#### intel foundry

# Challenges in Using LLVM as a Quantum Intermediate Representation

Andrew Litteken, Albert Schmitz, Xin-Chuan Wu, Anne Matsuura

LLVM Developer's Meeting - October 24, 2024

#### Using LLVM as a Quantum IR

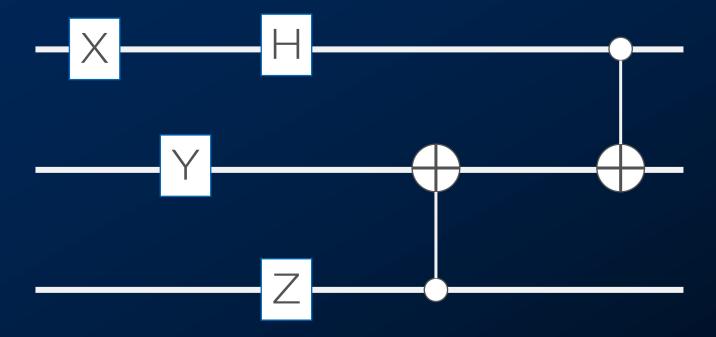
Quantum Compilation has specific constraints

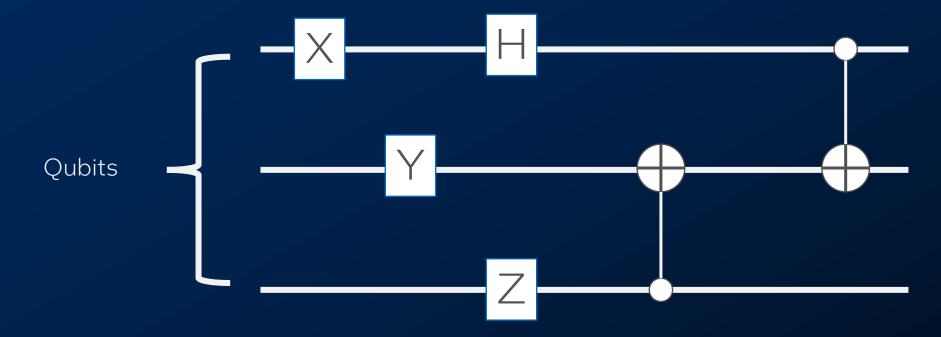
High-Level Programmers need to tools to express themselves

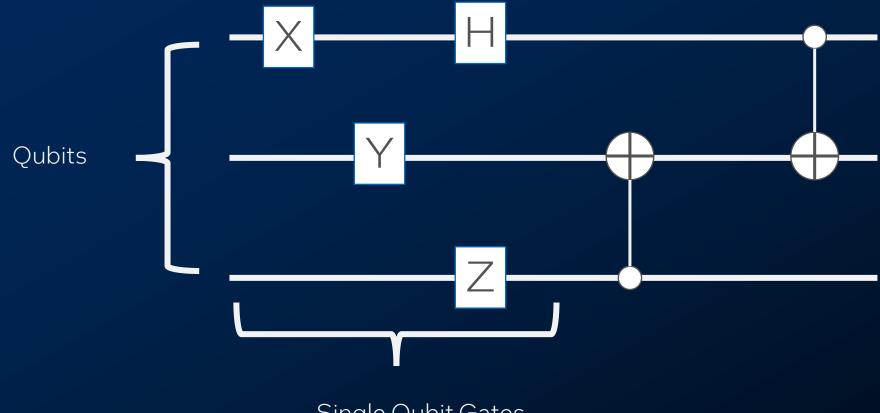
Non-Compiler Users need familiar tools for Optimization

