

# Enabling Accuracy-Aware Quantum Compilers using Symbolic Resource Estimation

GIULIA MEULI, EPFL, Switzerland

MATHIAS SOEKEN, Microsoft, Switzerland

MARTIN ROETTELER, Microsoft, United States of America

THOMAS HÄNER, Microsoft, Switzerland

Approximation errors must be taken into account when compiling quantum programs into a low-level gate set. We present a methodology that tracks such errors automatically and then optimizes accuracy parameters to guarantee a specified overall accuracy while aiming to minimize the implementation cost in terms of quantum gates. The core idea of our approach is to extract functions that specify the optimization problem directly from the high-level description of the quantum program. Then, custom compiler passes optimize these functions, turning them into (near-)symbolic expressions for (1) the total error and (2) the implementation cost (e.g., total quantum gate count). All unspecified parameters of the quantum program will show up as variables in these expressions, including accuracy parameters. After solving the corresponding optimization problem, a circuit can be instantiated from the found solution. We develop two prototype implementations, one in C++ based on Clang/LLVM, and another using the Q# compiler infrastructure. We benchmark our prototypes on typical quantum computing programs, including the quantum Fourier transform, quantum phase estimation, and Shor's algorithm.

CCS Concepts: • **Software and its engineering** → **Compilers**; • **Hardware** → **Quantum computation**.

Additional Key Words and Phrases: quantum computing, quantum programming, quantum algorithms, approximation errors, resource estimation

## ACM Reference Format:

Giulia Meuli, Mathias Soeken, Martin Roetteler, and Thomas Häner. 2020. Enabling Accuracy-Aware Quantum Compilers using Symbolic Resource Estimation. *Proc. ACM Program. Lang.* 4, OOPSLA, Article 130 (November 2020), 26 pages. <https://doi.org/10.1145/3428198>

## 1 INTRODUCTION

There exists a wide range of quantum algorithms that promise asymptotic speed-ups with respect to their classical counterparts. Application domains include cryptography [Bennett and Brassard 2014; Pirandola et al. 2020], machine learning [Havlíček et al. 2019], material science [McArdle et al. 2020], and quantum chemistry [Babbush et al. 2017; Low and Chuang 2019]. However, concrete resource estimates for problems and corresponding problem sizes at which quantum computers are expected to outperform their classical counterparts remain scarce. To carry out such resource estimates, several quantum programming languages and toolchains have been developed such as Q# [Svore et al. 2018], Quipper [Green et al. 2013], Scaffold/ScaffCC [Javadi-Abhari et al. 2014], Qiskit [Aleksandrowicz and et al. 2019], ProjectQ [Steiger et al. 2018], and QuRE [Suchara

---

Authors' addresses: Giulia Meuli EPFL, Lausanne, Switzerland, giulia.meuli@epfl.ch; Mathias Soeken Microsoft, Zurich, Switzerland, masoeken@microsoft.com; Martin Roetteler Microsoft, Redmond, WA, United States of America, martinro@microsoft.com; Thomas Häner Microsoft, Zurich, Switzerland, thhaner@microsoft.com.

---

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2020 Copyright held by the owner/author(s).

2475-1421/2020/11-ART130

<https://doi.org/10.1145/3428198>

et al. 2013]. Some of these frameworks provide domain-specific languages with the necessary abstractions, libraries, simulators, and cloud access to small-scale quantum computers. Despite the availability of these languages, there is still a significant amount of manual work involved in resource estimation [Reiher et al. 2017; Scherer et al. 2017]—one reason being the lack of built-in support for handling approximation errors.

Why do approximation errors occur in quantum programs in the first place? We discuss three main sources of errors in this paper, however, we note that the framework is extensible to other sources of errors.

*1. Synthesis Errors.* Due to the discrete nature of fault-tolerant instruction sets (and indeed, it is known that any universal fault-tolerant instruction set necessarily must be discrete [Nielsen and Chuang 2000]), it cannot be avoided to introduce approximation errors for general target operations. For instance, consider a rotation around the  $Z$ -axis such as

$$R_Z(\theta) = \begin{bmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{bmatrix},$$

which can be defined for any  $\theta \in [0, 4\pi)$ . These rotations can only be implemented exactly for a discrete subset of the interval  $[0, 4\pi)$ , as gates have to be expressed as words of finite length over any universal set of generators. It should be noted that there is a mathematical function that expresses the length of the approximating word in terms of an approximation error, which we call  $\varepsilon_R$  in this paper. This mathematical function depends on the concrete synthesis algorithm used to perform the factorization into fault-tolerant instructions. State-of-the-art synthesis algorithms lead to a cost (e.g., number of  $T$  gates, where  $T = e^{i\pi/8}R_Z(\pi/4)$ ) that is proportional to  $\log_2(\varepsilon_R^{-1})$  [Kliuchnikov et al. 2013; Ross and Selinger 2016].

*2. Phase Estimation Errors.* An important technique in quantum computing is to extract estimates of an eigenvalue  $\lambda$  of an operator  $U$  [Kitaev 1995; Shor 1994] to  $k$  bits of precision. A common method to achieve this is to prepare an eigenstate  $|\psi_\lambda\rangle$  of  $U$  and to then apply powers  $U^{2^i}$ , for  $i = 0, \dots, k-1$ , to the eigenstate  $|\psi_\lambda\rangle$ . This application is done conditionally on the value of a reference system and allows us to extract the  $k$  most significant bits of the eigenvalue. As  $\lambda$  can in principle be any complex number of the form  $\lambda = e^{i\alpha}$ , where  $\alpha \in [0, 2\pi)$ , the particular choice of  $k$  introduces an approximation error and limits how precisely we can estimate  $\lambda$ . We call the resulting approximation error  $\varepsilon_{QPE}$  in this paper.

*3. Algorithmic Errors.* Some quantum programs are part of a parametric family of programs that gracefully degenerate with a reduction of the parameters. A concrete example for such a family of programs is the quantum Fourier transform [Shor 1994], or QFT for short. While the transformation itself can be implemented exactly and with no approximation error over a gate set that includes continuous rotations such as  $R_Z(\theta)$  for arbitrary  $\theta \in [0, 4\pi)$ , it is possible to approximate the transformation by selectively dropping some of the rotations that occur, in particular by “pruning” the values of  $\theta$  that are very close to 0. One such pruning method is well known [Coppersmith 2002] and allows us to drop many of the  $O(n^2)$  rotations that a simple implementation of the Fourier transform requires and just retain  $O(n \log n)$  rotations, while still maintaining a sufficient approximation. We call the resulting approximation error  $\varepsilon_{QFT}$  in this paper. Another example of algorithmic error comes from formulas that are known to converge to a target program when taking a suitable limit, e.g., of alternations of other, typically smaller and simpler, programs. An example for the latter is the so-called Trotter formula, a well-known identity to implement an approximation to  $e^{i(A+B)}$  for Hermitian matrices  $A$  and  $B$ , from the knowledge of implementations for  $e^{iA}$  and  $e^{iB}$ . The resulting approximation error (the “Trotter error”) is called  $\varepsilon_{TE}$  in this paper.

## 1.1 Overview of Our Approach

A quantum program can be expressed at a high level of abstraction using any of the languages and frameworks mentioned above. Instead of specifying quantum circuits at the level of single- and two-qubit quantum gates, these languages provide high-level abstractions such as *quantum Fourier transform* and *quantum phase estimation*, which can be used to express quantum algorithms. Once the target machine has been specified, a compiler is used to translate these high-level abstractions into lower-level gates, such as the gates from the Clifford+ $T$  gate set [Amy et al. 2013], which can be implemented fault-tolerantly on several scalable quantum computer architectures [Bravyi and Kitaev 2005].

Various approximations, such as the ones described above, may be necessary during compilation and with existing languages, programmers must manually keep track of all the introduced errors. Furthermore, programmers must tune the parameters of their implementation to keep the total error within a target budget. To guarantee this, they must derive the resulting error bounds manually—a task that is tedious and error prone.

To address this issue, our methodology introduces language support into existing quantum programming languages, allowing programmers to deal independently with the approximation errors in each subroutine. The job of inferring how all introduced approximation errors interact is thus transferred from the programmer to the compiler. Our methodology automatically infers an error bound for the overall quantum program and then selects appropriate values for each of the program's accuracy parameters to simultaneously (1) satisfy a user-specified overall tolerance and (2) reduce the required quantum resources.

More specifically, our methodology supports:

- (1) given the desired approximation error, determining the assignment of accuracy parameters that guarantees the given approximation error while aiming to minimize the number of operations;
- (2) given a maximal operation count, determining the assignment of accuracy parameters that yield at most the given operation count while aiming to reduce the total approximation error.

The automatic optimization of accuracy parameters is carried out by solving an optimization problem before the quantum circuit is generated. The constraint and cost functions describing the optimization problem are extracted by the compiler directly from the source code of the quantum program. Then, before execution, the optimized accuracy parameters are fed into the main program.

Finding a suitable assignment of accuracy parameters using, e.g., simulated annealing, requires hundreds of evaluations of both constraint and cost functions. Hence, our methodology has to lean on a fast method to estimate resource requirements and the total approximation error. Available methods, e.g., in Q#, estimate resources by actually generating quantum circuits from completely specified programs and then counting the generated gates. Hence, their runtime will increase with increasing problem size, making them ill-suited for use as cost functions in an optimization procedure, as we will show in Section 8.

Instead, we propose fast symbolic methods that extract a symbolic expression for the desired cost or constraint function (total approximation error or number of gates) directly from the source code of the quantum program. The resulting expressions feature variables that correspond to the various parameters of the program, including accuracy parameters. The symbolic approach does not need to execute the complete control flow of the quantum program to get an estimate, hence it provides a much faster solution that is viable even for application-scale programs.

Our two prototypes both implement a symbolic approach for resource and error estimation. Since the resulting expressions may theoretically still contain some residual code that must be executed (e.g., certain if-else statements), we refer to them as being (*near-*)*symbolic*. The parentheses indicate

that, for most applications, the resulting expressions would be fully symbolic. However, there are examples where this is not the case, e.g., a quantum program that executes one of two different algorithms depending on a runtime parameter. We note that our prototypes both generate fully symbolic expressions for all the examples in this paper.

## 1.2 Contributions

To the best of our knowledge, we are the first to present a quantum programming framework that provides built-in support for automatic accuracy management. Our methodology automatically selects accuracy parameters such that the overall error is at most equal to a user-specified value while aiming to reduce the quantum resource requirements, or vice versa. As the interplay between the various approximation errors can be quite complicated, it can be difficult for a human to find the best trade-offs.

In addition, our approach for extracting *symbolic* resource estimates from the quantum program, and using them to specify and solve the optimization problem of tuning accuracy parameters, appears to be new. Some state-of-the-art methods exist to automate resource estimation. Examples are methods embedded in Q#, ProjectQ, and Quipper [Smith et al. 2014], which are based on circuit-description languages, and QuRE, which is capable of evaluating different technologies and error-correcting codes. Nevertheless, all these frameworks require *a priori* specified quantum circuits. Thus, they cannot be used to derive asymptotic estimates as a function of the input parameters of the algorithm, in contrast to the symbolic approach we propose.

