The 7th Asian Tier Center Forum, 2023

Driving massively scalable simulations of quantum circuits in supercomputers

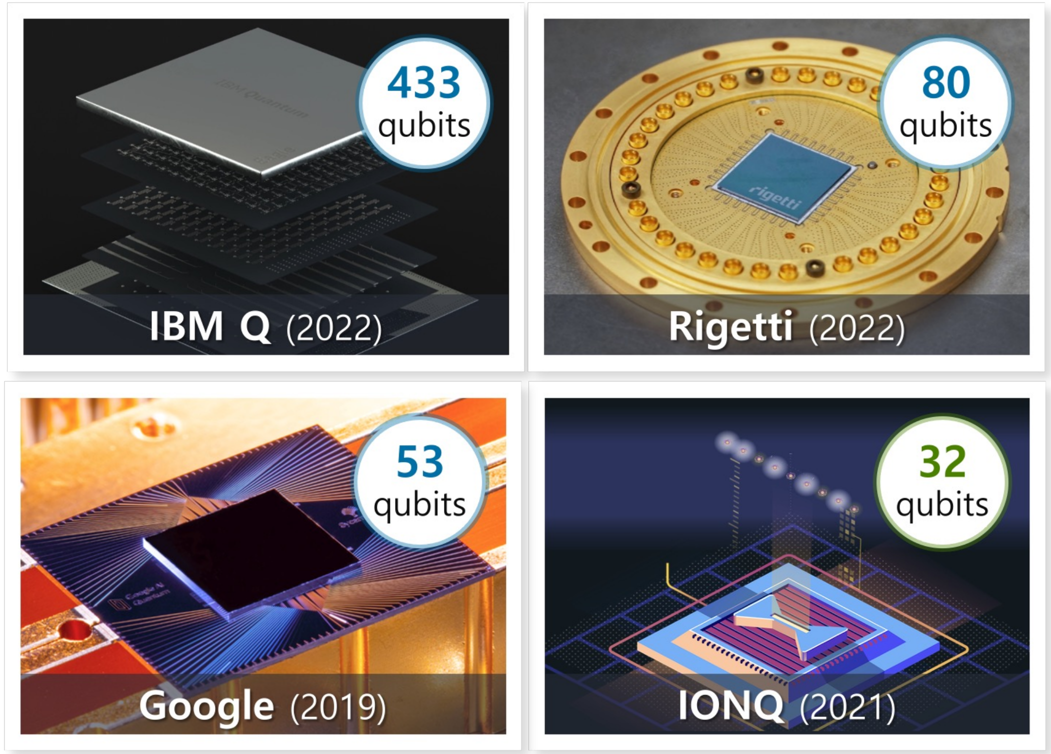
Hoon Ryu, Ph.D. (E: elec1020@kisti.re.kr)

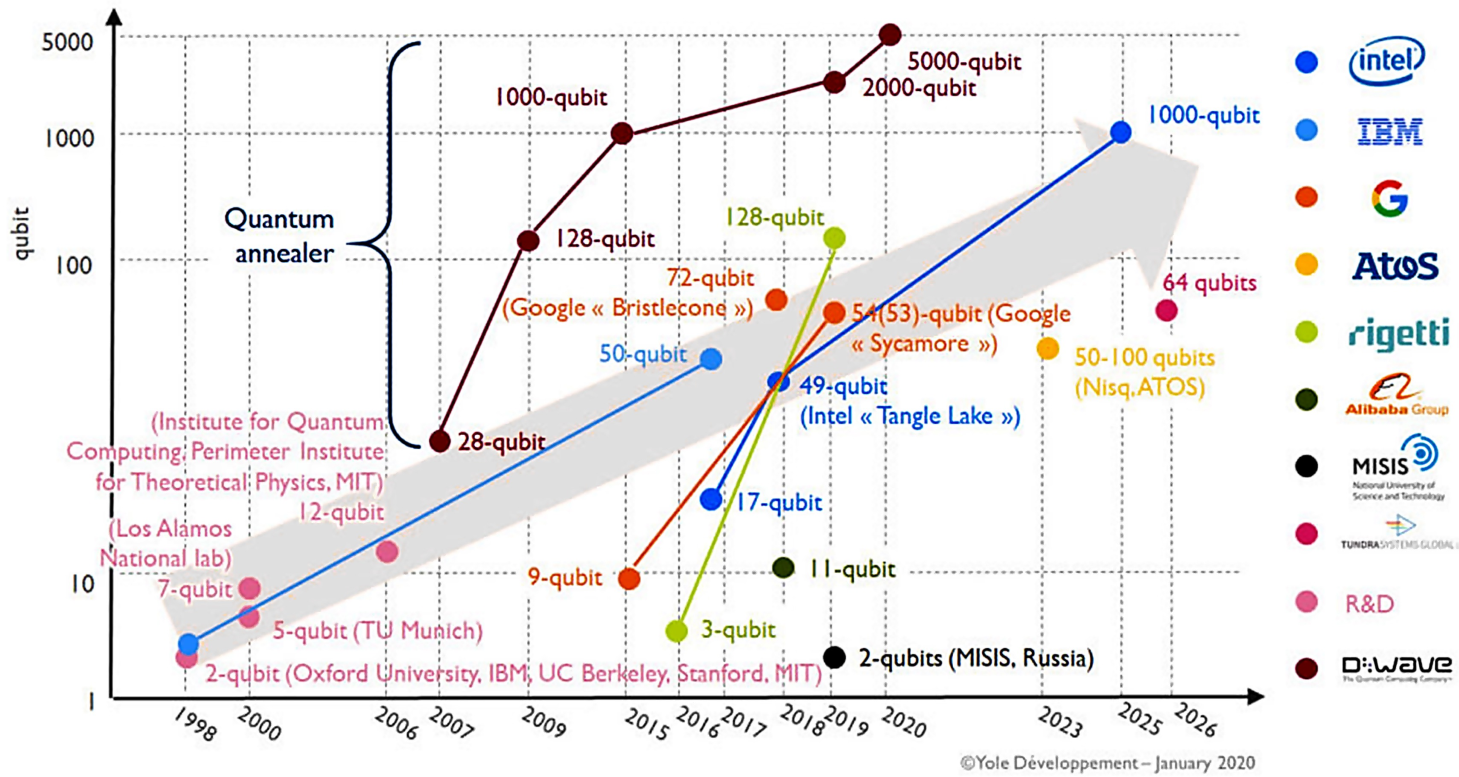
Principal Researcher
Lead of Quantum Information R&D Group
Korea Institute of Science and Technology Information

Existing Platforms of Physical Qubits

Current status of "circuit-based" quantum computers

Up-to-date Status of "Universal" Quantum Computers & Technical Roadmap





- Superconductor & trapped ions lead the industry
- Cloud-based service available for processors having physical qubits > 400
- Near-future processors will have over-1000 physical qubits

In Terms of Utilization

Quantum volume & Algorithmic qubits

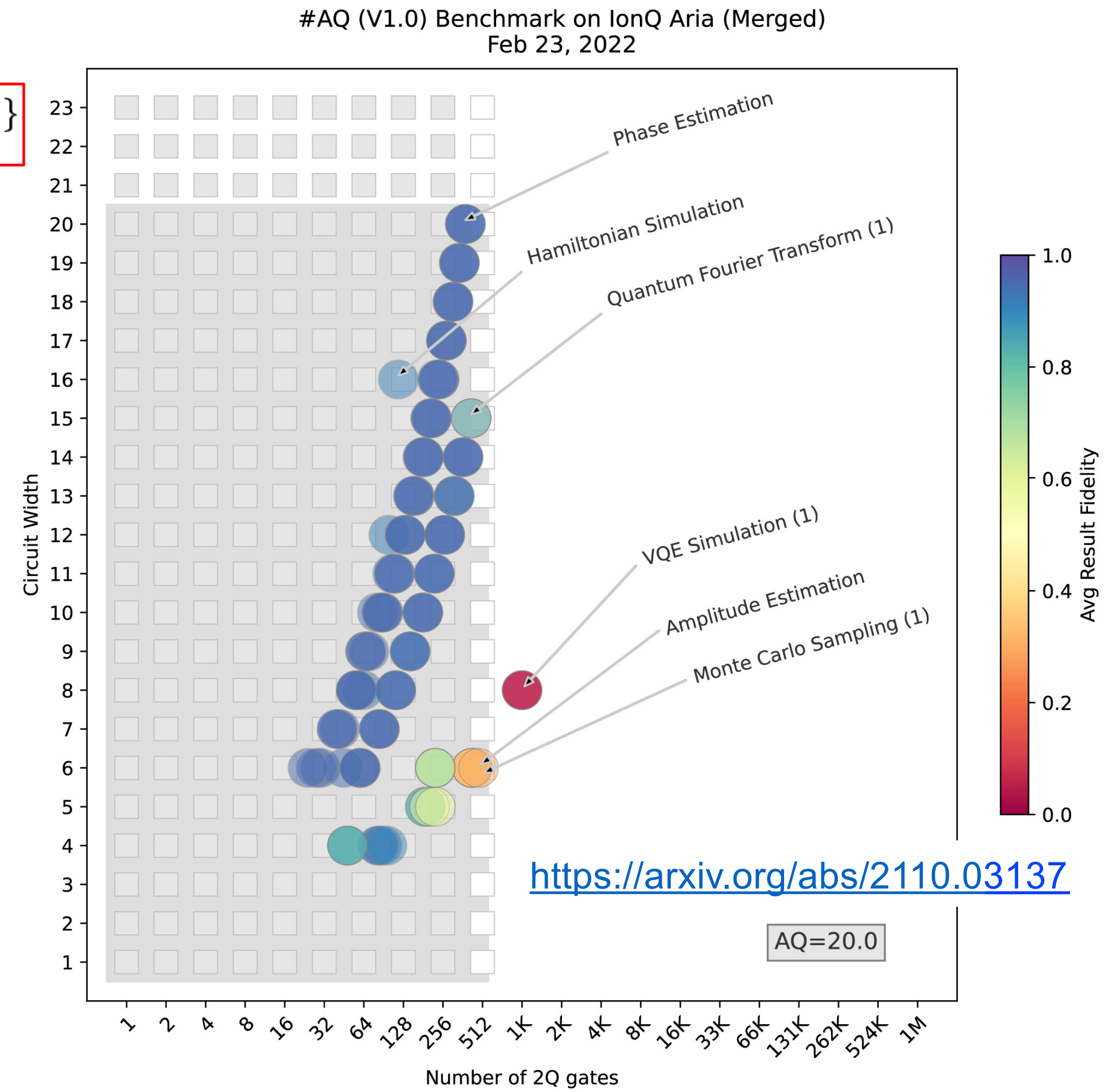
of Physical Qubits ≠ the Real Capability

