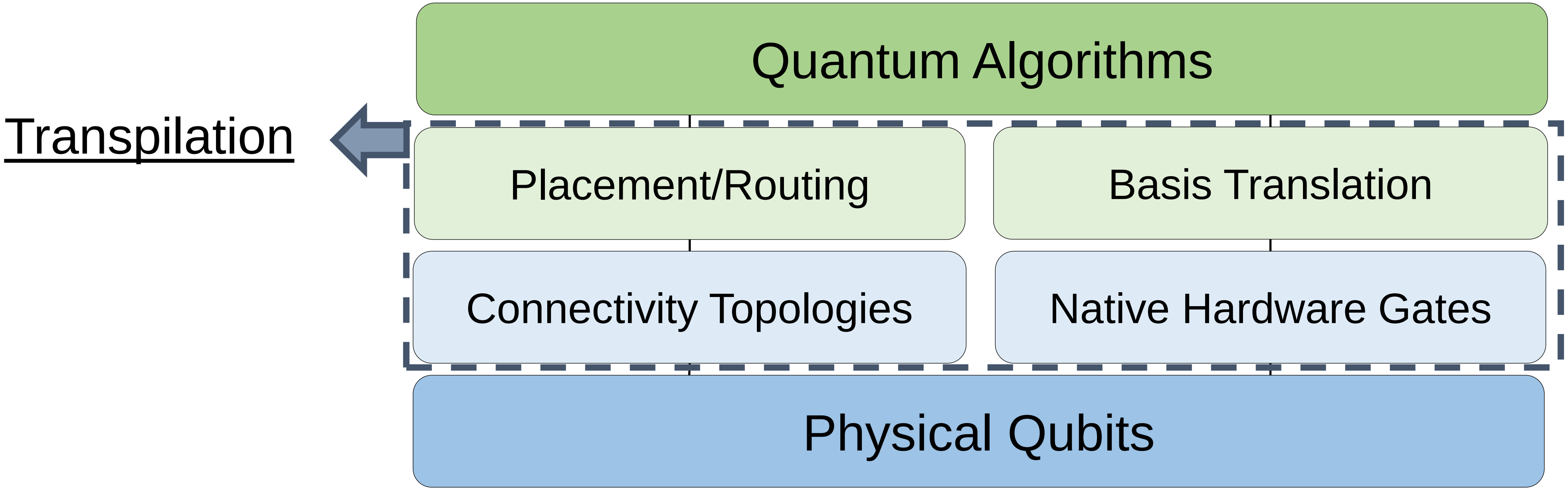
Using Co-Designed Architectures for Modular Superconducting Quantum Computers

Evan McKinney[†],
M. Xia[§], C. Zhou[§], P. Liu[§], M. Hatridge[§], A.K. Jones[†]

[†]Department of Electrical and Computer Engineering, University of Pittsburgh
[§]Department of Physics and Astronomy, University of Pittsburgh

HPCA, 2023

Quantum computer co-design



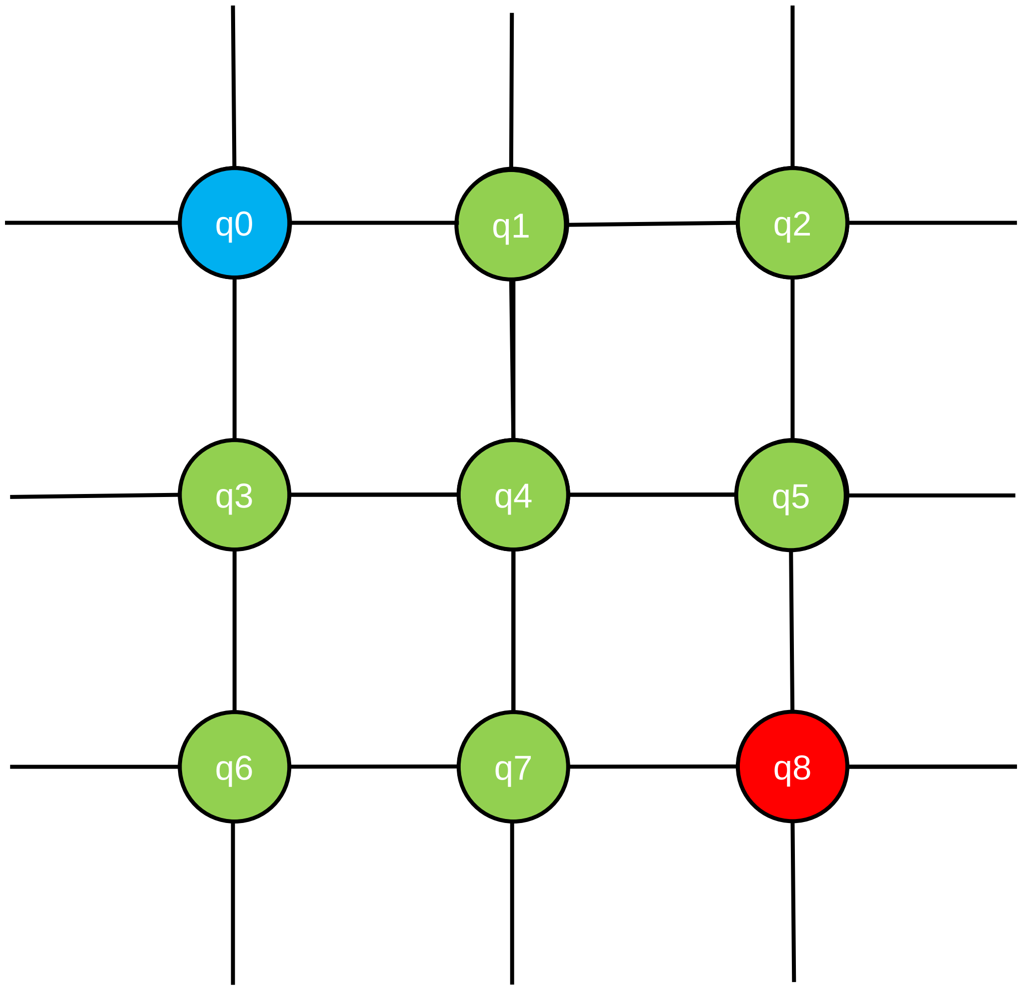
- Physics constrains possible topologies and basis gates
- Prioritize improving qubit and gate fidelities