Indeed, a salient feature is that we acquire symbolic estimates for the total error and gate count directly from the high-level description of a quantum program, without executing the control-flow. It is only after the final resource requirements and total approximation error are known that we instantiate a circuit using the determined accuracy parameters. This allows us to automatically evaluate the resource requirements of a given quantum program in an accuracy-aware fashion, without the overhead of having to execute the entire control-flow of the program.

We implement a prototype of our framework in C++ for evaluation purposes. To demonstrate integration into an existing quantum programming language, we then present a prototype implementation into the Q# compiler. In both cases we implement new compiler passes to enable the new functionality. Finally, we validate and benchmark our methodology on quantum programs such as *quantum phase estimation* and *Shor's algorithm* [Shor 1994]. We show that the runtime gap between the best previous methods and our symbolic method for resource and error estimation is *unbounded* as a function of the problem size.

Our methodology can be integrated into any of the cited quantum programming frameworks. We give details on the language and compiler features that are required to achieve such integration. Furthermore, our methodology is fully extensible: (1) more algorithmic errors can be defined at any point in the quantum program and will be handled automatically by our implementation; (2) more gate errors can be defined by adding a new type of error to each primitive gate; (3) custom implementations of library functions can be added.

## 1.3 Related Work

In the work by Hung et al. [2019], a theoretical framework is presented to reason about the robustness of quantum programs when executed on noisy quantum hardware. Specifically, the authors develop a logic with which it is possible to characterize the distance between an ideal quantum program and its noisy counterpart, given a noise model of the target hardware. Hung et al. [2019] present several case studies consisting of small quantum circuits (between 1 and 6 qubits). Computing the  $(Q, \lambda)$ -diamond norm, which measures the distance between the ideal and the erroneous program, involves solving a semidefinite programming problem (as is the case for

the regular diamond norm [Watrous 2009]). This becomes computationally intractable for large systems (and their corresponding noise models) due to the exponential scaling of the dimension with the number of qubits.

In contrast, our framework is concerned with approximation errors that occur at the algorithmic level. In particular, such errors may be decreased by using more quantum resources. Our framework thus aims to select appropriate accuracy parameters balancing this trade-off. Furthermore, we present two prototype implementations of our framework and demonstrate that it is capable of handling application-scale quantum programs featuring thousands of qubits and billions of operations.

Our methodology also relates to the automatic approach presented by Panchekha et al. [2015] for handling floating-point rounding errors in classical computing. Their approach, called *Herbie*, is capable of locating sources of errors in the code and proposing candidate rewrites to improve the overall precision. Similarly, our approach locates and adapts accuracy parameters to reduce the total error. While *Herbie* must be combined with other methods [Barr et al. 2013; Boldo 2009; Darulova and Kuncak 2014] to provide worst-case guarantees, our methodology proposes parameter assignments that guarantee the specified worst-case upper bound, assuming a correct implementation. We are not concerned with verifying actual correctness of the quantum program, but refer the reader to, e.g., the works by Amy et al. [2017]; Paykin et al. [2017]; Ying [2011].

Symbolic execution, which is widely used for code testing, is another approach related to our work. The method translates the given program into a logical formula in order to check some input properties. A relevant example is veritesting [Avgerinos et al. 2014], which alternates dynamic (DSE) and static (SSE) symbolic execution. Similarly, we extract from a quantum program symbolic expressions for upper bounds on the total error and the gate count.

## 1.4 Paper Outline

This paper is organized as follows.

Section 2 provides the necessary background information on quantum computing. In particular, Section 2.3 provides a detailed description of how approximation errors compose in quantum programs.

Section 3 illustrates three typical quantum algorithms, with an emphasis on the trade-offs between implementation cost and approximation error.

Section 4 motivates the need for language support when implementing large-scale quantum algorithms in an accuracy-aware fashion. Furthermore, it illustrates how the proposed language features improve the readability and simplicity of the code, thus simplifying the implementation process.

Section 5 describes the proposed procedure to extract expressions for cost and constraint functions from a given quantum program. These expressions define an optimization problem, the solution of which yields appropriate values for all accuracy parameters. Moreover, the section details how the extracted expressions are optimized in order to arrive at (near-)symbolic expressions that may be evaluated much faster.

Section 6 provides a list of all the features that a quantum programming language and its compiler should support to integrate our proposed methodology. It also contains a description of how these features have been implemented in our LLVM prototype and in our Q#-based implementation.

Section 7 shows the pseudo-code of our sample quantum programs and the corresponding cost and constraint functions that our framework automatically deduces from the code.

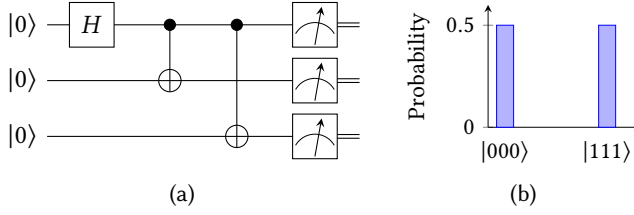


Fig. 1. Example of a quantum circuit computing an entangled state (a) and the corresponding expected measurement outcomes (b).

Section 8 reports the results of our runtime evaluation, which compares solving the optimization problem using a simulated annealing procedure with or without extracting (near-)symbolic expressions. It shows that our approach of extracting (near-)symbolic expressions is capable of reducing the time to solution by many orders of magnitude.

## 2 QUANTUM COMPUTING BACKGROUND

### 2.1 Quantum States

Quantum computers process information encoded in qubits. The quantum state of a qubit  $|\psi\rangle$  (using *Dirac* or *bra-ket* notation) can be written as

$$|\psi\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle,$$

where  $\alpha_0, \alpha_1 \in \mathbb{C}$  are the complex probability amplitudes corresponding to the *computational basis states*  $|0\rangle$  and  $|1\rangle$ , respectively, and  $|\alpha_0|^2 + |\alpha_1|^2 = 1$ . The computational basis states  $|0\rangle$  and  $|1\rangle$  may be associated with two-dimensional basis vectors  $|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  and  $|1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ . Single-qubit quantum gates (or quantum operations) can be written as  $2 \times 2$ -dimensional complex unitary matrices. A matrix  $U$  is unitary if its conjugate transpose corresponds to its inverse:  $UU^\dagger = U^\dagger U = \mathbf{1}$ . When applying a quantum gate to a single qubit, its new quantum state can be computed via a matrix-vector multiplication of the gate matrix and the two-dimensional amplitude vector  $|\psi\rangle = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix}$ .

Composition of vector spaces is achieved using the *tensor product*. Suppose  $V$  and  $W$  are Hilbert spaces of dimension  $m$  and  $n$  respectively. Then  $V \otimes W$  is an  $n \times m$ -dimensional vector space. To see the connection to quantum states, let  $V$  denote a two-dimensional vector space with basis vectors  $|0\rangle$  and  $|1\rangle$  (a single-qubit state), then  $|0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle$  is an element of  $V \otimes V$  (i.e., a 2-qubit state). The notation  $|\psi\rangle^{\otimes k}$  means  $|\psi\rangle$  tensored with itself  $k$  times. An analogous notation is also used for operators on tensor product spaces. A more detailed description of the tensor product and its properties can be found in [Nielsen and Chuang \[2000, Section 2.1.7\]](#).

Therefore, an  $n$ -qubit quantum state may be described by a vector containing the  $2^n$  amplitudes corresponding to all possible bitstrings of length  $n$ , or, using the bra-ket notation and two shorter alternatives:

$$\sum_{x \in \mathbb{B}^n} a_x |x_1\rangle \otimes \cdots \otimes |x_n\rangle = \sum_{x \in \mathbb{B}^n} a_x |x_1 \cdots x_n\rangle = \sum_{x \in \mathbb{B}^n} a_x |x\rangle,$$

where  $x_i$  denotes the  $i$ -th bit of  $x$  and  $\sum_x |a_x|^2 = 1$ . Measuring all  $n$  qubits of the  $n$ -qubit state above results in a *collapse* of the superposition onto one of the computational basis states  $|x\rangle$  with probability  $|a_x|^2$ .



## 2.2 Quantum Programs and Gates

A quantum program consists of classical *and* quantum instructions [Knill 1996]. While classical instructions are performed on the (classical) controller, the latter are performed on the quantum processing unit (QPU). In each step of the computation, the controller sends a sequence of quantum instructions to the QPU. After executing each sequence, the QPU returns classical measurement results to the controller, which may use them to decide on the quantum instructions to perform next (if any).

The mentioned sequences of quantum instructions can be visualized using circuit diagrams, so-called *quantum circuits*.

**EXAMPLE 1.** *An example of a quantum circuit is shown in Fig. 1. This circuit generates a Greenberger–Horne–Zeilinger state on three qubits, i.e., the state*

$$\frac{1}{\sqrt{2}}(|000\rangle + |111\rangle).$$

*In a quantum circuit, each qubit is represented by a horizontal line. Operations are denoted by boxes or other symbols on the qubit(s) they are being applied to. Time advances from left to right. The initial state is  $|0\rangle^{\otimes 3}$ . The first operation is a Hadamard gate ( $H$ ) applied to the first qubit. In matrix notation,*

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

*Applying  $H$  to the first qubit maps  $|000\rangle$  to  $\frac{1}{\sqrt{2}}(|000\rangle + |100\rangle)$ . The next gate is a controlled-NOT or CNOT,*

$$\text{CNOT} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

*which entangles the first with the second qubit by flipping the latter if the first qubit is  $|1\rangle$ . After the two CNOT gates, the three qubits are in the state  $\frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$ . Finally, all three qubits are measured. There is a 50% probability of measuring all 0s and a 50% probability of measuring all 1s.*

Quantum programs can be written in one of the various languages and software frameworks for quantum computing, e.g., ProjectQ, Q#, Qiskit. All of these languages allow programmers to specify the quantum program in terms of high-level operations. They also provide methods to decompose such operations into the native gate set of the QPU.

Different quantum architectures support different native operations. A common low-level instruction set is the so-called Clifford+ $T$  gate set [Amy et al. 2013]. This instruction set includes the Clifford gates, i.e., the CNOT gate and Hadamard gate  $H$ , as well as the phase gate  $S$ , plus the  $T$  gate, which is a non-Clifford gate required to achieve universality:

$$S = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}, T = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{bmatrix}.$$

In a fault-tolerant setting, it is particularly expensive to apply the  $T$  gate. Consequently, the  $T$ -count (number of  $T$  gates) is a good measure for the cost of a fault-tolerant implementation of a given quantum program [Campbell and Howard 2017; Fowler et al. 2012]. In contrast, IBM's quantum computers, which are examples of noisy-intermediate-scale-quantum (NISQ) systems, natively support the  $U$  gate,  $U(\theta, \psi, \lambda) = R_x(\psi)R_y(\theta)R_z(\lambda)$ , which is parameterized over 3 continuous variables, and the CNOT gate. Usually, 2-qubit gates are more error-prone than single qubit gates. For this reason, a good measure for the cost of a quantum circuit synthesized for NISQ machines is the number of CNOT gates. The same is also valid for the Controlled- $Z$  gate that is used, e.g., in Rigetti's NISQ systems.

