isQ: Towards a Practical Software Stack for Quantum Programming

Jingzhe Guo*, Huazhe Lou†, Riling Li‡, Wang Fang‡, Junyi Liu‡, Peixun Long‡, Shenggang Ying‡, and Mingsheng Ying‡*
*Department of Computer Science and Technology
Tsinghua University
Beijing, China.
Email: gjz20@mails.tsinghua.edu.cn, yingmsh@tsinghua.edu.cn
†Arclight Quantum Computing Inc.
Beijing, China.
Email: louhz@arclightquantum.com
‡State Key Laboratory of Computer Science
Institute of Software, Chinese Academy of Sciences
Beijing, China.
Email: {lirl,fangw,liuij,longpx,yingsg,yingms}@ios.ac.cn

Abstract—We introduce isQ, a new software stack for quantum programming in an imperative programming language, also named isQ. The aim of isQ is to make the programmers write quantum programs as conveniently as possible. In particular:
1) The isQ language and its compiler contain many features, including some not well supported by (most) other quantum programming platforms, e.g. classical control flow such as recursion; decomposition of self-defined unitary gates; and oracle programming and its circuit realization.
2) To make it flexible, an isQ program can be compiled into several kinds of intermediate representation, including OpenQASM 3.0, QIR and QCIS (specially tailored for the superconducting quantum hardware at USTC).
3) Besides interfacing isQ with true superconducting hardware, a QIR simulator is also developed for demonstration and testing of isQ programs.

Index Terms—quantum programming, compiler, quantum circuit.

I. INTRODUCTION

Recent progress in quantum hardware convinces people of the possibility that quantum computers outperform their classical predecessors in solving some important problems [1-4]. We have entered a noise-intermediate-scale-quantum (NISQ) era as the quantum hardware keeps advancing. However, just like classical computers, quantum hardware cannot unleash its full power unless equipped with a series of quantum computing software, including but not limited to:

- quantum programming language(s) for writing quantum programs;
- quantum compilers for optimization and transformation of quantum programs, as well as compiling quantum programs to different hardwares, e.g. trapped-iron and superconducting devices;
- quantum simulators for debugging small-scale quantum programs, etc.

In this paper, we present isQ, a software stack for quantum programming towards practical use.

isQ was first proposed as an experimental and educational language [5]. The first version of its compiler was implemented in Python using PLY [6] module, converting abstract syntax tree (AST) from the input isQ programs directly into a kind of modified Quil IR [7]. Whereas the original isQ language has limited features, and the compiler is not extensible, e.g. it’s hard to implement sophisticated compiler optimizations [8]. To design a truly practical and extensible software stack that permits more language features and various compiler optimizations, we extended isQ language and reimplemented the compiler based on MLIR [9] infrastructure. Specifically, we designed isQ-IR, an MLIR dialect as our compiler frontend target, and implemented various transformation passes that are useful for compiling quantum programs, e.g. gate decomposition and qubit mapping, as well as code generators that convert isQ-IR to different kinds of low-level intermediate representation (IR) including QIR [10], OpenQASM 3.0 [11] and QCIS [12].

Main Contributions: More explicitly, the contributions of this paper are:

- We propose isQ, an imperative quantum program-
TABLE I

Comparison of several quantum programming languages/compiler projects. We list five features. Some of the features belong to the language specification, and others belong to the compiler functions. Specifically, ‘complete compiler’ means whether the developers have designed a complete compiler for such language; ‘auto decomposition’ indicates the ability to decompose from high-level gate definitions like unitary matrices down to basic gates like $U_3$ and CNOT; ‘hardware connection’ means whether such language could be compiled into a practical hardware instruction set or interfaced with some hardware backends; ‘oracle programming’ means whether it permits defining quantum oracles on the language level. The question mark ‘?’ means the feature might have been already implemented/in progress in corresponding languages, but we find no open and rigorous materials as evidence.

| Languages/Projects      | Complete compiler | Hardware connection | Auto decomposition | Classical control flow | Oracle programming |
|-------------------------|-------------------|---------------------|--------------------|------------------------|--------------------|
| Q#                      | ✓                 | ?                   | Partial            | ✓                      | ✓                  |
| Scaffold                | ✓                 | ?                   |                     |                        |                    |
| Qiskit/OpenQASM 2.0†    | ✓                 | ✓                   |                     |                        | ✓                  |
| OpenQASM 3.0‡           | ?                 | ?                   | ✓                  | ✓                      | ✓                  |
| Silq                    | ?                 | ✓                   | ✓                  | ✓                      |                    |
| Quingo                  | ✓                 | ✓                   | ✓                  | ✓                      | ✓                  |
| QCOR                    | ✓                 | ✓                   | ✓                  | ✓                      | ✓                  |
| isQ                     | ✓                 | ✓                   | ✓                  | ✓                      | ✓                  |

We built a versatile software stack around isQ and isQ-IR, and then generate the target codes such as QCIS, OpenQASM 3.0 and QIR. Optimizations are permitted throughout the whole compilation process. Finally the output QCIS and QIR codes could be executed on corresponding hardware and simulators.

Fig. 1. The structure of isQ compilation stack. The compiler will first compile the input programs into isQ-IR, and then generate the target codes such as QCIS, OpenQASM 3.0 and QIR. Optimizations are permitted throughout the whole compilation process. Finally the output QCIS and QIR codes could be executed on corresponding hardware and simulators.