McKinney, et al. ***arXiv***:*2302.01252* (2023)

McKinney, et al. ***arXiv***:*2302.01252* (2023)

Qubit routing with SWAPs

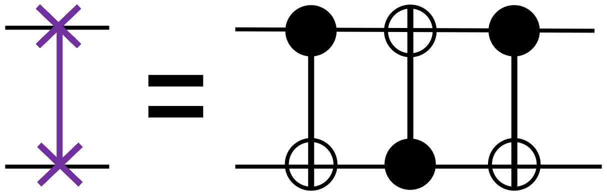
- Example: Square-Lattice Topology



- Inducing SWAP gates on a circuit

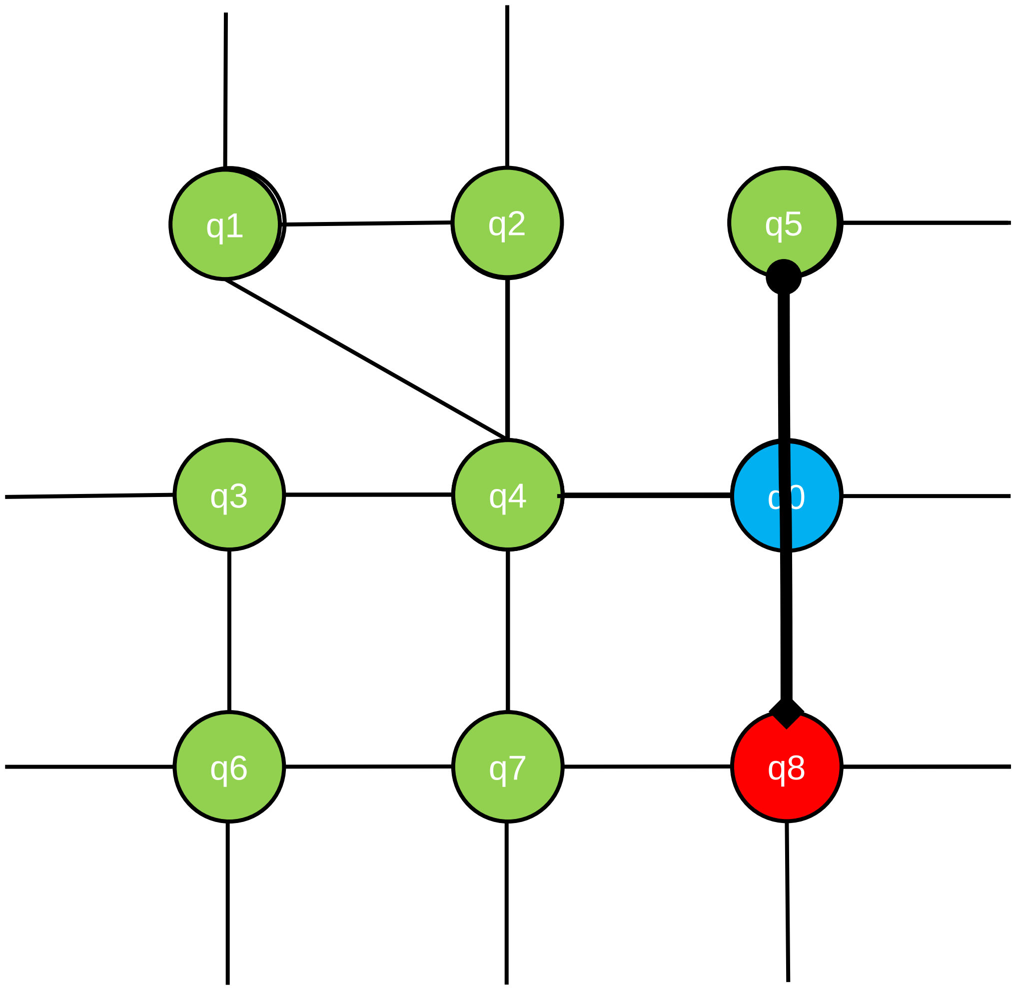
SWAP-minimization is NP-complete

- **SWAPs are expensive**, we desire connectivities which minimize the need for data movement