### 2.3 Approximation Errors in Quantum Circuits

As mentioned in Section 1, our framework addresses approximation errors, which may be reduced at an increased implementation cost. For the example of *synthesis errors*, which occur due to the discrete nature of fault-tolerant instruction sets, the trade-off between the number of  $T$  gates and the approximation error is logarithmic [Kliuchnikov et al. 2013]. That is, to achieve an approximation error  $\varepsilon$ ,  $O(\log(1/\varepsilon))$   $T$  gates are sufficient. In a typical quantum program, multiple different sources of such errors are present and thus the question arises: *How do the approximation errors of individual operations compose?*

In this section, we prove that it is possible to derive an upper bound on the total approximation error by summing up all the individual approximation errors. Consequently, our methodology produces quantum circuits with accuracy guarantees for quantum programs without measurement feedback. Note that this excludes programs that rely on repeat-until-success statements, i.e., loops that iterate until a certain measurement outcome is observed (see, e.g., Paetznick and Svore [2014]). Such cases can be handled separately using, e.g., an upper bound on the number of iterations. In the last section, we propose a possible way of enabling accuracy management for such programs as future work. In all other cases, branching on measurement results may be addressed using an expression for the total error of the form  $\varepsilon_b = \max(\varepsilon_1, \varepsilon_2)$ , where  $\varepsilon_1$  and  $\varepsilon_2$  denote the errors of each branch. This follows from the *Deferred Measurement Principle* [Nielsen and Chuang 2000, Section 4.4], which says that measurements may be delayed until the end of the computation, transforming all quantum operations that are executed conditionally on the measurement result into quantum-controlled operations.

Quantum circuits corresponding to feedback-free programs consist of a sequence of gates  $U_1, \dots, U_m$ , followed by a sequence of measurements  $M_1, \dots, M_k$  that produce a measurement outcome  $m = x_i$  for a final state

$$|\psi\rangle = U_m \cdots U_1 |0\rangle^{\otimes N}$$

with probability

$$P(m = x_i) = |\langle x_i | \psi \rangle|^2,$$

where  $\langle x_i | \psi \rangle$  denotes the overlap of  $|\psi\rangle$  with  $|x_i\rangle$ .

Therefore, our methodology must ensure that the actual final state  $|\tilde{\psi}\rangle$  is close to the desired final state  $|\psi\rangle$  after all decompositions have been applied to  $U_1, \dots, U_m$ .

Let  $V_1, \dots, V_n$  be an approximate decomposition of the quantum program in terms of the gates supported by the target hardware, i.e.,

$$\| \underbrace{U_m \cdots U_1}_U - \underbrace{V_n \cdots V_1}_V \| \leq \varepsilon,$$

where  $\|\cdot\|$  denotes the spectral norm as defined in Nielsen and Chuang [2000, Section 2.1.4]. Then,  $\| |\psi\rangle - |\tilde{\psi}\rangle \| = \| U |0\rangle^{\otimes N} - V |0\rangle^{\otimes N} \| \leq \varepsilon$ , which guarantees that, with  $|\psi\rangle = \sum_i a_i |x_i\rangle$  and  $|\tilde{\psi}\rangle = \sum_i \tilde{a}_i |x_i\rangle$ ,

$$\begin{aligned} |a_i - \tilde{a}_i| &= \sqrt{|a_i - \tilde{a}_i|^2} \\ &\leq \sqrt{\sum_i |a_i - \tilde{a}_i|^2} \\ &\leq \varepsilon. \end{aligned}$$

Therefore, it is sufficient that our methodology guarantees

$$\|U - V\| \leq \varepsilon.$$



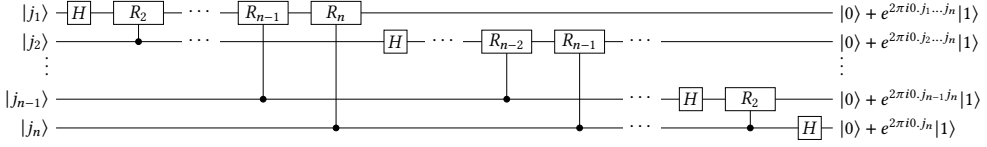


Fig. 2. Efficient circuit computing the quantum Fourier transform.

For more details on approximated unitary operators, we refer to Box 4.1 “*Approximating quantum circuits*” of Nielsen and Chuang [2000].

In the process of translating the quantum program to the native gate set, several decompositions are applied that introduce approximation errors. Let  $U$  be a quantum operation being approximated by the decomposition into  $W_1, \dots, W_t$ . The decomposition introduces at most  $\varepsilon_U$  if  $\|U - (W_t \cdots W_1)\| \leq \varepsilon_U$ , assuming that all  $W_i$  are implemented exactly. Now, combining multiple such approximate implementations  $\tilde{U}_i$  of  $U_i$  such that  $\|U_i - \tilde{U}_i\| \leq \varepsilon_i$  yields a total error of at most  $\sum_i \varepsilon_i$  [Bernstein and Vazirani 1997].

Therefore, it is possible to derive recursively-defined expressions for the error  $E(U, \varepsilon_U)$  and the total gate count  $T(U, \varepsilon_U)$  [Häner et al. 2018]:

$$E(U, \varepsilon_U) = \varepsilon_U + \sum_{W \in D(U, \varepsilon_U)} E(W, \varepsilon_W) f_W(\varepsilon_U),$$

$$T(U, \varepsilon_U) = \sum_{W \in D(U, \varepsilon_U)} T(W, \varepsilon_W) f_W(\varepsilon_U),$$

where  $D(U, \varepsilon_U)$  is the set of gates in the  $\varepsilon_U$ -approximate decomposition of  $U$  and  $f_W(\varepsilon_U)$  denotes the number of  $W$  operations in the decomposition. Looking at these expressions, it is clear that an upper-bound on the total error can be computed very similarly to counting gates.

In conclusion, by applying this reasoning to the main entry point of a quantum program, we can choose all  $\varepsilon_{(\cdot)}$  in the expression for

$$E(U_{Main}, \varepsilon_{U_{Main}})$$

such that  $E(U_{Main}, \varepsilon_{U_{Main}}) \leq \varepsilon$ . This ensures that the measurement probability amplitude for a given bit-string changes by at most  $\varepsilon$ .

### 3 SAMPLE QUANTUM PROGRAMS

#### 3.1 Quantum Fourier Transform

The quantum Fourier transform (QFT) is an algorithm that performs the Fourier transform of quantum mechanical amplitudes. QFT is implemented as a linear operator that applies the following unitary transformation on a basis state  $|j\rangle$  [Nielsen and Chuang 2000]:

$$|j\rangle \mapsto \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} e^{2\pi i jk/N} |k\rangle$$

The effect of the transform on an arbitrary state can be described using a product representation that maps  $|j_1, \dots, j_n\rangle$  to

$$\frac{(|0\rangle + e^{2\pi i 0.j_n} |1\rangle)(|0\rangle + e^{2\pi i 0.j_{n-1}j_n} |1\rangle) \cdots (|0\rangle + e^{2\pi i 0.j_1j_2 \dots j_n} |1\rangle)}{2^{n/2}}$$

where  $0.j_lj_{l+1} \dots j_n$  is the binary expansion  $j_l/2 + j_{l+1}/4 + \dots + j_n/2^{n-l+1}$ . This representation has a direct correspondence to the circuit implementation of the quantum Fourier transform shown

in Fig. 2. The circuit is composed of  $n$  steps, one for each qubit. In each step, for each qubit, a Hadamard gate is applied, followed by a series of rotation gates controlled by all remaining qubits.

In this work, we will largely refer to an approximate version of QFT (AQFT) in which the number of rotations is reduced according to the desired approximation error  $\varepsilon_{QFT}$ . This is done by pruning the rotations with small angles. In particular, for each qubit  $j_i$  with  $1 \leq i < n$  a maximum of

$$l = \lceil \log_2(n/\varepsilon_{QFT}) \rceil + 3$$

controlled-rotations is applied [Coppersmith 2002].

The quantum Fourier transform enables the *quantum phase estimation* algorithm and has a key role in the solution of many relevant problems, e.g., the *integer factorization* problem.

### 3.2 Simulating Time-Evolution of Operators

Being a quantum system, a quantum computer can be programmed to simulate other quantum systems. This, for example, can be used to elucidate chemical reaction mechanisms [Reiher et al. 2017].

Given the quantum-mechanical Hamiltonian  $\mathcal{H}$  which describes the system being studied, the time evolution of the system can be simulated by implementing the *time-evolution operator*  $U = e^{-i\mathcal{H}t}$ . The time-evolved quantum state can then be obtained by applying  $U$  to the initial state  $|\psi(0)\rangle$ :

$$|\psi(t)\rangle = e^{-i\mathcal{H}t}|\psi(0)\rangle$$

In order to implement the time-evolution operator on a quantum computer, it needs to be decomposed into the native gate set, e.g., Clifford+T. Different decomposition methods are available, each one with a different scaling with respect to the targeted precision  $\varepsilon_{TE}$ : polynomial for the Trotter decomposition method, logarithmic for the linear combination of unitaries (LCU) [Babbush et al. 2017; Poulin et al. 2015].

EXAMPLE 2. Consider the Hamiltonian of a 1D transverse-field Ising model (TFIM)

$$\mathcal{H} = -J \underbrace{\sum_{\langle i,j \rangle} Z^i Z^j}_{\mathcal{H}_1} - h \underbrace{\sum_i X^i}_{\mathcal{H}_2},$$

where  $\mathcal{H}_1$  defines the interaction of adjacent spins, denoted by  $\langle i, j \rangle$ , with periodic boundary conditions,  $\mathcal{H}_2$  defines the interaction of the system with the external transverse field, and  $X^i, Z^i$  denote the application of  $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$  and  $Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ , respectively, to spin  $i$ .

Using a second-order Trotter-Suzuki decomposition, the time-evolution operator under this Hamiltonian can be written as

$$e^{-i\mathcal{H}t} \approx (e^{-i\mathcal{H}_1 \frac{t}{2M}} e^{i\mathcal{H}_2 \frac{t}{M}} e^{-i\mathcal{H}_1 \frac{t}{2M}})^M.$$

The number of Trotter steps  $M$  will be chosen according to the desired accuracy  $\varepsilon_{TE}$ . In particular, for this second-order Trotter decomposition we have that  $M$  is proportional to  $1/\sqrt{\varepsilon_{TE}}$  [Reiher et al. 2017]. Each Trotter step can be implemented using CNOT and  $R_Z(\theta)$  gates. The latter being a gate that applies a rotation equal to the angle  $\theta$  around the z-axis. Considering Clifford+T as the native gate set, each rotation has to be synthesized or decomposed in terms of these gates. As not every rotation can be realized exactly using this gate library, rotation synthesis also introduces an error. Given a target approximation error  $\varepsilon_R$ , the number of T gates per rotation will be proportional to  $\log_2\left(\frac{1}{\varepsilon_R}\right)$ .

Thus, to express the time-evolution operator we need to take into account two inter-dependent approximation errors, namely  $\varepsilon_{TE}$  and  $\varepsilon_R$ .