- Quantum volume & Algorithmic qubit $\log_2 V_Q = \arg\max_{n \le N} \{\min[n, d(n)]\}$

  → Indicators to represent the largest complexity of algorithm circuits that a QPU device can run

Date	VQ	Notes
2020 Aug	64 = 2^6 (6 qubits)	Falcon R4 “Montreal” (27 physical qubits) [1]
2020 Dec	128 = 2^7 (7 qubits)	Falcon R4 “Montreal” (27 physical qubits) [2]
2022 Apr	256 = 2^8 (8 qubits)	Falcon R10 “Prague” (32 physical qubits) [3]
2022 May	512 = 2^9 (9 qubits)	Falcon R10 “Prague” (32 physical qubits) [4]

[1] https://www.zdnet.com/article/ibm-hits-new-quantum-computing-milestone/
[2] https://twitter.com/jaygambetta/status/1334526177642491904
[3] https://research.ibm.com/blog/quantum-volume-256
[4] https://twitter.com/jaygambetta/status/1529489786242744320



https://arxiv.org/abs/2110.03137

Large-scale Logic, e.g., Ones in the NISQ Region?

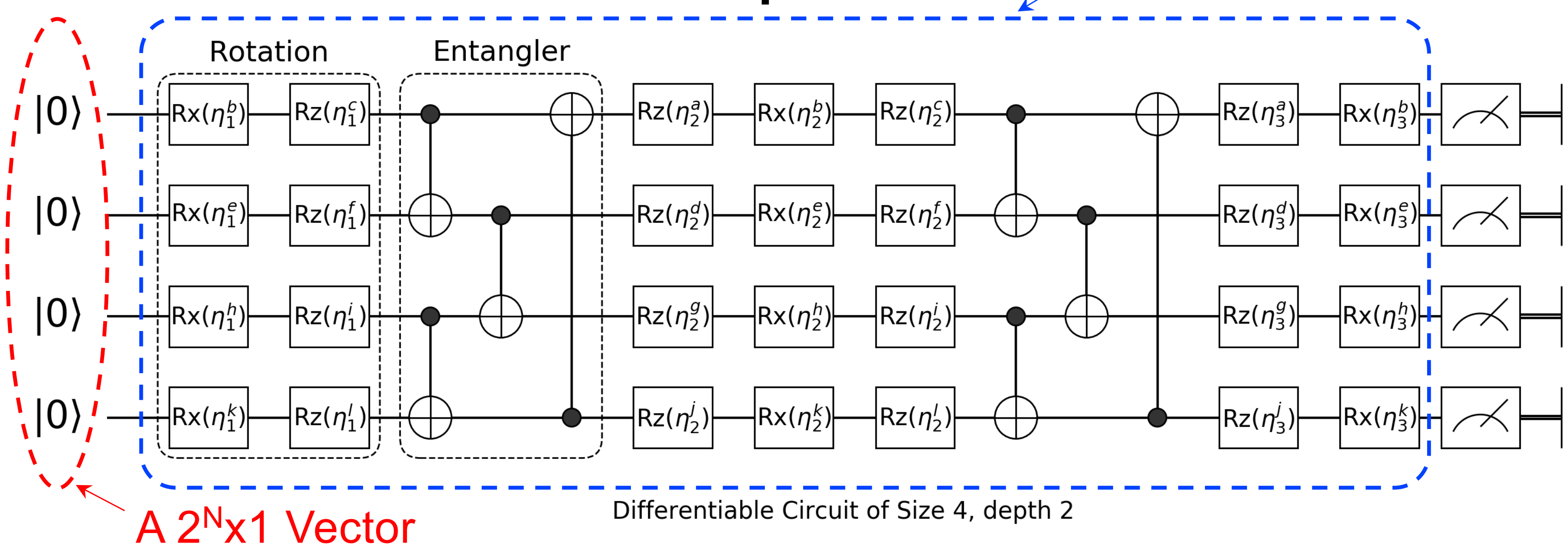
- Needs for simulations of quantum circuits in a very huge computing environment → *a.k.a.* Supercomputer

Classical Treatment of Quantum Logic Operations

Classical representation of gate-based quantum circuits

In General: Matrix-vector Multiplier

A 2^Nx2^N Unitary Matrix



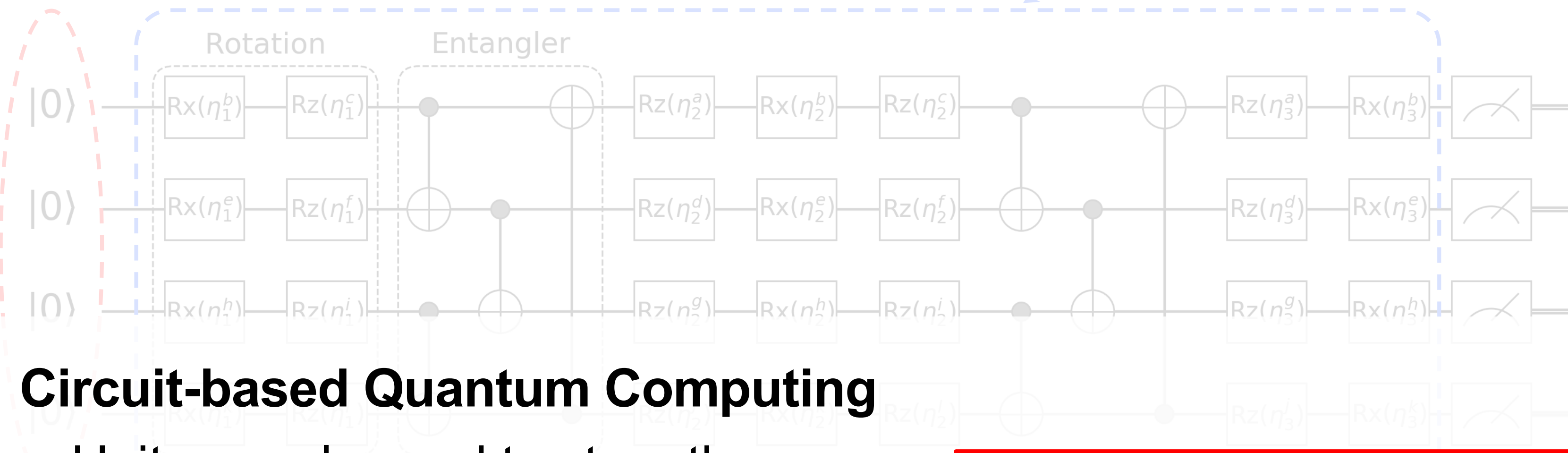
- All complex-valued
- Huge memory consumption for representation of the unitary
  - → N = 20 → 16 TB
  - → Reduction can be done for specific cases; (e.g.) indices for nonzeros are known in advance)