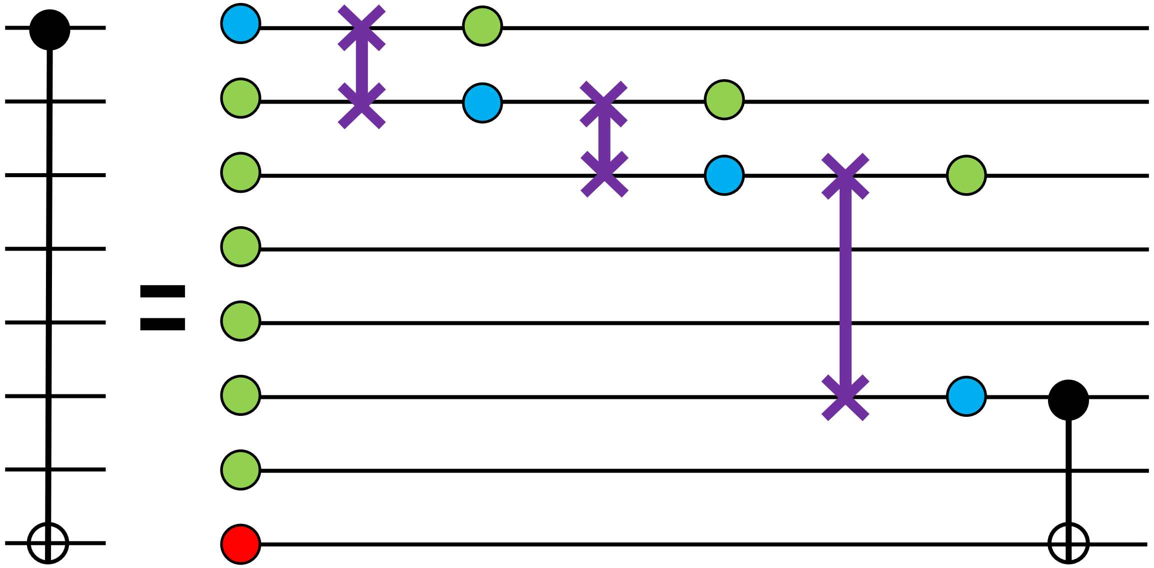
Qubit routing with SWAPs

➢ Example: Square-Lattice Topology

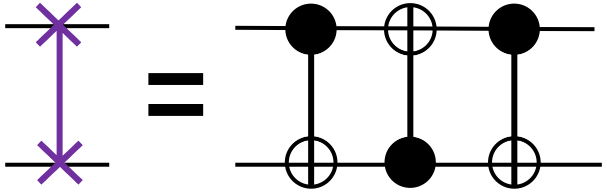


➢ Inducing SWAP gates on a circuit

SWAP-minimization is NP-complete



➢ **SWAPs are expensive**, we desire connectivities which minimize the need for data movement

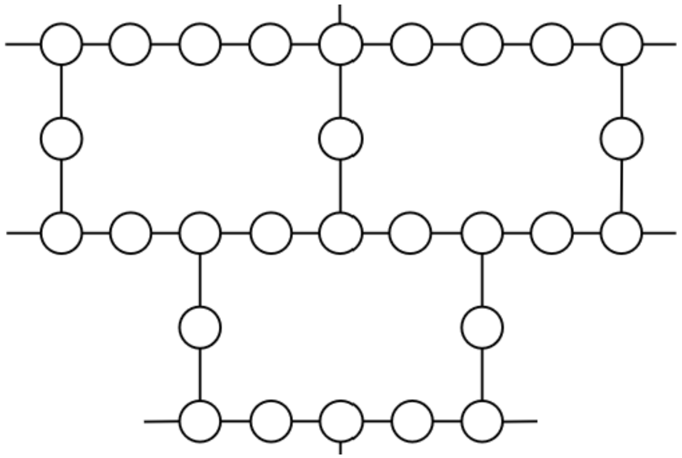


Motivating co-design

The choice of gate type and coupling topology are not independent, as they are both determined by the choice of *modulator*.

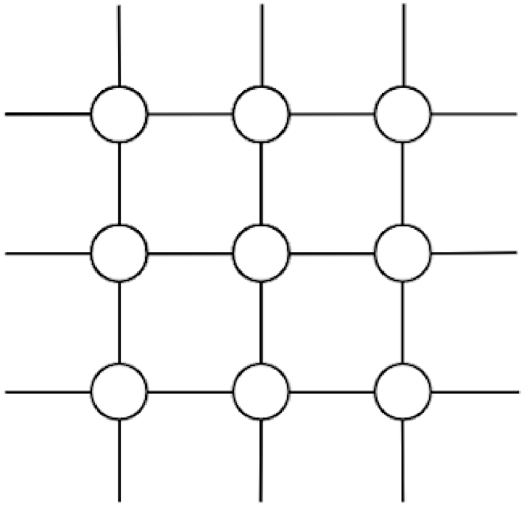
IBM

$$\mathtt{ZX}(\theta) = \begin{bmatrix} \cos\theta/2 & 0 & -i\sin\theta/2 & 0 \\ 0 & \cos\theta/2 & 0 & i\sin\theta/2 \\ -i\sin\theta/2 & 0 & \cos\theta/2 & 0 \\ 0 & i\sin\theta/2 & 0 & \cos\theta/2 \end{bmatrix}$$



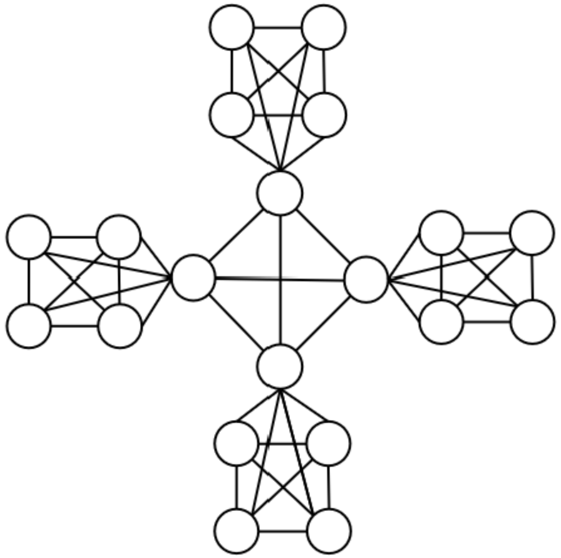
Google

$$\mathtt{FSIM}(\theta,\phi) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & -i\sin\theta & 0 \\ 0 & -i\sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & e^{-i\phi} \end{bmatrix}$$