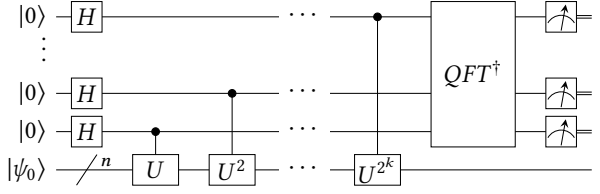


Fig. 3. Quantum circuit performing quantum-phase estimation on an  $n$ -qubit system with an accuracy of  $k + 1$  bits.

### 3.3 Quantum Phase Estimation

Once time-evolution under the Hamiltonian  $\mathcal{H}$  is implemented, one may perform measurements similar to experiments with the actual system. Quantum computing, however, allows us to achieve a quadratic advantage over repeated measurement and sampling via quantum phase estimation (QPE). Given a state with large overlap with the ground state  $|\psi_0\rangle$  of the Hamiltonian, this algorithm allows us to determine the ground state energy  $E_0$

$$\mathcal{H}|\psi_0\rangle = E_0|\psi_0\rangle.$$

Fig. 3 shows one of several possible implementations of QPE [Nielsen and Chuang 2000]. The measurement outcomes of the top  $k + 1$  qubits yield a  $k + 1$ -bit approximation to the phase due to time-evolution. More precisely, the number of qubits to choose  $n_{QPE}$  depends on the desired accuracy *and* the probability  $p$  of a successful measurement as

$$n_{QPE} = n + \left\lceil \log \left( 2 + \frac{1}{2(1-p)} \right) \right\rceil,$$

where  $n$  is the desired accuracy in number of bits.

The number of controlled time-evolution unitaries required for QPE to succeed with  $p = 0.5$  and accuracy  $\epsilon_{QPE}$  may thus be bounded by  $2^{n_{QPE}} - 1 \leq 16\pi/\epsilon_{QPE}$ .

Not only does QPE allow one to infer the ground state energy if the ground state is known, but it also collapses a non-eigenstate input  $|\phi\rangle$  to the  $i$ -th eigenstate  $|\psi_i\rangle$  of the Hamiltonian  $H$  with probability  $|\langle\psi_i|\phi\rangle|^2$ .

In terms of accuracy, it is important to distinguish between the different applications of QPE. If QPE is used to determine only the energy, i.e.,  $|E_0 - \tilde{E}_0| \leq \epsilon$  is required, then it is sufficient to implement time-evolution such that  $\|U - \tilde{U}\| \leq \epsilon - \epsilon_{QPE}$ . However, if the goal is to prepare the ground state, i.e.,  $\|\psi_0\rangle - |\tilde{\psi}_0\rangle\| \leq \epsilon$ , then  $\|U - \tilde{U}\| \leq \frac{\epsilon - \epsilon_{QPE}}{2^{n_{QPE}} - 1}$  is sufficient (both via triangle inequality). Distinguishing these cases clearly has a great impact on the resulting resource requirements.

## 4 ADDING LANGUAGE SUPPORT FOR ACCURACY MANAGEMENT

Since large-scale quantum computers are not yet available, resource estimation is a crucial feature of any software framework for quantum computing. Typically, resource estimation is performed by compiling the quantum program into the chosen target gate set and then executing the resulting circuit on a classical simulator that counts native operations (instead of executing them). For this to be possible, however, all the parameters of the program, including accuracy parameters for each subroutine, must be determined.

Existing quantum programming languages do not offer built-in support for accuracy management. Consequently, it is very cumbersome to implement large-scale quantum algorithms in an accuracy-aware fashion. Thus, despite the availability of a wide range of quantum programming languages,

resource estimates are still computed (semi-)manually, taking care of accuracy parameters using pen and paper [Reiher et al. 2017; Scherer et al. 2017].

The main difficulty when selecting appropriate accuracy parameters is that parameters at a higher level of abstraction have an effect on the ones at lower levels, as illustrated by the following example:

**EXAMPLE 3.** Consider QPE on  $U = R_Z(\alpha) := e^{-i\frac{\alpha}{2}Z}$  and the target gate set Clifford+T. The number of phase-estimation qubits  $n_{QPE}$  depends on the desired precision of the phase (and the probability of success). At the lowest level, the various  $U^{2^i} = R_Z(\alpha_i)$  are decomposed into a sequence of Clifford+T gates featuring  $O(\log \frac{1}{\epsilon_r})$  T gates, where  $\epsilon_r$  is the error of a single rotation. To achieve an overall target accuracy  $\epsilon$ ,  $\epsilon_r$  must be chosen such that  $\epsilon_{QPE} + \epsilon_R \leq \epsilon$ , where  $\epsilon_R$  denotes the error introduced by all rotations in the quantum circuit. If the same accuracy is chosen for all the rotations, then  $\epsilon_R = N_{rot}\epsilon_r$ , where  $N_{rot}$  is the number of rotations. Since  $\epsilon_{QPE}$  affects the number of rotations in the circuit,  $\epsilon_r$  must be chosen as a function of  $\epsilon_{QPE}$ .

In general, it would be possible to adapt all values in the code manually on a case-by-case basis—however, this defeats the purpose of having a high-level programming language. The lack of language support forces programmers to manually handle accuracy parameters by passing all such parameters to the main routine, which forwards them to each subroutine. In case of the QPE algorithm, this might result in code as follows:

```

function QPE( $\epsilon_{QPE}, \epsilon_{TE}, \epsilon_{QFT}, \epsilon_{R_{TE}}, \epsilon_{R_{QFT}}, U$ )
   $reg\_size \leftarrow f(\epsilon_{QPE})$ 
  for  $i \leftarrow 0$  to  $reg\_size$  do
    for  $j \leftarrow 0$  to  $n\_iter(i)$  do
       ${}^cU(\epsilon_{TE}, \epsilon_{R_{TE}})$ 
  AQFT $^\dagger(\epsilon_{QFT}, \epsilon_{R_{QFT}})$ 

```

Here,  ${}^cU$  is the controlled version of  $U$ ; we also omitted all qubit variables for better readability.

This programmer-unfriendly approach does not allow code reuse for resource estimation, as implementations of subroutines need to be adapted to the context in which they are used and, in particular, to the choice of accuracy parameters. For example,  ${}^cU(\epsilon_{TE}, \epsilon_{R_{TE}})$  is implemented using a number of Clifford+T gates that depends on the chosen accuracy parameters.

When using our methodology, programmers need to worry about accuracy parameters only in subroutines where the corresponding errors are introduced. The compiler will take care of extracting all dependencies. Specifically, this allows us to express the pseudo-code for phase estimation as follows:

<pre> <b>function</b> <math>{}^cR</math>   declare <math>\epsilon_R</math>   ... <b>function</b> <math>{}^cU</math>   declare <math>\epsilon_{TE}</math>   ... <b>function</b> AQFT<math>^\dagger</math>   declare <math>\epsilon_{QFT}</math>   <math>l \leftarrow g(\epsilon_{QFT})</math>   <b>for</b> <math>i \leftarrow 0</math> to <math>g'(l)</math> <b>do</b>     ...     <math>{}^cR()</math>     ... </pre>	<pre> <b>function</b> QPE(<math>U</math>)   declare <math>\epsilon_{QPE}</math>   <math>reg\_size \leftarrow f(\epsilon_{QPE})</math>   <b>for</b> <math>i \leftarrow 0</math> to <math>reg\_size</math> <b>do</b>     <b>for</b> <math>j \leftarrow 0</math> to <math>n\_iter(i)</math> <b>do</b>       <math>{}^cU()</math>   AQFT<math>^\dagger()</math> </pre>
---	--

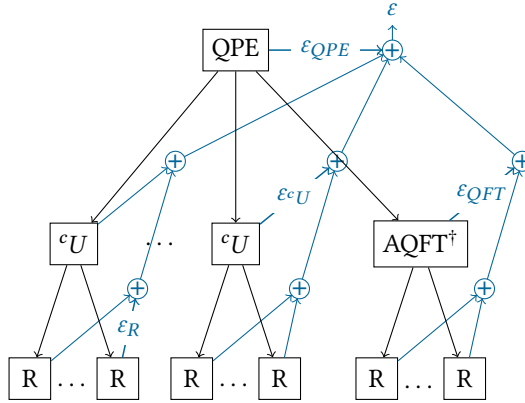


Fig. 4. Flow diagram explaining how the code of the QPE algorithm is transformed into a code evaluating the overall approximation error  $\epsilon$ .

Using abstract syntax tree (AST) transformations, our methodology is able to handle various levels of granularity: from using the same value for all accuracy parameters to using a different value for every instance that is created during runtime (via an accuracy parameter data structure that mirrors the call graph). This is crucial as there is a substantial trade-off between the number of accuracy parameters being considered and the resulting gate count [Häner et al. 2018]. This can be illustrated with the following example:

**EXAMPLE 4.** Consider Beauregard’s implementation of Shor’s algorithm [Beauregard 2002]. In addition to the (semi-classical) inverse Fourier transform of phase estimation, every addition circuit requires two (approximate) QFTs [Coppersmith 2002] (one inverted, one regular) [Draper 2000]. While the number of additions (and thus the number of QFTs) varies with the bit-size  $n$  of the number to factor, phase estimation always requires a single QFT. Therefore, it is natural to choose a different accuracy parameter for the (approximate) QFT of the phase estimation than for the (approximate) QFTs of the  $O(n^2)$ -many  $n$ -bit additions.

Besides facilitating accuracy-aware implementations of quantum programs and providing various levels of granularity for assigning accuracy parameters, our methodology allows us to automatically deduce the number of contexts in which a given (approximate) decomposition is applied. This enables automatic selection of the number of accuracy parameters and thus removes the need to perform this task manually.

## 5 AUTOMATING ACCURACY MANAGEMENT

In this section, we describe the proposed procedure to automatically determine accuracy parameters. The optimization problem is solved using a simulated annealing procedure that iteratively changes the parameters and evaluates the corresponding total approximation error. The procedure terminates as soon as accuracy parameters have been found that guarantee a user-specified overall accuracy. To further improve the parameter selection, we also evaluate the circuit cost in terms of number of expensive gates and pass the information to the optimization procedure. The result is a valid distribution of the available approximation error that in addition aims to minimize the circuit cost.

Our approach is to extract a (near-)symbolic expression for the total error and gate count from the algorithm and to use the obtained expressions in the annealing procedure. In general, it is possible

to use the annealing procedure with any available resource estimation method, the difference being that the runtime of each evaluation depends on the problem size if non-symbolic methods are used.

In the following sections, we use the  $T$ -count as a measure of the implementation cost of a quantum circuit, as this closely captures the cost in a fault-tolerant setting. We also envision that our framework is most beneficial in this setting, since larger programs lead to larger improvements over non-symbolic approaches. Nevertheless, we note that our approach may be employed with any cost function. For example, one may want to select accuracy parameters while aiming to minimize the number of CNOT or CZ gates when targeting NISQ devices.

### 5.1 Cost/Constraint Functions: Extraction

Our methodology proceeds by automatically generating two pieces of code that compute (an upper bound on) (1) the number of costly quantum gates ( $T$ ) and (2) the overall approximation error as a function of the different approximation errors ( $E$ ). We denote the two functions by

$$T(\varepsilon_1, \dots, \varepsilon_n) \quad \text{and} \quad E(\varepsilon_1, \dots, \varepsilon_n).$$

The automatic generation of these two functions can be achieved via a few simple transformations of the program's AST. Specifically, to generate  $T$ , all calls to native operations are removed from the AST, except those corresponding to costly gates that are replaced by counter-increments. The program computing a bound on the overall approximation error ( $E$ ) can be generated in a similar fashion, where increments are added for every epsilon declaration (see Sec. 2.3).