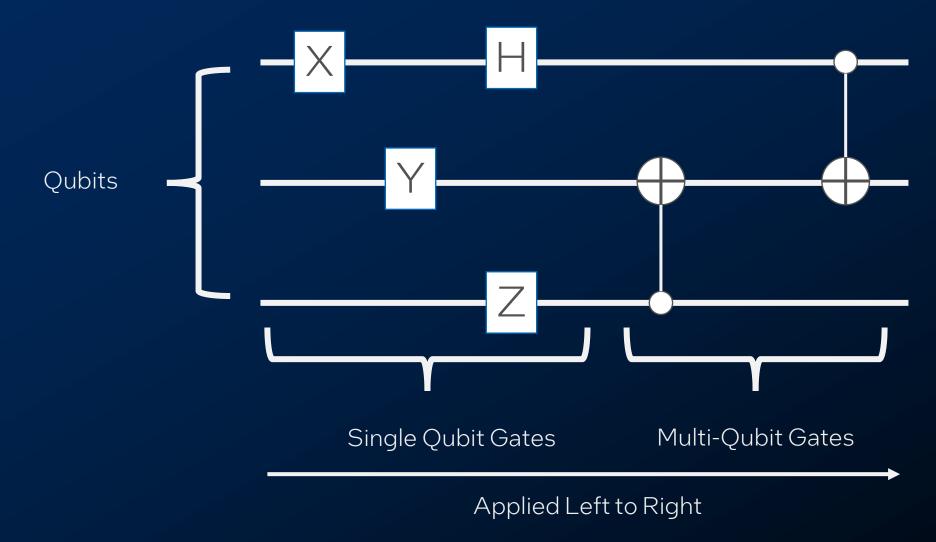
### Processing Quantum Programs

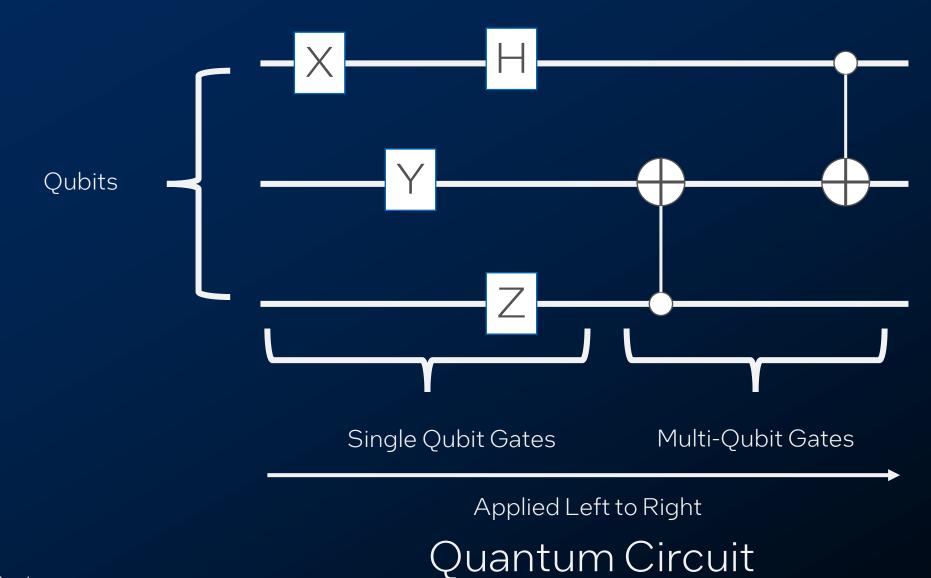
intel foundry



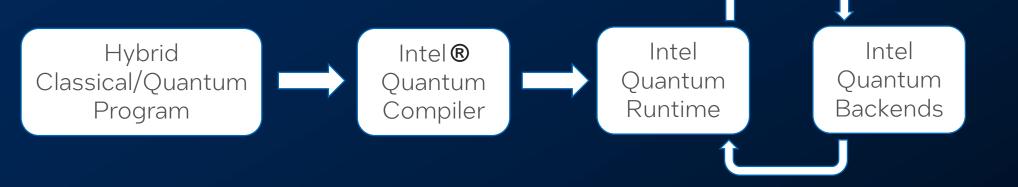


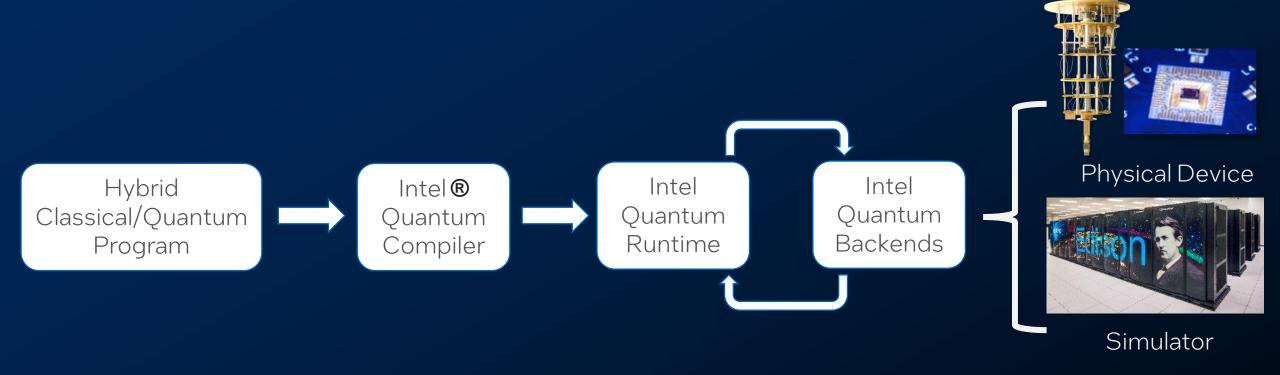


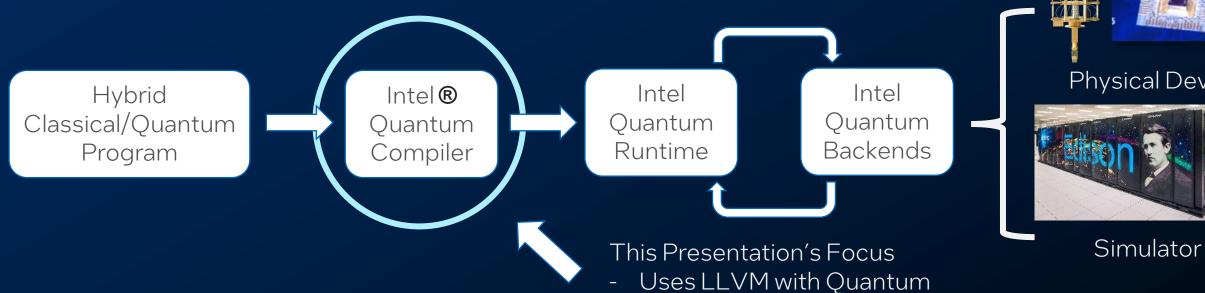




**intel** foundry







Intrinsics

frontend

frontends

Main frontend is SDK's C++ style

Could be targeted with different

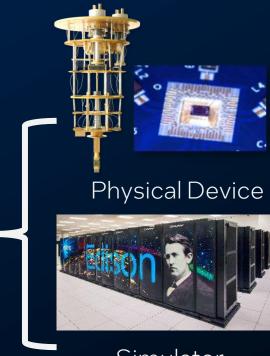


Hybrid Classical/Quantum Program



m L

Intel Quantum Runtime Intel Quantum Backends



Simulator

alongside each other
- Results in a module with
only classical operations,
and one with only quantum
operations

quantum compilation

Must perform classical and

- Deploys quantum programs to the device
- Runs classical instructions in between quantum components

Hybrid Classical/Quantum Program Intel® Quantum Compiler

Intel Quantum Runtime Intel Quantum Backends



Simulator

- Must perform classical and quantum compilation alongside each other

- Results in a module with only classical operations, and one with only quantum operations

- Deploys quantum programs to the device
- Runs classical instructions in between quantum components