1) Decorated gates and gate deriving notation: Controlled and/or adjoint quantum gates have been proved useful in the construction of quantum circuits. isQ supports using quantum gates in their decorated form, i.e. in (multi-)controlled or adjoint (inverse) version:
- “$\text{ctrl}(N)$” can be added before gate calling to add $N$ controller bits to the gate. ($N$ can be omitted for $N = 1$).
- “$\text{nctrl}(N)$” can be added to add $N$ negated-controller bits to the gate, i.e. appending $X$ gates

2The control flow in Qiskit is mainly on the Python level. In OpenQASM 2.0 the control flow is not well supported.

To our knowledge, Qiskit has not fully supported OpenQASM 3.0 yet.

II. isQ Language Specification

In this section, we introduce quantum programming language isQ, the basis of our software stack. isQ is an imperative quantum programming language that supports oracle definition, arbitrary unitary gate definition, classical control flow and other underlying features.

A. Grammar

To give an overall picture of isQ syntax, we summarize it in Backus-Naur form (BNF) in Figure. Let us further describe notable features of isQ in detail:

- For debugging purposes, we implemented our own QIR simulator that can execute QIR programs compiled by isQ compiler. The simulator provides an interface for supporting different backends, as well as a built-in CPU backend and a built-in CUDA backend.
- isQ programs without feedback control can be compiled to QCIS assembly, which can be executed on the superconducting quantum hardware at USTC (University of Science and Technology of China).
- isQ-IR can also be compiled into OpenQASM 3.0, while most high-level structures in isQ can be preserved and converted to OpenQASM 3.0 control flow.

Fig. 2. The overall isQ syntax defined in Backus-Naur form (BNF).
\begin{verbatim}
(program) ::= (topDefStmt) ;
  | (topDefStmt) ; (program)
(topDefStmt) ::= (varDefStmt) | (gateDefStmt) | (oracleDefStmt) | (procedure)
(varDefStmt) ::= (type) IDENTIFIER (varDefInit)
(varDefInit) ::= = (expr) | [ (expr) ] | e
(gateDefStmt) ::= defgate IDENTIFIER = [ (matrix) ]
(matrix) ::= (list) | (matrix) ; (list)
(list) ::= (expr) | (list) , (expr)
(oracleDefStmt) ::= oracle IDENTIFIER ( NATURAL , NATURAL ) = [ (list) ]
(procReturnType) ::= (type) | procedure
(procedure) ::= (procReturnTyp e) IDENTIFIER ( (procArgs) ) { (procBody) } (procDerive)
(procArgs) ::= (procArg) | (procArgs) , (procArg) | e
(procArg) ::= (type) IDENTIFIER | (type) IDENTIFIER [ ] | (type) IDENTIFIER [ expr ]
(procDerive) ::= e | deriving gate
(procBody) ::= e | (procStmt) ; (procBody)
(procStmt) ::= (ifStmt) | (forStmt) | (whileStmt) | (varDefStmt) | (callStmt)
  | (assignStmt) | (returnStmt) | (measureStmt) | (resetStmt) | (printStmt)
  | continue | break | pass
(ifStmt) ::= if (expr) { (procBody) } | if (expr) { (procBody) } else { (procBody) }
(forStmt) ::= for IDENTIFIER in (range) { (procBody) }
(range) ::= (expr) : (expr) | (expr) : (expr) : (expr)
(whileStmt) ::= while (expr) { (procBody) }
(callStmt) ::= (gateDecoration) IDENTIFIER ( (listMaybe) )
(gateDecoration) ::= e | (gateDecoration) ctrl ( NATURAL ) | (gateDecoration) ctrl
  | (gateDecoration) nctrl ( NATURAL ) | (gateDecoration) nctrl
(assignStmt) ::= (expr) = (expr)
(returnStmt) ::= return (expr)
(measureStmt) ::= M ( (expr) )
(resetStmt) ::= (expr) = [0]
(printStmt) ::= print (expr)
  | (expr) ::= NATURAL | IDENTIFIER | REALFLOAT | IMAGFLOAT | BOOLEAN
  | (expr) ::= (bop) (expr) | (bop) (expr) | (uop) (expr) | ( (expr) ) | (expr) | (expr) |
  | (expr) ::= (expr) | (expr) | (resetStmt)
(type) ::= int | qbit | double
(bop) ::= + | - | * | / | ≧ | ≤ | > | < = | != | ∗*
(uop) ::= + | −
IDENTIFIER ::= [a-zA-Z][a-zA-Za-zA-Z0-9_]*
NATURAL ::= 0 | [1-9][0-9]*
REALFLOAT ::= NATURAL (\.| [0-9]+)?([eE][\-\+]?[0-9]+)?
IMAGFLOAT ::= REALFLOAT j
\end{verbatim}

Fig. 2. The grammar structure of isQ in BNF. Here the productions of IDENTIFIER, NATURAL, REALFLOAT and IMAGFLOAT are not in strict BNF (they are regular expressions instead), but we put them into the overall grammar structure for reading convenience.
before and after controller bits. \( N \) can be omitted for \( N = 1 \).
- “\text{inv}” can be added to use the adjoint (inverse) version of the gate.

