A Verified Optimizer for Quantum Circuits

Kesha Hietala    Robert Rand    Shih-Han Hung
Xiaodi Wu    Michael Hicks
University of Maryland, College Park, USA

Abstract
We present voqc, the first fully verified compiler for quantum circuits, written using the Coq proof assistant. Quantum circuits are expressed as programs in a simple, low-level language called squire, which is deeply embedded in Coq. Optimizations and other transformations are expressed as Coq functions, which are proved correct with respect to a semantics of squire programs. We evaluate voqc’s verified optimizations on a series of benchmarks, and it performs comparably to industrial-strength compilers. voqc’s optimizations reduce total gate counts on average by 17.7% on a benchmark of 29 circuit programs compared to a 10.7% reduction when using IBM’s Qiskit compiler.

Keywords Formal Verification, Quantum Computing, Optimization, Certified Compilation, Programming Languages

1 Introduction
Programming quantum computers will be challenging, at least in the near term. Qubits will be scarce, and gate pipelines will need to be short to prevent decoherence. Fortunately, optimizing compilers can transform a source algorithm to work with fewer resources. Where compilers fall short, programmers can optimize their algorithms by hand.

Of course, both compiler and by-hand optimizations will inevitably have bugs. For evidence of the former, consider that higher optimization levels of IBM’s Qiskit compiler are known to have mistakes (as is evident from its issue tracker\(^1\)). Kissinger and van de Wetering [24] discovered mistakes in the optimized outputs produced by the circuit compiler by Nam et al. [30]. And Nam et al. themselves found that the optimization library they compared against sometimes produced incorrect results. Making mistakes when optimizing by hand is also to be expected: as put well by Zamdzhiev [49], quantum computing can be frustratingly unintuitive.

Unfortunately, the very factors that motivate optimizing quantum compilers make it difficult to test their correctness. Comparing runs of a source program to those of its optimized version may be impractical due to the indeterminacy of typical quantum algorithms and the substantial expense involved in executing or simulating them. Indeed, resources may be too scarce, or the qubit connectivity too constrained, to run the program without optimization!\(^1\)

\(^1\)https://github.com/Qiskit/qiskit-terra/issues/2752

Draft paper, November 22, 2019,
on a particular target architecture. These transformations were reasonably straightforward to prove correct thanks to sqir’s design. (Sections 4 and 5)

We find that voqc performs comparably to unverified, state-of-the-art compilers when run on a benchmark of 29 circuit programs developed by Amy et al. [5]. These programs range from 45 and 61,629 gates and use between 5 and 192 qubits. voqc reduced total gate counts on average by 17.7% compared to 10.7% by IBM’s Qiskit compiler [2]. There is still room for improvement: Nam et al. [30] produced reductions of 26.5% using additional optimizations we expect we could verify. (Section 6.)

voqc is the first fully verified circuit optimizer for a realistic quantum circuit language. Amy et al. [6] developed a verified optimizing compiler from source Boolean expressions to reversible circuits, but did not handle general quantum programs. Rand et al. [34] developed a similar compiler for quantum circuits but without optimizations. Other low-level quantum languages [12, 42] have not been developed with verification in mind, and prior circuit-level optimizations [5, 20, 30] have not been formally verified. In concurrent work, Shi et al. [40] developed CertiQ, which uses symbolic execution and SMT solving to verify some circuit transformations in the Qiskit compiler. CertiQ is limited to verifying correct application of local equivalences, rather than more general circuit transformations, and sometimes verification fails in which case CertiQ must validate the optimized ("translated") circuits on-line. The PyZX compiler [24] likewise uses translation validation to check its rewrites of circuits using the equational theory of the ZX-Calculus [8]. (Section 7.)

Our work on voqc and sqir constitutes a step toward developing a full-scale verified compiler toolchain. Next steps include developing certified transformations from high-level quantum languages to sqir and implementing optimizations with different objectives, e.g., that aim to reduce the probability that a result is corrupted by quantum noise. All code we reference in this paper can be found online at https://github.com/inQWIRE/SQIRE.

2 Overview

We begin with a brief background on quantum programs, and then provide an overview of voqc and sqir.

2.1 Preliminaries

Quantum programs operate over quantum states, which consist of one or more quantum bits (aka, qubits). A single qubit is represented as a vector of complex numbers (α, β) such that |α|^2 + |β|^2 = 1. The vector (1, 0) represents the state |0⟩ while vector (0, 1) represents the state |1⟩. A state written |ψ⟩ is called a ket, following Dirac’s notation. We say a qubit is in a superposition of |0⟩ and |1⟩ when both α and β are non-zero. Just as Schrodinger’s cat is both dead and alive until the box is opened, a qubit is only in superposition until it is measured, at which point the outcome will be 0 with probability |α|^2 and 1 with probability |β|^2. Measurement is not passive: it has the effect of collapsing the state to match the measured outcome, i.e., either |0⟩ or |1⟩. As a result, all subsequent measurements return the same answer.

Operators on quantum states are linear mappings. These mappings can be expressed as matrices, and their application to a state expressed as matrix multiplication. For example, the Hadamard operator H is expressed as a matrix \[ \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \]. Applying H to state |0⟩ yields \( \left\{ \frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}} \right\} \), also written as |+⟩.

Many quantum operators are not only linear, they are also unitary—the conjugate transpose (or adjoint) of their matrix is its own inverse. This ensures that multiplying a qubit by the operator preserves the qubit’s sum of norms squared. Since a Hadamard is its own adjoint, it is also its own inverse: hence H|+⟩ = |0⟩.

A quantum state with N qubits is represented as vector of length 2^N. For example a 2-qubit state is represented as a vector (α, β, γ, δ) where each component corresponds to (the square root of) the probability of measuring |00⟩, |01⟩, |10⟩, and |11⟩, respectively. Because of the exponential size of the complex quantum state space, it is not possible to simulate a 200-qubit quantum computer using even the most powerful classical computer!

N-qubit operators are represented as 2^N × 2^N matrices. For example, the CNOT operator over two qubits is expressed as the matrix shown at the right. It expresses a controlled not operation—if the first qubit is |0⟩ then both qubits are mapped to themselves, but if the first qubit is |1⟩ then the second qubit is negated, e.g., CNOT|00⟩ = |00⟩ while CNOT|10⟩ = |11⟩.

N-qubit operators can be used to create entanglement, which is a situation where two qubits cannot be described independently. For example, while the vector (1, 0, 0) can be written as (1, 0) ⊗ (1, 0) where ⊗ is the tensor product, the state \( \frac{1}{\sqrt{2}}, 0, 0, \frac{-1}{\sqrt{2}} \) cannot be similarly decomposed. We say that \( \frac{1}{\sqrt{2}}, 0, 0, \frac{-1}{\sqrt{2}} \) is an entangled state.

An important non-unitary quantum operator is projection onto a subspace. For example, |0⟩⟨0| (in matrix notation \[ \frac{1}{\sqrt{2}} \]) projects a qubit onto the subspace where that qubit is in the |0⟩ state. Projections are useful for describing quantum states after measurement has been performed. We sometimes use |i⟩_q⟨i| as shorthand for applying the projection |i⟩⟨i| to qubit q and an identity operation to every other qubit in the state.

2.2 Quantum Circuits

Quantum programs are typically expressed as circuits; an example is shown in Figure 1(a). In these circuits, each horizontal wire represents a qubit and boxes on these wires indicate unitary quantum operators, i.e., gates. For multiple-qubit gates, the inputs are often distinguished as either
a target or a control. In software, these circuits are often represented using lists of instructions that describe the different gate applications. For example, Figure 1(b) is the Quil [42] representation of the circuit in Figure 1(a).

In the QRAM model [25] quantum computers are used as co-processors to classical computers. The classical computer generates descriptions of circuits to send to the quantum computer, and then processes the returned results. High-level quantum computing languages are designed to follow this model. For example, Figure 1(c) shows a program in PyQuil [36], a quantum language/framework embedded in Python. The ghz_state function takes an array of qubits and constructs a circuit that prepares the Greenberger-Horne-Zeilinger (GHZ) state [19], which is an \( n \)-qubit entangled quantum state of the form

\[
|\text{GHZ}^n\rangle = \frac{1}{\sqrt{2}} (|0\rangle^\otimes n + |1\rangle^\otimes n).
\]

Calling ghz_state([0, 1, 2]) would return the Quil program in Figure 1(b), which the quantum computer could subsequently execute. The high-level language may provide facilities to optimize constructed circuits, e.g., to reduce gate count, circuit depth, and qubit usage. It may also perform transformations to account for hardware-specific details like the number of qubits, available set of gates, or connectivity between physical qubits.

2.3 voqc: A Verified Optimizer for Quantum Circuits

The structure of voqc is summarized in Figure 2.

Source programs are written in sqir, whose syntax and semantics we give in Section 3. sqir is a simple circuit-oriented language deeply embedded in Coq, similar in style to PyQuil. sqir programs are given a formal semantics in Coq, which is the basis for proving properties about them. For example, we can prove that the GHZ program prepares the expected quantum state.

3 sqir: A Small Quantum Intermediate Representation

This section presents the syntax and semantics of sqir programs. We begin with the core of sqir, which describes unitary circuits. We then describe the expanded language, which allows measurement and initialization.
\[
\begin{align*}
\|U_1; U_2\|_d &= \|U_2\|_d \times \|U_1\|_d \\
\|G_1 \ q\|_d &= \begin{cases}
\text{apply}_y(G_1, \ q, \ d) & \text{well-typed} \\
0_d & \text{otherwise}
\end{cases} \\
\|G_2 \ q_1 \ q_2\|_d &= \begin{cases}
\text{apply}_y(G_2, \ q_1, \ q_2, \ d) & \text{well-typed} \\
0_d & \text{otherwise}
\end{cases}
\end{align*}
\]

Figure 3. Semantics of unitary \texttt{sqr} programs, assuming a global register of dimension \(d\). The \texttt{apply}_y function maps a gate name to its corresponding unitary matrix and extends the intended operation to the given dimension by applying an identity operation on every other qubit in the system. For example, \texttt{apply}_y(X, \ q, \ d) = I_{2q} \otimes \sigma_x \otimes I_{d(4-q-1)}\) where \(\sigma_x\) is the matrix interpretation of the X gate.