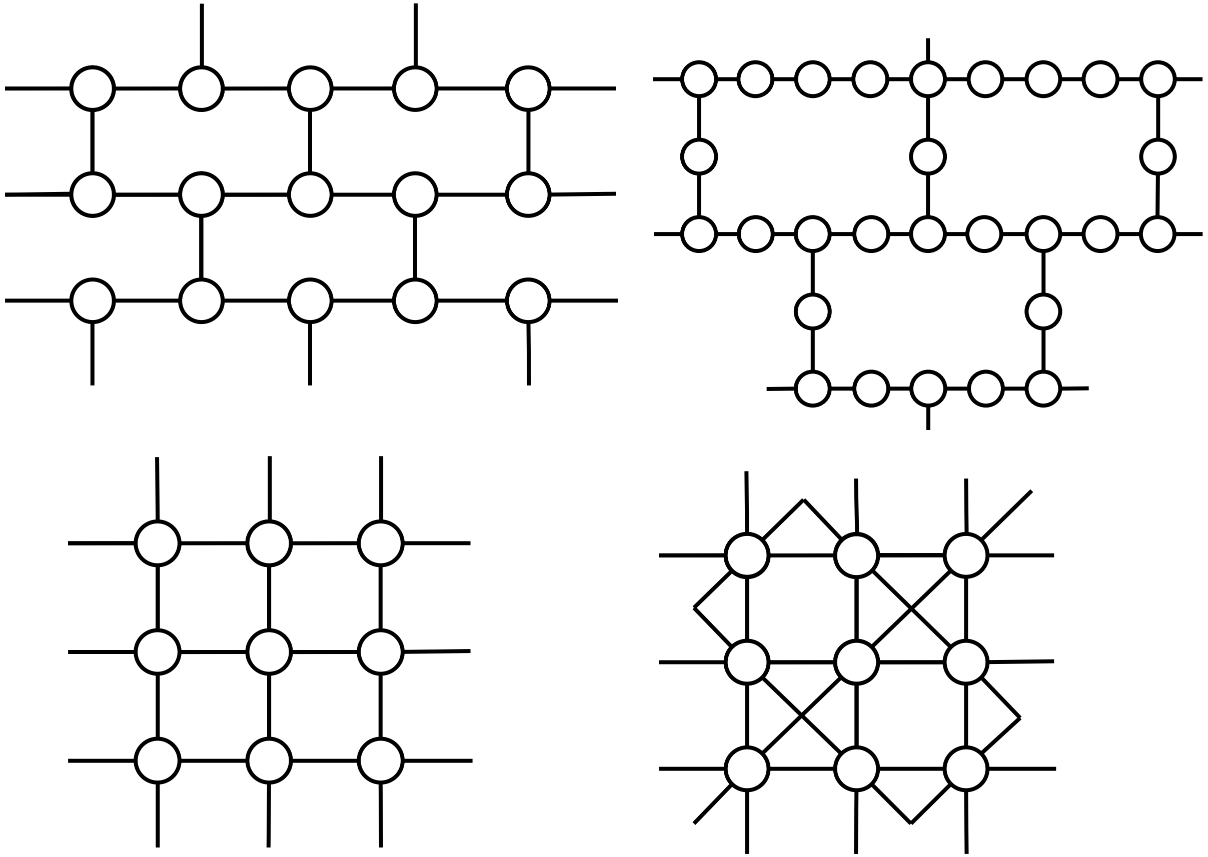
Hatlab

$$\sqrt[n]{\mathtt{iSWAP}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\pi/2n) & i\sin(\pi/2n) & 0 \\ 0 & i\sin(\pi/2n) & \cos(\pi/2n) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$



Coupling topologies

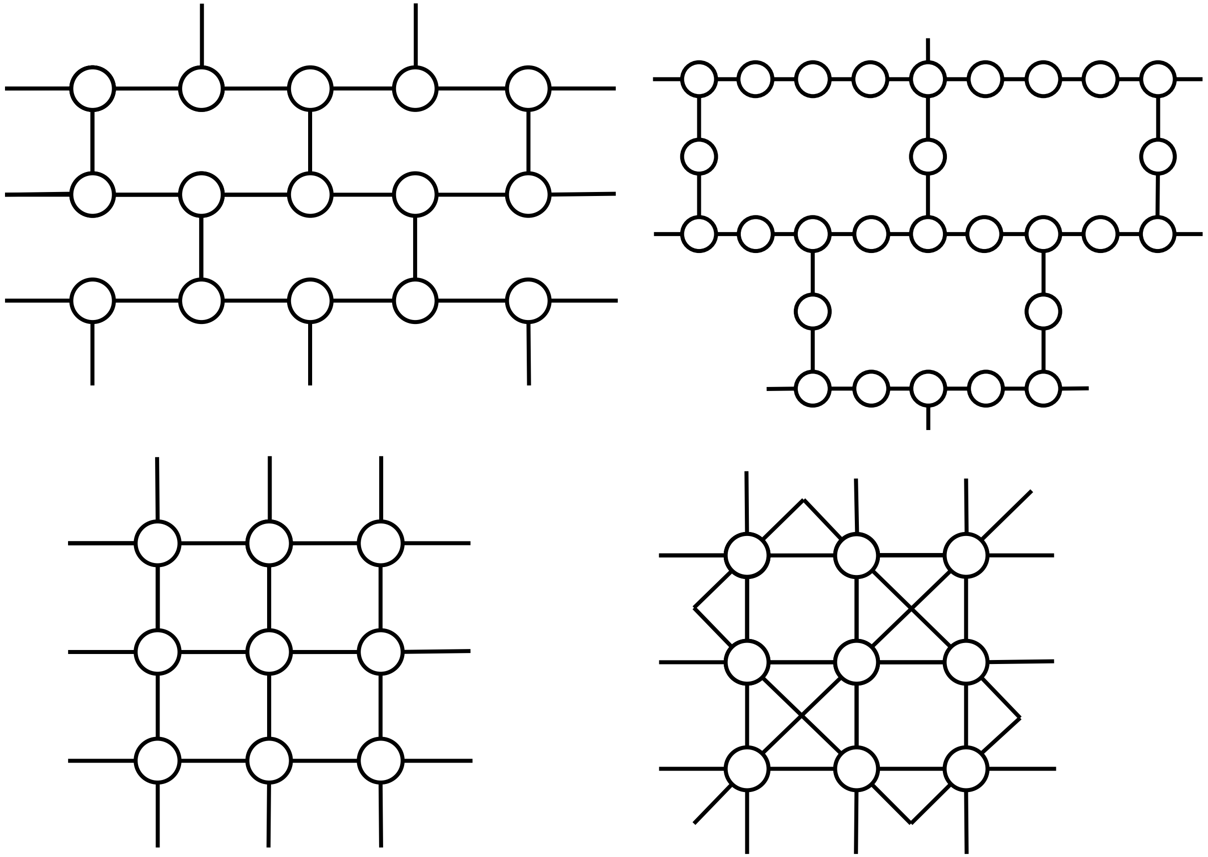
- Common Planar Topologies [1,2]