Specially, isQ allows user-defined gates by adding “\text{deriving gate}” notation to a procedure definition. The procedure must satisfy the following conditions:

- The parameters of the procedure must start with zero or more classical (i.e. boolean, floating, integer) parameters, and then zero or more qbit parameters. No qbit arrays are allowed. Thus all user-defined gates are fixed-sized.
- The procedure should have no return value.
- If an adjoint version of the gate is used, no classical control flow statements should appear in the procedure body.
- No measurement or side effects are allowed in the procedure body. While ancilla qubit allocation is allowed, the programmer should guarantee that all ancilla qubits are reset to \( |0\rangle \) when released.

When decorated forms of user-defined gates are used, isQ will automatically generate controlled and adjoint versions of these gates. For controlled gates, new qubit parameters are inserted after the classical parameters and before original qbit parameters. For example, using \text{ctrl } R_Z \text{ gate}, which has argument types (\text{double, qbit}), requires parameters to be passed in as (\text{double, qbit, qbit}) where controller qubit is the second argument.

2) \text{Classical computation support: } isQ provides support for classical computation. Useful integer and floating-point arithmetic are provided for programmers to accomplish common classical computing tasks, e.g. calculating gate parameter and counting measurement outcomes. isQ also provides full classical control flow support, including if-statement, for-and-while-loops, and procedure calls. These classical computation infrastructures allow programmers to write quantum-classic hybrid programs easily without having to use another host language.

3) \text{Oracle support: } Quantum oracles are important constructs in quantum algorithms like Grover search [13] or recursive Fourier sampling [14]. Currently, isQ language permits oracle definition by writing out the truth table of oracle functions with type \( f : \{0,1\}^n \to \{0,1\}^n \). For example,

```plaintext
oracle F(2,1) = [0,0,1,0];

// In procedure:
qbit a[2], b;
F (a[0],a[1],b)
```

In the code fragment above, an oracle \( F \) with two input bits and one output bit is defined. The oracle \( F \) is defined to be \( F |a_0\rangle a_1 \rangle b \rangle = |a_0\rangle a_1 \rangle (b \oplus (a_0 \land \neg a_1)) \).

B. Operational Semantics

In this subsection, as a theoretical basis, we formally define the operational semantics of isQ language. To simplify the analysis, we focus on an imperative quantum programming model with the language being a subset of the complete syntax of isQ. Since the language permits recursion, let us introduce the notions of quantum program schemes and declaration first.

We assume a finite (or countably infinite) set \( V \) of classical variables, and a finite (or countably infinite) set \( QV \) of quantum variables. Quantum program schemes are defined in the following:

**Definition II.1.** Quantum program schemes are generated by the syntax:

\[
P ::= X \mid \text{skip} \mid (x_1,...,x_n) := (t_1,...,t_n) \mid P_1; P_2 \mid \text{if } \varphi \text{ then } P_1 \text{ else } P_2 \mid \text{while } \varphi \text{ do } P \text{ od} \mid q := |0\rangle \mid \bar{q} := U[\bar{q}] \mid x := M[\bar{q}]
\]

where \( X \) is a procedure identifier, \( x,x_1,...,x_n \in V \) are classical variables, \( t_1,...,t_n \) are (classical) expressions, \( q \in QV \) is a quantum variable, and \( \bar{q} \subseteq QV \) is a quantum register, i.e.

\[
\text{a family of distinct quantum variables.}
\]

The statement \text{skip}, assignment \( (x_1,...,x_n) := (t_1,...,t_n) \), sequential composition \( P_1; P_2 \), conditional if \( \varphi \) then \( P_1 \) else \( P_2 \) fi and loop while \( \varphi \) do \( P \) od are the same as in a classical programming language. Initialisation \( q := |0\rangle \mid \bar{q} := U[\bar{q}] \mid x := M[\bar{q}] \) sets quantum variable \( q \) to a basis state \( |0\rangle \). Statement \( \bar{q} := U[\bar{q}] \) means that unitary transformation \( U \) is performed on quantum register \( \bar{q} \), leaving the states of the variables not in \( \bar{q} \) unchanged. In executing statement \( x := M[\bar{q}] \), quantum measurement \( M = \{M_m\} \) is performed on \( \bar{q} \), and an outcome is stored into classical variable \( x \).

Now we have some procedure identifiers in a program scheme. Given these procedure identifiers, we can define a declaration to specify their procedure bodies.

**Definition II.2.** Let \( X_1,...,X_n \) be different procedure identifiers. A declaration for \( X_1,...,X_n \) is a system of equations:

\[
D : \begin{cases} X_1 \leftarrow P_1, \\ \hspace{1cm} \ldots \\ X_n \leftarrow P_n, \end{cases}
\]

where for every \( 1 \leq i \leq n \), \( P_i \equiv P_i[X_1,...,X_n] \) is a quantum program scheme.

With Definition II.1 and Definition II.2 together, we give a formal definition for quantum programs.

**Definition II.3.** A quantum program (possibly with recursion) consists of:

1) a quantum program scheme \( P \equiv P[X_1,...,X_n] \), called the main statement; and
2) a declaration \( D \) for \( X_1,...,X_n \).
For each quantum variable $q \in QV$, let $\mathcal{H}_q$ be its Hilbert space. Then the tensor product $\mathcal{H} = \bigotimes_{q \in QV} \mathcal{H}_q$ is the state space of all quantum variables. A classical-quantum state is a pair $\langle \sigma, \rho \rangle$, where $\sigma$ is a state of classical variables, and $\rho$ is a density operator in $\mathcal{H}_q$, denoting a state of quantum variables. We use $\downarrow$ to denote termination. A classical-quantum configuration is a triple $\langle P, \sigma, \rho \rangle$, where $P$ is a program scheme or $P = \downarrow$, and $\langle \sigma, \rho \rangle$ is a classical-quantum state. The operational semantics describes the probabilistic transition between configurations.