Hybrid Classical/Quantum Program

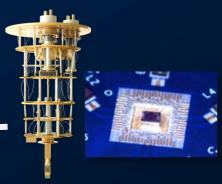




Intel Quantum Backends

- Must perform classical and quantum compilation alongside each other
- Results in a module with only classical operations, and one with only quantum operations

- Uses runtime instructions for execution
- Currently cannot perform conditional execution
- Returns execution information to the run time
- Limited resources



Physical Device



Simulator

#### The Compiler has many Quantum Challenges

- No Branching
- Match Device Constraints
- Reduce Quantum Operations
- No Classical Operations in Quantum Code

#### The Compiler has many Quantum Challenges

- No Branching
- Match Device Constraints
- Reduce Quantum Operations
- No Classical Operations in Quantum Code

These are constraints based on today's technology

### Handling Constraints with LLVM

intel foundry

#### Quantum Kernel (Can be nested)

```
#include <clang/Quantum/quintrinsics.h>
/// Quantum Runtime Library APIs
#include <quantum full state simulator backend.h>
const int total qubits = 2;
qbit qubit register[total qubits];
quantum_kernel void ghz_total_qubits() {
 for (int i = 0; i < total qubits; i++) {</pre>
    PrepZ(qubit register[i]);
 H(qubit_register[0]);
 for (int i = 0; i < total qubits - 1; i++) {
    CNOT(qubit register[i], qubit register[i + 1]);
```

#### Quantum Kernel (Can be nested)

```
#include <clang/Quantum/quintrinsics.h>
/// Quantum Runtime Library APIs
#include <quantum full state simulator backend.h>
const int total qubits = 2;
                                       Kernel Annotation
qbit qubit_register[total_qubits];
_quantum_kernel__oid ghz_total_qubits() {
  Tor (int 1 = 0; i < total qubits; i++) {
   PrepZ(qubit register[i]);
 H(qubit register[0]);
 for (int i = 0; i < total qubits - 1; i++) {
   CNOT(qubit register[i], qubit register[i + 1]);
```

quantum\_kernel adds section attribute

Quantum

Quantum Kernel (Can be nested)

```
Type
```

```
#include <clang/Quantum/quintrinsics.h>
/// Ouantum Runtime Library APIs
#include <quantum full state simulator backend.h>
const int total qubits = 2;
                                       Kernel Annotation
qbit oubit register[total qubits];
_quantum_kernel__oid ghz_total_qubits() {
  Tor (int I = 0; i < total qubits; i++) {
    PrepZ(qubit register[i]);
  H(qubit register[0]);
  for (int i = 0; i < total qubits - 1; i++) {
    CNOT(qubit register[i], qubit register[i + 1]);
```

- quantum\_kernel adds section attribute
- qbit and cbit are typedefs over particular integer types

```
Quantum Kernel
                      (Can be nested)
Ouantum
  Type
      #include <clang/Quantum/quintrinsics.h>
       /// Ouantum Runtime Library APIs
       #include <quantum full state simulator backend.h>
       const int total qubits = 2;
                                            Kernel Annotation
      qbit oubit register[total qubits];
      _quantum_kernel__oid ghz_total_qubits() {
               1 = 0; i < total qubits; i++) {</pre>
          PrepZ/qubit register[i]);
        H(qu it_register[0]);
        for (int i 0; i < total qubits - 1; i++) {
          CNOT(pubit_register[i], qubit_register[i + 1]);
                                Ouantum
                               Operations
```

- quantum\_kernel adds section attribute
- qbit and cbit are typedefs over particular integer types
- quintrinsics.h defines quantum functions

## Users Can Write Code that Mixes Classical And Quantum Computation

Quantum Kernel
Quantum (Can be nested)
Type

A #include < clang (Quantum (quint pinsies)

Classical Component (calls quantum components)

```
#include <clang/Quantum/quintrinsics.h>
/// Ouantum Runtime Library APIs
#include <quantum full state simulator backend.h>
const int total qubits = 2;
                                      – Kernel Annot<mark>at</mark>ion
qbit pubit register[total qubits];
_quantum_kernel__loid ghz_total_qubits() {
         1 = 0; i < total qubits; i++) {</pre>
   PrepZ/qubit register[i]);
 H(qu it_register[0]);
  for (int i 0; i < total qubits - 1; i++) {
   CNOT(pubit_register[i], qubit_register[i + 1]);
                           Ouantum
                          Operations
```

```
int main() {
 iqsdk::DeviceConfig qd_config("QD_SIM");
 iqsdk::FullStateSimulator quantum 8086(qd config);
 if (iqsdk::QRT ERROR SUCCESS != quantum 8086.ready())
   return 1;
 // get references to qbits
 std::vector<std::reference wrapper<qbit>> qids;
 for (int id = 0; id < total qubits; ++id) {</pre>
   qids.push back(std::ref(qubit register[id]));
 ghz_total_qubits();
 iqsdk::QssMap<double> probability map;
 probability map = quantum 8086.getProbabilities(qids, bases);
```

## Users Can Write Code that Mixes Classical And Quantum Computation

```
Quantum Kernel
                      (Can be nested)
Quantum
  Type
      #include <clang/Quantum/quintrinsics.h>
      /// Ouantum Runtime Library APIs
       #include <quantum full state simulator backend.h>
       const int total qubits = 2;
                                            Kernel Annotation
      qbit pubit register[total qubits];
      _quantum_kernel__loid ghz_total_qubits() {
               1 = 0; i < total qubits; i++) {</pre>
          PrepZ/qubit register[i]);
        H(qu it_register[0]);
        for (int i 0; i < total qubits - 1; i++) {
          CNOT(pubit_register[i], qubit_register[i + 1]);
                                Ouantum
                               Operations
```