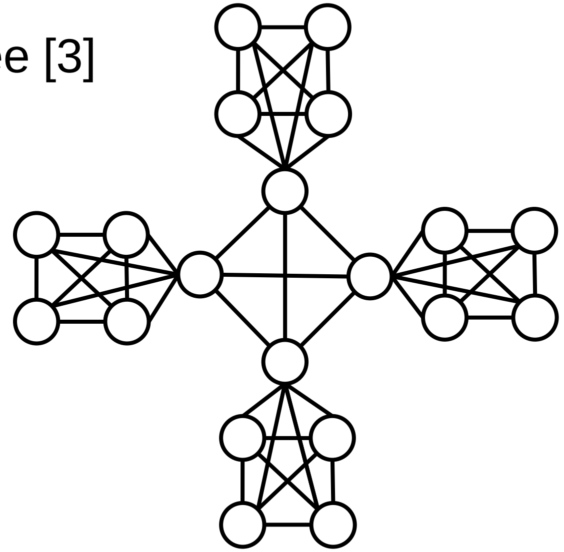
[1] Nation, et al. **IBM** (2021)
[2] Arute, et al. **Nature** (2019)
[3] Zhou, et al. **arXiv**: 2109.06848 (2021)

Coupling topologies

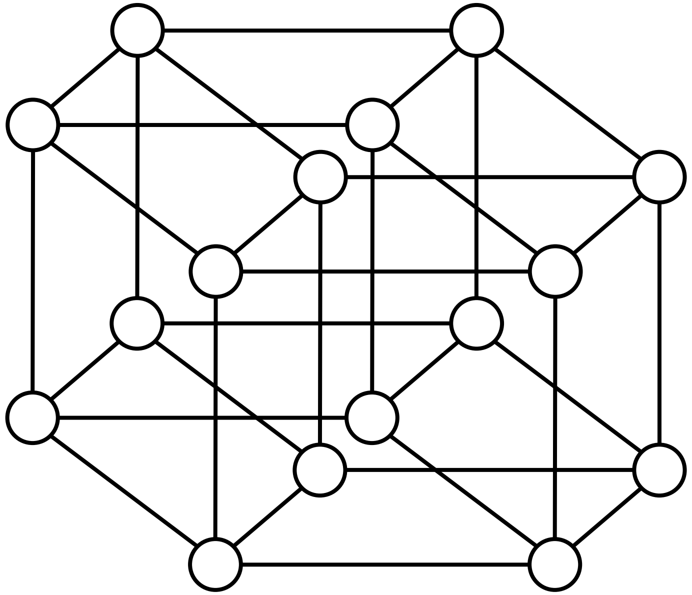
- Common Planar Topologies [1,2]



- Modular Tree [3]