Classical Treatment of Quantum Logic Operations

Classical representation of gate-based quantum circuits

In General: Matrix-vector Multiplier

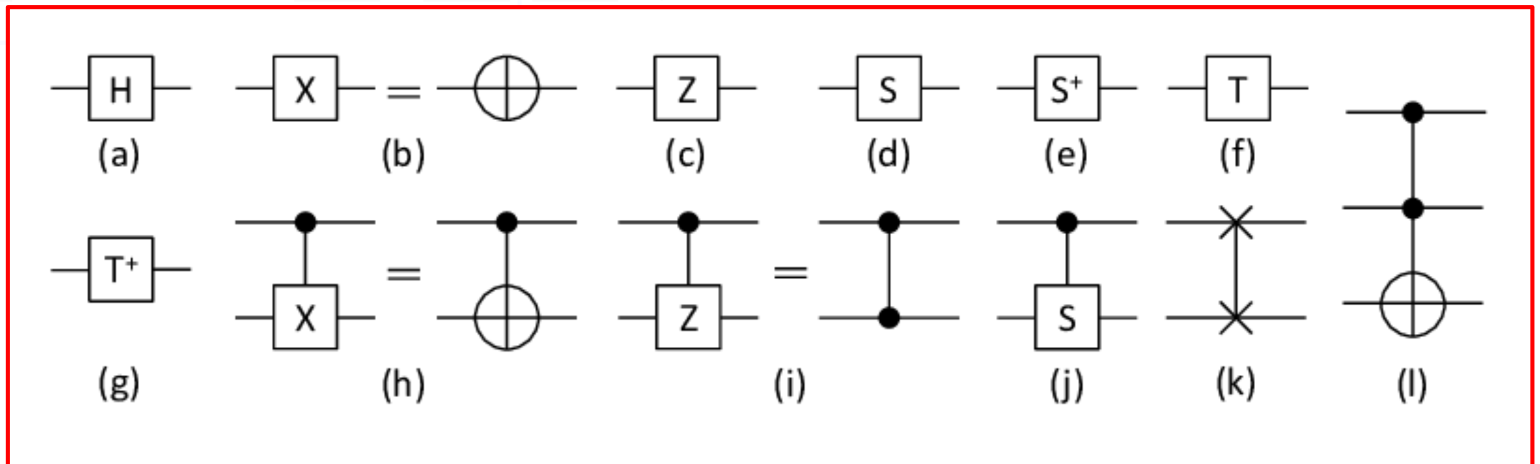
A $2^N x 2^N$ Unitary Matrix



- All complex-valued
- Huge memory consumption for representation of the unitary
  - → N = 20 ~ 16 TBytes
  - → Reduction can be done for specific cases; (e.g.) indices for nonzeros are known in

Circuit-based Quantum Computing

- Unitary: only need to store those for universal gates
- State vectors (in principle) must be fully stored
- Cares must be put for conduction of matrix-vector multiplication
  - → The size of unitary is not equal to that of a state vector



Classical Treatment of Quantum Logic Operations

Classical representation of gate-based quantum circuits

Matrix-vector Multiplier: Circuit-based Quantum Computing

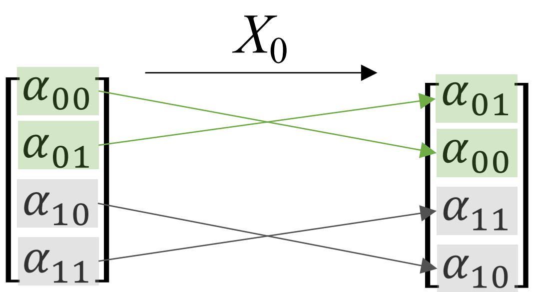
- A simple example: Conduction of X (Pauli-X) gating

X gating against a two-qubit state

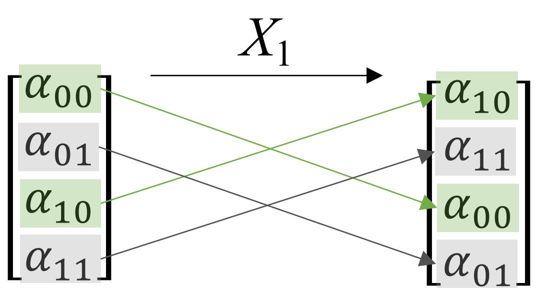
$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

X gating against a N-qubit state

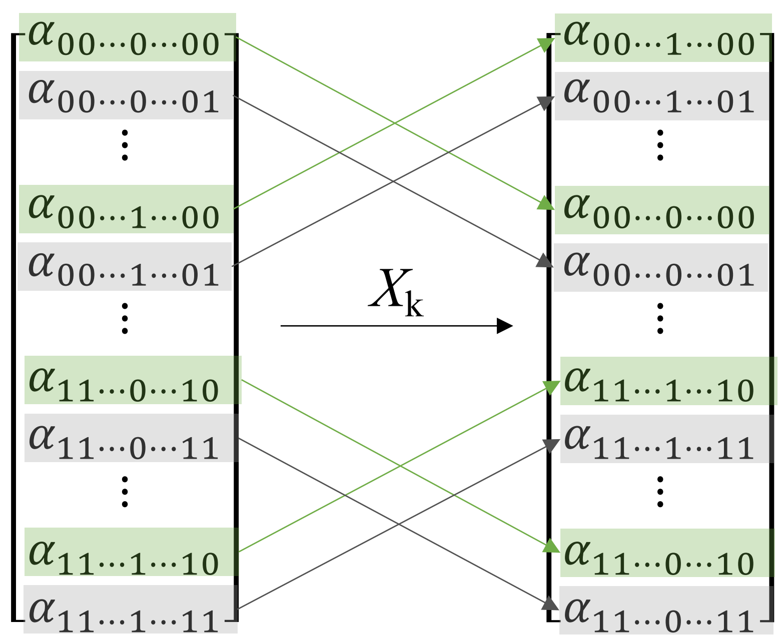
On the right qubit (0th)



On the left qubit (1st)



On k-th qubit



- Mapping of state indices where corresponding elements need to be updated according to the logic operation

  → Details depend on the type (category) of universal gates

- Size of circuits to be simulated

  → Memory consumption required by a quantum state

Large-scale Circuit Simulations

Objective of this talk

Memory Consumption of State Representation: BIG DEAL!

- A single 30-qubit state: 2^{30} elements (amplitudes) x 16 Bytes = 16 GB
  → A 40-qubit one: 16 TB & A 50-qubit one: 16 PB
  → Total memory (8,305 nodes) of the National Supercomputer of Korea: ~778.6 TB (~0.76 PB)

Large-scale Circuit Simulations in Classical Computers?

- A distributed computing system: physically separated nodes that are connected with network
  → Can use the whole memory with communications (**distributed computing**)
- Can use storage & partially load the state vector as needed

What we cover in this talk…

- A SW package for classical simulations with a distributed computing
- A brief overview of the cloud-based service framework currently under development: the gateway for public service of the code package

Workload Parallelization

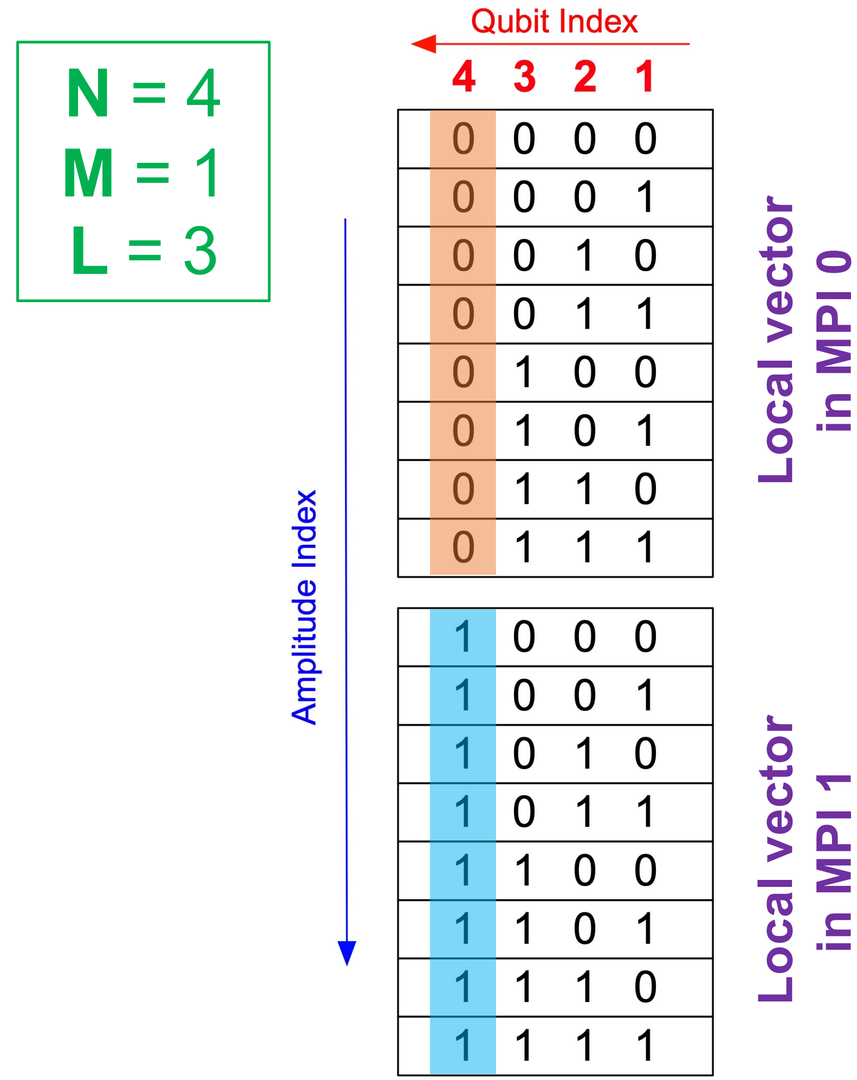
Distributed computing with Message Passing Interface (MPI)