### 3.1 Unitary Core

\texttt{sqr} is a language for describing quantum programs that is deeply embedded in the Coq proof assistant. In \texttt{sqr}, a qubit is referred to by a natural number that indices into a global register of quantum bits. Unitary \texttt{sqr} programs allow sequencing and unitary gate application to one or two qubits, drawing from a fixed set of gates.

\[U := U_1; U_2 \mid G \ q \mid G \ q_1 \ q_2\]

Each \texttt{sqr} program is parameterized by a unitary gate set (from which \(G\) is drawn) and the dimension of the global register (i.e., the number of available qubits).

A unitary program \(U\) is well-typed if every gate application is valid. A gate application is valid if all of its arguments are in-bounds indices into the global register, and no index is repeated. This second requirement enforces linearity and thereby quantum mechanics’ no-cloning theorem.

The semantics of unitary \texttt{sqr} programs is shown in Figure 3. If a program is not well-typed, its denotation is the zero matrix. The advantage of this definition is that it allows us to reference the denotation of a program without explicitly assuming or (re)proving that the program is well-typed, thus removing clutter from theorems and proofs.

When a program is well-typed, computing its denotation requires a matrix interpretation for every unitary gate \(G\). In our development we define the semantics of \texttt{sqr} programs over the gate set \(\{R_{\theta, \phi, \lambda}, \ CNOT\}\) where \(R_{\theta, \phi, \lambda}\) is a general single-qubit rotation parameterized by three real-valued rotation angles and \(CNOT\) is the standard two-qubit controlled-not gate. We refer to this gate set as the base set.

It is the same as the underlying set used by OpenQASM [12] and is universal, meaning that it can approximate any unitary operation to within arbitrary error. The matrix interpretation of the single-qubit \(R_{\theta, \phi, \lambda}\) gate is

\[
\begin{bmatrix}
\cos(\theta/2) & -e^{i\lambda} \sin(\theta/2) \\
e^{i\phi} \sin(\theta/2) & e^{i(\phi+\lambda)} \cos(\theta/2)
\end{bmatrix}
\]

The semantics of unitary \texttt{sqr} programs to within arbitrary error. The matrix interpretation and is globally set to a controlled-not gate. We refer to this gate set as the base set. When a program is well-typed, computing its denotation requires a matrix interpretation for every unitary gate \(G\). In our development we define the semantics of \texttt{sqr} programs over the gate set \(\{R_{\theta, \phi, \lambda}, \ CNOT\}\) where \(R_{\theta, \phi, \lambda}\) is a general single-qubit rotation parameterized by three real-valued rotation angles and \(CNOT\) is the standard two-qubit controlled-not gate. We refer to this gate set as the base set.

It is the same as the underlying set used by OpenQASM [12] and is universal, meaning that it can approximate any unitary operation to within arbitrary error. The matrix interpretation of the single-qubit \(R_{\theta, \phi, \lambda}\) gate is

\[
\begin{bmatrix}
\cos(\theta/2) & -e^{i\lambda} \sin(\theta/2) \\
e^{i\phi} \sin(\theta/2) & e^{i(\phi+\lambda)} \cos(\theta/2)
\end{bmatrix}
\]

\texttt{Figure 4.} \texttt{sqr} density matrix semantics, assuming a global register of size \(d\).

and the matrix interpretation of the \(CNOT\) gate is given in Section 2.1.

Common single-qubit gates can be defined in terms of \(R_{\theta, \phi, \lambda}\). For example, the identity \(I\) is \(R_{0,0,0}\); the Hadamard \(H\) gate is \(R_{\pi/2,0,0}\); the Pauli \(X\) gate is \(R_{\pi,0,0}\) and the Pauli \(Z\) gate is \(R_{0,0,\pi}\). We can also define more complex operations as \texttt{sqr} programs. For example the SWAP operation, which swaps two qubits, is a sequence of three \(CNOT\) gates.

We say that two unitary programs are equivalent, written \(U_1 \equiv U_2\), if their denotation is the same, i.e., \(\|U_1\|_d = \|U_2\|_d\). For verifying equivalence of quantum programs, however, we will often want something more general since \(|\psi\rangle\) and \(e^{i\theta} |\psi\rangle\) (for \(\theta \in \mathbb{R}\)) represent the same physical state. We therefore say that two circuits are equivalent up to a global phase, written \(U_1 \equiv U_2\), when there exists a \(\theta\) such that \(\|U_1\|_d = e^{i\theta} \|U_2\|_d\).

### 3.2 Adding Measurement

To describe general quantum programs \(P\), we extend unitary \texttt{sqr} with a \texttt{branching measurement} operation.

\[
P := \text{skip} \mid P_1 \mid P_2 \mid U \mid \text{meas} \ q \ P_1 \ P_2
\]

The command \texttt{meas} \(q \ P_1 \ P_2\) (inspired by a similar construct in QPL [39]) measures the qubit \(q\) and either performs program \(P_1\) or \(P_2\) depending on the result. We define non-branching measurement and resetting a qubit to \(|0\rangle\) in terms of branching measurement:

\[
\text{measure} \ q = \text{meas} \ q \ \text{skip} \ \text{skip} \\
\text{reset} \ q = \text{meas} \ q \ (X \ q) \ \text{skip}
\]

Figure 4 defines the semantics of non-unitary programs in terms of density matrices, following the approach of several previous efforts [31, 48]. The density matrix semantics encodes different measurement outcomes as a probability distribution. We also provide a non-deterministic semantics in Appendix A.4, which is sometimes more convenient.

Since the density matrix semantics denotes programs as functions over matrices, we say that two programs \(P_1\) and \(P_2\) are equivalent if for every input \(\rho\), \(\|P_1\|_d(\rho) = \|P_2\|_d(\rho)\).

### 3.3 Example

Recall the GHZ preparation example from Section 2. The following Coq function \texttt{ghz} recursively constructs an \(n\)-qubit
\texttt{qir} program in the base set (i.e., a value of type \texttt{ucom base n}) that prepares the GHZ state. It is similar to the PyQuil program in Figure 1(c).

\texttt{Fixpoint GHZ \(n : \mathbb{N}\) : \texttt{ucom base n} :=
\begin{align*}
| 0 \Rightarrow & \ 1 \ 0 \\
| 1 \Rightarrow & \ H \ 0 \\
| S \ n' \Rightarrow & \ \text{GHZ} \ n'; \ \text{CNOT} \ (n'-1) \ n' \\
\end{align*}
end.

When \(n = 0\), the result is just the identity \(I\) on wire \(0\). When \(n = 1\), the result is the Hadamard gate applied to wire \(0\). (Here, \(I\) and \(H\) are the notations for \(R_\theta, \phi, \lambda\) gates presented in Section 3.1.) When \(n > 1\), it constructs the program \(U_1; U_2\), where \(U_1\) is the GHZ circuit on \(n-1\) qubits, and \(U_2\) is the appropriate \texttt{CNOT} gate. The result of \texttt{GHZ 3} is equivalent to the circuit shown in Figure 1(a).

### 3.4 Designing \texttt{qir}

\texttt{qir} was designed to be expressive while facilitating mechanically checked proofs of quantum programs. We conclude this section by discussing key features of \texttt{qir}’s design.

**Expressiveness** As a deeply embedded, domain-specific circuit language, \texttt{qir} is similar to PyQuil [36] and Cirq [46], both in terms of programming experience and expressive power. As such, it makes a reasonable source programming language. However, unlike these languages, we can use \texttt{qir}’s host language, Coq, to prove properties about its programs. For example, we can prove that every circuit generated by \texttt{GHZ n}, above, produces the corresponding state \(|\text{GHZ}^n\rangle\) given in Section 2.2 when applied to \(|0 \ldots 0\rangle\).

As a demonstration of \texttt{qir}’s expressiveness as a source language, Appendix A presents proof of this property along with other examples of \texttt{qir} programs and corresponding formal properties about them. In particular, we present quantum teleportation and proof that it indeed transports the intended qubit; superdense coding and proof that it prepares the expected state; and the Deutsch-Jozsa algorithm with proof that it correctly distinguishes between a constant and balanced Boolean oracle.

**Verification** \texttt{qir}’s design has two key features that facilitate verification. First, unlike most quantum languages, \texttt{qir} features a distinct core language of unitary operators; the full language adds measurement to this core. The semantics of a unitary program is expressed directly as a matrix, which means that proofs of correctness of unitary optimizations (the bulk of \texttt{voqc}) involve reasoning directly about matrices. Doing so is far simpler than reasoning about functions over density matrices, as is required for the full language.

Second, a \texttt{qir} program uses concrete (numeric) indices into a global register to refer to wires. As such, the semantics can simply map concrete indices to rows and columns in the denoted matrix. In addition, wire disjointness in the \texttt{qir} program is obvious—\(G_1 m\) operates on a different wire than \(G_2 n\) when \(m \neq n\). Both elements are important for easily proving equivalences, e.g., that gates acting on disjoint qubits commute (a property that lets us reason about gates acting on different parts of the circuit in isolation).

The alternative, used by Quipper [17] and QWire [31], is to use variables to refer to abstract wires, which are later allocated to concrete wires. These languages are embedded in Haskell and Coq, respectively, and take advantage of their host languages’ variable bindings. This approach eases programmability—larger circuits can be built by composing smaller ones, connecting abstract outputs to abstract inputs. However, we find that this approach complicates formal proof. The semantics of a program that uses abstract wires must convert those wires into concrete indices into the denoted matrix. Reasoning about this conversion can be laborious, especially for recursive circuits and those that allocate and deallocate wires (entailing de Bruijn-style index shifting [33]). Moreover, notions like disjointness are no longer obvious—\(G_1 x\) and \(G_2 y\) for variables \(x \neq y\) may not be disjoint if \(x\) and \(y\) could be allocated to the same concrete wire. Appendix B has more details.