**Definition II.4 (Operational Semantics).** Let $D$ be a given declaration. The operational semantics of quantum programs with respect to $D$ is the probabilistic transition relation $\xrightarrow{p}_D$ between configurations defined by the transition rules presented in Figure 3 where probability $p$ in transition $\langle P, \sigma, \rho \rangle \xrightarrow{p}_D \langle P', \sigma', \rho' \rangle$ is dropped whenever $p = 1$.

The denotational semantics of our quantum programming model can be defined based on the above operational semantics by generalizing the techniques developed in Section 3.4 of [15] where only purely quantum programs without classical variables are considered. We omit it here due to the limitation of space.

### III. **isQ Compiler Architecture**

In this section, we describe the design of our compiler. The compiler is based on MLIR [9] infrastructure, a highly-extensible compiler framework that supports representation, transformation and code generation for domain-specific computing, e.g. neural network, circuit logic, etc. Specifically for our purpose:

- We defined isQ-IR, an MLIR dialect supporting certain quantum operations, as our intermediate representation for representing and transforming quantum programs. Our isQ frontend can generate isQ-IR directly.
- Based on MLIR’s powerful general IR transformation framework, we utilized both existing general transformation passes provided by MLIR, as well as our own quantum-specific transformation passes, to transform and optimize quantum programs.
- By utilizing MLIR’s code generation and lowering infrastructure, we are able to generate different types of output code, including both high-level representations like QIR and OpenQASM 3.0 [11], and low-level real-device instruction sets like QCIS.

#### A. **Dialect definition of isQ-IR**

The principles for defining our isQ-IR dialect include:

- **Allowing easy reuse of compilation infrastructures** provided by MLIR, including compilation infrastructures originally designed for transforming classical programs. We designed our dialect in integration with existing MLIR dialects (scf, affine, memref, etc.) and only added minimum number of required operations enough for quantum programs, while preserving necessary properties for quantum-agnostic transformation passes to transform correctly.

- **Facilitating gate-level optimizations.** We adopted an SSA form of representing qubit fragments, which allows local quantum-circuit fragments to be exposed as def-use chains by a memory dependency analysis. MLIR’s powerful rewrite framework also simplifies the definition of gate identities and transformations.

- **Preserving high-level program structures.** isQ-IR allows high-level program information, e.g. decorated gates, built-in gate optimization hints, gate definition by matrices and oracles, to be represented. We can both perform high-level optimizations (e.g. canceling out user-defined UU†) and finally transform them into lower-level gate primitives easily.

- **Allowing extension for future features.** Our dialect itself is designed to be extensible so that we can add more features easily, e.g. new compilation passes and new ways of defining gates.

1) **New MLIR types:** isQ-IR defined two new types for describing quantum programs: \texttt{!isq.qstate} and \texttt{!isq.gate(N, hints)}.

- \texttt{!isq.qstate} represents an intermediate state of a single qubit, which can be seen as an open wire in a quantum circuit fragment. We do not define new types for representing qubits or qubit arrays; instead, we model them using \texttt{memref(n \times !isq.qstate)}: qubit allocation/deallocation are represented by \texttt{memref.alloc} and \texttt{memref.free}; for quantum operations, we need to use \texttt{memref.load} operation (or \texttt{affine.load}) to extract the \texttt{!isq.qstate} value out, perform operations to obtain a new value, and use \texttt{memref.store} to store it back.

Using quantum SSA representation poses additional requirements for legality:

- Every \texttt{!isq.qstate} value must be stored in the exact memory location where it was loaded from. This is guaranteed by our input IR.

- Two \texttt{!isq.qstate} values that are both “alive” at a certain point of a program must belong to two different qubits. They can be seen as distinct “open wires” in a quantum circuit, on which we can safely perform multi-qubit gates or freely switch two gates on disjoint qubits.

- Passes should not introduce \texttt{!isq.qstate} values that are not finally stored back to its \texttt{memref}, or store an invalid qstate back. This condition ensures that there are no extra unused \texttt{!isq.apply} statements. To satisfy this, our gate rewrite passes remove redundant SSA values as well as store operations using them.

- \texttt{!isq.gate(N, hints)} represents a pure quantum N-qubit gate as an SSA value that can be decorated, applied
to qubits, and passed around. **hints** on the gate type describes special properties about the gate. Currently supported hints include:

- **hermitian**, indicating the gate is Hermitian, e.g. CNOT, H.
- **diagonal**, indicating the gate has a diagonal matrix form, e.g. CZ, RZ.
- **antidiagonal**, indicating the single-qubit gate is anti-diagonal, e.g. X.
- **symmetric**, indicating the order of gate operands does not matter, e.g. SWAP and CZ gates.
- **phase**, indicating the gate is in the form $U_θ = \sum_{i<j=1}^{2^n-1} |i⟩⟨j| + e^{iθ} |11\cdots1⟩⟨11\cdots1|$. All phase gates are naturally diagonal and symmetric.

2) **New operations**: Table [II] lists isQ-IR-defined MLIR operations.