Decomposition of State Vectors

- Decomposed blocks are stored in different memory locations
  → Local to each MPI process
- (e.g.) Let's say that 2^N amplitudes of a **N**-qubit state are distributed over 2^M MPI processes
  → Each MPI has a local vector of $2^L = 2^{(N-M)}$ amplitudes, where **L** indicates the qubit size of a local state

Index-dependent Parallel Operations of Universal Gates

- SU(4): U(**a**,**b**) where **a** & **b** are qubit-indices against which a gating operation is conducted
  → If both **a** & **b** <= **L**, then no communication is needed among inter-MPI processes: Embarrassingly Parallel (EP)
  → MPI communications must happen otherwise: operation against a local vector may update another local vectors allocated in other MPI processes. (e.g.) SWAP(2,4)



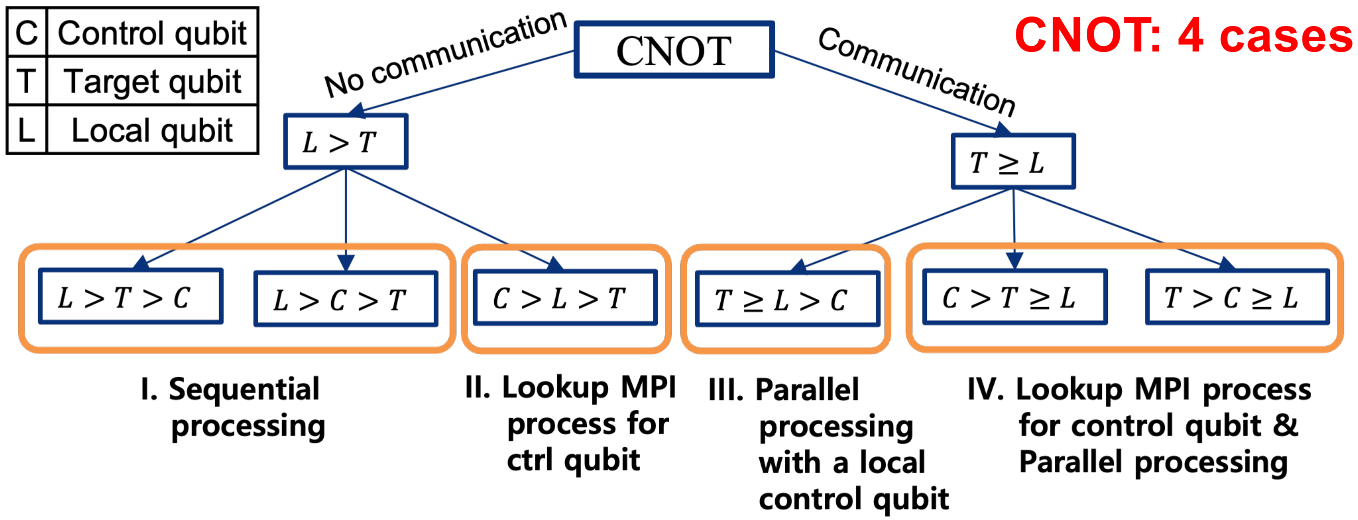
Workload Parallelization

Distributed computing with Message Passing Interface (MPI)

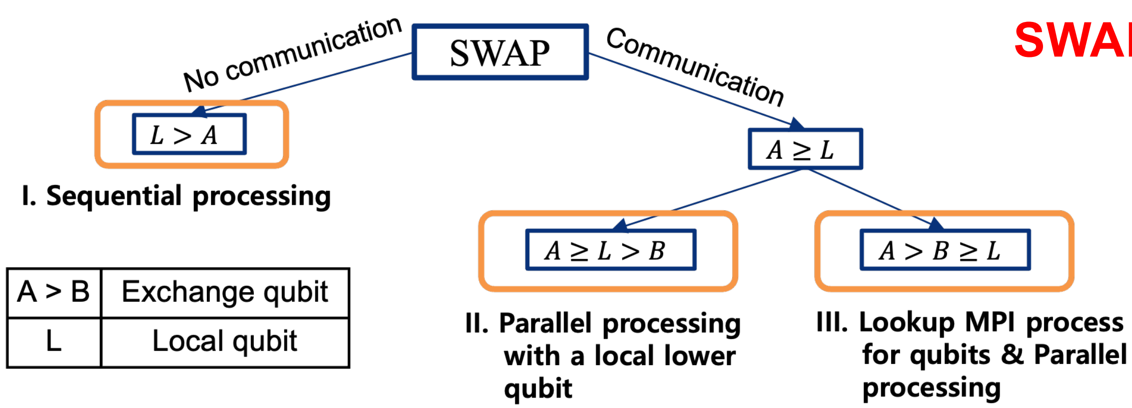
Index-dependent Parallel Operations of Gates (Cont.)

C	Control qubit
T	Target qubit
L	Local qubit

CNOT: 4 cases



SWAP: 3 cases



A > B	Exchange qubit
L	Local qubit

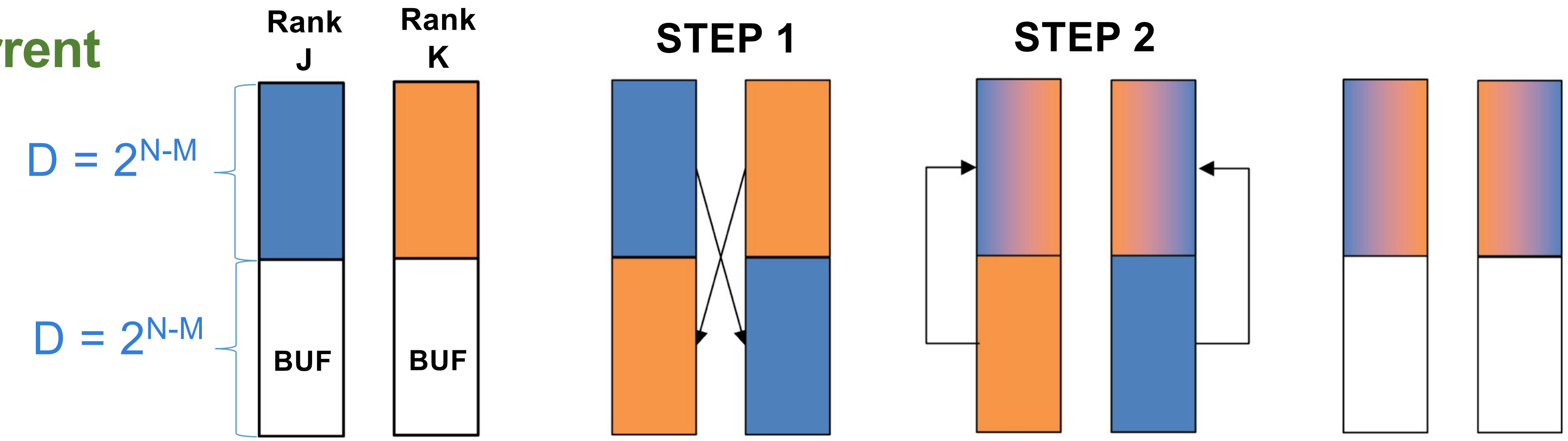
Ops. supporting a parallel computing (so far)

BasisState	Prepares a single computational basis state.
CNOT	The controlled-NOT operator
CROT	The controlled-Rot operator
CRX	The controlled-RX operator
CRY	The controlled-RY operator
CRZ	The controlled-RZ operator
Hadamard	The Hadamard operator
PauliX	The Pauli X operator
PauliY	The Pauli Y operator
PauliZ	The Pauli Z operator
PhaseShift	Arbitrary single qubit local phase shift
ControlledPhaseShift	The controlled phase shift.
QubitStateVector	Prepare subsystems using the given ket vector in the computational basis.
ROT	Arbitrary single qubit rotation
RX	The single qubit X rotation
RY	The single qubit Y rotation
RZ	The single qubit Z rotation
S	The single-qubit phase gate
T	The single-qubit T gate