### 4 Optimizing Unitary \texttt{qir} Programs

The \texttt{voqc} compiler takes as input a \texttt{qir} program and attempts to reduce its total gate count by applying a series of optimizations. This section describes optimizations on unitary \texttt{qir} programs. The next section discusses how \texttt{voqc} optimizes full \texttt{qir} programs and maps them to a connectivity-constrained architecture.

#### 4.1 Overview

\texttt{voqc}’s unitary optimizations are defined as Coq functions that map an input program to an optimized one. A program is represented as a list of gate applications. Sequences of gate applications are flattened so that a \texttt{qir} program like \(\langle G_1 p; G_2 q \rangle\); \(G_3 r\) is represented as the Coq list \([G_1 p; G_2 q; G_3 r]\). This representation simplifies finding patterns of gates.

The optimization functions expect a program’s gates \(G\) to be drawn from the set \(\{H, X, R_{\pi/4}, \text{CNOT}\}\) where \(R_{\pi/4}(k)\) describes rotation about the \(z\)-axis by \(k \cdot \pi/4\) for \(k \in \mathbb{Z}\). Either the parser must produce input programs using this gate set, or \texttt{voqc} must convert the program to use it before optimizations can be applied. This gate set is universal, and is consistent with previous circuit optimizers, e.g., Amy et al. [5]. Rotations are not parameterized by arbitrary reals, which would make verification unsound if these were extracted to OCaml floating point numbers (which are used in the gate sets used by Nam et al. [30] and Qiskit [2]). It would be easy to support \(R_{\pi/2^n}\) for higher \(n\) if finer-grained rotations are needed.

Programs in the list representation are deemed equivalent if their back-converted \texttt{qir} programs are equivalent, per
the definition in Section 3.1. Conversion translates voqc’s gates $H$, $X$, and $R_{\pi/4}(k)$ into base gates $R_{\pi/2,0,\pi}$, $R_{\pi,0,\pi}$, and $R_{0,0,k,\pi/4}$, respectively. (CNOT translates to itself.)

Most of voqc’s optimizations are inspired by the state-of-the-art circuit optimizer by Nam et al. [30]. There are two basic kinds of optimizations: replacement and propagation and cancellation. The former simply identifies a pattern of gates and replaces it by an equivalent pattern. The latter works by commuting sets of gates when doing so produces an equivalent quantum program—often with the effect of “propagating” a particular gate rightward in the program—until two adjacent gates can be removed because they cancel each other out.

### 4.2 Proving Circuit Equivalences

All of voqc’s optimizations use circuit equivalences to justify local rewrites. Proof that an optimization is correct thus relies on proofs that the circuit equivalences it uses are correct. Many of our circuit equivalence proofs have a common form, which we illustrate by example.

Suppose we wish to prove the equivalence

$$X n \equiv CNOT m n \equiv CNOT m n; X n$$

for arbitrary $n$, $m$, and dimension $d$. Applying our definition of equivalence, this amounts to proving

$$apply_1(X, n, d) \times apply_2(CNOT, m, n, d) =$$

$$apply_2(CNOT, m, n, d) \times apply_1(X, n, d),$$

per Figure 3. Suppose both sides of the equation are well typed ($m < d$ and $n < d$ and $m \neq n$), and consider the case $m < n$ (the $n < m$ case is similar). We expand $apply_1$ and $apply_2$ as follows with $p = n - m - 1$ and $q = d - n - 1$:

$$apply_1(X, n, d) = I_2^p \otimes \sigma_x \otimes I_2^q$$

$$apply_2(CNOT, m, n, d) = I_2^m \otimes |1\rangle \otimes I_2^p \otimes \sigma_x \otimes I_2^q$$

$$+ I_2^m \otimes |0\rangle \otimes I_2^p \otimes I_2^q \otimes I_2^q$$

Here, $\sigma_x$ is the matrix interpretation of the $X$ gate and $|1\rangle \otimes |1\rangle \otimes \sigma_x + |0\rangle \otimes I_2$ is the matrix interpretation of the CNOT gate (in Dirac notation). We complete the proof of equivalence by normalizing and simplifying each side of Equation (1), and showing both sides to be the same.

**Automation** Matrix normalization and simplification are almost entirely automated in voqc. We wrote a Coq tactic called gridify for proving general equivalences correct. Rather than assuming $m < n < d$ as above, the gridify tactic does case analysis, immediately solving all cases where the circuit is ill-typed (e.g., $m = n$ or $d \leq m$) and thus has the zero matrix as its denotation. In the remaining cases ($m < n$ and $n < m$ above), it puts the expressions into their “grid normal” form and applies a set of matrix identities.

In grid normal form, each arithmetic expression has addition on the outside, followed by tensor product, with multiplication on the inside, i.e., $((\ldots x \ldots) \otimes (\ldots x \ldots)) + ((\ldots x \ldots) \otimes (\ldots x \ldots)).$

The gridify tactic rewrites an expression into this form by using the following rules of matrix arithmetic (where all the dimensions are appropriate):

- $I_m = I_m \otimes I_n$
- $A \times (B + C) = A \times B + A \times C$
- $(A + B) \times C = A \times B + B \times C$
- $A \otimes (B + C) = A \otimes B + A \otimes C$
- $(A + B) \otimes C = A \otimes B + B \otimes C$
- $(A \otimes B) \times (C \otimes D) = (A \times B) \otimes (C \times D)$

The first rule is applied to facilitate application of the other rules. (For instance, in the example above, $I_{2n}$ would be replaced by $I_{2m} \otimes I_2 \otimes I_{2p}$ to match the structure of the apply$_2$ term.) After expressions are in grid normal form, gridify simplifies them by removing multiplication by the identity matrix and rewriting simple matrix products (e.g. $\sigma_x \sigma_x = I_2$).

In our example, after normalization and simplification by gridify, both sides of the equality in Equation (1) become

$$I_{2m} \otimes |1\rangle \otimes I_{2p} \otimes I_2 \otimes I_{2q} + I_{2m} \otimes |0\rangle \otimes I_{2p} \otimes I_2 \otimes I_{2q}$$

proving that the two expressions are equal.

We use gridify to verify most of the equivalences used in the optimizations given in Sections 4.3 and 4.4. The tactic is most effective when equivalences are small: The equivalences used in gate cancellation and Hadamard reduction apply to patterns of at most five gates on up to three qubits. For equivalences over larger, non–concrete circuits like the one used in rotation merging, we do not use gridify directly, but still rely on our automation for matrix simplification.

### 4.3 Optimization by Propagation and Cancellation

Our propagate-cancel optimizations have two steps. First we localize a set of gates by repeatedly applying commutation rules. Then we apply a circuit equivalence to replace that set of gates. In voqc most optimizations of this form use a library of code patterns, but one—not propagation—is different, so we discuss it first.

**Not Propagation** The goal of not propagation is to remove cancelling $X$ (‘not”) gates. Two $X$ gates cancel when they are adjacent or they are separated by a circuit that commutes with $X$. We find $X$ gates separated by commuting circuits by repeatedly applying the propagation rules in Figure 5. These
rules switch between propagating $X$ and $Z (= R_{\pi/4}(4))$ gates. In particular, an $X$ gate can “propagate” past an $H$ gate by becoming a $Z$, and likewise a $Z$ can propagate past an $H$ by becoming an $X$. An example application of the not propagation algorithm is shown in Figure 6.

This implementation may introduce extra $X$ and $Z$ gates at the end of a circuit. However, redundant $Z$ gates will be removed by the single-qubit gate cancellation optimization, and moving $X$ gates to the end of a circuit makes the rotation merging optimization more likely to succeed.

We note that our version of this optimization is more general than Nam et al.’s, which is specialized to a three-qubit TOFF gate. The TOFF gate can be decomposed into a $\{H, R_{\pi/4}, CNOT\}$ program, and Nam et al.’s propagation rules can be written in terms of the rules in Figure 5.

**Gate Cancellation**  The single- and two-qubit gate cancellation optimizations rely on the same propagate-cancel pattern used in not propagation, except that gates are returned to their original location if they fail to cancel. To support this pattern, we provide a general propagate function in voqc. This function takes as input (i) an instruction list, (ii) a gate to propagate, and (iii) a set of rules for commuting and cancelling that gate. At each iteration, propagate performs the following actions:

1. Check if a cancellation rule applies. If so, apply that rule and return the modified list.
2. Check if a commutation rule applies. If so, commute the gate and recursively call propagate on the remainder of the list.
3. Otherwise, return the gate to its original position.

We have verified that our propagate function is sound when provided with valid commutation and cancellation rules.

Each commutation or cancellation rule is implemented as a partial Coq function from an input circuit to an output circuit. A common pattern in these rules is to identify one gate (e.g., an $X$ gate), and then to look for an adjacent gate it might commute with (e.g., $CNOT$) or cancel with (e.g., $X$). For commutation rules, we use the rewrite rules shown Figure 7. For cancellation rules, we use the fact that $H$, $X$, and $CNOT$ are all self-cancelling and $R_{\pi/4}(k)$ and $R_{\pi/4}(k’)$ combine to become $R_{\pi/4}(k + k’)$.

**4.4 Circuit Replacement**

We have implemented two optimizations—Hadamard reduction and rotation merging—that work by replacing one pattern of gates with an equivalent one; no preliminary propagation is necessary. These aim either to reduce the gate count directly, or to set the stage for additional optimizations.

**Hadamard Reduction**  The Hadamard reduction routine employs the equivalences shown in Figure 8 to reduce the number of $H$ gates in the program. Removing $H$ gates is useful because $H$ gates limit the size of the $\{X, R_{\pi/4}, CNOT\}$ sub-circuits used in the rotation merging optimization.
**Rotation Merging**

The rotation merging optimization allows for combining \( Rz_{\pi/4} \) gates that are not physically adjacent in the circuit. This optimization is more sophisticated than the previous optimizations because it does not rely on small structural patterns (e.g., that adjacent X gates cancel), but rather on more general (and non-local) circuit behavior.