Instead of defining a builtin basic gate set, all gates are defined by **isq.defgate**. Both non-parametric gate and parametric gate families can be defined. The operation is attached with an attribute "definition", allowing specifying multiple ways of defining the operation:

- "qir", definition by a QIR function.
- "unitary", specifying a unitary matrix definition.
- "oracle", specifying an oracle truth-table definition.
- "decomposition", specifying a gate decomposition by a **builtin.func**.

A **builtin.func** can be used to describe a decomposition of a gate only if:

- The function accepts gate parameters and n qstates as arguments exactly.
- The function returns an n-qstate tuple exactly.
- Returned qstates correspond to argument qstates in exact order.
- For auto-generation of adjoint version, the function should only consist of one basic block and contains no operations with sub-regions.

These requirements above allow us to automatically generate a controlled (and adjoint) version of a gate.

A defined gate can be referenced using **isq.use** operation. If the gate has classical parameters, they need to be specified as operands of **isq.use**. A gate can then be applied to qstates using **isq.apply** operation, returning new SSA qstate values that can be applied on by the next gates or stored back.

**isq.decorate** is added to represent decorated (controlled and adjoint) operations. An **isq.decorate** operation is attached with two attributes, **ctrl** and **adjoint**. The **ctrl** attribute is a list of boolean values, where "false" indicates corresponding controller bit is negated. The operand of **isq.decorate** is an **isq.qstate**, and the result is also an **isq.qstate** with correct size and hints:

- Adding controller bits increases the gate size.
- Adjoint version of Hermitian gate is equivalent to non-adjoint version.
- Adjoint version of **diagonal** is still diagonal. The same is true for **antidiagonal** and **symmetric**.
- Controlled diagonal gates are still diagonal gates. The same is true for positively-controlled phase gates.

**isq.declare_qop** represents external quantum operations other than pure gates (e.g. measurements) that can be called by **isq.call_qop** that operates on zero or more qubits, zero or more classical input, interacts with external environment, and return zero or more classical output.

We use signature $[n](input) → output$ to represent a quantum operation that works on n qstates and classical values $input$, yielding n qstates and classical values $output$. For example, (computational basis) measurement is an operation that operates on one qubit and yields an $|1⟩$ value, and thus has signature $[1](\cdot) → |1⟩$. This is equivalent to traditional function signature $(\text{isq.qstate}) → \text{...}$.
Global operation of a gate that can be used.

Example

```cpp
func @error_1(%q: !isq.qstate) -> !isq.qstate { %q = memref.alloc(): memref<1x!isq.qstate>
  %q = affine.load [%0]: memref<1x!isq.qstate>
  %X = isq.use @X : !isq.gate<1>
  %Z = isq.use @Z : !isq.gate<1>
  %q2 = isq.apply %q(%q): !isq.gate<1>
  %q3 = isq.apply %q(%q2): !isq.gate<1>
  %q4 = isq.apply %q(%q3): !isq.gate<1>
  affine.store %q4, %0: memref<1x!isq.qstate>
  memref.dealloc(): memref<1x!isq.qstate>
  return %q: !isq.qstate
}
```

Example III.1. Consider the following function definition. The gates on ancilla qubits %Q introduces global phase (−1) to the system and can be optimized out. However, if we try to generate a controlled version of this function, eliminating them early will result in error in relative phase.

The gates on ancilla qubits %Q introduces global phase (−1) to the system and can be optimized out. However, if we try to generate a controlled version of this function, eliminating them early will result in error in relative phase.

| Operation        | Usage                                                                 | Example                                                                 |
|------------------|----------------------------------------------------------------------|------------------------------------------------------------------------|
| isq.defgate      | Global definition of a gate that can be used.                        | isq.defgate @X [definition = #X_DEF]: !isq.gate<1, hermitian>           |
| isq.use          | Use a globally defined gate as a value.                              | %X = isq.use @X : !isq.gate<1, hermitian>                               |
| isq.apply        | Apply a gate value onto qubit states.                                | %b = isq.apply %X(%a): !isq.gate<1, hermitian>                          |
| isq.decorate     | Extend control qubits onto a gate or inverting a gate.               | %CNOT = isq.decorate (%X: !isq.gate<1, hermitian>: !isq.gate<2>         |
| isq.downgrade    | Auxiliary operation for removing hints.                              | %H2 = isq.downgrade (%H: !isq.gate<1, hermitian>: !isq.gate<1>          |
| isq.declare_qop  | Define an external quantum operation.                                | isq.declare_qop @measure : [1]()→i1                                    |
| isq.call_qop     | Calls an external quantum operation.                                 | %b, %outcome = isq.call_qop @measure (%a): [1]()→i1                    |
| isq.apply_gphase | Applying global phase.                                              | isq.apply_gphase %g : !isq.gate<0>                                    |
| isq.accumulate_gphase | Auxiliary operation for “moving” global phase out of local system. | isq.accumulate_gphase %q : memref<1x!isq.qstate>                       |

TABLE II
New MLIR operations defined by isQ-IR.
Canonicalization: several useful local peephole optimization patterns are added to MLIR’s canonicalizer, which is executed after every transformation pass.

- **Decorate-op folding**: a rewriter pattern that folds two consecutive isq.decorate operations into one.
- **Gate cancellation**: rewrite pattern that cancels out pairs of Hermitian gates and $UU^\dagger$ gate pairs.
- **Symmetric operand rearranging**: when all inputs of a symmetric gate are from outputs of one gate, reorder the operands to match the output order of previous gate. This is useful for cancelling two CZ gates in “reverse” directions.