Workload Parallelization

Communication between MPI processes: State vectors

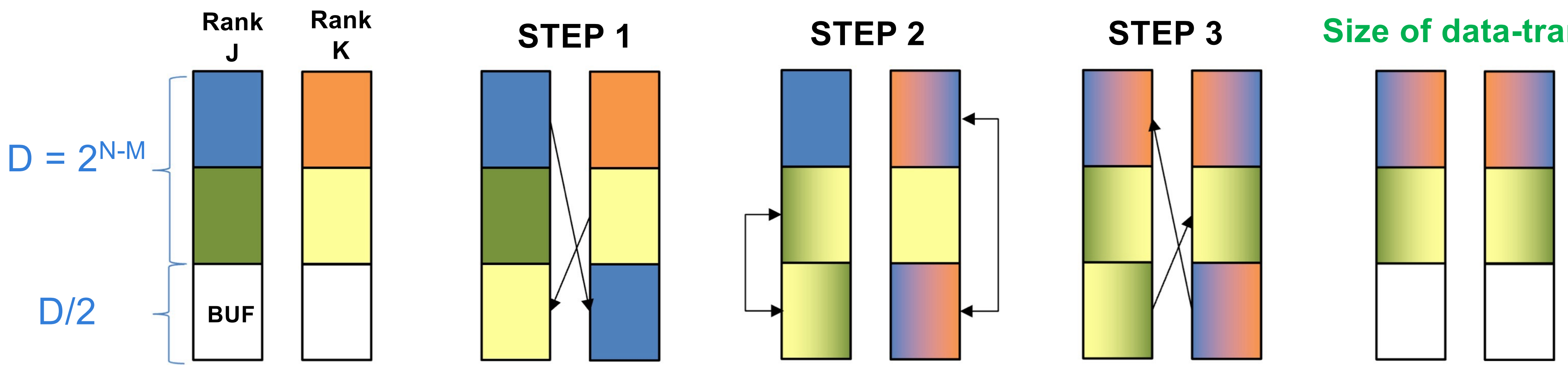
Current



1. N-qubit (# of amp. = 2^N)
2. 2^M MPI processors
3. Size of local vector = D = 2^L = 2^{N-M}

Size of data-transfer = D

Buffer-saving Approach



Size of data-transfer = D/2

Scalability: Element-gate Operation

Index-dependent performance

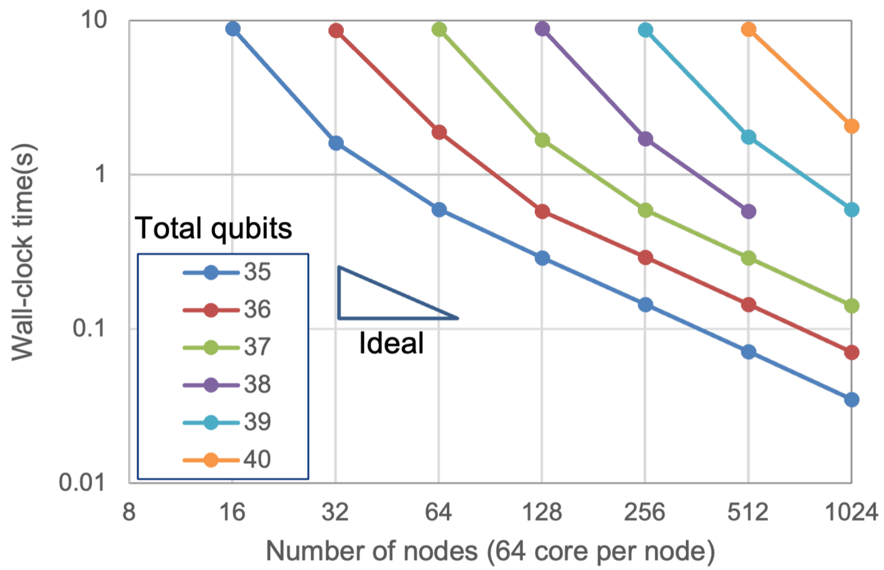
Nat'l Supercomputer of ROK
(The NURION System)

Node Spec.
- Intel® Xeon Phi KNL 7250
- Single processor / 68 cores
- 96GB DDR4

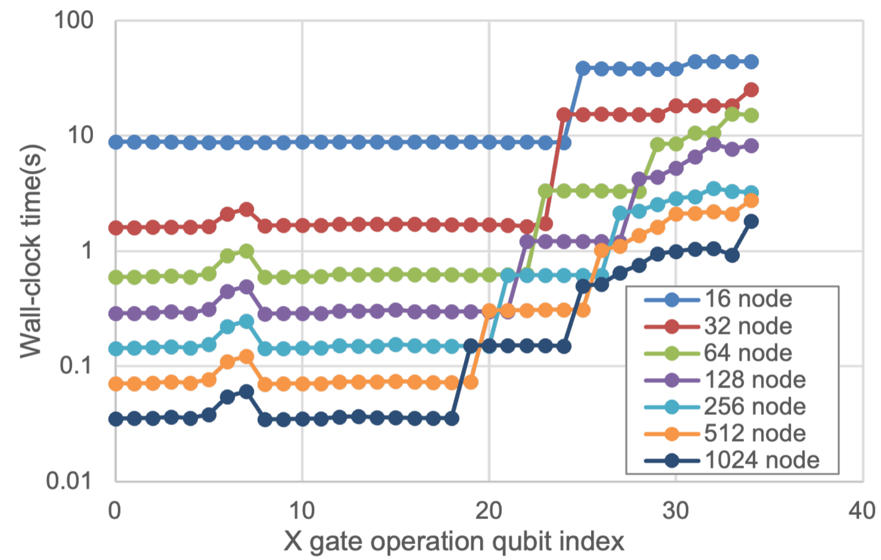
8,305 Computing Nodes
- Total DRAM ~ 0.76PB
- Up to 44-qubit circuits

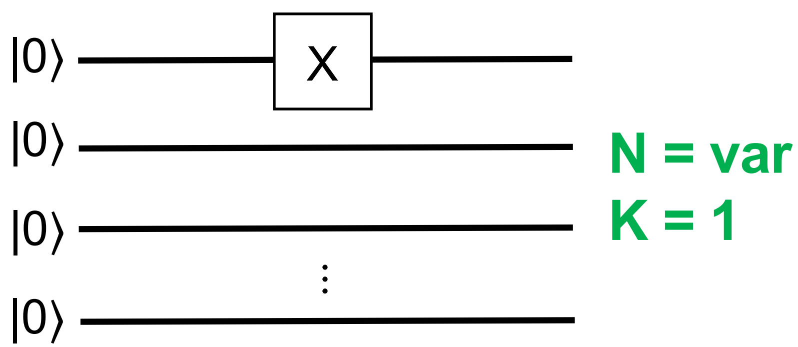
Compiler & Setup
- MVAPICH2 2.3.6
- GNU 10.2
- MPI-only (64 procs/node)