The argument for the correctness of this optimization relies on the phase polynomial representation of a circuit. Let \( C \) be a circuit consisting of \( X \) gates, CNOT gates, and rotations about the \( z \)-axis. Then on basis state \( |x_1, \ldots, x_n \rangle \), \( C \) will produce the state

\[
e^{i p(x_1, \ldots, x_n)} |h(x_1, \ldots, x_n)\rangle
\]

where \( h : \{0,1\}^n \rightarrow \{0,1\}^n \) is an affine reversible function and

\[
p(x_1, \ldots, x_n) = \sum_{i=1}^l (\theta_i \mod 2\pi) f_i(x_1, \ldots, x_n)
\]

is a linear combination of affine boolean functions. \( p(x_1, \ldots, x_n) \) is called the phase polynomial of circuit \( C \). Each rotation gate in the circuit is associated with one term of the sum and if two terms of the phase polynomial satisfy \( f_i(x_1, \ldots, x_n) = f_j(x_1, \ldots, x_n) \) for some \( i \neq j \), then the corresponding \( i \) and \( j \) rotations can be merged.

As an example, consider the two circuits shown below.

\[
\begin{array}{c}
R_{\pi/4}(k') \\
\end{array}
\begin{array}{c}
\Rightarrow \\
\end{array}
\begin{array}{c}
R_{\pi/4}(k + k') \\
\end{array}
\]

To prove that these circuits are equivalent, we can consider their behavior on basis state \( |x_1, x_2\rangle \). Recall that applying \( R_{\pi/4}(k) \) to the basis state \( |x\rangle \) produces the state \( e^{i(k\pi/4)x} |x\rangle \) and CNOT \( |x,y\rangle \) produces the state \( |x, x \oplus y\rangle \) where \( \oplus \) is the xor operation. Evaluation of the left-hand circuit proceeds as follows:

\[
|x_1, x_2\rangle \rightarrow e^{i(k\pi/4)x_1} |x_1, x_2\rangle \\
\rightarrow e^{i(k\pi/4)x_1} |x_1, x_1 \oplus x_2\rangle \\
\rightarrow e^{i(k\pi/4)x_2} |x_2, x_1 \oplus x_2\rangle \\
\rightarrow e^{i(k\pi/4)x_1} e^{i(\pi/4)x_2} |x_2, x_1 \oplus x_2\rangle.
\]

Whereas evaluation of the right-hand circuit produces

\[
|x_1, x_2\rangle \rightarrow |x_1, x_1 \oplus x_2\rangle \\
\rightarrow |x_2, x_1 \oplus x_2\rangle \\
\rightarrow e^{i((k+k')\pi/4)x_2} |x_2, x_1 \oplus x_2\rangle.
\]

The two resulting states are equal because \( e^{i(k\pi/4)x_2} e^{i(\pi/4)x_2} = e^{i((k+k')\pi/4)x_2} \). This implies that the unitary matrices corresponding to the two circuits are the same. We can therefore replace the circuit on the left with the one on the right, removing one gate from the circuit.

Our rotation merging optimization follows the logic above for arbitrary \( \{X, R_{\pi/4}, \text{CNOT}\} \) circuits. For every gate in the program, it tracks the Boolean function associated with every qubit (the Boolean functions above are \( x_1, x_2, x_1 \oplus x_2 \), and merges rotations \( Rz_{\pi/4}(k) \) when they are applied to qubits associated with the same Boolean function. To prove equivalence over \( \{X, R_{\pi/4}, \text{CNOT}\} \) circuits, we show that the original and optimized circuits produce the same output on every basis state. We have found evaluating behavior on basis states is useful for proving equivalences that are not as direct as those listed in Figures 7 and 8.

Although our merge operation is identical to Nam et al.’s, we apply it to smaller circuits. For ease of verification, we only consider continuous \( \{X, R_{\pi/4}, \text{CNOT}\} \) sub-circuits within the larger program. Nam et al. allow some intervening \( H \) gates, provided that those \( H \) gates do not impact computation of the phase polynomial. This is a restriction that we plan to relax.

### 4.5 Scheduling

voqc applies all of these optimizations with its optimize function. It applies them one after the other, in the following order (due to Nam et al. [30]):

\[
0, 1, 3, 2, 3, 1, 2, 4, 3, 2
\]

where 0 is not propagation, 1 is Hadamard reduction, 2 is single-qubit gate cancellation, 3 is two-qubit gate cancellation, and 4 is rotation merging. The rationale for this ordering is that removing \( X \) and \( H \) gates \((0, 1)\) allows for more effective application of the gate cancellation (2,3) and rotation merging (4) optimizations. In our experiments (Section 6), we observed that single-qubit gate cancellation and rotation merging were the most effective at reducing gate count.

### 5 Other Verified Transformations

We have also implemented verified optimizations of non-unary programs in voqc (inspired by optimizations in IBM’s Qiskit compiler [2]) and verified a transformation that maps a circuit to a connectivity-constrained architecture.

#### 5.1 Non-unitary Optimizations

We have implemented two non-unitary optimizations: removing pre-measurement \( z \) rotations, and classical state propagation. For these optimizations, a non-unitary program \( P \) is represented as a list of blocks. A block is a binary tree, where a leaf is unitary program (in list form), and a node is a measurement \( m \langle q \ P_1 \ P_2 \rangle \) whose children \( P_1 \) and \( P_2 \) are lists of blocks. Equivalence is defined in terms of the density matrix semantics of the \( \text{sqin} \) representation (per Section 3.2).

**z-rotations Before Measurement**

\( z \)-axis rotations (or, more generally, diagonal unitary operations) before a measurement will have no effect on the measurement outcome, so they can safely be removed from the program. We have implemented and verified an optimization that locates \( Rz_{\pi/4} \) gates before measurement operations and removes them. This
 optimization was inspired by the RemoveDiagonalGatesBeforeMeasure pass implemented in Qiskit.

**Classical State Propagation** Once a qubit has been measured, the subsequent branch taken provides information about the qubit’s (now classical) state, which may allow pre-computation of some values. For example, in the branch where qubit q has been measured to be in the $|0\rangle$ state, any CNOT with q as the control will be a no-op and any subsequent measurements of q will still produce zero.

In detail, given a qubit $q$ in classical state $|i\rangle$, our analysis applies the following rules:

- $R_{\pi/4}(k)q$ preserves the classical state of $q$.
- $Xq$ flips the classical state of $q$.
- If $i = 0$ then CNOT $q q'$ is removed, and if $i = 1$ then CNOT $q q'$ becomes $Xq'$.
- $\text{meas} P_0 P_1$ becomes $P_i$.
- $Hq$ and CNOT $q'q$ destroy the classical state and terminate analysis.

Our statement of correctness for one round of propagation says that if qubit $q$ is in a classical state in the input, then the optimized program will have the same denotation as the original program. We express the requirement that qubit $q$ be in classical state $i \in \{0, 1\}$ with the condition

$$|i\rangle_q (i|\rho| \times |i\rangle_q |i\rangle = \rho, $$

which says that projecting state $\rho$ onto the subspace where $q$ is in state $|i\rangle$ results in no loss of information.

This optimization is not implemented directly in Qiskit, but Qiskit contains passes that have a similar effect. For example, the RemoveResetInZeroState pass removes adjacent reset gates, as the second has no effect.

### 5.2 Circuit Mapping

Similar to how optimization aims to reduce qubit and gate usage to make programs more feasible to run on near-term machines, circuit mapping aims to address the connectivity constraints of near-term machines [38, 50]. Circuit mapping algorithms take as input an arbitrary circuit and output a circuit that respects the connectivity constraints of some underlying architecture.

For example, consider the connectivity of IBM’s five-qubit Tenerife machine, shown in Figure 9. This is a representative example of a modern superconducting qubit system, where qubits are laid out in a 2-dimensional grid and possible interactions are described by directed edges between the qubits. The direction of the edge indicates which qubit can be the control of a two-qubit gate and which can be the target. Thus, no two-qubit gate is possible between physical qubits Q4 and Q1 on the Tenerife. A CNOT gate may be applied with Q4 as the control and Q2 as the target, but not the reverse.

We have implemented a simple circuit mapper for sqir programs and verified that it is sound and produces programs that satisfy the relevant hardware constraints. Our circuit mapper is parameterized by functions describing the connectivity of an architecture (in particular, one function that determines whether an edge is in the connectivity graph and another function that finds an undirected path between any two nodes). We map a program to this architecture by adding SWAP operations before and after every CNOT so that the target and control are adjacent when the CNOT is performed, and are returned to their original positions before the next operation. This algorithm inserts more SWAPs than the optimal solution, but our verification framework could be applied to optimized implementations as well. To handle directionality of edges in the connectivity graph, we make use of the equivalence $H a; H b; $ CNOT a b; $ H a; H b \equiv \text{CNOT} b a.$

We have implemented and verified mapping functions for the Tenerife architecture pictured in Figure 9 and the linear nearest neighbor (LNN), LNN ring, and 2D nearest neighbor architectures pictured in Figure 10.

### 6 Experimental Evaluation

We evaluate the performance of voqc’s verified optimizations against IBM’s Qiskit transpiler [2] and the optimizer presented in Nam et al. [30]. We find that voqc’s performance is comparable to these state-of-the-art compilers.