**Recognize famous gates**: this pass inserts definition of isQ builtin gates, including Paulis, Pauli rotations, $U_3$, CNOT and Toffoli.

- **Pauli-rotation to $U_3$**: this pass converts all parametric Pauli rotations into parametric $U_3$ gates.
- **Fold decorate gates**: this pass folds all isq.decorate operations:
  - Negated controller bits are eliminated by inserting $X$ gates.
  - For matrix-defined gates, a new matrix definition of the decorated gate is generated.
  - For decomposition-defined gates, the builtin.func is cloned and modified by adding controller bits and/or inverting the gate sequence.

After this pass, global phase auxiliary operations can be removed.

- **Decompose controlled $U_3$**: this pass decomposes controlled, parametric $U_3$ gate using $AXBXC$ rule [16], resulting in controlled-$X$, controlled-$GPhase$ (controlled phase shift), $R_2$ and $R_Y$ operations. Controlled-$X$ and controlled-$GPhase$ are decomposed recursively according to [16]. This pass eliminates all controlled parametric gates.

- **Decompose known matrices**: this pass decomposes unitary-matrix-defined gates into a bunch of basic gates using quantum Shannon decomposition [17]. Till now all gates are basic gates.

- **Remove trivial gates**: this pass removes constant gates that are very close to the identity.

5) **Lower to QIR**: After these passes, we defined lowering passes to convert isQ-IR to QIR.

- **Expand decomposition**: replace all gates defined by decomposition with builtin.call to decomposition function or QIR stub.

- **Reg2mem**: finally we convert isq.qstate to auxiliary type isq.qubit and gate applications to QIR calls. Measurements, resets and print operations, represented by isq.call_qop, are converted to corresponding QIR functions.

- **Lower to LLVM**: lower auxiliary type isq.qubit to llvm.struct(“Qubit”, opaque), the corresponding type in QIR. Applying lowering rules of other dialects (affine, scf, etc.) results in legal QIR.

6) **CodeGen to QCIS**: Our isQ compiler supports generating QCIS assembly that can execute on superconducting hardware. Challenges for targeting QCIS superconducting hardware include:

- QCIS hardware uses a different instruction set: for single-qubit parametric gate QCIS supports $R_z$, for two-qubit gate, QCIS provides CZ instead of CNOT. Retargeting gates in the program is required.
- QCIS hardware does not support classical control flow or feedback control. The “quantum” part of our program needs to be extracted and flattened.
- Superconducting hardware has qubit connectivity constraints. Qubit mapping is required to map logical qubits in isQ program onto real qubits.

To solve these problems, we proposed new passes to convert isQ program into valid QCIS assembly that can execute on real devices.

- We first check input isQ-IR to make sure there are no reset statements or uses of measurement results. This prevents feedback control in our program.
- We retarget gates in the program into QCIS instruction set, by converting CNOT into $H$-$CZ$-$H$ triples and $U_3$ into $R_Z$, $X2P$, $X2M$ [12, 18].
- We can extract all quantum gates by executing the program on our simulator and collecting every gate call. This effectively flattens high-level control flow structures in the program (loops and conditional branches) into a gate list.
- We perform qubit mapping on the collected gate list by our qubit mapper based on [19], generating QCIS assembly executable on superconducting hardware.

7) **CodeGen to OpenQASM 3.0**: We implemented direct code generator from isQ-IR to logical-level OpenQASM 3.0, a high-level quantum assembly language with control flow support. Since the control flow primitives in OpenQASM3 are high-level ones (e.g. if-statement, while-statements) instead of low-level ones (e.g. goto statements, basic blocks), we map structured MLIR control-flow operations directly to OpenQASM3, e.g. scf.if onto if-statements, affine.for onto for-statements, etc.

IV. Examples

In this section, we present several illustrative implementations of some interesting quantum algorithms that can clearly manifest the main features of isQ.

A. A pure quantum program

We use a simple example, Fourier sampling, also called Bernstein-Vazirani algorithm, to show the main workflow of isQ software stack, from isQ language to hardware-supported QCIS assembly. It is a pure quantum program without classical variables.

**Example IV.1 (Bernstein-Vazirani algorithm)**: We first give the implementation of a small-scale Bernstein-Vazirani algorithm on 5 qubits, with secret string “1101”. 

...
First we perform loop unrolling to obtain unrolled loop kernel, and obtain the snippet:

```
%2 = affine.load %0[4] : memref<5x!isq.qstate>
%3 = isq.apply %%2 : !isq.gate<1> // X(q[4])
affine.store %3, %0[4] : memref<5x!isq.qstate>
/* omitted */
%13 = affine.load %0[4] : memref<5x!isq.qstate>
```

Applying load-store forwarding is able to recover the dataflow on q[4] and other qubits:

```
%2 = affine.load %0[4] : memref<5x!isq.qstate>
%3 = isq.apply %%2 : !isq.gate<1> // X(q[4])
%13 = isq.apply %%3 : !isq.gate<1> // H(q[4])
```