Classical Component (calls quantum components)

```
int main() {
   iqsdk::DeviceConfig qd_config("QD_SIM");
   iqsdk::FullStateSimulator quantum_8086(qd_config);
   if (iqsdk::QRT_ERROR_SUCCESS != quantum_8086.ready())
        return 1;

On

// get references to qbits
   std::vector<std::reference_wrapper<qbit>> qids;
   for (int id = 0; /d < total_qubits; ++id) {
        qids.push_back(std::ref(qubit_register[id]));
   }

   ghz_total_qubits();

   iqsdk::QssMap<double> probability_map;
   probability_map = quantum_8086.getProbabilities(qids, bases);
}
```

#### LLVM Provides a Leg Up to Compilation

Quantum Kernel (Can be nested)

```
#include <clang/Quantum/quintrinsics.h>
/// Ouantum Runtime Library APIs
#include <quantum full state simulator backend.h>
const int total qubits = 2;
qbit qubit register[total qubits];
quantum kernel void ghz total qubits() {
 for ()nt i = 0; i < total qubits; i++) {
    PrepZ(qubit register[i]);
 H(qubit_register[0]);
 for () nt i = 0: i < total qubits - 1; i++) {
    cNOT(qubit register[i], qubit register[i +
```

Classical Component (calls quantum components)

```
int main() {
  iqsdk::DeviceConfig qd_config("QD_SIM");
  iqsdk::FullStateSimulator quantum_8086(qd_config);
  if (iqsdk::QRT_ERROR_SUCCESS != quantum_8086.ready())
    return 1;

// get references to qbits
  std::vector<std::reference_wrapper<qbit>> qids;
  for (int id = 0; id < total_qubits; ++id) {
    qids.push_back(std::ref(qubit_register[id]));
  }

  ghz_total_qubits();

  iqsdk::QssMap<double> probability_map;
  probability_map = quantum_8086.getProbabilities(qids, bases);
}
```

Conditional Structures with Quantum Operations

#### Structures are Found in Classical and Quantum

Quantum Kernel (Can be nested)

Classical Component (calls quantum components)

```
#include <clang/Quantum/quintrinsics.h>
/// Ouantum Runtime Library APIs
#include <quantum full state simulator backend.h>
const int total qubits = 2;
qbit qubit register[total qubits];
quantum kernel void ghz total qubits() {
 for ()nt i = 0; i < total qubits; i++) {
    PrepZ(qubit register[i]);
 H(qubit_register[0]);
 for () nt i = 0: i < total qubits - 1; i++) {
    cNOT(qubit register[i], qubit register[i +
```

```
int main() {
 iqsdk::DeviceConfig qd_config("QD_SIM");
 iqsdk::FullStateSimulator quantum 8086(qd config);
 if (igsdk::QRT ERROR SUCCESS != quantum 8086.ready())
   return 1;
 // get references to qbits
 std::vector<std::reference wrapper<qbit>> qids;
 for int id = 0; id < total qubits; ++id) {
   qids.push back(std::ref(qubit register[id]));
 ghz_total_qubits();
  qsdk::QssMap<double> probability map;
  probability map = quantum 8086.getProbabilities(qids, bases);
```

Conditional Structures with Quantum Operations and classical

Inlining
Kernels and
Adding
Intrinsics

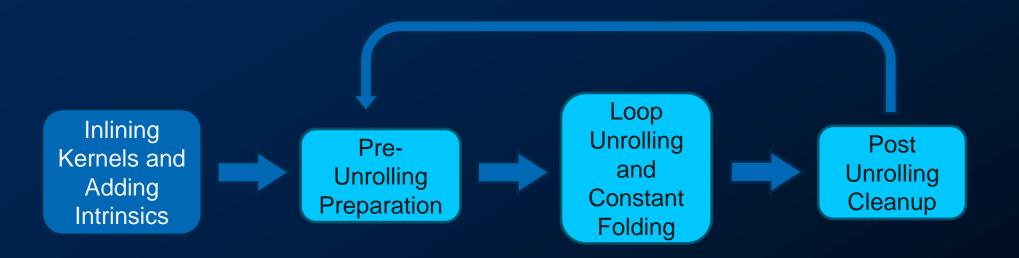
Pre-Unrolling Preparation Loop
Unrolling
and
Constant
Folding

Post Unrolling Cleanup

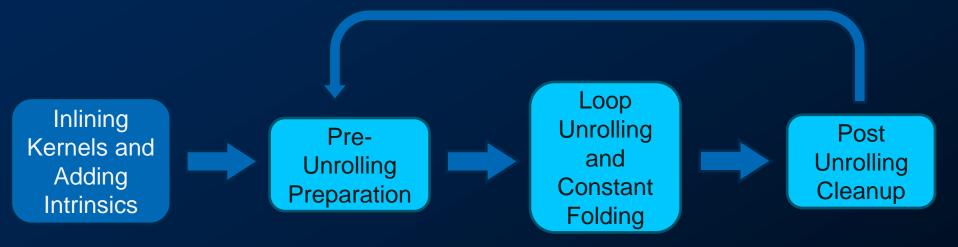
Inlining
Kernels and
Adding
Intrinsics

Pre-Unrolling Preparation Loop
Unrolling
and
Constant
Folding

Post Unrolling Cleanup

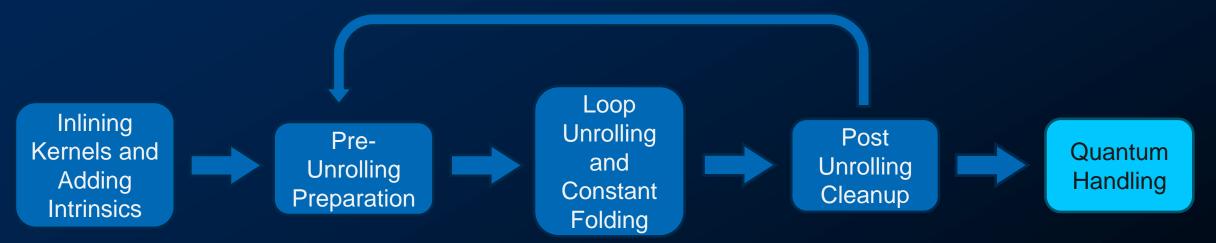


Re-runs until all marked kernels valid, or attempt number exceeded.