**Benchmarks** We run on a set of benchmarks developed by Amy et al. [3] and evaluate performance by measuring the reduction in gate counts. The benchmarks consist of arithmetic circuits and implementations of multiple-control Toffoli gates. Each circuit contains between 45 and 61,629 gates and uses between 5 and 192 qubits. These benchmarks
only contain unitary circuits, so they only serve to evaluate our unitary circuit optimizations. Some of the benchmarks contain the three-qubit \(\text{TOFF}\) gate. Before applying optimizations we convert \(\text{TOFF}\) gates to \text{voqc}'s gate set using the following standard decomposition, where \(R_{\pi/4}(1)\) is the familiar \(T\) gate and \(R_{\pi/4}(7)\) is its inverse \(T^\dagger\).

\[
\text{TOFF} = a \ b \ c \ := \\
\begin{bmatrix}
H \ c ; \ \\
\text{CNOT} \ b \ c ; \ \\
R_{\pi/4}(1) \ c ; \ \\
\text{CNOT} \ a \ c ; \\
R_{\pi/4}(7) \ c ; \ \\
\text{CNOT} \ a \ b ; \\
R_{\pi/4}(1) \ a ; \ \\
R_{\pi/4}(1) \ b ; \ \\
R_{\pi/4}(1) \ c ; \ \\
H \ c.
\end{bmatrix}
\]

**Baseline** We compare \text{voqc}'s performance with that of Nam et al. and Qiskit version 0.13.0 when run on the same programs. We do not include the results from Amy et al. because their optimization is aimed at reducing a particular type of gate, and often results in a higher total count.

As mentioned in Section 4, \text{voqc}'s verified optimizations are inspired by (unverified) ones implemented by Nam et al. [30], and these overlap with (also unverified) optimizations present in IBM's Qiskit compiler [2]. As such, our goal is not to beat the performance of these compilers, but rather to show that \text{voqc}'s suite of verified optimizations can achieve performance comparable to the state of the art.

Table 1 performs a direct comparison of functionality. For the Qiskit optimizations, \(L_i\) indicates that a routine is used by optimization level \(i\). For Nam et al., \(P\) stands for "preprocessing" and \(L\) and \(H\) indicate whether the routine is in the "light" or "heavy" versions of the optimizer. \text{voqc} provides the complete and verified functionality of the routines marked with \(\checkmark\); we write \(\checkmark^*\) to indicate that \text{voqc} contains a verified optimization with similar, although not identical, behavior. Compared to Nam et al.'s rotation merging, \text{voqc} performs a less powerful optimization (as discussed in Section 4.4). Conversely, \text{voqc}'s not propagation routine generalizes Nam et al.'s; and \text{voqc}'s one- and two-qubit cancellation routines generalize Qiskit's \text{Optimize1qGates} and \text{CXCancellation} when using \text{voqc}’s gate set. For \text{CommutativeCancellation}, Qiskit’s routine follows the same pattern as our gate cancellation routines, but uses matrix multiplication to determine whether gates commute while we use a rule-based approach; neither is strictly more effective than the other.

In our experiment we evaluate two settings of Qiskit—\textit{Qiskit A} and \textit{Qiskit B}. The latter is Qiskit with all of its unitary optimizations enabled (i.e., up to optimization level 3). However, the circuits produced by \textit{Qiskit B} are not guaranteed to be (and are most likely not) in \text{voqc}'s gate set, so the two are not entirely comparable. \textit{Qiskit B} uses the gate set \(\{u_1, u_2, u_3, \text{CNOT}\}\) where \(u_3\) is \(R_{\phi, \theta, \lambda}\) from \text{voqc}'s base set and \(u_1\) and \(u_2\) are \(u_3\) with certain arguments fixed.

\textit{Qiskit A} is more similar to \text{voqc} in that it uses the gate set \(\{H, X, u_1, \text{CNOT}\}\) where \(u_1\) corresponds to rotation about the \(z\)-axis by an arbitrary angle (which ends up as a multiple of \(\pi/4\) in our benchmark). \textit{Qiskit A} includes all unitary program optimizations used up to optimization level 2; it does not include optimizations from level 3 because these optimizations produce circuits with gates outside \text{voqc}'s gate set.

**Results** The results are shown in Table 2. In each row, we have marked in bold the gate count of the best-performing optimizer. The average gate count reduction for each optimizer is given in the last row, although performance varies substantially between benchmarks.

On average, Nam et al. [30] heavy optimization reduces the total gate count by 26.5%, \textit{Qiskit A} reduces the total gate count by 4.6%, \textit{Qiskit B} reduces the total gate count by 10.7%, and \text{voqc} reduces the total gate count by 17.7%. \text{voqc} outperforms \textit{Qiskit A} on all benchmarks and outperforms or matches the performance of \textit{Qiskit B} on all benchmarks except two. In 8 out of 29 cases \text{voqc} outperforms Nam et al.'s heavy optimization.

The gap in performance between \text{voqc} and Qiskit is primarily due to \text{voqc}'s rotation merging optimization, which has no analogue in Qiskit. The gap in performance between Nam et al. and \text{voqc} is due in part to the fact that we have not yet implemented all their optimization passes (per Table 1). We see no fundamental difficulties in implementing these, but we expect the biggest performance boost will come from generalizing our rotation merging optimization to consider larger sub-circuits, which will require some additional verification effort.

These results are encouraging evidence that \text{voqc} supports useful and interesting verified optimizations.

## 7 Related Work

**Verified Quantum Programming** We designed \texttt{sqir} primarily as the intermediate language for \text{voqc}'s verified optimizations, but we find it adequate for verified source programming as well (per Section 3.4 and Appendix A).
Several lines of work have explored formally verifying aspects of a quantum computation. The earliest attempts to do so in a proof assistant were Green et al. [18] implementation of the Quantum IO Monad and a small Coq quantum library by Boender et al. [7]. These were both proofs of concept, and neither developed beyond verifying basic protocols.

The high-level Qwire programming language, like sqir, embedded in the Coq proof assistant, and has been used to verify a variety of simple programs, assertions regarding ancilla qubits, and its own meta-theory. VOQC and Sqir reuse parts of Qwire's Coq development, and take-inspiration and lessons from its design. However, as discussed in Section 3.4 and Appendix B, Qwire's higher-level abstractions complicate verification. Moreover, such abstractions do not reflect the kind of quantum programming we can expect to do in the near future. For example, a key element of Qwire is dynamic lifting, which permits measuring a qubit and using the result as a Boolean value in the host language to compute the remainder of a circuit [17]. Today's quantum computers cannot reliably exchange information between a (typically supercooled) quantum chip and a classical computer before qubits decohere. Thus, practically-minded languages like IBM's OpenQASM [12] only allow for a limited form of branching that is close to sqir's.

Another line of work, pioneered by D'Hondt and Panangaden [15] and Ying [48], uses program logics to reason about quantum programs. These logics allow proof of a variety of program properties inside a formal deductive system. Liu et al. [27] implemented Ying's quantum Hoare logic inside the Isabelle proof assistant and used it to prove the correctness of Grover's algorithm, and Unruh [47] developed a relational quantum Hoare logic and built an Isabelle-based tool to prove the security of quantum cryptosystems. Implementing these kinds of logics in Coq and proving them correct with respect to sqir's denotational semantics may prove useful (though we have proved several interesting sqir programs correct directly).

**Verified Quantum Compilation** Quantum compilation is an active area. In addition to Qiskit and Nam et al. [30]
(discussed in Section 6), other recent compiler efforts include \(\ddagger\)ket \([10, 11]\), quilc \([37]\), ScaffCC \([22]\), and Project Q \([44]\). Due to resource limits on near-term quantum computers, most compilers for quantum programs contain some degree of optimization, and nearly all place an emphasis on satisfying architectural requirements, like mapping to a particular gate set or qubit topology. None of the optimization or mapping code in these compilers is formally verified.

However, voqc is not the only quantum compiler to which automated reasoning or formal verification has been applied. Amy et al. \([6]\) developed a certified optimizing compiler from source Boolean expressions to reversible circuits, but did not handle general quantum programs. Rand et al. \([34]\) developed a similar compiler for quantum circuits but without optimizations (using the Qwire language).

The problem of quantum program optimization verification has previously been considered in the context of the ZX-calculus \([8]\), which is a formalism for describing quantum computation based on categorical quantum mechanics \([1]\). The ZX-calculus is characterized by a small set of rewrite rules that allow translation of a diagram to any other diagram representing the same computation. The Quantomatic tool \([16]\) automatically rewrites ZX diagrams according to this set of rules, but suffers from two limitations as a quantum compiler: Its rewriting procedures are not guaranteed to terminate and not every ZX diagram corresponds to a valid quantum circuit. PyZX \([24]\) addresses both of these limitations, using the ZX-calculus as an intermediate representation for compiling quantum circuits, and generally achieving performance comparable to leading compilers. While PyZX is not verified in a proof assistant like Coq (the “Py” stands for Python), it does rely on a small, well-studied equational theory. Additionally, PyZX performs translation validation on its compiled circuits, checking (where feasible) that the compiled circuit is equivalent to the original.

A recent paper from Smith and Thornton \([41]\) presents a compiler with built-in translation validation via QMDD equivalence checking \([29]\). However, the optimizations they consider are much simpler than ours and the QMDD approach scales poorly with increasing number of qubits. Our optimizations are all verified for arbitrary dimension.

Concurrently with our work, Shi et al. \([40]\) developed CertiQ, an approach to verifying properties of circuit transformations in the Qiskit compiler, which is implemented in Python. Their approach has two steps. First, it uses matrix multiplication to check that the unitary semantics of two concrete gate patterns are equivalent. Second, it uses symbolic execution to generate verification conditions for parts of Qiskit that manipulate circuits. These are given to an SMT solver to verify that pattern equivalences are applied correctly according to programmer-provided function specifications and invariants. That CertiQ can analyze Python code directly in a mostly automated fashion is appealing. However, it is limited in the optimizations it can verify. For example, equivalences that range over arbitrary indices, like \(\text{CNOT} m x = \text{CNOT} n x \equiv \text{CNOT} n x\); \(\text{CNOT} n x\) cannot be verified by matrix multiplication; CertiQ checks a concrete instance of this pattern and then applies it to more general circuits. More complex optimizations like rotation merging (the most powerful optimization in our experiments) cannot be generalized from simple, concrete circuits. CertiQ can also fail to prove an optimization correct, e.g., because of complicated control code; in this case it falls back to translation validation, which adds extra cost and the possibility of failure at run-time. By contrast, every optimization in voqc has been proved correct.

