of each segment of the mesh. An approximation to the function of the form

$$f(x) \approx \sum_{j=1}^N a_j S_j(x), \tag{E.1}$$

is then sought where each $S_j(x)$ is a square wave with period T_j . We achieve this by demanding that the approximant equals $f(x)$ at each midpoint x_i

$$f(x_i) = \sum_{j=1}^N a_j S_j(x_i). \tag{E.2}$$

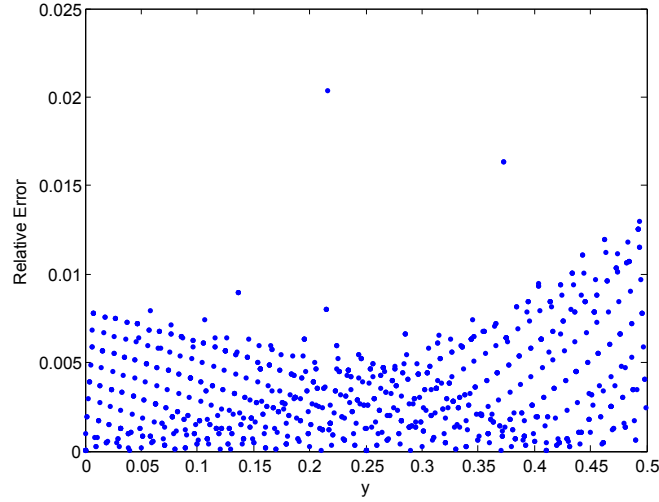


Fig. E.1. Plot of the relative error in square wave synthesis of $1/(1 - y)$ for $k = 8$ and $N = 71$ for values of y corresponding to $a \in [1, 1024]$.

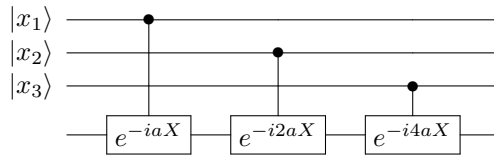
The values of a_j are then found by solving the resultant system of equations. The error in this piecewise approximation is therefore at most

$$\left| f(x) - \sum_{j=1}^N a_j S_j(x) \right| \leq \max_x |f'(x)|(x_{\max} - x_{\min})/(2N). \tag{E.3}$$

Now let us consider a quantization of this problem. Assume that we have a qubit string, $|x\rangle$, that encodes the number x and we want to implement

$$|x\rangle |0\rangle \mapsto |x\rangle e^{-if(x)X} |0\rangle. \tag{E.4}$$

As an intermediate step, note that if $|x\rangle$ is given then $|x\rangle e^{-iaxX} |0\rangle$ can be implemented for any fixed $a \in \mathbb{R}$ using traditional circuit synthesis methods and rescaling the rotations used in implementing the rotations that are conditioned on the value of x . For example, if x is represented using three qubits then $|x\rangle e^{-iaxX} |0\rangle$ can be implemented as



This observation will be useful because it shows that we can rescale inputs by a constant factor without using RUS arithmetic.

There are a few slight complications that arise when using GB to approximate the square waves that appear in the synthesis method. Firstly, instead of square waves we use

$$\frac{4}{\pi}(\text{GB}^{\circ k}(x\pi + \pi/4) - \pi/4) \approx S(x), \tag{E.5}$$

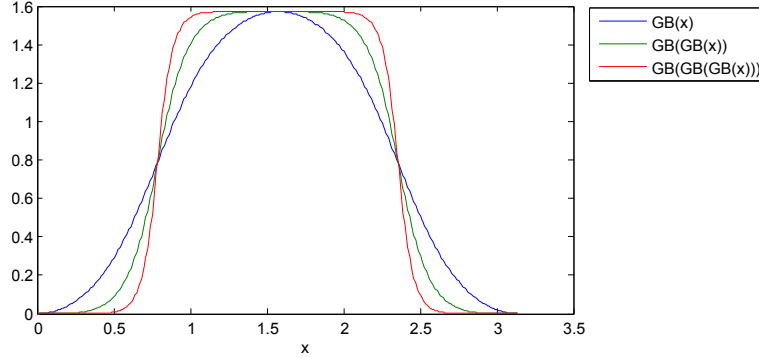


Fig. E.2. $\text{GB}^{\circ k}(x)$ for $k = 1, 2$, and 3 . The functions approach a square wave as $k \rightarrow \infty$.

where S is the square wave function with period 1, and replace each term in (E.1) with

$$\text{GB} \left(\arctan \left(\sqrt{\tan(2a_j)} \right), \text{GB}^{\circ k}(x\pi/T_j + \pi/4) \right) - a_j, \quad (\text{E.6})$$

for $a_j \leq \pi/8$. This can be verified by observing that when $\text{GB}^{\circ k}(x\pi + \pi/4)$ evaluates to $\pi/2$ the function evaluates to a_j as required. Larger values of a_j can also be straight forwardly implemented by summing several gearbox circuits that use smaller a_j .

The approximation procedure then works as follows. We first linearize the expressions and find a_j such that for each midpoint x_i ,

$$f(x_i) = \sum_{j=1}^N a_j \left(\frac{4}{\pi} (\text{GB}^{\circ k}(x_i\pi/T_j + \pi/4) - \pi/4) \right). \quad (\text{E.7})$$

The product of the a_j with each of the square waves could be approximated using the multiplication circuits given previously. A much better approach is to approximate the near-linear function of a_j given in (E.8) as

$$f(x) \approx \sum_{j=1}^N \text{GB} \left(\arctan \left(\sqrt{\tan(2a_j)} \right), \text{GB}^{\circ k}(x\pi/T_j + \pi/4) \right) - a_j, \quad (\text{E.8})$$

where the periods for each square wave obey $T_j = 2(x_{\max} - x_{\min})j/N$. The resultant circuit does not require time slicing in order to increase the approximation accuracy, unlike the previous methods. A formal description of the algorithm is given in Figure 20. It is worth noting that square waves can also be generated using the analogous property of PAR, but we favor GB in this construction because it is naturally an RUS circuit.

There are some issues that arise from this approximation. The first is results of composed gearbox circuits do not approximate the square wave well near each discontinuity. In particular, $\text{GB}^{\circ k}(x + \pi/4) \approx \pi/4 + 2^k x$ whereas $\text{GB}^{\circ k}(x) \approx x^{2^k}$. Therefore we expect the square wave to jump to its maximum value over a range of 2^{-k} radians whereas the gearbox circuit will be in near perfect agreement with the square wave near $x = 0$. This means that $k \geq 8$ may be needed to appreciably reduce the errors due to non-instantaneous ramp up near $\pi/4$.

Since the expected number of rotations for $\text{GB}^{\circ k}$ is roughly $2^{k-1/2}$ [17] this implies that these errors are can be reduced at linear cost.

Another issue facing square wave synthesis is that each of the approximate square waves achieves a value of 0 near $x = 0$. If the function is continuous then there exists an ϵ neighborhood about 0 such that for any finite k the approximation error is $O(1)$, unless the function to be approximated also obeys $f(0) = 0$. Similar problems can occur at the end of the approximation interval. Both such issues can be addressed by padding the approximation interval from $[x_{\min}, x_{\max}]$ to $[x_{\min} - \Delta, x_{\max} + \Delta]$ for some $\Delta > 0$.