- Only runs on marked quantum kernels
- Only runs Loop optimizations on loops that contain quantum operations

Re-runs until all marked kernels valid, or attempt number exceeded.



- Only runs on marked quantum kernels
- Only runs Loop optimizations on loops that contain quantum operations

#### Declaration of Quantum Operation in Intermediate Representation

```
define dso_local void @_Z1HRt(ptr noundef nonnull align 2 dereferenceable(2) %q) #0 {
  entry:
    call void @llvm.quantum.qubit(ptr noundef nonnull %q)
    ret void
}
```

#### Declaration of Quantum Operation in Intermediate Representation

### Declaration of Quantum Operation in Intermediate Representation

```
define dso_local void @_Z1HRt(ptr noundef nonnull align 2 dereferenceable(2) %q) #0 {
  entry:
    call void @llvm.quantum.qubit(ptr noundef nonnull %q)
    ret void
}
```



Marking as a qubit reference

```
define dso_local void @_Z1HRt(ptr noundef nonnull align 2 dereferenceable(2) "qubit_ref" 3q) #0 {
  entry:
    call void @llvm.quantum.qubit(ptr noundef nonnull %q)
    ret void
}
```

## Providing Tools to Users



#### Writing Code Within these Constraints is Hard

#### Standard SDK

```
quantum kernel void qft() {
 // qft non-recursive implementation
 // Apply H and rotations
 // Starting from qubit 0
 for (int index = 0; index < N; index++) {</pre>
   H(QubitReg[index]);
   for (int index r = 1; index r < N - index; index r++) {
      double angle = 2 * (1 / M 1 PI) / std::pow(2, index r
+ 1);
      CPhase(QubitReg[index + index r], QubitReg[index],
angle);
 // Add SWAP gates
 for (int q_index = 0; q_index < std::floor(N / 2);</pre>
q index++) {
    SWAP(QubitReg[q index], QubitReg[N - q index - 1]);
```

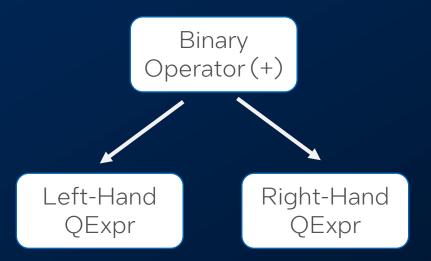
# Functional Language Extension for Quantum (FLEQ) Makes it Easier

Standard SDK

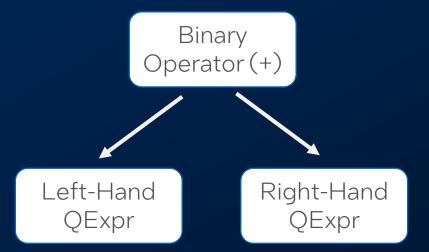
```
quantum kernel void qft() {
 // qft non-recursive implementation
 // Apply H and rotations
 // Starting from qubit 0
 for (int index = 0; index < N; index++) {</pre>
   H(QubitReg[index]);
   for (int index r = 1; index r < N - index; index r++) {
      double angle = 2 * (1 / M 1 PI) / std::pow(2, index r
+ 1);
      CPhase(QubitReg[index + index r], QubitReg[index],
angle);
 // Add SWAP gates
 for (int q_index = 0; q_index < std::floor(N / 2);</pre>
q index++) {
    SWAP(QubitReg[q index], QubitReg[N - q index - 1]);
```

#### FLEQ

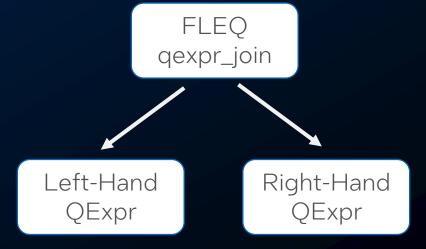
```
PROTECT QExpr qftCPhaseLadder(qbit& q, qlist::QList
reg){
 int sz = reg.size();
 double theta = - M PI / pow(2, sz);
 return qexpr::cIfTrue(sz > 0,
            qexpr::_CPhase(q, reg[0], theta)
            + qftCPhaseLadder(q, reg+1));
PROTECT QExpr qftHelper(qlist::QList reg){
 int sz = reg.size();
 return gexpr::cIfTrue(sz > 0,
     qexpr:: H(reg[sz-1]) +
       qftCPhaseLadder(reg[sz-1],
                       reg<<1)
        + qftHelper(reg<<1));
QExpr qft(qlist::QList reg){
  return qftHelper(reg) + reverseRegister(reg);
```

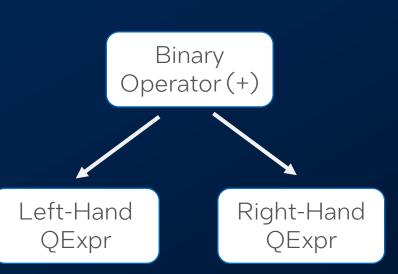






```
PROTECT QExpr ghz(qlist::QList qs){
   int len = qs.size();
   return (
        qexpr::join(
            qexpr::_PrepZ, qs),
        qexpr::join(
            qexpr::_H(qs[0]),
            qexpr::map(
                 qexpr::_CNOT,
                 qs(0,len-1), qs(1,len))
        ));
}
```

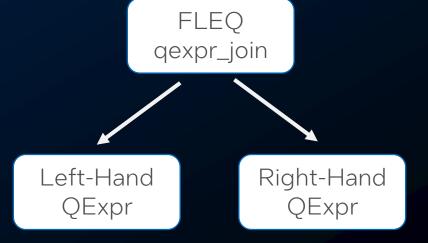






#### Replaced Operands

- Binary Add (join)
- Binary Multiply (join)
- ~,!,-(invert)
- ^(power)
- << (bind)
- >> (bind)