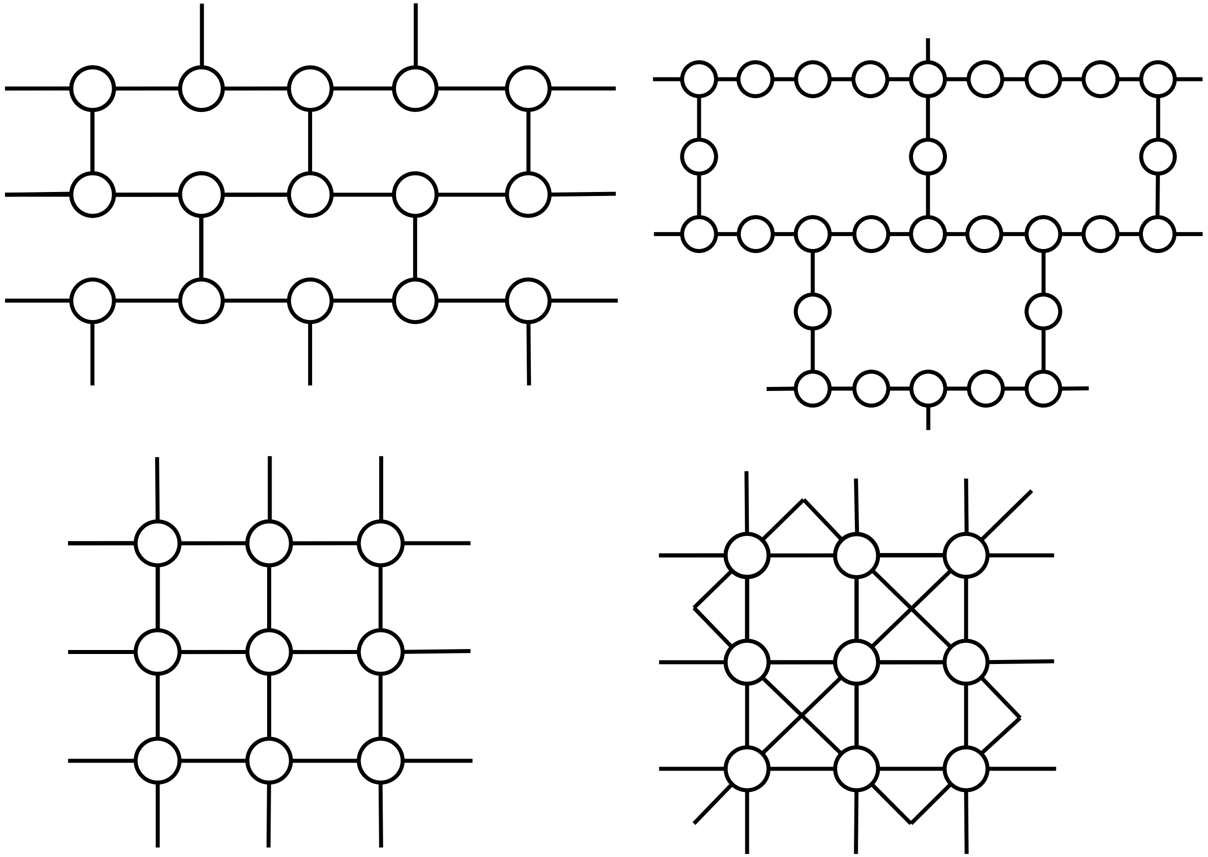
- 4D-Hypercube



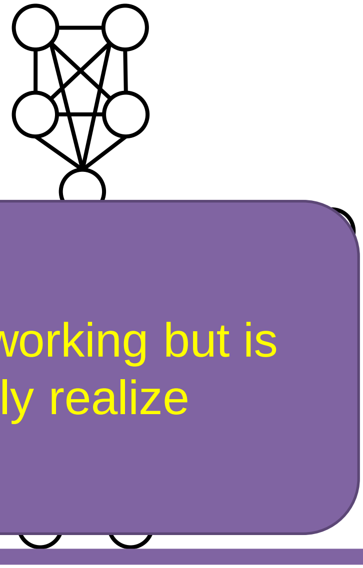
[1] Nation, et al. **IBM** (2021)
[2] Arute, et al. **Nature** (2019)
[3] Zhou, et al. **arXiv**: 2109.06848 (2021)

Coupling topologies

- Common Planar Topologies [1,2]

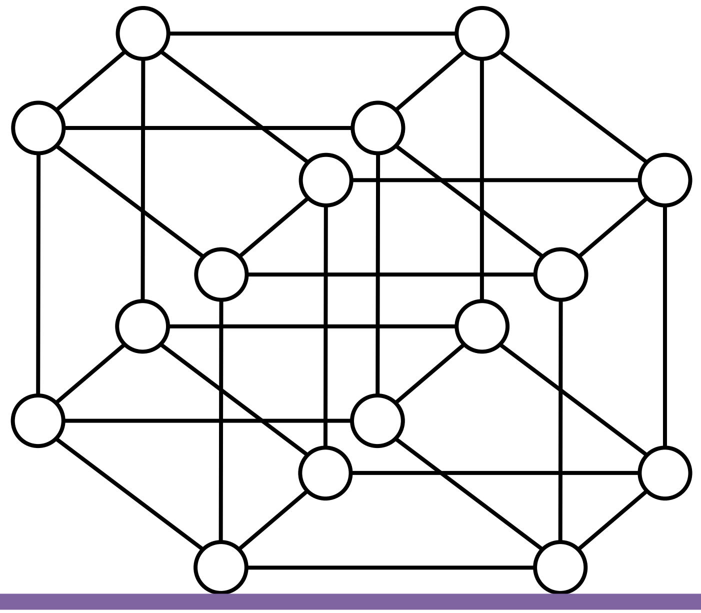


- Modular Tree [3]



Useful in classical networking but is difficult to physically realize

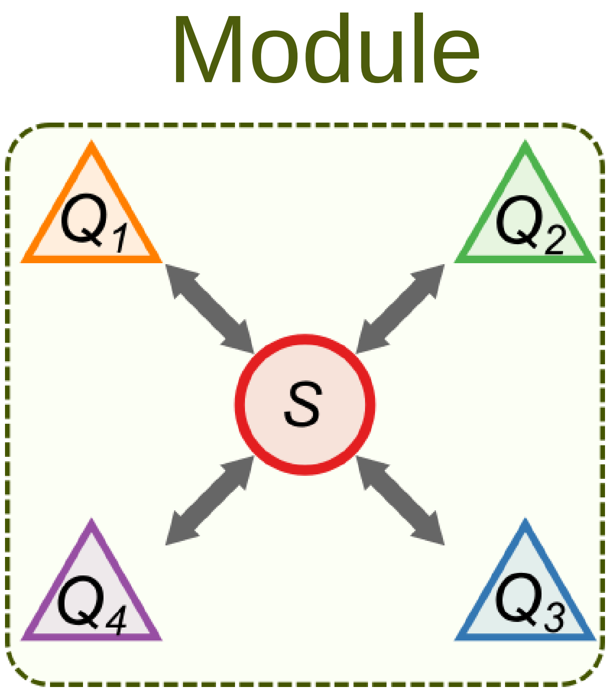
- 4D-Hypercube

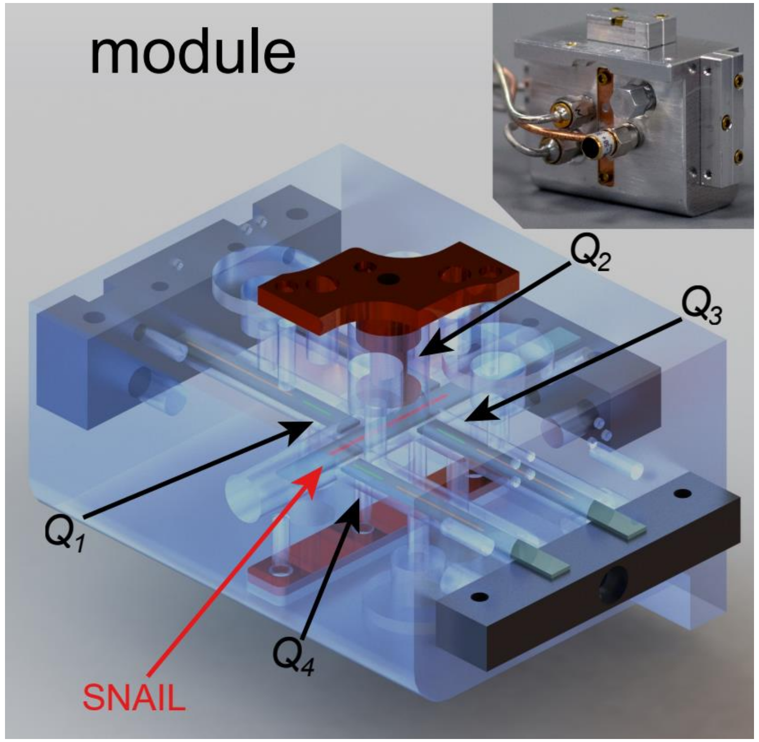


[1] Nation, et al. **IBM** (2021)
[2] Arute, et al. **Nature** (2019)
[3] Zhou, et al. **arXiv**: 2109.06848 (2021)