**EXAMPLE 5.** *In Fig. 4 we show the different decomposition levels of the QPE algorithm. The standard coherent QPE requires  $2^{n_{QPE}} - 1$  controlled time-evolution unitaries  ${}^cU$ , followed by an inverse QFT. At the next decomposition level, each unitary (including the QFT) is decomposed into rotation gates. In turn, those rotations will be fed into rotation synthesis, which outputs a sequence of  $O(\log \frac{1}{\varepsilon_R})$  Clifford+T gates for each rotation, where  $\varepsilon_R$  denotes the target accuracy of rotation synthesis (per rotation). Since errors accumulate at most linearly due to being unitary (see Sec. 2.3), an upper bound on the overall approximation error can be computed by adding all the  $\varepsilon_i$  introduced by the various decomposition steps.*

### 5.2 Cost/Constraint Functions: Optimization

Once the two pieces of code evaluating the total approximation error  $E(\varepsilon_1, \dots, \varepsilon_n)$  and the cost  $T(\varepsilon_1, \dots, \varepsilon_n)$  have been generated, they could be fed into the simulated annealing procedure. While this would allow us to perform accuracy management *automatically*, the resulting code will take substantial time to execute: typical quantum applications require on the order of  $10^{15}$  or more operations [Reiher et al. 2017] and the optimization loop is executed hundreds of times until suitable accuracy parameters are found.

As a remedy, we employ custom compiler optimization passes that significantly reduce the time required to evaluate gate counts and error bounds. Specifically, our methodology aims to infer symbolic and loop-free expressions for (upper bounds on) gate count and overall approximation error. The following example demonstrates how beneficial the use of our symbolic approach is:

**EXAMPLE 6.** *Consider the example of the approximate quantum phase estimation algorithm and a two-mode simulated annealing procedure. Even with an optimized annealing schedule, it will require a minimum of about 200 evaluations to guarantee an overall approximation error of at most  $10^{-2}$ . As accuracy parameters  $\varepsilon_1 \dots \varepsilon_n$  approach the optimal values (minimizing the  $T$ -count), one evaluation of the non-optimized  $T(\varepsilon_1, \dots, \varepsilon_n)$  function on 8 qubits takes 9m 10s, while evaluating the inferred symbolic expression takes  $0.1\mu\text{s}$ . If the number of qubits grows to 16, then we have 34m 14s for the non-optimized case, while evaluating the expression still takes  $0.1\mu\text{s}$ .*



The transformations that we propose to implement at the level of the intermediate-representation of the compiler are shown in the following table:

Table 1. Compiler transformations that our approach uses to optimize the cost and constraint functions.

	Original Code	Symbolic expression
1	<b>for</b> $i \leftarrow 0$ to $N$ <b>do</b> $v \ += \ const$	$v \ += \ const \cdot N$
2	<b>for</b> $i \leftarrow 0$ to $N$ <b>do</b> $v \ += \ f(i)$	$v \ += \ \sum_i f(i)$
	<b>for</b> $i \leftarrow 0$ to $N$ <b>do</b> $v \ += \ \min(g(i), h(i))$	$v \ += \ \sum_i \min(g(i), h(i))$ $\leq$ $v \ += \ \min(\sum_i g(i), \sum_i h(i))$
	<b>for</b> $i \leftarrow 0$ to $N$ <b>do</b> $v \ += \ i^p$	$v \ += \ \sum_i i^p$ $(p + 1)^{\text{th}}$ degree polynomial derived from Faulhaber's formula
3	<b>if</b> (...) <b>then</b> $expr1$ <b>else</b> $expr2$	$\max(expr1, expr2)$

In particular, our optimization routine would:

- (1) Check if there is an addition between a variable  $v$  initialized outside the loop and a loop invariant, and if  $v$  is not used elsewhere in the loop. If so, apply transformation 1, where  $N$  is the number of iterations of the loop.
- (2) Check if there is an addition between a variable  $v$  initialized outside the loop and a function  $f$  only depending on the inductive variable and other loop invariants. If so, apply transformation 2. In the particular case where  $f(i)$  is  $\min()$ , the expression can be upper bounded as shown in Table 1. In addition, polynomial expressions can be derived from some finite series using Faulhaber's formula [Conway and Guy 1996].
- (3) Transform generic branching instructions into  $\max(if, else)$  instructions, where the branch that gives the largest contribution to the cost function is selected.

The following example shows the described transformations applied to the AQFT quantum algorithm.

**EXAMPLE 7.** *The pseudo-code of the function AQFT\_T, which computes the total number of T gates required for the AQFT algorithm, obtained after source-to-source transformation is shown in Alg. 1. As our implementation uses three rotation gates for each controlled-rotation, the function takes as input three accuracy parameters. This is why there are three innermost loops in Alg. 1. Since we want to extract a symbolic expression for the variable  $T_{count}$ , we have to get rid of as many loops as possible. The if statement in the inner-loop may be hoisted, resulting in the following expression:*

$$\mathbf{for} \ j \leftarrow 0 \ \text{to} \ \min(n - 1 - i, l) \ \mathbf{do}$$

*Then all the loops are optimized by applying the transformations in Table 1. Finally, the code in Alg. 2 is obtained, which shows the closed-form expression for the  $T_{count}$  with respect to the algorithm's parameters.*

**Algorithm 1** Cost function for the AQFT algorithm

---

```

function N_ROT( $\epsilon_{rot}$ ) return  $1.5 * \log_2(1./\epsilon_{rot})$ 
function AQFT_T( $\epsilon_{QFT}, \epsilon_{R_1}, \epsilon_{R_2}, \epsilon_{R_3}$ )
   $T_{count} \leftarrow 0$ 
   $l \leftarrow \lceil \log_2(n/\epsilon_{QFT}) \rceil + 3$ 
  for  $i \leftarrow 0$  to  $n$  do
    for  $j \leftarrow 0$  to  $n - 1 - i$  do
      if  $j \leq l$  then
        for  $k \leftarrow 0$  to  $N\_ROT(\epsilon_{R_1})$  do  $T_{count}++$ 
        for  $k \leftarrow 0$  to  $N\_ROT(\epsilon_{R_2})$  do  $T_{count}++$ 
        for  $k \leftarrow 0$  to  $N\_ROT(\epsilon_{R_3})$  do  $T_{count}++$ 
  return  $T_{count}$ 

```

---

**Algorithm 2** Cost function for the AQFT algorithm after loop optimization

---

```

function AQFT_T( $\epsilon_{QFT}, \epsilon_{R_1}, \epsilon_{R_2}, \epsilon_{R_3}$ )
   $T_{count} \leftarrow 0$ 
   $l \leftarrow \lceil \log_2(n/\epsilon_{QFT}) \rceil + 3$ 
   $T_{count} = T_{count} + \min(\frac{n(n-1)}{2}, nl) \cdot (N\_ROT(\epsilon_{R_1}) + N\_ROT(\epsilon_{R_2}) + N\_ROT(\epsilon_{R_3}))$ 
  return  $T_{count}$ 

```

---

While our methodology succeeds at extracting closed-form expressions for all our examples, we note that this is not necessary for our methodology to work: the remaining control flow would not affect applicability or correctness, but merely cause an increase in runtime. Indeed, there are cases in which some residual code remains in the resulting expressions. This happens, for example, when some of the program parameters are read from a file. Consider the program performing phase estimation of the time evolution of a TFIM Hamiltonian described in Section 3, but where the parameters  $J$ ,  $h$  and  $n$  are read from a file. The program may check whether the input Hamiltonian is valid, e.g., whether  $n \geq 0$ , before instantiating the circuit. As a consequence, such an if/else statement would remain in the final expression.

## 6 COMPILER AND LANGUAGE REQUIREMENTS

To add our methodology for automatic accuracy management to any quantum programming language, a few features must be added to the compiler if not already supported. Here we list all the features that the compiler must support to bring such an integration into fruition.

We have identified such features by working on our initial prototype. The prototype has been developed using the LLVM project, which we chose for its modular infrastructure and libraries. The strategy that we identified while implementing our methodology in LLVM can be applied to any other quantum programming language. We illustrate this by developing a second prototype in Q#. Furthermore, our LLVM prototype shows that support for our framework may also be added to quantum programming languages that are embedded in a classical host language.

*Don't Cares.* Our methodology requires that the compiler identifies subroutine parameters that do not or only negligibly affect the total error and the cost function. We call such function parameters *don't cares*. The corresponding arguments will be replaced by a default value in all calls to that subroutine. This allows the compiler to optimize repeated calls to the same function.

**EXAMPLE 8.** Consider AQFT, which includes many calls to the rotation gate with different rotation angles. Normally, these calls will have to be evaluated several times by the compiler, even if the angle will have no impact on the approximation error selected to decompose the rotation. We address this problem by annotating the angle parameter as a *don't care*. This will result in many calls to the same function with identical arguments that hence may be removed by the compiler.

*Epsilon Declarations.* To provide language support (see Section 4) the compiler must be capable of locating all introduced accuracy parameters. This can be done by matching against a specific class or by using annotations.

*AST Transformation.* The compiler must provide access to the AST (or to a reasonably high-level IR) and allow rewriting and copying. In particular, we need to generate 3 different versions of the entire program: one that computes the total approximation error, one that computes the total cost, and the original quantum program, which will ultimately be invoked using optimized accuracy parameters.

The programs obtained after the described AST modifications could already be used to estimate the resource requirements of the quantum program. In this case, the only advantage with respect to using state-of-the-art methods would be that our approach provides language support by keeping track of all the introduced approximation errors, work that otherwise would have to be performed manually. In addition, the estimation would be too slow to be used in an optimization procedure (see results in Fig. 7). As a remedy, we introduce specific rewrites that reduce the time to evaluate these expressions (see next paragraph).

*Rewrites to Make Evaluation More Efficient.* In order to speed up the optimization process, we need fast evaluations of the cost and error functions. The two functions extracted from most quantum algorithms will feature many loops performing simple counter increments or floating-point additions. To achieve our goal of extracting a symbolic expression from the control flow structure, we implement and employ several compiler optimizations, see Table 1.

## 6.1 LLVM Prototype

Having described the features that are necessary to equip a programming language with automatic accuracy management, we now provide implementation details for our LLVM prototype.

*Don't Cares.* In our LLVM implementation, we use compiler annotations to introduce additional information in the source code. In particular, we attach a *don't care* annotation to a parameter declaration if it has negligible effect on the cost/constraint functions.

*Epsilon Declarations.* Epsilon declarations are matched with a custom type. Our prototype framework provides a macro that allows the programmer to quickly define new types of error. This can be seen in the source code for the approximate QFT in Fig. 5, where `epsilon_QFT` is registered as a new type of error in line 4 and then used in lines 8–9.