## FLEQ Enables More Advanced Optimization

FLEQ
Modified IR
Module

Internal Graph
Representation

Graph Based Quantum
Optimizations

IR Module with
Reduced Quantum
Resources

#### FLEQ Enables More Advanced Optimization

FLEQ Modified IR Module



Internal Graph Representation



Graph Based Quantum Optimizations



IR Module with Reduced Quantum Resources

#### Original Hybrid Program

```
quantum_kernel void qft() {
    . . . .
}

quantum_kernel void qft_inverse() {
    . . . .
}

quantum_kernel void qft_all() {
    prepZAll();
    qft();
    qft_inverse();
}
```

#### FLEQ with inferred inversion

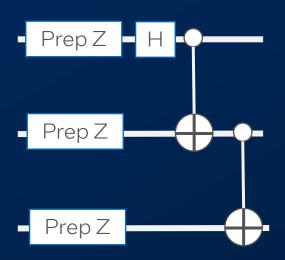
**Explicit Inverse** 

Implicit Inverse

# Simplifying Optimization Development

intel foundry

#### LLVM Doesn't Match Quantum Abstractions



Circuit Form

```
quantum_kernel void ghz_total_qubits() {
   for (int i = 0; i < total_qubits; i++) {
      PrepZ(q[i]);
   }

   H(q[0]);

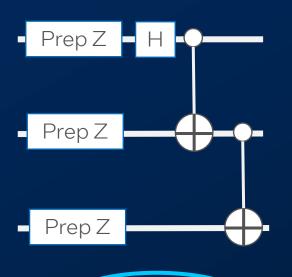
   for (int i = 0; i < total_qubits - 1;
   i++) {
      CNOT(q[i], q[i + 1]);
   }
}</pre>
```

SDK Form

```
agcc.quantum:
 %arrayidx34 = getelementptr inbounds
[12 x i16], ptr @Qumem, i64 0, i64 0
  call void @ Z5PrepZRt(ptr %arrayidx34)
 %arrayidx33 = getelementptr inbounds
[12 x i16], ptr @Qumem, i64 0, i64 1
  call void @ Z5PrepZRt(ptr %arrayidx33)
 %arrayidx32 = getelementptr inbounds
[12 x i16], ptr @Qumem, i64 0, i64 2
  call void @ Z1HRt(ptr %arrayidx22)
 %arrayidx20 = getelementptr inbounds
[12 x i16], ptr @Qumem, i64 0, i64 0
 %arrayidx21 = getelementptr inbounds
[12 x i16], ptr @Qumem, i64 0, i64 1
  call void @ Z4CNOTRtS (ptr %arrayidx20,
ptr %arrayidx21)
 %arrayidx18 = getelementptr inbounds
[12 x i16], ptr @Qumem, i64 0, i64 1
 %arrayidx19 = getelementptr inbounds
[12 x i16], ptr @Qumem, i64 0, i64 2
  call void @ Z4CNOTRtS (ptr %arrayidx18,
ptr %arrayidx19)
  br label %agcc.meas.move.end
```

IR Form (Post Unrolling)

#### LLVM Doesn't Match Quantum Abstractions



Circuit Form

```
quantum_kernel void ghz_total_qubits() {
   for (int i = 0; i < total_qubits; i++) {
      PrepZ(q[i]);
   }

   H(q[0]);

   for (int i = 0; i < total_qubits - 1;
   i++) {
      CNOT(q[i], q[i + 1]);
   }
}</pre>
```

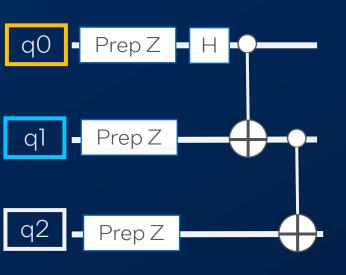
SDK Form

Typical Abstraction for Optimization

```
agcc.quantum:
 %arrayidx34 = getelementptr inbounds
[12 x i16], ptr @Qumem, i64 0, i64 0
  call void @ Z5PrepZRt(ptr %arrayidx34)
 %arrayidx33 = getelementptr inbounds
[12 x i16], ptr @Qumem, i64 0, i64 1
  call void @ Z5PrepZRt(ptr %arrayidx33)
 %arrayidx32 = getelementptr inbounds
[12 x i16], ptr @Qumem, i64 0, i64 2
  call void @ Z1HRt(ptr %arrayidx22)
 %arrayidx20 = getelementptr inbounds
[12 x i16], ptr @Qumem, i64 0, i64 0
 %arrayidx21 = getelementptr inbounds
[12 x i16], ptr @Qumem, i64 0, i64 1
  call void @ Z4CNOTRtS (ptr %arrayidx20,
ptr %arrayidx21)
 %arrayidx18 = getelementptr inbounds
[12 x i16], ptr @Qumem, i64 0, i64 1
 %arrayidx19 = getelementptr inbounds
[12 x i16], ptr @Qumem, i64 0, i64 2
  call void @ Z4CNOTRtS (ptr %arrayidx18,
ptr %arrayidx19)
  br label %agcc.meas.move.end
```

IR Form (Post Unrolling)