8 Conclusions and Future Work

This paper has presented voqc, the first verified optimizer for quantum circuits implemented within a proof assistant. A key component of voqc is qir, a simple, low-level quantum language deeply embedded in the the Coq proof assistant. Compiler passes are expressed as Coq functions which are proved to preserve the semantics of their input qir programs. voqc’s optimizations are mostly based on local circuit equivalences, implemented by replacing one pattern of gates with another, or commuting a gate rightward until it can be cancelled. Others, like rotation merging, are more complex. These were inspired by, and in some cases generalize, optimizations in industrial compilers, but in voqc are proved correct. When applied to a benchmark suite of 29 circuit programs, we found voqc performed comparably to state of the art compilers, reducing gate counts on average by 17.7% compared to 10.7% for IBM’s Qiskit compiler, and 26.5% for the cutting-edge research compiler of Nam et al. \([30]\).

Moving forward, we plan to incorporate voqc into a full-featured verified compilation stack for quantum programs, following the vision of a recent Computing Community Consortium report \([28]\). We can implement validated parsers \([23]\) for languages like OpenQASM and verify their translation to qir (e.g., using metaQASM’s semantics \([4]\)). We can also add support for hardware-specific transformations that compile to a particular gate set. Indeed, most of the sophisticated code in Qiskit is devoted to efficiently mapping programs to IBM’s architecture, and IBM’s 2018 Developer Challenge centered around designing new circuit mapping algorithms \([43]\). We leave it as future work to incorporate optimizations and mapping algorithms from additional compilers into voqc. Our experience so far makes us optimistic about the prospects for doing so successfully.

Acknowledgments

We thank Leonidas Lampropoulos and Kartik Singhal for comments on drafts of this paper. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research,
Quantum Testbed Pathfinder Program under Award Number DE-SC0019040.

References


with continuous parameters. npj Quantum Information 4, 1 (2018), 23. https://doi.org/10.1038/s41534-018-0072-4


A Verified Optimizer for Quantum Circuits

The Greenberger-Horne-Zeilinger (GHZ) state \([19]\) is an \(n\)-qubit entangled quantum state of the form
\[
|GHZ^n\rangle = \frac{1}{\sqrt{2}}(|0\rangle \otimes \cdots \otimes |0\rangle + |1\rangle \otimes \cdots \otimes |1\rangle).
\]

This vector can be defined in Coq as follows:

Definition ghz (n : \(\mathbb{N}\)) : Matrix (2 ^ n) 1 :=
match n with
| 0 => I 1
| \(\leq\) n' => \(\frac{1}{\sqrt{2}}\) * nket n \(|0\rangle + \frac{1}{\sqrt{2}}\) * nket n \(|1\rangle\)
end.

Above, nket \(n\) \(|i\rangle\) is the tensor product of \(n\) copies of the basis vector \(|i\rangle\). The GHZ state can be prepared by a circuit that begins with all qubits initialized to the \(|0\rangle\) state, applies an \(H\) to the first qubit (yielding \(|+\rangle\)), and then sequentially applies a \(CNOT\) from each qubit to the next. A circuit that prepares the 3-qubit GHZ state is shown in Figure 1(a) and the \(\text{qir}\) description of this circuit can be produced by the recursive function in Figure 1(d).

The function \(\text{ghz}\) describes a family of \(\text{qir}\) circuits: For every \(n\), \(GHZ\ n\) is a valid \(\text{qir}\) program and quantum circuit.\(^{4}\)

We aim to show via an inductive proof that every circuit \(\text{ghz}\ n\) is well-typed and sound.

\(^{4}\)For the sake of readability, Figure 1(d) elides the coercion from \(\text{ucom base} n\) to \(\text{ucom base} n'\) in the recursive case.

A.1 Superdense Coding

Superdense coding is a protocol that allows a sender to transmit two classical bits, \(b_1\) and \(b_2\), to a receiver using a single quantum bit. The circuit for superdense coding is shown in Figure 11. The \(\text{qir}\) program corresponding to the unitary portion of the superdense b1 b2 circuit is shown in Figure 12. Note that in the \(\text{qir}\) program, encode is a Coq function that takes two Boolean variables and returns a circuit.

\[
\begin{array}{c}
|0\rangle \\
|0\rangle
\end{array}
\begin{array}{c}
X \\
Z
\end{array}
\begin{array}{c}
H \\
H
\end{array}
\begin{array}{c}
\not{X} \\
\not{H}
\end{array}
\begin{array}{c}
b_2 \\
b_1
\end{array}
\]

Figure 11. Circuit for superdense coding.

We can prove that the result of evaluating the program superdense b1 b2 on an input state consisting of two qubits initialized to zero is the state \(|b_1, b_2\rangle\).

Lemma superdense_correct : \(\forall\ b_1\ b_2,\ \|\text{superdense}\ b_1\ b_2\|_2 \times |0, 0\rangle = |b_1, b_2\rangle\).

A.2 GHZ State Preparation

The Greenberger-Horne-Zeilinger (GHZ) state \([19]\) is an \(n\)-qubit entangled quantum state of the form
\[
|GHZ^n\rangle = \frac{1}{\sqrt{2}}(|0\rangle \otimes \cdots \otimes |0\rangle + |1\rangle \otimes \cdots \otimes |1\rangle).
\]
The correct property for this program says that for any (well-formed) density matrix $\rho$, teleport takes the state $\rho \otimes |0\rangle \otimes |0\rangle$ to the state $|0\rangle \otimes |0\rangle \otimes |0\rangle \otimes \rho$. Formally,

**Lemma teleport_correct**: $\forall \rho : \text{Density } 2, \quad \text{WF\_Matrix } \rho \rightarrow \langle \text{teleport} | \rho \otimes |0\rangle \otimes |0\rangle \otimes |0\rangle = |0\rangle \otimes |0\rangle \otimes |0\rangle \otimes \rho$.

The proof for the density matrix semantics is simple: We perform (automated) arithmetic to show that the output matrix has the desired form.

### A.4 Nondeterministic Semantics

The proof of the correctness of teleport for the density matrix semantics is simple but not useful for understanding why the protocol is correct. A more illuminating proof can be carried out with an alternative nondeterministic semantics in which evaluation is given as a relation. Given a state $\psi$, unitary program $u$ will (deterministically) evaluate to $[u]_d \times \psi$. However, measure $p_1 p_2$ may evaluate to $|1\rangle \otimes |1\rangle \times \psi$ or $p_2$ applied to $|0\rangle \otimes |0\rangle \times \psi$. We have found the nondeterministic semantics simpler to work with for certain types of proofs.

However, because we do not rescale the output of measurement (to avoid reasoning about matrix norms in Coq), the nondeterministic semantics is only useful for proving properties for which measurement outcome does not matter. The correctness of the teleport protocol above is an example of such a property, because the values measured by Alice do not impact the final output state. The nondeterministic semantics cannot be used for verifying soundness of non-unitary transformations in Coq because equivalence between programs requires equality between output probability distributions.

For the non-deterministic semantics the proof of teleport is more involved, but also more illustrative of the inner workings of the algorithm. Under the non-deterministic semantics, we aim to prove the following:

**Lemma teleport_correct**:

$\forall \psi : \text{Vector } (2^1), \psi' : \text{Vector } (2^3), \text{WF\_Matrix } \psi \rightarrow \langle \text{teleport} | (\psi \otimes |0\rangle \otimes |0\rangle) \downarrow \psi' \rightarrow \psi' \propto |0\rangle \otimes \psi$.

This says that on input $|\psi\rangle \otimes |0\rangle \otimes |\psi\rangle$, teleport will produce a state that is proportional to $\langle\psi| \otimes |0\rangle \otimes |\psi\rangle$. Note that this statement is quantified over every outcome $\psi'$ and hence all possible paths to $\psi'$. If instead we simply claimed that

$$\text{teleport } / (\psi \otimes |0\rangle \otimes |\psi\rangle) \downarrow 1/2 \times (|\psi\rangle \otimes |\psi\rangle),$$

where the $1/2$ factor reflects the probability of each measurement outcome ($(1/2)^2 = 1/4$), we would only be stating that some such path exists.

The first half of the circuit is unitary, so we can simply compute the effect of applying a $H$ gate, two CNOT gates and another $H$ gate to the input state. We can then take both measurement steps, leaving us with four different cases to prove correct. In each of the four cases, we can use the outcomes of measurement to correct the final qubit, putting it into the state $|\psi\rangle$. Finally, resetting the already-measured qubits is deterministic, and leaves us in the desired state.

### A.5 The Deutsch-Jozsa Algorithm

In the quantum query model, we are given access to a Boolean function $f : \{0,1\}^n \rightarrow \{0,1\}$ through an oracle defined by the map $U_f : |y, x\rangle \mapsto |y \oplus f(x), x\rangle$. For a function $f$ on $n$ bits, the unitary matrix $U_f$ is a linear operator over a $2^n+1$ dimensional Hilbert space. In order to describe the Deutsch-Jozsa algorithm in Coq, we must first give a $\text{sqrt}$ definition of oracles.

To begin, note that any $n$-bit Boolean function $f$ can be written as

$$f(x_1, \ldots, x_n) = \begin{cases} f_0(x_1, \ldots, x_{n-1}) & \text{if } x_n = 0 \\ f_1(x_1, \ldots, x_{n-1}) & \text{if } x_n = 1 \end{cases},$$

where $f_0(x_1, \ldots, x_{n-1}) = f(x_1, \ldots, x_{n-1}, b)$ is a Boolean function on $(n-1)$ bits for $b \in \{0,1\}$. Similarly, an oracle can be written as $U_f = U_{f_0} \otimes |0\rangle \langle 0| + U_{f_1} \otimes |1\rangle \langle 1|$. For $U_f : |y, x_1, \ldots, x_{n-1}, b\rangle = U_{f_0} |y, x_1, \ldots, x_{n-1}\rangle |b\rangle$. In the base case ($n = 0$), a Boolean function is a constant function of the form $f(\perp) = 0$ or $f(\perp) = 1$ and an oracle is either the identity matrix, i.e., $|y\rangle \mapsto |y\rangle$, or a Pauli-X matrix, i.e., $|y\rangle \mapsto |y \oplus 1\rangle$. As a concrete example, consider the following correspondences between the 1-bit Boolean functions and $4 \times 4$ unitary matrices:

$$\begin{align*}
    & f_0(x) = 0 & U_{f_0} &= I \otimes |0\rangle \langle 0| + I \otimes |1\rangle \langle 1|, \\
    & f_1(x) = 1 - x & U_{f_1} &= X \otimes |0\rangle \langle 0| + I \otimes |1\rangle \langle 1|, \\
    & f_0(x) = x & U_{f_0} &= I \otimes |0\rangle \langle 0| + X \otimes |1\rangle \langle 1|, \\
    & f_1(x) = 1 & U_{f_1} &= X \otimes |0\rangle \langle 0| + X \otimes |1\rangle \langle 1|. 
\end{align*}$$

The observation above enables the following inductive definition of an oracle.