*AST Transformation.* We approach the problem of modifying the AST using a source-to-source transformation. We implement a *ClangTool* and run an *ASTFrontendAction*: a routine that has access to the AST and allows us to interface with the source code. Our *ClangTool* outputs files containing the function to compute the  $T$ -count, e.g., the one in Alg. 1, and the function computing the total approximation error  $E$ . Our action exploits the *ASTMatcher* library, which allows us to match nodes in the AST that have some specific properties. The library provides a concise way of describing patterns and is implemented as a *domain-specific language* (DSL). In addition, matchers allow us to access the source code by running a callback function on the matched AST nodes. Once the locations of interest in the code have been identified, we can use an instance of the *Rewriter* class

```

1 #include <qadf/epsilon.h>
2 #include <qadf/operations.h>
3
4 REGISTER_EPSILON(epsilon_QFT)
5
6 inline void Q_FUNC AQFT(int qubits[], const int n)
7 {
8     epsilon_QFT eps_QFT;
9     int lim = ceil(log2(n / eps_QFT))+3;
10    for(int i = 0; i < n; i++)
11    {
12        H(qubits[i]);
13        for (int j = 0; j < min(n-1-i, lim); j++)
14        {
15            CR(qubits[j + i + 1], M_PI/(1 << (j + 1)), qubits[i]);
16        }
17    }
18 }

```

Fig. 5. C++ code for the approximate QFT as consumed by our LLVM prototype.

to modify the code accordingly. As the AST is constant by design, we must generate a new file containing our transformed source code.

Our tool also makes use of header files in which we define basic quantum operations, such as the ones in the Clifford+ $T$  gate set. Those header files can be adjusted according to the specific application. For example, in addition to the  $T$ -gate, we might want to consider other expensive operations, e.g., two-qubit gates for NISQ devices.

The result of running the Clang tool is a new source file computing the total error or gate count as a function of all the accuracy parameters defined in the source code.

*Rewrites to Make Evaluation More Efficient.* Our compiler optimization passes eliminate loops in the expressions for the cost (or the total error) by replacing them with additions and multiplications. For example, the first loop optimization described in Table 1 is supported for both integral and floating-point numbers. We use a custom LLVM *loop pass* to perform this optimization.

**EXAMPLE 9.** Consider the following code that computes the total approximation error of  $n$  quantum operations, characterized by the same approximation error  $val$ :

```

1 double Eps = 0.00;
2 double val = 0.02;
3 for (int i = 0; i < n; i++) {
4     Eps += val;
5 }

```

Once this function is compiled into Intermediate Representation (IR) code,  $val$  will be identified as a loop-invariant variable, while  $Eps$  will be assigned to a so-called PHI node. PHI nodes assign a variable with a different value, depending on the predecessor of the current block, where blocks are groups of instructions. In our example, the PHI node would have two incoming values: 0.00 (if the predecessor block is outside the loop) and the temporary value containing the addition result (if the predecessor was the previous loop iteration). PHI nodes are defined in the loop header block.

The loop pass traverses the code from the innermost to the outermost loop in the IR and checks whether:

- (1) it contains an instruction performing the addition operation between a loop-invariant operand and a variable defined through a PHI node in the loop header,
- (2) the PHI node is only used in the addition operation inside the loop or in the loop latch block,
- (3) the result of the addition is only used as the incoming value of the PHI node.

If the described conditions apply, the loop is removed. Referring to the code in Example 9, the operations  $\%1 = val * n$  and  $\%2 = Eps + \%1$  would be added to the pre-header loop block, i.e., outside the loop. In addition, the result  $\%2$  would replace all uses of the original addition operation, which can then be erased. All the other transformations in Table 1 are implemented in a similar fashion.

*Extraction of symbolic Representation.* Our LLVM prototype also has a method that navigates the fully optimized IR code and extracts symbolic expressions for the two estimates. The symbolic expression that we obtained by running our sample quantum algorithms are listed in Table 2.

The pass to extract the symbolic expression from the IR is implemented as an LLVM *function pass*. Given the main function, it starts from the return instruction and recursively visits all instruction's operands annotating the respective functionality. The recursion terminates when the operand is a constant or, in general, not an instruction.

The extraction pass supports the following instructions: casting, PHI nodes, selects, truncations, zero extensions, call instructions, compare instructions, shifts, addition, multiplication, division, and subtraction. The expression is written in the Wolfram language, such that Mathematica [Wolfram Research Inc. 2019] can be used for conversion to  $\LaTeX$  and further expression simplifications.

## 6.2 Q# Integration

Q# is a standalone quantum programming language developed by Microsoft to facilitate the description of hybrid quantum-classical programs. In Fig. 6, we show a snippet of Q# code that implements the approximate quantum Fourier transform operation, as described in Section 3.

Qubits are represented in Q# using the type `Qubit` and they are treated as opaque items that can be passed to both functions and operations, but that can only be interacted with by passing them to intrinsic (built-in) operations. Q# also uses namespaces to group definitions together, and elements from other namespaces may be referenced. Q# distinguishes functions from operations. Functions are pure and free of side effects, whereas operations can have side effects, such as the application of an intrinsic operation to a qubit or register. Q# can perform type-safe symbolic computations to automatically derive the adjoint (inverse) and the controlled variants of an operation, enforced by providing the `is Adj+Ctl` declaration in line 13 of Fig. 6. Line 15 shows a declaration of an approximation parameter, i.e., `eps_QFT`, which is possible by adding language support for approximation errors. In the remainder of this section, we provide details on how this and other features have been implemented in the Q# compiler.

*Don't Cares.* Currently, our Q# prototype does not support *don't cares*. We implement operations such as rotation gates—in which *don't cares* can help to declare that the rotation angle affects the gate cost only negligibly—as intrinsic operations that introduce an epsilon variable and increment the gate counter explicitly by a value depending on the accuracy parameter. For more general cases, we would require Q# to support parameter-level annotations to declare *don't cares* explicitly.

*Epsilon Declarations.* We use an intrinsic function for epsilon declarations. Being intrinsic, it does not require an implementation inside the Q# program to be used, but we can locate it inside the AST in our transformation passes. In the AQFT example, the function is declared in lines 9-11 and allows the programmer to declare and use accuracy parameters, e.g., as done in lines 15-16.

```

1 namespace Accuracy {
2
3   open Microsoft.Quantum.Arrays;
4   open Microsoft.Quantum.Convert;
5   open Microsoft.Quantum.Intrinsic;
6   open Microsoft.Quantum.Math;
7
8   // intrinsic function for epsilon declaration
9   function EpsilonValue() : Double {
10    body intrinsic;
11  }
12
13  operation AQFT(qs : Qubit[]) : Unit is Adj+Ctl {
14    let nQubits = Length(qs);
15    let eps_QFT = EpsilonValue(); // epsilon-declaration
16    let lim = Ceiling(Lg(IntAsDouble(nQubits) / eps_QFT)) + 3;
17    for ((i, q) in Enumerated(qs)) {
18      H(q);
19      for (j in 0 .. MinI(nQubits - i - 1, lim)) {
20        (Controlled R1Frac)([q], (1, j + 1, qs[j + i + 1]));
21      }
22    }
23  }
24 }

```

Fig. 6. Q# implementation of the approximate quantum Fourier transform with epsilon declarations.

*AST Transformation.* We have implemented an AST transformation pass to detect all `EpsilonValue` declarations, remove them, and add them as arguments to the operation signature. This step is performed before producing the two pieces of code that compute the number of costly quantum gates and the upper bound on the overall approximation error. Note that the intrinsic `EpsilonValue` function is never called in the resulting programs.

There exist no global variables and no call-by-reference parameters in Q#. However, it is possible to declare an intrinsic operation in Q# and implement it in C#. In order to count the number of  $T$  gates, we introduce an operation `IncrementCounter(id, value)` whose implementation in C# increments a global counter, called `id`, by `value`.

We use a similar technique as used for counting gates to accumulate the bound on the overall approximation error. Each declaration of an epsilon value is replaced by a call to an operation `IncrementValue(id, value)` whose implementation in C# increments a global variable called `id` by `value`.

*Rewrites to Make Evaluation More Efficient.* The Q# compiler already contains some transformation passes, e.g., for operation inlining, propagating constants, or removing unused code. We have added two additional transformation passes that optimize the use of `IncrementCounter` and `IncrementValue` calls. Without loss of generality we explain the transformation passes by means of the `IncrementCounter` operation, remarking that they work analogously for the `IncrementValue` operation.

The first transformation pass collects all `IncrementCounter` calls inside a scope level that have the same `id` and do not contain values incorporating mutable variables. These calls can be merged



into a single call by accumulating all values, benefiting from further optimization, e.g., constant propagation. The second transformation pass lifts `IncrementCounter` calls inside a for-loop. If the call is the only statement in the body of the for-loop and does not use the loop variable to compute the value, the for-loop can be removed when multiplying the value in the `IncrementCounter` call by the number of loop iterations.

## 7 QUALITATIVE EVALUATION

In this section, we evaluate our prototypes using different quantum algorithms. Starting with a simple quantum Fourier transform, we increase the complexity of our examples. As a highlight, our prototype automatically extracts a symbolic expression for the phase estimation of a Trotter-decomposed time-evolution under a transverse-field Ising model Hamiltonian, where the phase estimation features an approximate QFT. The C++ code for the approximate QFT can be found in Fig. 5 and the corresponding Q# code is shown in Fig. 6. In addition, we also tested our prototype on Shor’s algorithm [Shor 1994], to provide the reader with a large-scale example. We used the period finding quantum routine as implemented in the open source programming framework ProjectQ [Steiger et al. 2018]. All the obtained expressions for the  $T$ -count and the total approximation error are reported in Table 2.

*Exact QFT.* The first example is an implementation of the exact quantum Fourier transform. Our prototype is able to directly optimize all loops, including the outermost loop which yields a sum of the form  $\sum_i^{n-1} i = \frac{n(n-1)}{2}$ . It finds the correct closed-form expression for the total error and it correctly identifies the number of  $T$  gates to be  $O(n^2 \log(1/\epsilon_R))$ . For the detailed output, we refer to Table 2.

*Approximate QFT (AQFT).* Next, we consider the approximate quantum Fourier transform. The C++ source code that serves as the input to our prototype is depicted in Fig. 5. Our prototype upper bounds an intermediate expression of the form  $c + \sum_i \min(f(i), g(i))$  by choosing one of the arguments to the min-function and successfully derives a closed-form expression for both the  $T$  gate count and the total approximation error, see Table 2.

*Quantum Phase Estimation (QPE).* We combine the time evolution of a TFIM with QPE, to find the ground state of the TFIM. In this first QPE example, we make the simplifying assumption that the inverse QFT can be performed natively. As can be seen in Table 2 (labeled *QPE simplified*), our methodology is capable of removing all loops and outputs two closed-form expressions for the  $T$ -count and the total error.

In a next step, we drop the simplifying assumption that the inverse QFT can be performed natively. We implement the inverse QFT as discussed in Section 3.1 and run our prototype on this larger example. The closed-form expressions for this case can be found in Table 2 labeled *QPE with QFT*.

Finally, we replace the exact QFT by an approximate QFT (see Fig. 5 for the C++ code). As in the AQFT example, our optimization pass upper bounds an intermediate expression of the form  $c + \sum_i \min(f(i), g(i))$  by choosing one of the arguments to the min-function. We refer to Table 2 for the detailed output, which consists of two fully-symbolic expressions for the  $T$ -count and for the approximation error.