## The Quantum Circuit Object is an LLVM Wrapper

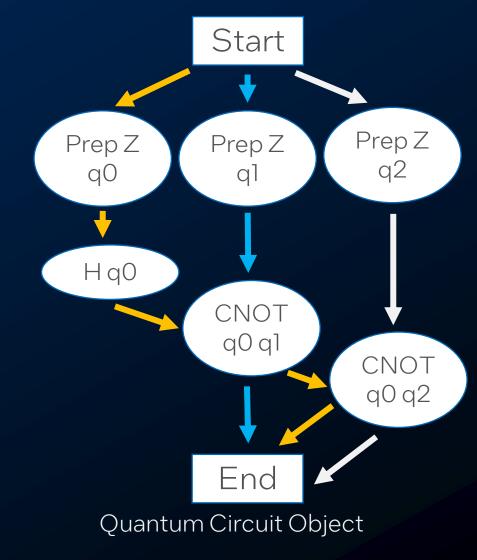


Circuit Form

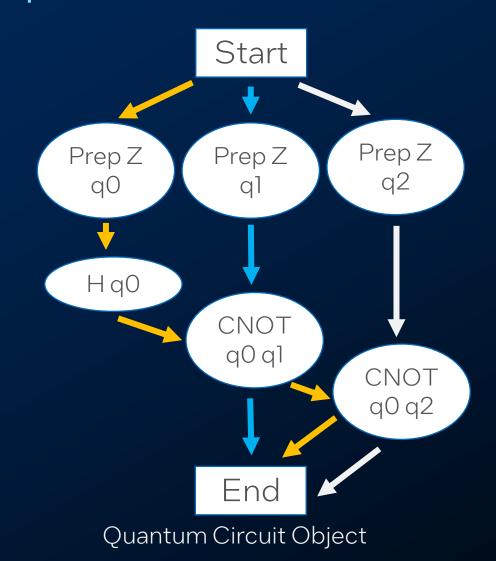
```
agcc.quantum:
 %arrayidx34 = getelementptr inbounds
[12 x i16], ptr @Qumem, i64 0, i64 0
  call void @ Z5PrepZRt(ptr %arrayidx34)
 %arrayidx33 = getelementptr inbounds
[12 x i16], ptr @Qumem, i64 0, i64 1
  call void @ Z5PrepZRt(ptr %arrayidx33)
 %arrayidx32 = getelementptr inbounds
[12 x i16], ptr @Qumem, i64 0, i64 2
  call void @ Z1HRt(ptr %arrayidx22)
 %arrayidx20 = getelementptr inbounds
[12 x i16], ptr @Qumem, i64 0, i64 0
  %arrayidx21 = getelementptr inbounds
[12 x i16], ptr @Qumem, i64 0, i64 1
  call void @ Z4CNOTRtS (ptr %arrayidx20,
ptr %arravidx21)
 %arrayidx18 = getelementptr inbounds
[12 x i16], ptr @Qumem, i64 0, i64 1
 %arrayidx19 = getelementptr inbounds
[12 x i16], ptr @Qumem, i64 0, i64 2
  call void @ Z4CNOTRtS (ptr %arrayidx18,
ptr %arrayidx19)
```

IR Form (Post Unrolling)

br label %agcc.meas.move.end

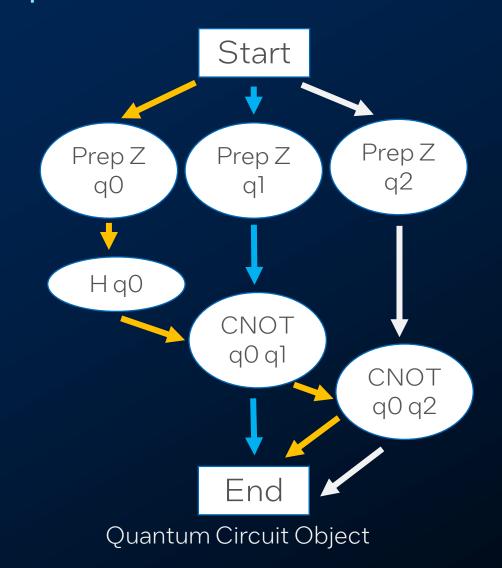


```
aqcc.quantum:
    ...
    %arrayidx33 = getelementptr inbounds [12
x i16], ptr @Qumem, i64 0, i64 1
    ...
    %arrayidx20 = getelementptr inbounds [12
x i16], ptr @Qumem, i64 0, i64 0
    %arrayidx21 = getelementptr inbounds [12
x i16], ptr @Qumem, i64 0, i64 1
    call void @_Z4CNOTRtS_(ptr %arrayidx20, ptr %arrayidx21)
    ...
```



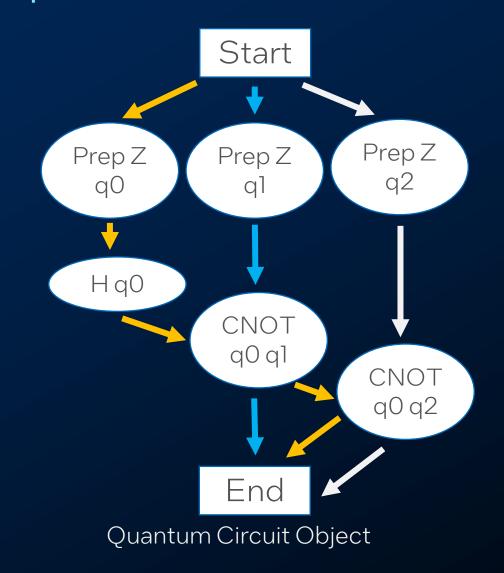
#### Supported Options

- Deletion
- Insertion
- Moving Operations
- New Operations
- New Qubits
- Consistent Iteration



#### Supported Options

- <u>Deletion</u>
- Insertion
- Moving Operations
- New Operations
- New Qubits
- Consistent Iteration