Now qubit data dependency is exposed by use-def chain upon which our quantum passes can take effects: a conversion pass converts CNOT into CZ and H supported by QCIS, while the canonicalizer eliminates consecutive H gate pairs. After all conversions, isQ-IR is lowered to auxiliary QIR form:

```
%0 = memref.alloc() : memref<5x!isq.qir.qubit>
scf.for %arg0 = 0 to 5 step 1 {
  %5 = call @_quantum.rt.qubit.allocate() : () -> !isq.qir.qubit
  memref.store %5, %0[0] : memref<5x!isq.qir.qubit>
}
%1 = memref.load %0[4] : memref<5x!isq.qir.qubit>
call @_quantum.qis_x_body(%1) : (!isq.qir.qubit) -> ()
%2 = memref.load %0[0] : memref<5x!isq.qir.qubit>
call @_quantum.qis_h_body(%2) : (!isq.qir.qubit) -> ()
```

The snippet will be further lowered to QIR, compiled by LLVM, and linked with our simulator. We implemented a special backend in the simulator that converts to code-generation calls, e.g. `_quantum.qis_x_body(i)` generating a “X q,” QCIS instructions. These instructions will be collected and mapped to real hardware using our qubit mapping pass:

```
q[1] ——— H ——— q[1]
q[4] ——— X ——— H ——— H ——— q[0]
q[0] ——— H ——— H ——— H ——— H ——— q[4]
q[3] ——— H ——— q[3]
```

B. Recursion and oracles

Next we give an example including both recursion and oracle definitions.

Example IV.2 (Recursive Fourier Sampling). Recursive Fourier sampling (RFS) [14, 20] is the recursive version of Fourier sampling. The height-h RFS problem is defined to be a h-depth tree with each subtree of the root corresponding to a height-\((h - 1)\) RFS. The RFS problem could be solved recursively. Below is an instance for RFS of height-2.
oracle A(4,1) = \[0,1,1,0,0,0,0,0,0,1,0,1,0,1,0,1\];
oracle g(2,1) = \[1,0,1,0\];

int a;
qbit q[4], p[3];

procedure recursive_fourier_sampling(int k){
  if (k == 2){
    A(q[0], q[1], q[2], q[3], p[2]);
  } else {
    H(q[2*k]);
    H(q[2*k+1]);
    X(p[k+1]);
    H(p[k+1]);
    recursive_fourier_sampling(k+1);
    H(q[2*k]);
    H(q[2*k+1]);
    g(q[2*k],q[2*k+1], p[k]);
    H(q[2*k]);
    H(q[2*k+1]);
    recursive_fourier_sampling(k+1);
    H(q[2*k]);
    H(q[2*k+1]);
    H(p[k+1]);
    X(p[k+1]);
  }
}

procedure main(){
a = 0;
  recursive_fourier_sampling(a);
  int g = M(p[0]);
  print g;
}

The secret string \(s\) for this instance is "11", thus the expected print output should be \(g(s) = 0\), which could be validated in our simulator.

C. A quantum-classical hybrid program

We further give another complicated example, iterative phase estimation, a quantum-classical hybrid algorithm that requires real-time feedback control.

Example IV.3 (Iterative phase estimation). Iterative phase estimation (IPE) is a kind of phase estimation algorithm requiring only one ancilla qubit. IPE requires real-time intermediate measurement and feedback control: for every iteration, the ancilla qubit is operated on, measured and reset; the parameter of R\(_z\) gate in each iteration is determined by previous measurement outcomes; all gates and measurements must be finished within decoherence time. Below is our isQ program for a two-qubit, 20-bit precision IPE instance:

```c
/* the phase angle is 2*\pi*867893/(2**20) = 2*\pi*0.8276872 */
int x=867893;
double theta(){
  double y = 2**20;
  return 2*\pi*x / y;
}
procedure U(double theta, qbit q){
  nctrl GPhase(theta, q);
} deriving gate
```

```
procedure pow2_ctrlU(int n, qbit anc, qbit ev){
  double t = theta() * (2**n);
  ctrl U(t, anc, ev);
}
double ipe_U(int n, qbit ev){
  qbit anc;
  double phi = 0;
  for i in 0:n{
    H(anc);
    pow2_ctrlU(n - i - 1, anc, ev);
    Rz(- phi * 2 * pi, anc);
    H(anc);
    int res = M(anc);
    phi = (phi + (res / 2.0)) / 2.0;
    anc = |0>;
  }
  return (phi * 2.0);
}
procedure main(){
  qbit ev;
  ev = |0>;
  // eigen vector of U
  double phi = ipe_U(20, ev);
  print (phi * (2**20)); // should be x=867893.
}
```

We created two examples, IPE the example above, and IPE-9, an IPE instance but \(U\) is a 9-qubit gate. We write two examples in both isQ and Q#, compile and simulate them to compare the compilation and simulation performance. All experiments are carried out on a same PC with an Intel i9-11900K CPU.

| Time/s     | IPE          | IPE-9        |
|------------|--------------|--------------|
| Q# compile | 4.31 ± 0.14  | 5.24 ± 0.09  |
| Q# simulate| 1.25 ± 0.01  | 1.41 ± 0.05  |
| isQ compile| 0.40 ± 0.02  | 1.70 ± 0.06  |
| isQ simulate| 0.003 ± 0.000 | 2.11 ± 0.02  |

Table III shows that our toolchain has much smaller overhead for both compilation and execution for IPE. While Q# toolchain compiles the controlled-\(U\) in IPE-9 into mutli-qubit controlled gates and run such controlled gates directly on the simulator, we are able to compile the program into finer-grain CNOT and \(U_3\) gates and simulate them with a low overhead while preserving overall performance.