*Shor’s Algorithm.* As last example, we present the results for Beauregard’s implementation of Shor’s algorithm [Beauregard 2002], which defines two approximation errors, one for the rotation gates  $\epsilon_R$  and one for the approximate QFT ( $\epsilon_{QFT}$ ). In this implementation, each controlled unitary in the phase-estimation procedure is a modular multiplication (by a constant). Please note that in

Table 2. Symbolic expressions for the  $T$ -count ( $T$ ) and total approximation error ( $E$ ) as extracted by our prototype from the source code of various example programs.

Algorithm (selected function)	Symbolic expression
QFT	$E = \frac{3}{2}\varepsilon R n(n-1)$ $T \approx 3.246n(n-1) \log\left(\frac{1}{\varepsilon R}\right)$
AQFT	$E \approx 3\varepsilon R n \left( \left\lfloor \frac{\log(\frac{\varepsilon QFT}{\log 2})}{\log 2} \right\rfloor + 3 \right) + \varepsilon QFT$ $T \approx 6.492n \log\left(\frac{1}{\varepsilon R}\right) \left( \left\lfloor \frac{\log(\frac{\varepsilon QFT}{\log 2})}{\log 2} \right\rfloor + 3 \right)$
QPE simplified	$E \approx \frac{\left( \varepsilon R \left( n - \frac{1}{2} \right) + \frac{\varepsilon T E}{4} \right) \left( 2^{\lceil 2.652 - 1.443 \log(\varepsilon QPE) \rceil + 4} \right)^{-4} + \varepsilon QPE}{34.625 \left( n - \frac{1}{2} \right) \log\left(\frac{1}{\varepsilon R}\right) \left( 2^{\lceil 2.652 - 1.443 \log(\varepsilon QPE) \rceil - \frac{1}{4}} \right) \sqrt{\varepsilon T E}}$ $T \approx \frac{\sqrt{\varepsilon T E}}{34.625 \left( n - \frac{1}{2} \right) \log\left(\frac{1}{\varepsilon R}\right) \left( 2^{\lceil 2.652 - 1.443 \log(\varepsilon QPE) \rceil - \frac{1}{4}} \right) \sqrt{\varepsilon T E}}$
QPE with QFT	$E \approx \frac{16 \left( \varepsilon R \left( n - \frac{1}{2} \right) + \frac{\varepsilon T E}{4} \right) \left( 2^{\lceil 2.652 - 1.443 \log(\varepsilon QPE) \rceil - \frac{1}{4}} \right)}{\sqrt{\varepsilon T E}} + \frac{3}{2} \varepsilon R \left( \lceil 2.652 - 1.443 \log(\varepsilon QPE) \rceil + 1 \right) \left( \lceil 2.652 - 1.443 \log(\varepsilon QPE) \rceil + 2 \right) + \varepsilon QPE$ $T \approx \log\left(\frac{1}{\varepsilon R}\right) \left( \left( n - \frac{1}{2} \right) \left( 34.625 \right)^{\lceil 2.652 - 1.443 \log(\varepsilon QPE) \rceil - 8.656} + \lceil 2.652 - 1.443 \log(\varepsilon QPE) \rceil \left( 3.246 \lceil 2.652 - 1.443 \log(\varepsilon QPE) \rceil + 9.738 \right) + 6.492 \right)$
QPE with AQFT	$E \approx 3\varepsilon R \left( \lceil 2.652 - 1.443 \log(\varepsilon QPE) \rceil + 2 \right) \left( \left\lfloor \frac{\log\left(\frac{\varepsilon QFT}{\log 2}\right) \lceil 2.652 - 1.443 \log(\varepsilon QPE) \rceil + 2}{\log 2} \right\rfloor + 3 \right) + \frac{16 \left( \varepsilon R \left( n - \frac{1}{2} \right) + \frac{\varepsilon T E}{4} \right) \left( 2^{\lceil 2.652 - 1.443 \log(\varepsilon QPE) \rceil - \frac{1}{4}} \right)}{\sqrt{\varepsilon T E}} + \varepsilon QFT + \varepsilon QPE$ $T \approx \log\left(\frac{1}{\varepsilon R}\right) \left( 6.492 \left( \lceil 2.652 - 1.443 \log(\varepsilon QPE) \rceil + 2 \right) \left( \left\lfloor \frac{\log\left(\frac{\varepsilon QFT}{\log 2}\right) \lceil 2.652 - 1.443 \log(\varepsilon QPE) \rceil + 2}{\log 2} \right\rfloor + 3 \right) + \frac{34.625 \left( n - \frac{1}{2} \right) \left( 2^{\lceil 2.652 - 1.443 \log(\varepsilon QPE) \rceil - \frac{1}{4}} \right)}{\sqrt{\varepsilon T E}} \right)$
Shor's algorithm [Beauregard 2002]	$E \approx 4n^2 \left( \varepsilon R (n+1) \left( \left\lfloor \frac{\log\left(\frac{n+1}{\log 2}\right)}{\log 2} \right\rfloor + 109 \right) + 10\varepsilon QFT \right)$ $T \approx 2n^2 \left( 129.843(n+1) \log\left(\frac{1}{\varepsilon R}\right) \left( \left\lfloor \frac{\log\left(\frac{n+1}{\log 2}\right)}{\log 2} \right\rfloor + 471.761(n+1) \log\left(\frac{1}{\varepsilon R}\right) + 84n + 91 \right) \right)$

our QPE example, we perform a phase-estimation on the time-evolution operator that evolves the system according to the transverse-field Ising model Hamiltonian. Therefore, while both examples make use of phase estimation, their cost and error functions are vastly different because the phase estimation is performed on different unitaries.

In summary, our methodology successfully produces closed-form expressions for the total error and gate count for all our examples. In the next section, we show that access to symbolic expressions enables a significant reduction in the time required to optimize accuracy parameters.

## 8 QUANTITATIVE EVALUATION

In this section, we demonstrate how our prototype implementation enables faster evaluations of the cost and constraint functions by leveraging the transformations in Table 1. We compare the runtimes of our symbolic resource estimation method against a *non-symbolic* approach. The latter does not use symbolic estimations for resources and errors. The non-symbolic estimates are generated by our prototypes during the AST modification step (see Section 6).

We ran all experiments on a MacBook Pro with an Intel Core i5 processor with 3.1 GHz processor clock frequency and 16 GB main memory. All source codes have been compiled using Clang version 9.0.0 with level 3 optimization (`-O3`) and with the fast-math mode enabled (`-ffast-math`).

For these experiments, we choose our Clang/LLVM prototype. As previously described, it uses a two-mode annealing procedure to find suitable assignments for all accuracy parameters. We measure the runtime for performing one evaluation of the  $T$  and  $E$  functions using accuracy parameters provided by the annealer, guaranteeing an upper bound on the overall approximation equal to  $5 \cdot 10^{-3}$ . The runtime measurements are performed using different input sizes for Shor's algorithm and for our QPE example (with approximate QFT). The plots in Fig. 7 depict the sum of the runtimes required for evaluating the constraint and cost functions once, as they are always evaluated the same number of times in the simulated annealing procedure. While the non-symbolic approach exhibits a growing runtime as a function of the problem size, our symbolic method shows the expected constant behavior. In particular, the runtime of the non-symbolic approach grows linearly for the QPE example and (roughly) cubically for our implementation of Shor's algorithm.

Given that both examples are valid applications of quantum computers only for large problem sizes (e.g., the target number of bits for Shor's algorithm is  $n \approx 4000$ ), we show function extrapolations for the non-symbolic approach in Fig. 7. The two resulting functions are  $f_{QPE} = 1737.30x + 816.98$  and  $f_{Shor} = 2.38 \cdot 10^{-5}x^3$ .

To estimate the time it would take to optimize accuracy parameters using a *non-symbolic* approach, we multiply our runtime results in Fig. 7 by a lower bound on the number of function evaluations. For our examples, we find a loose lower bound of 100 evaluations (of each function). To determine suitable accuracy parameters for Shor's algorithm, we thus obtain a theoretical runtime of approximately 1890 days for  $n = 4096$  bits and an overall approximation error of at most  $5 \cdot 10^{-3}$ .

Therefore, we may conclude that non-symbolic approaches are not suitable for large-scale applications.

## 9 CONCLUSION AND OUTLOOK

We describe the first framework with the ability to automatically manage approximation errors and outputting (near-)symbolic resource estimates. Our methodology can be added to any quantum software framework, thereby greatly facilitating resource estimation of quantum programs. Such integration will allow even domain experts from, e.g., chemistry or machine learning, to write accuracy-aware quantum programs without having to manually derive and prove error bounds.

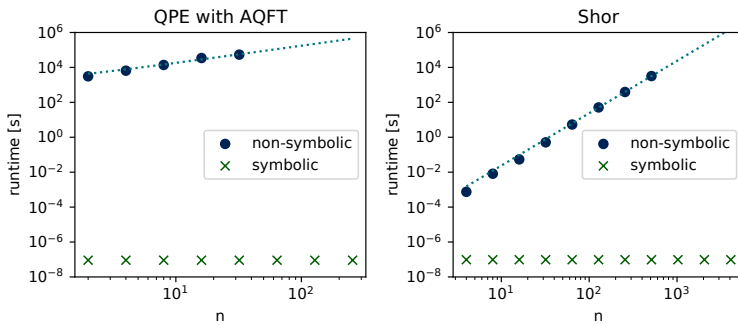


Fig. 7. Each data point marks the runtime required for a single evaluation of the  $T$  function plus a single evaluation of the  $E$  function for the QPE algorithm with approximate QFT and for Shor’s algorithm. Optimized accuracy parameters are used as input in order to assert that the total approximation error is at most  $5 \cdot 10^{-3}$ . A comparison between the runtimes of the non-symbolic approach and the symbolic approach developed in this paper is shown. We provide polynomial extrapolations from the collected runtimes for the non-symbolic approach as the runtime tops out for bit sizes above 512 for Shor and 32 for QPE.

In this work, we identify the features that a quantum programming language must support in order to enable our methodology. We develop two prototype implementations that are capable of handling several example programs, including an implementation of Shor’s algorithm.

Future work could implement improved handling of branching on measurement results and repeat-until-success-like structures. To handle such programs, our methodology requires additional input such as the maximal or expected iteration count (e.g., as a program annotation). For verification purposes, one could instrument the code in order to assert that the actual number of iterations does not deviate (too much) from the provided estimate. Future work could also compare upper bounds to actually achieved errors on example applications. Currently, our methodology does not take into account gate cancellations that may be performed by circuit optimization. This, in addition to the repeated use of the triangle inequality to bound the overall error, likely leads to pessimistic error bounds.

## ACKNOWLEDGMENTS

We thank the shepherd, Robert Rand, and the anonymous reviewers for their helpful suggestions and feedback, which significantly improved this paper. We are grateful to Vadym Kliuchnikov for enlightening discussions and comments.