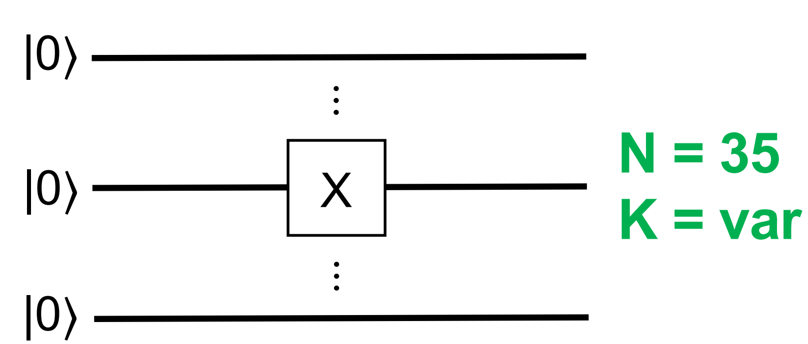
X on the first qubit



X on the K^{th} qubit & N = 35







Scalability: Element-gate Operation

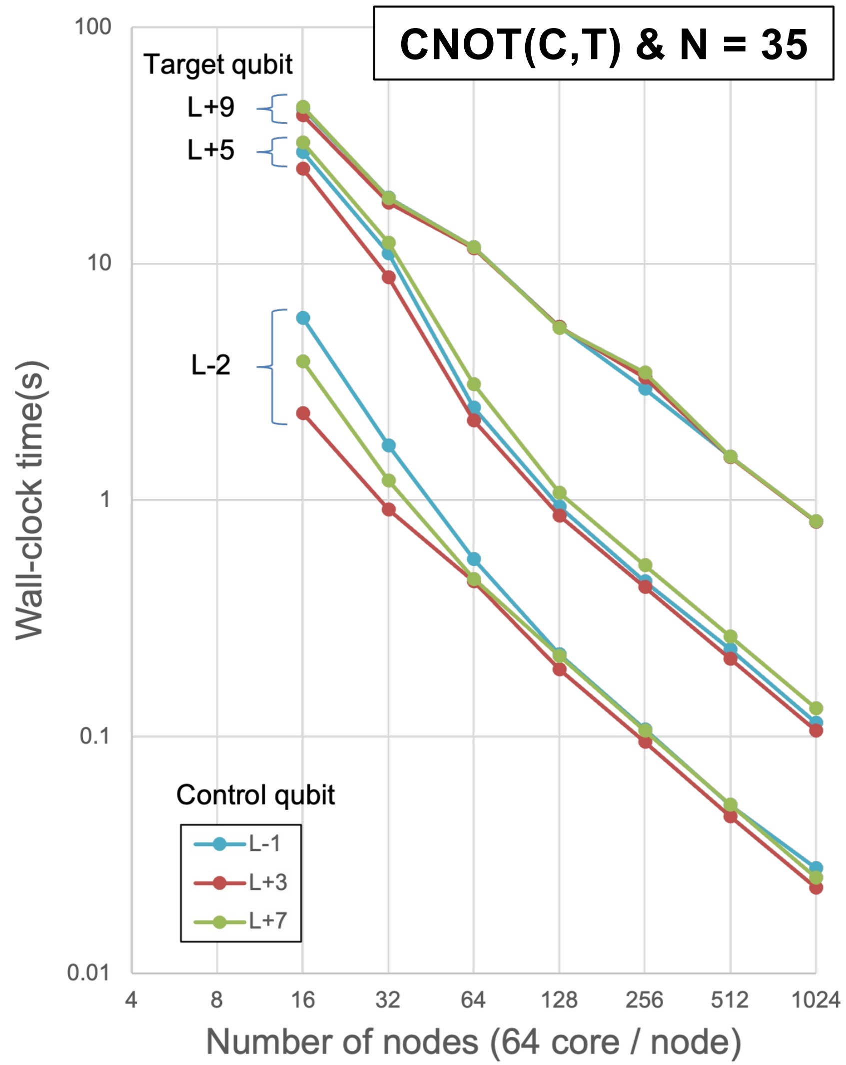
Index-dependent performance

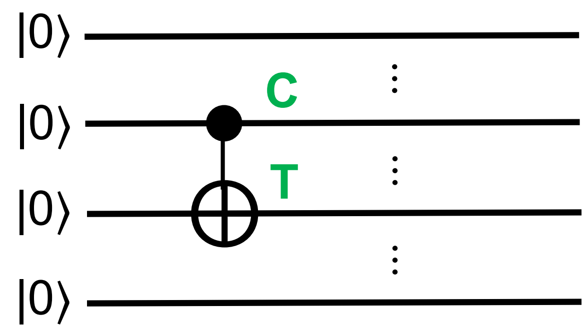
Nat'l Supercomputer of ROK
(The NURION System)

Node Spec.
- Intel® Xeon Phi KNL 7250
- Single processor / 68 cores
- 96GB DDR4

8,305 Computing Nodes
- Total DRAM ~ 0.76PB
- Up to 44-qubit circuits

Compiler & Setup
- MVAPICH2 2.3.6
- GNU 10.2
- MPI-only (64 procs/node)





N = 35
C & T = var
('C > T' is possible)
$2^L = 2^N/N_{MPI}$

Messages

- T (index of target qubit) is equal to or less than L (size of local qubit)
  → No data-transfer via MPI comm.
- T > L
  → Data-transfer via MPI comm.
  → Communication overhead increases as T >> L

Scalability: A Realistic Case

Universal quantum circuit for N-qubit quantum gate

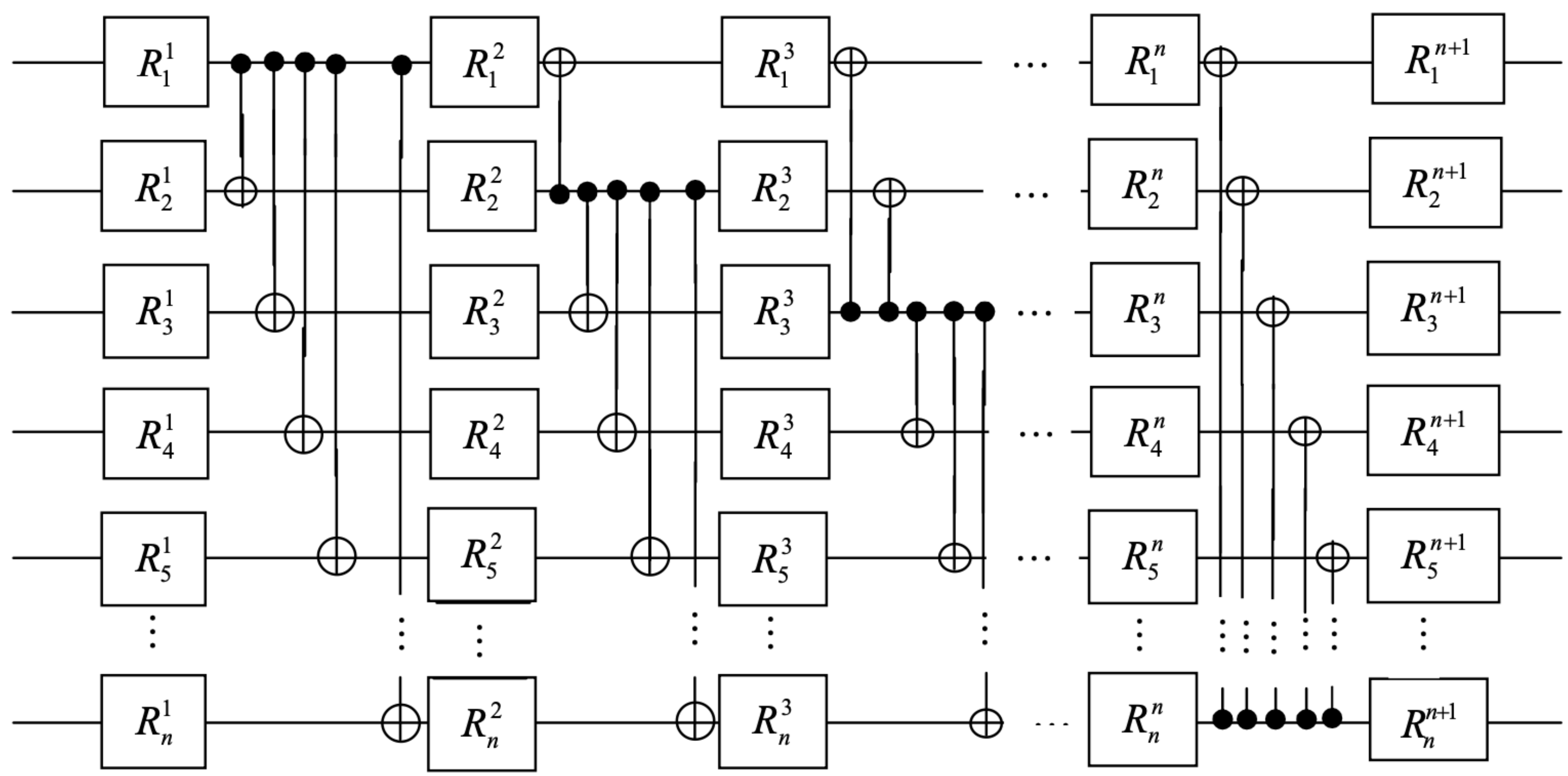
Nat'l Supercomputer of ROK
(The NURION System)

Node Spec.
- Intel® Xeon Phi KNL 7250
- Single processor / 68 cores
- 96GB DDR4

8,305 Computing Nodes
- Total DRAM ~ 0.76PB
- Up to 44-qubit circuits

Compiler & Setup
- MVAPICH2 2.3.6
- GNU 10.2
- MPI-only (64 procs/node)

P. Sousa *et al.*, *Quantum Information & Computation* **7**(3), 228-242



- 3*N*(N+1) + N*(N-1) parameters, where N = qubit size
- Single case: All the R's = X & All the CNOT's are employed