Experimental hardware design

Four qubit SNAIL-based quantum module





Rendered module and image

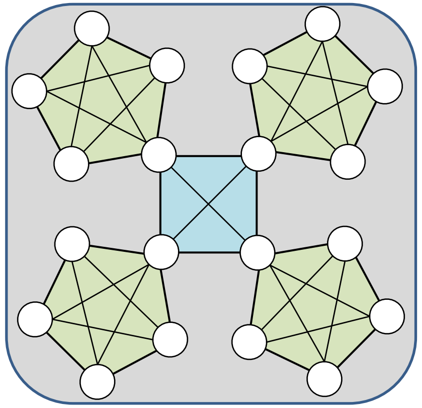
Assembled device

Zhou, et al. **arXiv**: 2109.06848 (2021)

Scaling SNAIL tree topologies

20 Qubit Trees

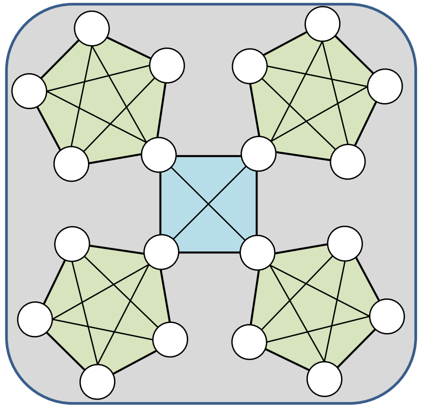
4-ary



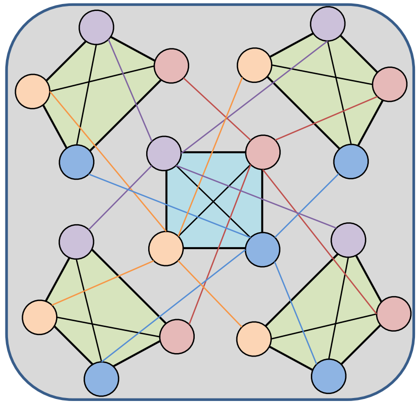
Scaling SNAIL tree topologies

20 Qubit Trees

4-ary