V. RELATED WORK

A. Quantum programming languages and compilers

There have already been a series of work in quantum programming languages and compilers targeting at different levels of abstraction. For example, Qiskit proposed by IBM is a popular Pythonic circuit construction framework for OpenQASM 2, which can run on IBM’s hardware backends but has only limited support for classical control flow. Languages like Q# and Silq support powerful language features; however, Q# can only run on limited hardware with limited features, while Silq can only run on its simulator.
A closely related previous work is QCOR \cite{28}, also aiming at building a practical and versatile quantum programming software stack. QCOR modified Clang so that quantum kernels could be interleaved with C++ source code, allowing existing C++ libraries to be imported out of convenience for quantum programmers, e.g. the gradient descent methods in variational quantum algorithms. Instead of depending on a host language, our work proposes to use one unified language for both classical and quantum parts of a hybrid program. Both QCOR and isQ are able to target multiple backends.

B. Quantum Intermediate Representations

Many quantum compilers employed their own IR for the transformation and optimization of quantum programs. For example, XACC \cite{30} defines a general IR for representing quantum programs compiled from various languages executing on various backends. ScaffCC \cite{31} extended LLVM IR to represent quantum operations to leverage LLVM framework for analyzing quantum programs, as well as relying on LLVM enable fast code generation (instrumentation-driven approach). QIR \cite{10}, one compilation target of our toolchain, is another LLVM-based IR that introduces quantum functionality by using LLVM’s opaque struct mechanism. Quantum MLIR dialect \cite{32} is proposed to fill the gap between quantum languages and QIR.

There have also been several works on using SSA for representing quantum programs. QIRO \cite{33} is proposed as an MLIR dialect that allows exposure of quantum dataflow as use-def chains, thus allowing quantum dataflow analysis and optimizations. QSSA \cite{34} further proposed a single-use analysis to statically check if no-cloning theorem is obeyed in the program. Compared with them, isQ-IR is designed with extensibility in mind: no built-in gates are defined in the dialect, allowing transformations based on “general properties” (gate hints) to be widely used. isQ-IR also seeks tighter integration with existing MLIR infrastructure, for example, by leveraging existing “memref” and “affine” dialects and corresponding built-in analysis passes for dependency analysis (instead of re-implementing array operations).

VI. Conclusion

This paper describes a software stack for quantum programming, including a high-level quantum programming language and tools to compile and simulate the language. The isQ programming language provides a series of high-level language features to facilitate writing quantum programs. By leveraging MLIR infrastructure, we defined isQ-IR, an MLIR dialect for representing quantum-classical hybrid programs, and built isQ compiler, allowing some powerful transformations and optimizations of isQ programs. Finally, backend codegenerators and simulators allow isQ programs to be finally compiled to and tested out on different backends, both simulated environment and real superconducting hardware, and also allow cooperation with lower-level compilation toolchains.

A. Future Work

Potential future works include further extending the isQ programming stack and adding more optimization passes, resource analysis and adding assertion functions, etc. Targeting isQ to more different kinds of quantum hardware, e.g. trapped-ion quantum devices, is also one of the future tasks.
[28] A. McCaskey, T. Nguyen, A. Santana, D. Claudino, T. Kharazi, and H. Finkel, “Extending c++ for heterogeneous quantum-classical computing,” ACM Transactions on Quantum Computing, vol. 2, no. 2, pp. 1–36, 2021.

[29] A. W. Cross, L. S. Bishop, J. A. Smolin, and J. M. Gambetta, “Open quantum assembly language,” arXiv preprint arXiv:1707.03429, 2017.

[30] A. J. McCaskey, D. I. Lyakh, E. F. Dumitrescu, S. S. Powers, and T. S. Humble, “XACC: A system-level software infrastructure for heterogeneous quantum-classical computing,” CoRR, vol. abs/1911.02452, 2019. [Online]. Available: http://arxiv.org/abs/1911.02452

[31] A. JavadiAbhari, S. Patil, D. Kudrow, J. Heckey, A. Lvov, F. T. Chong, and M. Martonosi, “Scaffcc: Scalable compilation and analysis of quantum programs,” Parallel Comput., vol. 45, pp. 2–17, 2015. [Online]. Available: https://doi.org/10.1016/j.parco.2014.12.001

[32] A. McCaskey and T. Nguyen, “A MLIR dialect for quantum assembly languages,” in IEEE International Conference on Quantum Computing and Engineering, QCE 2021, Broomfield, CO, USA, October 17-22, 2021, H. A. Müller, G. Byrd, C. Culhane, and T. Humble, Eds. IEEE, 2021, pp. 255–264. [Online]. Available: https://doi.org/10.1109/QCE52317.2021.00043

[33] D. Ittah, T. Hänner, V. Kliuchnikov, and T. Hoefler, “Enabling dataflow optimization for quantum programs,” CoRR, vol. abs/2101.11030, 2021. [Online]. Available: https://arxiv.org/abs/2101.11030

[34] A. Peduri, S. Bhat, and T. Grosser, “QSSA: an ssa-based IR for quantum computing,” in CC ’22: 31st ACM SIGPLAN International Conference on Compiler Construction, Seoul, South Korea, April 2 - 3, 2022, B. Egger and A. Smith, Eds. ACM, 2022, pp. 2–14. [Online]. Available: https://doi.org/10.1145/3497776.3517772