**Inductive boolean**: $\forall \dim, \text{dim} : \text{Set} \rightarrow \text{Set} :=$

| boolean_{I} : \forall \dim, \text{dim} \rightarrow \text{Set} |
| boolean_{X} : \forall \dim, \text{dim} \rightarrow \text{Set} |
| boolean_{U} : \forall \dim, \text{dim} \rightarrow \text{Set} |
We define balanced and constant oracles in $s$.

Definition constant $\{\text{dim : } \mathbb{N}\}$ $\text{U : ucom base \text{dim}}$ $:=$ $\text{X 0 ; cpar n H ; U ; cpar n H.}$

Figure 14. The Deutsch-Jozsa algorithm in $sqir$ and as a circuit.

$$
\text{boolean dim u2} \rightarrow \\
\begin{align*}
\text{u1+dim} & = \text{u1dim } \otimes |0\rangle \langle 0| + \\
\text{u2dim } \otimes |1\rangle \langle 1| \\
\text{u3dim} & = \text{u3dim } \otimes |0\rangle \langle 0| + \\
\text{u4dim} & = \text{u4dim } \otimes |1\rangle \langle 1| \\
\end{align*}
$$

boolean dim $\text{U}$ describes an oracle for a dim-1-bit Boolean function whose denotation is a $2^{dim} \times 2^{dim}$ unitary matrix.

A Boolean function is balanced if the number of inputs that evaluate to $1$ is exactly half of the domain size. A Boolean function is constant if for all inputs, the function evaluates to the same output, i.e., $\forall x. f(x) = 0$ or $\forall x. f(x) = 1$. Given an oracle, we can determine whether it describes a balanced or constant function by counting the number of inputs that evaluate to $1$.

Fixpoint count $\{\text{dim : } \mathbb{N}\}$ $\{\text{U : ucom base \text{dim}}\}$ $\{\text{P : boolean \text{dim} \text{U}}\}$ $: \mathbb{N} := \\
\text{match \text{P} with} \\
| \text{boolean_i _ _} \Rightarrow 0 \\
| \text{boolean_X _ _ _} \Rightarrow 1 \\
| \text{boolean_U _ _ _ _ P1 P2_} \Rightarrow \text{count P1 + count P2} \\
\text{end.}$

We define balanced and constant oracles in $sqir$ as follows.

Definition balanced $\{\text{dim : } \mathbb{N}\}$ $\{\text{U : ucom base \text{dim}}\}$ $\{\text{P : boolean \text{dim} \text{U}}\}$ $: \mathbb{P} := \\
\text{count P = 2 ^ (dim - 2).}$

Definition constant $\{\text{dim : } \mathbb{N}\}$ $\{\text{U : ucom base \text{dim}}\}$ $\{\text{P : boolean \text{dim} \text{U}}\}$ $: \mathbb{P} := \\
\text{count P = 0 \lor count P = 2 ^ (dim - 1).}$

In the Deutsch-Jozsa [1992] problem, we are promised that the function $f$ is either balanced or constant, and the goal is to decide which is the case by querying the oracle. The Deutsch-Jozsa algorithm begins with an all $|0\rangle$ state, and prepares the input state $|\_\rangle \otimes \text{nket \text{dim}} |+\rangle$. This state is prepared by applying an $X$ gate on the first qubit, and then applying a $H$ gate to every qubit in the program. Next, the oracle $U$ is queried, and a $H$ gate is again applied to every qubit in the program. Finally, all qubits except the first are measured in the standard basis. This algorithm is shown as a circuit and in $sqir$ in Figure 14. Note the use of Coq function $\text{cpar}$, which constructs a $sqir$ program that applies the same operation to every qubit in the program.

If measuring all the qubits after the first yields an all-zero string, then the algorithm outputs “accept,” which indicates that the function is constant. Otherwise the algorithm outputs “reject.” Instead of manually doing to the measurement, we will mathematically describe the output. Formally, the algorithm will output “accept” when the output state is supported on $\Pi = I \otimes |0\rangle \langle 0| @dim$ and output “reject” when the output state is orthogonal to $\Pi$. We can express this in Coq as follows:

Definition accept $\{\text{dim : } \mathbb{N}\}$ $\{\text{U : ucom base \text{dim}}\}$ $\{\text{P : boolean \text{dim} \text{U}}\}$ $: \mathbb{P} := \\
\forall (\psi : \text{Matrix 2 1}) , \\
(\psi \otimes \text{nket (dim-1) } |0\rangle) \vdash \text{[deutsch_jozsa dim \text{U}]u} \times (\text{nket dim } |0\rangle)) \otimes \text{0 } = 1.$

Definition reject $\{\text{dim : } \mathbb{N}\}$ $\{\text{U : ucom base \text{dim}}\}$ $\{\text{P : boolean \text{dim} \text{U}}\}$ $: \mathbb{P} := \\
\forall (\psi : \text{Matrix 2 1}) , \text{WF Matrix } \psi \rightarrow \\
((\psi \otimes \text{nket (dim-1) } |0\rangle) \vdash \text{[deutsch_jozsa dim \text{U}]u} \times (\text{nket dim } |0\rangle)) \otimes \text{0 } = \text{0.}$

We now prove the following theorems.

Theorem deutsch_jozsa_constant_correct $\forall (\text{dim : } \mathbb{N}) (\text{U : ucom base \text{dim}}) (\text{P : boolean \text{dim} \text{U}}), \\
\text{constant P } \rightarrow \text{accept P.}$

Theorem deutsch_jozsa_balanced_correct $\forall (\text{dim : } \mathbb{N}) (\text{U : ucom base \text{dim}}) (\text{P : boolean \text{dim} \text{U}}), \\
\text{balanced P } \rightarrow \text{reject P.}$

The key lemma in our proof states that the probability of outputting “accept” depends on the number of inputs that evaluate to $1$, i.e., $\text{count P}$. This lemma is proved by induction on $\text{P}$, which is the proof that $U$ is a Boolean oracle. We sketch the structure of the proof below, using mathematical notation for ease of presentation.

In the base case, either $\text{U } \equiv \text{I 0}$ or $\text{U } \equiv \text{X 0}$, the former of which is constant, and the latter is balanced. The lemma holds for the base case since $\langle \psi | \text{HIH} | 1\rangle = \langle \psi | 1\rangle$ for $\text{U } \equiv \text{I 0}$ and $\langle \psi | \text{HXH} | 1\rangle = - \langle \psi | 1\rangle$ for $\text{U } \equiv \text{X 0}$. Thus the factor can be written as $1 - 2 * \text{count P}$. For the inductive step, the inductive hypothesis says that, for any Boolean function of $\text{dim - 1}$ bits, the factor is $1 - 4 * \text{count P} / 2 ^ \text{dim}$. We observe that with probability $1/2$, the bit $x_{\text{dim}}$ input to the oracle of a dim-bit Boolean function $f$...
is either 0 or 1. Conditioned on the first bit being 0 (resp.
1), the function is \( f_0 \) (resp. \( f_1 \)). Let \( s_b \) the number of inputs
evaluating to 1 when \( x_{\text{dim}} = b \in \{0, 1\} \). Thus the sign can be calculated as
\[
\frac{1}{2} \left( 1 - \frac{s_0}{2^{\text{dim}-2}} \right) + \frac{1}{2} \left( 1 - \frac{s_1}{2^{\text{dim}-2}} \right) = 1 - \frac{s_0 + s_1}{2^{\text{dim}-1}}.
\]
Since the number of inputs evaluating to 1 is \( s_0 + s_1 \) for \( f \), we conclude the lemma.

For a balanced function, \( \text{count } P \) is equal to \( 2^{\text{dim} - 2} \). For a constant function, \( \text{count } P \) is either 0 or \( 2^{\text{dim} - 1} \). Thus the factor \( 1 - 4 \cdot \text{count } P / 2^{\text{dim}} \) is 0 for a balanced function and \( \pm 1 \) for a constant function. For a balanced function, no input state \( \psi \) can succeed with nonzero probability, and thus reject \( P \) is true, whereas for a constant function, one can present a state to maximize the probability to 1, and thus accept \( P \) can be proved.

**B \( \mathcal{Q}\text{wire vs. } \mathcal{S}\text{qir}**

When we first set out to build \( \mathcal{V} \text{oqc} \), we thought to do it using \( \mathcal{Q}\text{wire} \) [31], another formally verified quantum programming language embedded in Coq. However, we were surprised to find that we had tremendous difficulty proving that even simple transformations were correct. This experience led to the development of \( \mathcal{S}\text{qir} \). As discussed in Section 3.4, the fundamental difference between \( \mathcal{S}\text{qir} \) and \( \mathcal{Q}\text{wire} \) is that \( \mathcal{S}\text{qir} \) relies on a global register of qubits. Every operation is applied to an explicit set of qubits within the global register. By contrast, \( \mathcal{Q}\text{wire} \) uses Higher Order Abstract Syntax [32] to take advantage of Coq variable binding and function composition. This difference is most noticeable in how the two languages support composition.