## REFERENCES

- Gadi Aleksandrowicz and et al. 2019. Qiskit: An Open-source Framework for Quantum Computing. (Jan. 2019). <https://doi.org/10.5281/zenodo.2562110>
- Matthew Amy, Dmitri Maslov, Michele Mosca, and Martin Roetteler. 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, 6 (Jun 2013), 818–830. <https://doi.org/10.1109/tcad.2013.2244643>
- Matthew Amy, Martin Roetteler, and Krysta M. Svore. 2017. Verified Compilation of Space-Efficient Reversible Circuits. In *Computer Aided Verification - 29th International Conference, CAV*, 3–21. [https://doi.org/10.1007/978-3-319-63390-9\\_1](https://doi.org/10.1007/978-3-319-63390-9_1)
- Thanassis Avgerinos, Alexandre Rebert, Sang Kil Cha, and David Brumley. 2014. Enhancing symbolic execution with veritesting. In *36th International Conference on Software Engineering, ICSE '14*. 1083–1094. <https://doi.org/10.1145/2568225.2568293>
- Ryan Babbush, Dominic W. Berry, Yuval R. Sanders, Ian D. Kivlichan, Artur Scherer, Annie Y. Wei, Peter J. Love, and Alán Aspuru-Guzik. 2017. Exponentially more precise quantum simulation of fermions in the configuration interaction representation. *Quantum Science and Technology* 3, 1 (2017), 015006. <https://doi.org/10.1088/2058-9565/aa9463>

- Earl T. Barr, Thanh Vo, Vu Le, and Zhendong Su. 2013. Automatic detection of floating-point exceptions. *ACM SIGPLAN Notices* 48, 1, 549–560. <https://doi.org/10.1145/2480359.2429133>
- Stephane Beauregard. 2002. Circuit for Shor’s algorithm using  $2n+3$  qubits. *Quantum information & computation* 3 (06 2002).
- Charles H. Bennett and Gilles Brassard. 2014. Quantum cryptography: Public key distribution and coin tossing. *Theoretical Computer Science* 560 (dec 2014), 7–11. <https://doi.org/10.1016/j.tcs.2014.05.025>
- Ethan Bernstein and Umesh Vazirani. 1997. Quantum Complexity Theory. *SIAM J. Comput.* 26, 5 (oct 1997), 1411–1473. <https://doi.org/10.1137/s0097539796300921>
- S. Boldo. 2009. Kahan’s Algorithm for a Correct Discriminant Computation at Last Formally Proven. *IEEE Trans. Comput.* 58, 2 (2009), 220–225. <https://doi.org/10.1109/tc.2008.200>
- Sergey Bravyi and Alexei Kitaev. 2005. Universal quantum computation with ideal Clifford gates and noisy ancillas. *Physical Review A* 71 (Feb 2005), 022316. Issue 2. <https://doi.org/10.1103/PhysRevA.71.022316>
- Earl T. Campbell and Mark Howard. 2017. Unified framework for magic state distillation and multiqubit gate synthesis with reduced resource cost. *Physical Review A* 95, 2 (Feb 2017). <https://doi.org/10.1103/physreva.95.022316>
- John H. Conway and Richard K. Guy. 1996. *The Book of Numbers*. Springer New York. <https://doi.org/10.1007/978-1-4612-4072-3>
- Don Coppersmith. 2002. An approximate Fourier transform useful in quantum factoring. arXiv:quant-ph/0201067 [quant-ph]
- Eva Darulova and Viktor Kuncak. 2014. Sound compilation of reals. *ACM SIGPLAN Notices* 49, 1, 235–248. <https://doi.org/10.1145/2578855.2535874>
- Thomas G. Draper. 2000. Addition on a quantum computer. *arXiv preprint quant-ph/0008033* (2000). arXiv:quant-ph/0008033
- Austin G. Fowler, Matteo Mariantoni, John M. Martinis, and Andrew N. Cleland. 2012. Surface codes: Towards practical large-scale quantum computation. *Physical Review A* 86, 3 (Sep 2012). <https://doi.org/10.1103/physreva.86.032324>
- Alexander S. Green, Peter LeFanu Lumsdaine, Neil J. Ross, Peter Selinger, and Benoît Valiron. 2013. Quipper. *ACM SIGPLAN Notices* 48, 6, 333–342. <https://doi.org/10.1145/2499370.2462177>
- Thomas Häner, Martin Roetteler, and Krysta M. Svore. 2018. Managing approximation errors in quantum programs. (2018). arXiv:1807.02336 [quant-ph]
- Vojtěch Havlíček, Antonio D. Córcoles, Kristan Temme, Aram W. Harrow, Abhinav Kandala, Jerry M. Chow, and Jay M. Gambetta. 2019. Supervised learning with quantum-enhanced feature spaces. *Nature* 567, 7747 (Mar 2019), 209–212. <https://doi.org/10.1038/s41586-019-0980-2>
- Shih-Han Hung, Kesha Hietala, Shaopeng Zhu, Mingsheng Ying, Michael Hicks, and Xiaodi Wu. 2019. Quantitative robustness analysis of quantum programs. *Proceedings of the ACM on Programming Languages* 3, POPL (2019), 31:1–31:29. <https://doi.org/10.1145/3290344>
- Ali Javadi-Abhari, Shruti Patil, Daniel Kudrow, Jeff Hecke, Alexey Lvov, Frederic Chong, and Margaret Martonosi. 2014. ScaffCC: a framework for compilation and analysis of quantum computing programs. *Proceedings of the 11th ACM Conference on Computing Frontiers, CF 2014* (05 2014). <https://doi.org/10.1145/2597917.2597939>
- Alexei Kitaev. 1995. Quantum measurements and the Abelian Stabilizer Problem. *arXiv preprint arXiv:quant-ph/9511026* (1995). arXiv:quant-ph/9511026 [quant-ph]
- Vadym Kliuchnikov, Dmitri Maslov, and Michele Mosca. 2013. Fast and efficient exact synthesis of single qubit unitaries generated by Clifford and  $T$  gates. *Quantum Information & Computation* 13 (June 2013), 607–630.
- E. Knill. 1996. *Conventions for quantum pseudocode*. Technical Report. <https://doi.org/10.2172/366453>
- Guang Hao Low and Isaac L. Chuang. 2019. Hamiltonian Simulation by Qubitization. *Quantum* 3 (jul 2019), 163. <https://doi.org/10.22331/q-2019-07-12-163>
- Sam McArdle, Suguru Endo, Alán Aspuru-Guzik, Simon C. Benjamin, and Xiao Yuan. 2020. Quantum computational chemistry. *Rev. Mod. Phys.* 92 (Mar 2020), 015003. Issue 1. <https://doi.org/10.1103/RevModPhys.92.015003>
- Michael A. Nielsen and Isaac L. Chuang. 2000. *Quantum Computation and Quantum Information*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511976667>
- Adam Paetznick and Krysta M. Svore. 2014. Repeat-until-Success: Non-Deterministic Decomposition of Single-Qubit Unitaries. *Quantum Information & Computation* 14, 15–16 (2014), 1277–1301.
- Pavel Panchekha, Alex Sanchez-Stern, James R. Wilcox, and Zachary Tatlock. 2015. Automatically improving accuracy for floating point expressions. In *Proceedings of the 36th ACM SIGPLAN Conference on Programming Language Design and Implementation - PLDI 2015*. ACM Press. <https://doi.org/10.1145/2737924.2737959>
- Jennifer Paykin, Robert Rand, and Steve Zdancewic. 2017. QWIRE: A Core Language for Quantum Circuits. *SIGPLAN Not.* 52, 1, 846–858. <https://doi.org/10.1145/3093333.3009894>
- Stefano Pirandola, Ulrik L. Andersen, Leonardo Banchi, Mario Berta, Darius Bunandar, Roger Colbeck, Dirk R. Englund, Tobias Gehring, Cosmo Lupo, Carlo Ottaviani, Jason Pereira, Mohsen Razavi, Jesni S. Shaari, Marco Tomamichel, Vladyslav C. Usenko, Giuseppe Vallone, Paolo Villoresi, and Petros Wallden. 2020. Advances in Quantum Cryptography. *Advances in Optics and Photonics* (2020). <https://doi.org/10.1364/AOP.361502>

- David Poulin, Matthew B. Hastings, Dave Wecker, Nathan Wiebe, Andrew C. Doherty, and Matthias Troyer. 2015. The Trotter Step Size Required for Accurate Quantum Simulation of Quantum Chemistry. *Quantum Information & Computation* 15, 5–6 (2015), 361–384.
- Markus Reiher, Nathan Wiebe, Krysta M. Svore, Dave Wecker, and Matthias Troyer. 2017. Elucidating reaction mechanisms on quantum computers. *Proceedings of the National Academy of Sciences* 114, 29 (2017), 7555–7560. <https://doi.org/10.1073/pnas.1619152114>
- Neil J. Ross and Peter Selinger. 2016. Optimal ancilla-free Clifford+ $T$  approximation of  $z$ -rotations. *Quantum Information & Computation* 16, 11&12 (2016), 901–953.
- Artur Scherer, Benoît Valiron, Siun-Chuon Mau, Scott Alexander, Eric van den Berg, and Thomas E. Chapuran. 2017. Concrete resource analysis of the quantum linear-system algorithm used to compute the electromagnetic scattering cross section of a 2D target. *Quantum Information Processing* 16, 3 (Jan 2017). <https://doi.org/10.1007/s11128-016-1495-5>
- Peter W. Shor. 1994. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*. 124–134. <https://doi.org/10.1109/SFCS.1994.365700>
- Jonathan M. Smith, Neil J. Ross, Peter Selinger, and Benoît Valiron. 2014. Quipper: concrete resource estimation in quantum algorithms. arXiv:1412.0625
- Damian S. Steiger, Thomas Häner, and Matthias Troyer. 2018. ProjectQ: an open source software framework for quantum computing. *Quantum* 2 (Jan 2018), 49. <https://doi.org/10.22331/q-2018-01-31-49>
- Martin Suchara, John Kubitowicz, Arvin I. Faruque, Frederic T. Chong, Ching-Yi Lai, and Gerardo Paz. 2013. QuRE: The Quantum Resource Estimator toolbox. In *2013 IEEE 31st International Conference on Computer Design, ICCD 2013*. 419–426. <https://doi.org/10.1109/ICCD.2013.6657074>
- Krysta M. Svore, Alan Geller, Matthias Troyer, John Azariah, Christopher Granade, Bettina Heim, Vadym Kliuchnikov, Mariia Mykhailova, Andres Paz, and Martin Roetteler. 2018. Q#: Enabling Scalable Quantum Computing and Development with a High-level DSL. In *Real World Domain Specific Languages Workshop*. 7:1–7:10. <https://doi.org/10.1145/3183895.3183901>
- John Watrous. 2009. Semidefinite Programs for Completely Bounded Norms. *Theory of Computing* 5, 11 (2009), 217–238. <https://doi.org/10.4086/toc.2009.v005a011>
- Wolfram Research Inc. 2019. Mathematica, Version 12.0. <https://www.wolfram.com/mathematica> Champaign, IL.
- Mingsheng Ying. 2011. Floyd-hoare logic for quantum programs. *ACM Trans. Program. Lang. Syst.* 33, 6 (2011), 19:1–19:49. <https://doi.org/10.1145/2049706.2049708>