Scalability: A Realistic Case

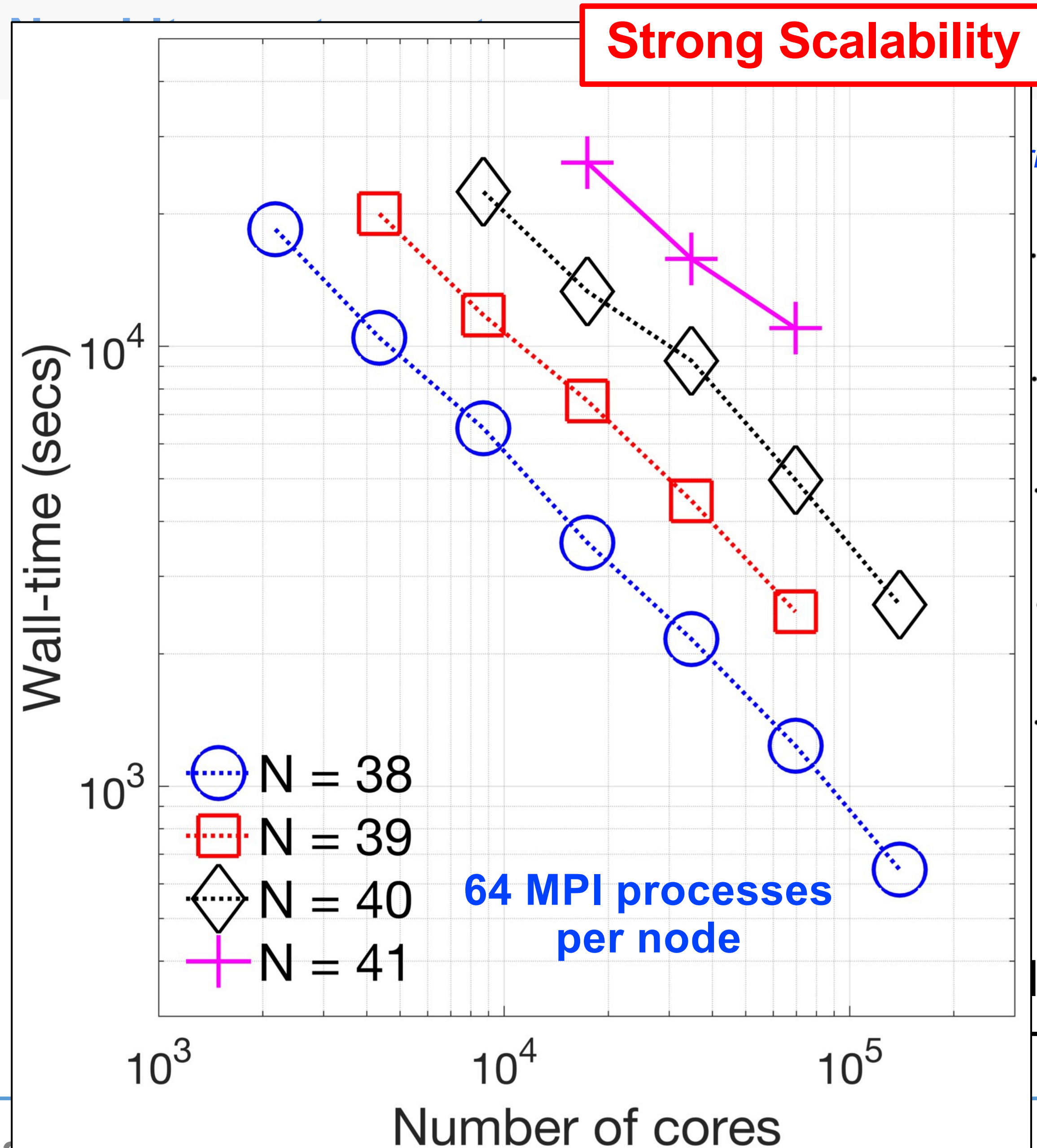
Universal quantum circuit for

Nat'l Supercomputer of ROK
(The NURION System)

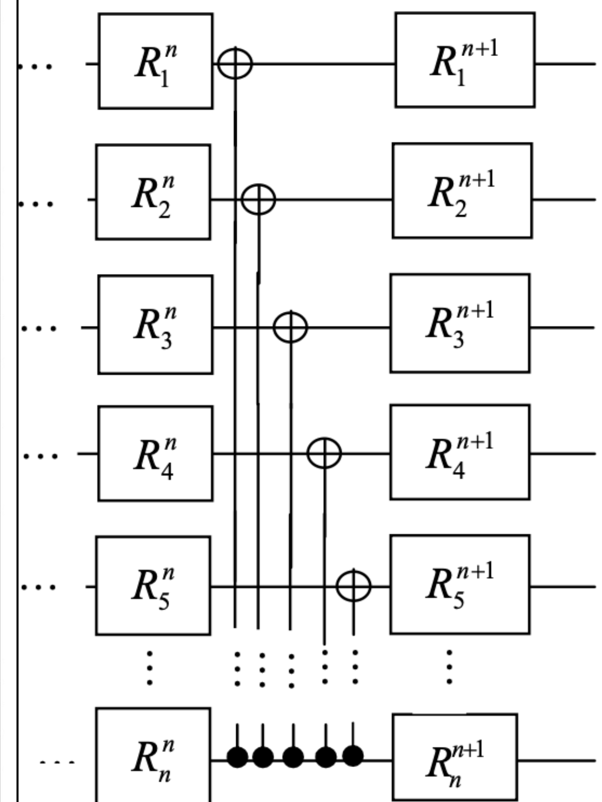
Node Spec.
- Intel® Xeon Phi KNL 7250
- Single processor / 68 cores
- 96GB DDR4

8,305 Computing Nodes
- Total DRAM ~ 0.76PB
- Up to 44-qubit circuits

Compiler & Setup
- MVAPICH2 2.3.6
- GNU 10.2
- MPI-only (64 procs/node)



mation & Computation **7**(3), 228-242



| = qubit size

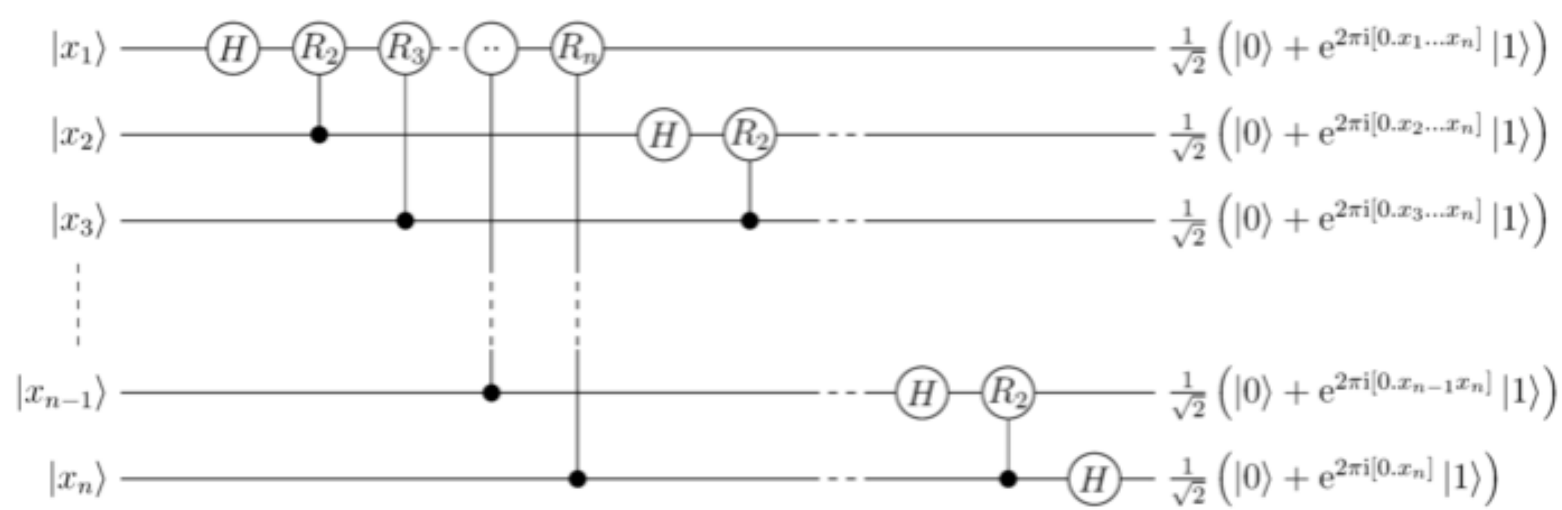
T's are employed

Scalability: A Realistic Case

Universal quantum circuit for N-qubit quantum gate

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad \text{and} \quad R_k = \begin{pmatrix} 1 & 0 \\ 0 & e^{i2\pi/2^k} \end{pmatrix}$$