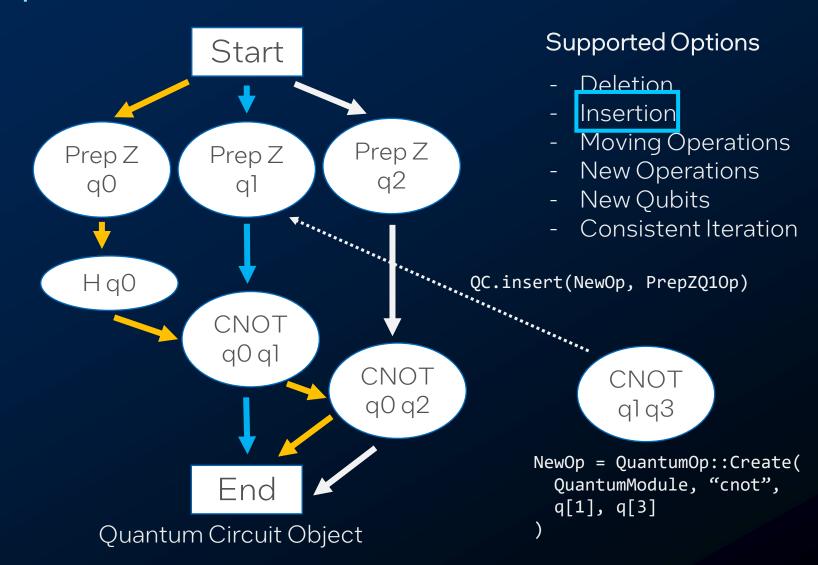
#### Supported Options

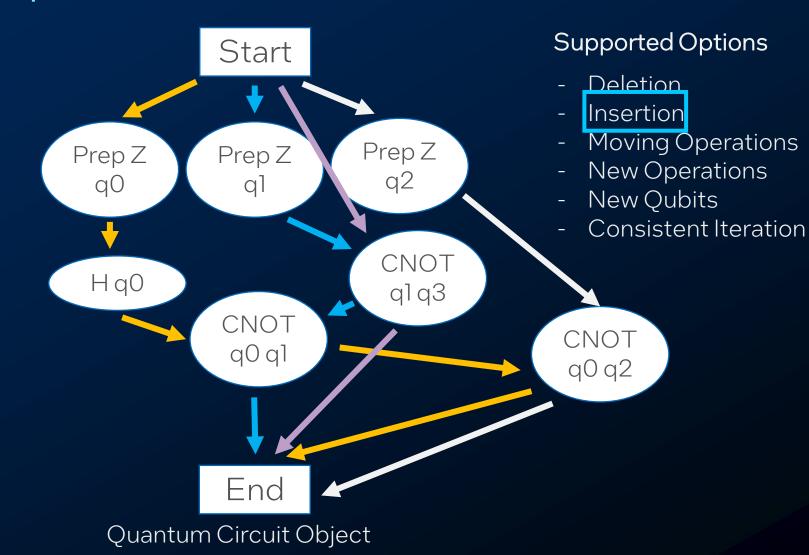
- <u>Deletion</u>
- Insertion
- Moving Operations
- New Operations
- New Qubits
- Consistent Iteration

```
CNOT
qlq3

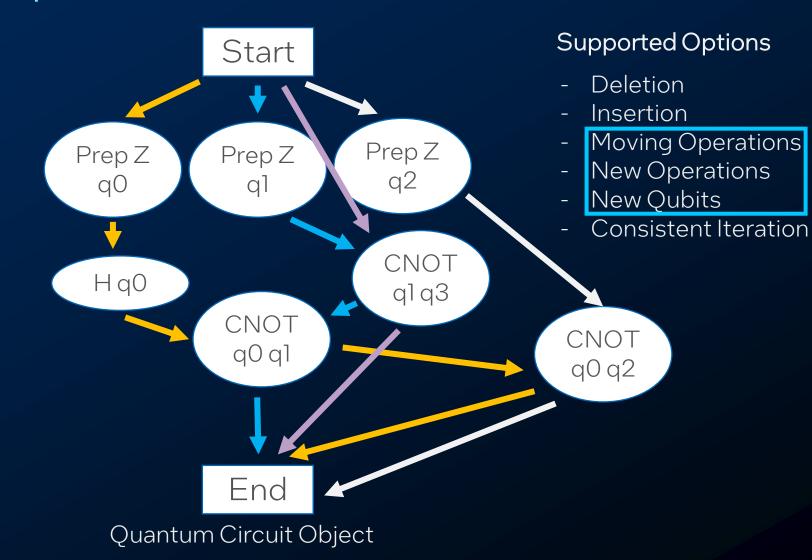
NewOp = QuantumOp::Create(
QuantumModule, "cnot",
q[1], q[3]
```

```
aqcc.quantum:
    ...
    %arrayidx33 = getelementptr inbounds [12
x i16], ptr @Qumem, i64 0, i64 1
    ...
    %arrayidx20 = getelementptr inbounds [12
x i16], ptr @Qumem, i64 0, i64 0
    %arrayidx21 = getelementptr inbounds [12
x i16], ptr @Qumem, i64 0, i64 1
    call void @_Z4CNOTRtS_(ptr %arrayidx20, ptr %arrayidx21)
    ...
```





```
aqcc.quantum:
    ...
    %arrayidx33 = getelementptr inbounds
[12 x i16], ptr @Qumem, i64 0, i64 1
    ...
% arrayidx21 = getelementptr inbounds [12 x i16], ptr @Qumem, i64 0, i64 1
    %arrayidx02 = getelementptr inbounds
[12 x i16], ptr @Qumem, i64 0, i64 3
    call void @_Z4CNOTRtS_(ptr %arrayidx01, ptr %arrayidx02)
    %arrayidx20 = getelementptr inbounds
[12 x i16], ptr @Qumem, i64 0, i64 0
    call void @_Z4CNOTRtS_(ptr %arrayidx20, ptr %arrayidx21)
    ...
```



## Conclusions



#### Conclusions

Using LLVM is a powerful part of the Intel Quantum Compiler

Flexibility and Plugins give powerful tools to high-level developers

LLVM is complex, the Quantum Circuit Object can alleviate some difficulties

#### Open Source

- Improvements in structure
- More efficient ways to use LLVM
- Quantum Optimizations from researchers



https://developer.intel.com/quantumsdk



https://github.com/intel/ quantum-passes



https://github.com/intel/ quantum-intrinsics



https://developer.intel.com/quantumsdk

# Questions?



https://github.com/intel/ quantum-passes



https://github.com/intel/ quantum-intrinsics

intel foundry