**Composition in \( \mathcal{Q}\text{wire}**

\( \mathcal{Q}\text{wire} \) circuits have the following form:

\[
\begin{align*}
\text{Inductive } & \text{ Circuit } (w : \text{ WType}) : \text{ Set } := \\
| & \text{ output : } \text{ Pat } w \rightarrow \text{ Circuit } w \\
| & \text{ gate : } \forall \text{ } (w1 w2), \text{ Gate } w1 w2 \rightarrow \text{ Pat } w1 \rightarrow \\
& \quad (\text{ Pat } w2 \rightarrow \text{ Circuit } w) \rightarrow \text{ Circuit } w \\
| & \text{ lift : } \text{ Pat Bit } \rightarrow (\exists \text{ } \rightarrow \text{ Circuit } w) \rightarrow \text{ Circuit } w.
\end{align*}
\]

Patterns \( \text{Pat} \) type the variables in \( \mathcal{Q}\text{wire} \) circuits and have a corresponding wire type \( w \), corresponding to some collection of bits and qubits. The definition of gate takes in a parameterized gate, an appropriate input pattern, and a continuation of the form \( \text{Pat } w2 \rightarrow \text{ Circuit } w \), which is a placeholder for the next gate to connect to. This is evident in the definition of the composition function:

\[
\begin{align*}
\text{Fixpoint } & \text{ compose } (w1 w2) (c : \text{ Circuit } w1) \\
(f : & \text{ Pat } w1 \rightarrow \text{ Circuit } w2) : \text{ Circuit } w2 := \\
\text{ match } & c \text{ with} \\
| & \text{ output } p \Rightarrow f \ p
\end{align*}
\]

\[\text{ In the gate case, the continuation is applied directly to the output of the first circuit.} \]

Circuits correspond to open terms; closed terms are represented by \( \text{boxed} \) circuits:

\[
\begin{align*}
\text{Inductive } & \text{ Box } w1 w2 : \text{ Set } := \\
& \text{ box : } \text{ (Pat } w1 \rightarrow \text{ Circuit } w2) \rightarrow \text{ Box } w1 w2.
\end{align*}
\]

\[\text{This representation allows for easy composition: Any two circuits with matching input and output types can easily be combined using standard function application. For example, consider the following convenient functions for sequential and parallel composition of closed terms:} \]

\[
\begin{align*}
\text{Definition } & \text{ inSeq } (w1 w2 w3) (c1 : \text{ Box } w1 w2) \\
& \quad (c2 : \text{ Box } w2 w3) : \text{ Box } w1 w3 := \\
& \quad \text{ box } p1 \Rightarrow \\
& \quad \text{ let } p2 \leftarrow \text{ unbox } c1 p1; \\
& \quad \text{ unbox } c2 p2.
\end{align*}
\]

\[
\begin{align*}
\text{Definition } & \text{ inPar } (w1 w2 w1' w2') (c1 : \text{ Box } w1 w2) \\
& \quad (c2 : \text{ Box } w1' w2') : \text{ Box } (w1 \otimes w1') (w2 \otimes w2') := \\
& \quad \text{ box } (p1,p2) \Rightarrow \\
& \quad \text{ let } p1' \leftarrow \text{ unbox } c1 p1; \\
& \quad \text{ let } p2' \leftarrow \text{ unbox } c2 p2; \\
& \quad (p1',p2').
\end{align*}
\]

\[\text{Unfortunately, proving useful specifications for these functions is quite difficult. Since the denotation of a circuit must be (in the unitary case) a square matrix of size } 2^n \text{ for some } n, \text{ we need to map all of our variables to } 0 \text{ through } n - 1, \text{ ensuring that the mapping function has no gaps even when we initialize or discard qubits. We maintain this invariant through compiling to a de Bruijn-style variable representation [13]. Reasoning about the denotation of our circuits, then, involves reasoning about this compilation procedure. In the case of open circuits (our most basic circuit type), we must also reason about the contexts that type the available variables, which change upon every gate application.} \]

**Composition in \( \mathcal{S}\text{qir}**

Composing two \( \mathcal{S}\text{qir} \) programs requires manually defining a mapping from the global registers of both programs to a new, combined global register. For example, consider the following code, which composes two \( \mathcal{S}\text{qir} \) programs in parallel.

\[
\begin{align*}
\text{Fixpoint } & \text{ map_qubits } (U \text{ dim}) (f : N \rightarrow N) \\
& \quad (c : \text{ ucom } U \text{ dim}) : \text{ ucom } U \text{ dim } := \\
& \quad \text{ match } c \text{ with} \\
& \quad \text{ | c1; c2 } \Rightarrow \text{ map_qubits } f c1; \text{ map_qubits } f c2 \\
& \quad \text{ | uapp1 } u n \Rightarrow \text{ uapp1 } u (f n) \\
& \quad \text{ | uapp2 } u m n \Rightarrow \text{ uapp2 } u (f m) (f n) \text{ end.}
\end{align*}
\]
A Verified Optimizer for Quantum Circuits

Figure 15. The patterns for the output of GHZ state preparation ghz on a list of 4 qubits and the controlled swap on a tree of 4 qubits fredkin_seq. The left part describes the output pattern (((q1,q2),(q3,q4)) of ghz, and the right part describes the pattern ((q1,q2),(q3,q4)) of fredkin_seq. In composition, the mismatching patterns require an extra gadget to transform the former into the latter.

Fixpoint cast (U dim) (c : ucom U dim) dim’
: ucom U dim’ :=
match c with
| c1; c2 ⇒ cast c1 dim’ ; cast c2 dim’
| uapp1 u n ⇒ uapp1 u n
| uapp2 u m n ⇒ uapp2 u m n
end.

Definition inPar (U dim1 dim2) (c1 : ucom U dim1) (c2 : ucom U dim2) :=
(cast c1 (dim1 + dim2));
(cast (map_qubits (fun q ⇒ dim1 + q) c2) (dim1 + dim2)).

The correctness property for inPar says that the denotation of inPar c1 c2 can be constructed from the denotations of c1 and c2.

Lemma inPar_correct : ∀ c1 c2 d1 d2,
uc_well_typed d1 c1 →
\[\|\text{inPar} c1 c2 d1\|_{d1+d2} = \|c1\|_{d1} \otimes \|c2\|_{d2}.\]

The inPar function is relatively simple, but more involved than the corresponding Qwire definition because it requires relabeling the qubits in program c2.

General composition in qir requires even more involved relabeling functions that are less straightforward to describe. For example, consider the composition expressed in the following Qwire program:

box (ps, q) ⇒
let (x, y, z) ← unbox c1 ps;
let (q, z) ← unbox c2 (q, z);
(x, y, z, q);

This program connects the last output of program c1 to the second input of program c2. This operation is natural in Qwire, but describing this type of composition in qir requires some effort. In particular, the programmer must determine the required size of the new global register (in this case, 4) and explicitly provide a mapping from qubits in c1 and c2 to indices in the new register (for example, the first qubit in c2 might be mapped to the fourth qubit in the new global register). When qir programs are written directly, this puts extra burden on the programmer. When qir is used as an intermediate representation, however, these mapping functions should be produced automatically by the compiler.

The issue remains, though, that any proofs we write about the result of composing c1 and c2 will need to reason about the mapping function used (whether produced manually or automatically).

As an informal comparison of the impact of Qwire’s and qir’s representations on proof, we note that while proving the correctness of the inPar function in qir took a matter of hours, there is no correctness proof for the corresponding function in Qwire, despite many months of trying. Of course, this comparison is not entirely fair: Qwire’s inPar is more powerful than qir’s equivalent. qir’s inPar function does not require every qubit within the global register to be used – any gaps will be filled by identity matrices. Also, qir does not allow introducing or discarding qubits, which we suspect will make ancilla management difficult.

Quantum Data Structures qir also lacks some other useful features present in higher-level languages. For example, in QIO [3] and Quipper [17] one can construct circuits that compute on quantum data structures, like lists and trees of qubits. In Qwire, this concept is refined to use more precise dependent types to characterize the structures; e.g., the type for the GHZ program indicates it takes a list of n qubits to a list of n qubits. More interesting dependently-typed programs, like the quantum Fourier transform, use the parameter n as an argument to rotation gates within the program.

Regrettably, these structures can make reasoning about programs difficult. For instance, as shown in Figure 15, the GHZ program written in Qwire emits a list of qubits while the fredkin_seq circuit takes in a tree of qubits. Connecting the wires from a GHZ to fredkin_seq circuit with the same arity requires an intermediate gadget. And if we want to verify a property of this composition, we need to prove that this gadget is an identity. In qir, which has neither quantum data structures nor typed circuits, this issue does not present itself.

Dynamic Lifting qir also does not support dynamic lifting, which refers to a language feature that permits measuring a qubit and using the result as a Boolean value in the host language to compute the remainder of a circuit Green et al. [17]. Dynamic lifting is used extensively in Quipper and Qwire. Unfortunately, its presence complicates the denotational semantics, as the semantics of any Quipper or Qwire program depends on the semantics of Coq or Haskell, respectively. In giving a denotational semantics to Qwire, Paykin et al. [31] assume an operational semantics for an arbitrary host language, and give a denotation for a lifted circuit only when both of its branches reduce to valid Qwire circuits.
Although sqir does not support dynamic lifting, its measure construct is a simpler alternative. Since the outcome of the measurement is not used to compute a new circuit, sqir does not need a classical host language to do computation: It is an entirely self-contained, deeply embedded language. As a result, we can reason about sqir circuits in isolation, though in practice we will often reason about families of circuits described in Coq.

**Other Differences** Another important difference between Qwire and sqir is that Qwire circuits cannot be easily decomposed into smaller circuits because output variables are bound in different places in the circuit. By contrast, a sqir program is an arbitrary nesting of smaller programs, and $c1:(c2:(c3:c4)):c5$ is equivalent to $c1:c2:c3:c4:c5$ under all semantics, whereas every Qwire circuit (only) associates to the right. As such, rewriting using sqir identities is substantially easier.

The differences between these tools stem from the fact that Qwire was developed as a programming language for quantum computers [31], and was later used as a verification tool [34, 35]. By contrast, sqir is mainly a tool for verifying quantum programs, ideally compiled from another language such as Q# [45], Quipper [17] or even Qwire itself.