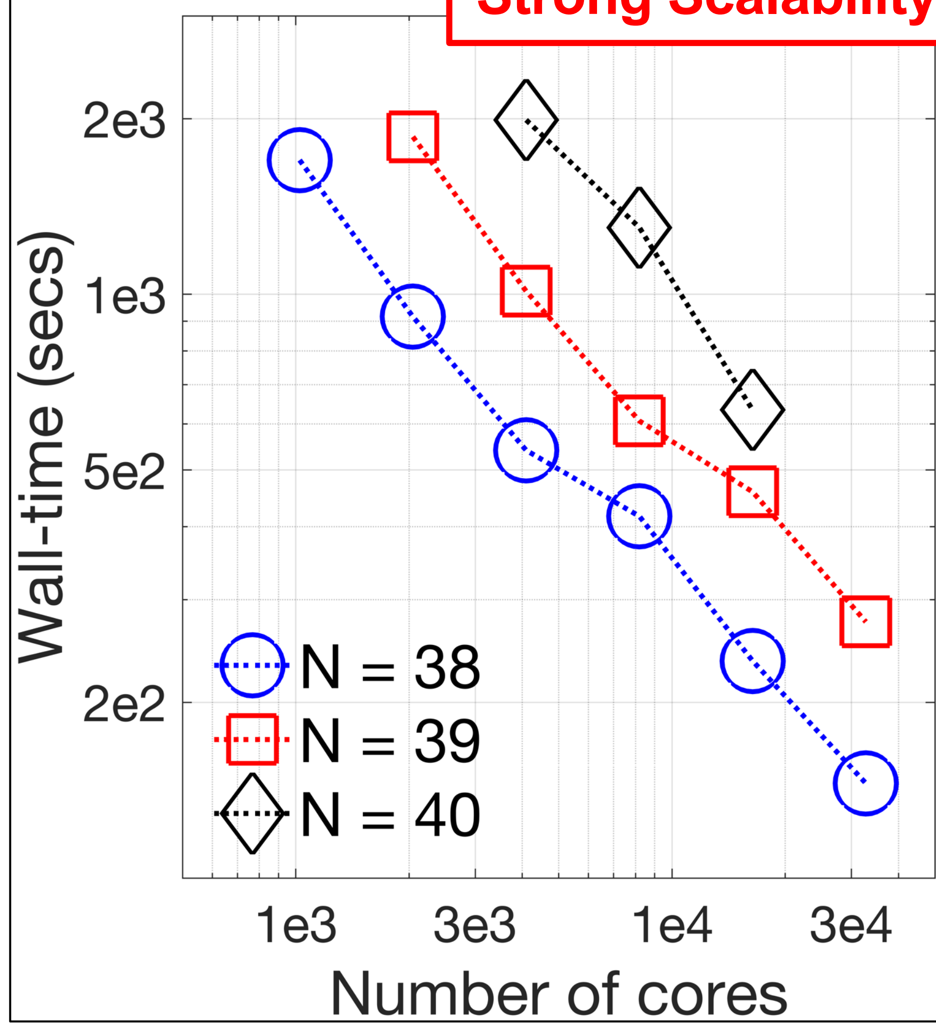
The circuit is composed of H gates and the controlled version of R_k:



- MVAPICH2 2.3.6
- GNU 10.2
- MPI-only (64 procs/node)

Quantum Fourier Transform (QFT)

Strong Scalability

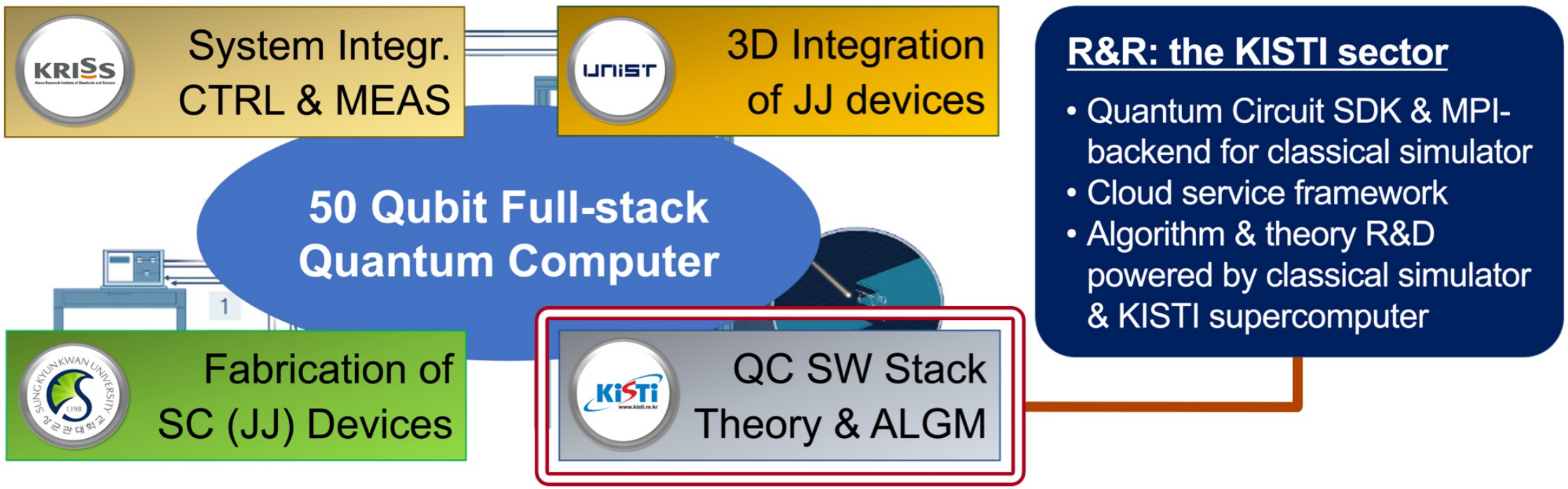


Strategies for Service

National flagship project ongoing in ROK

Project Overview

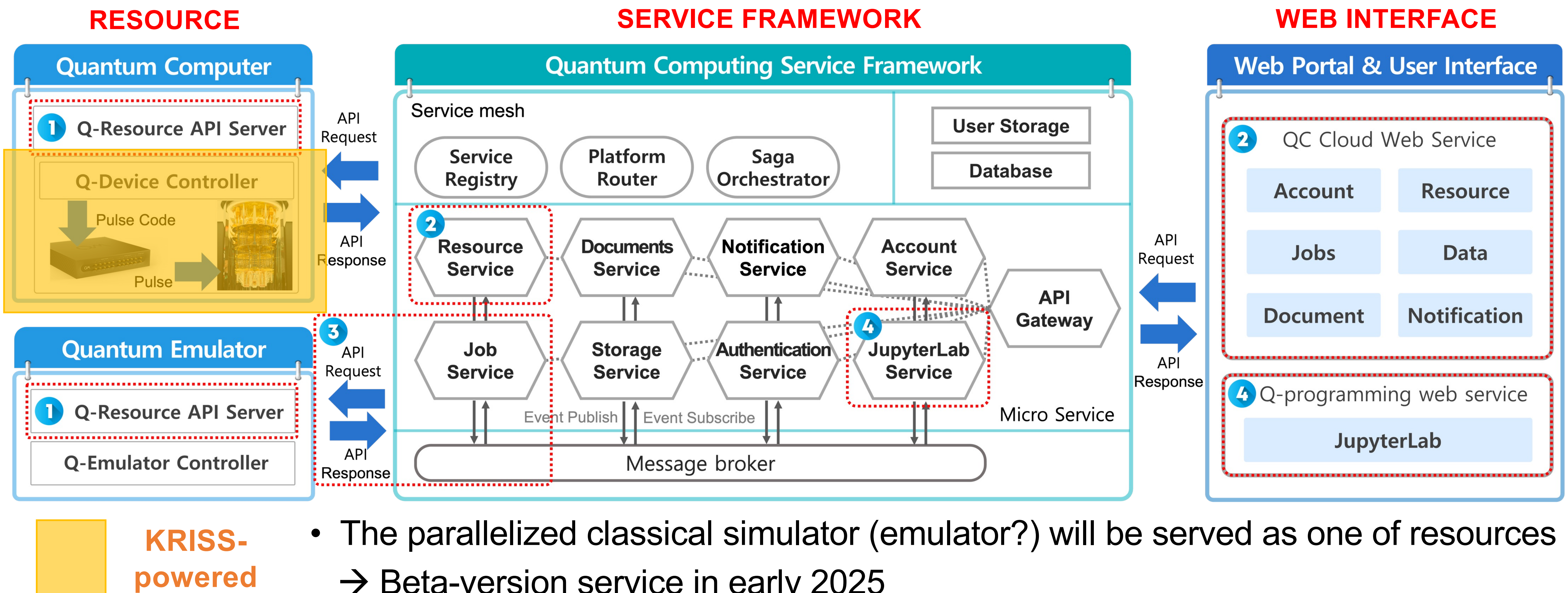
- A full-stack & superconductor-based 50-qubit quantum computer (circuit-based)
  → Project launched in 2022-Jun under support from NRF & MSIT of ROK
- Research consortium and KISTI R&R:



Strategies for Service

Cloud-based service framework

Overview: Technical Components & Flow of KISTI-powered Cloud Service Framework



- The parallelized classical simulator (emulator?) will be served as one of resources → Beta-version service in early 2025

Summary & Remarks

A Massively-scalable Classical Quantum Circuit Simulator

- Message Passing Interface (MPI) to support distributed computing in HPCs
  → A brief discussion on state-mapping & parallelization scheme
- Demonstration: simulations of up to 41-qubit circuits
  → The Universal quantum circuit for N-qubit quantum gates
  → Possible to handle up to 44-qubit circuits in the 5th national HPC of ROK

Overview: KISTI-powered Cloud-based Service Framework

- Target resources: Classical simulator & KRISS-powered quantum computer
- Beta-version service of the simulator through our framework: Early 2025

Thank You for Attention