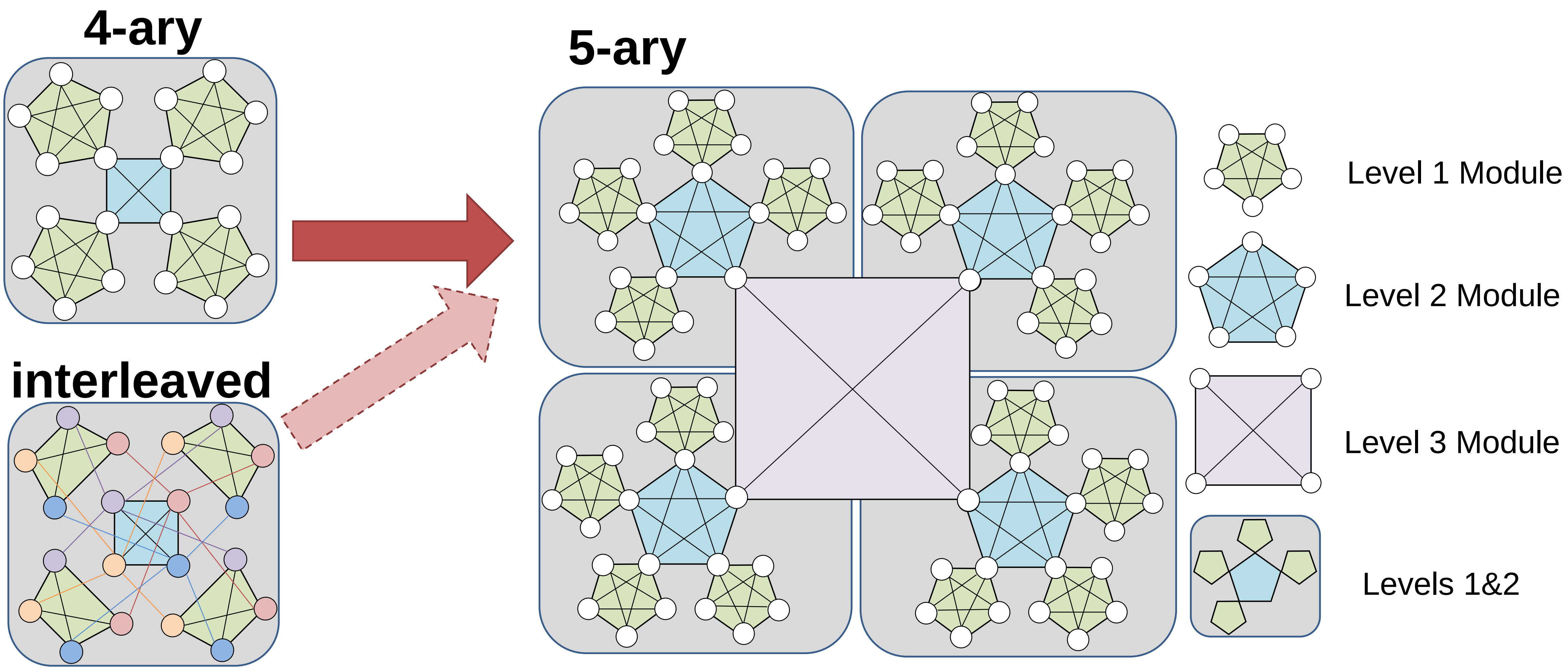
interleaved



Scaling SNAIL tree topologies

20 Qubit Trees

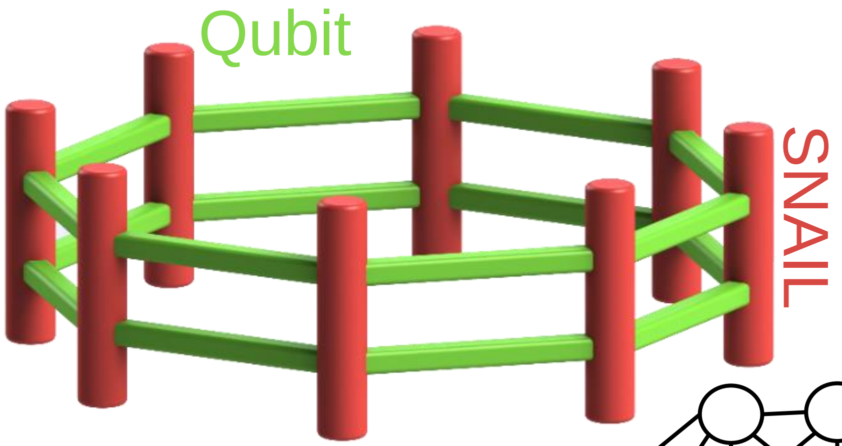
84 Qubit Tree

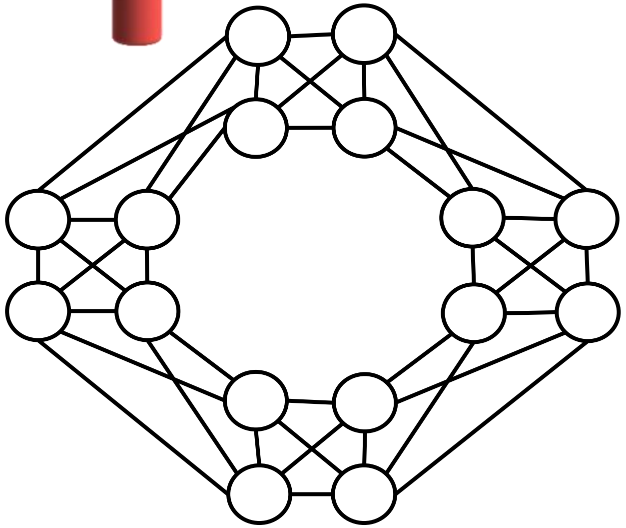


Co-designed SNAIL topologies

Objective: Maintain the low- diameter property of hypercubes without the poor dimensionality scaling.

16 Qubit Corral$_{11}$

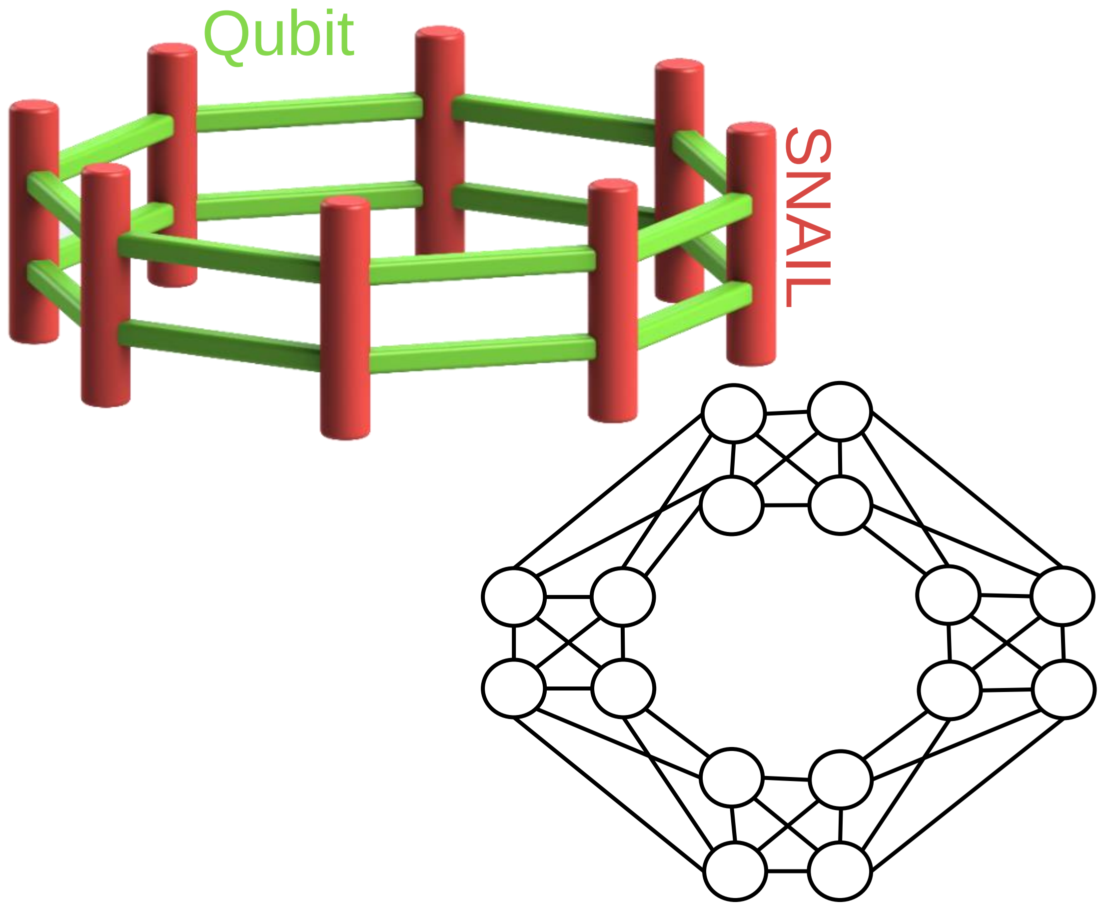




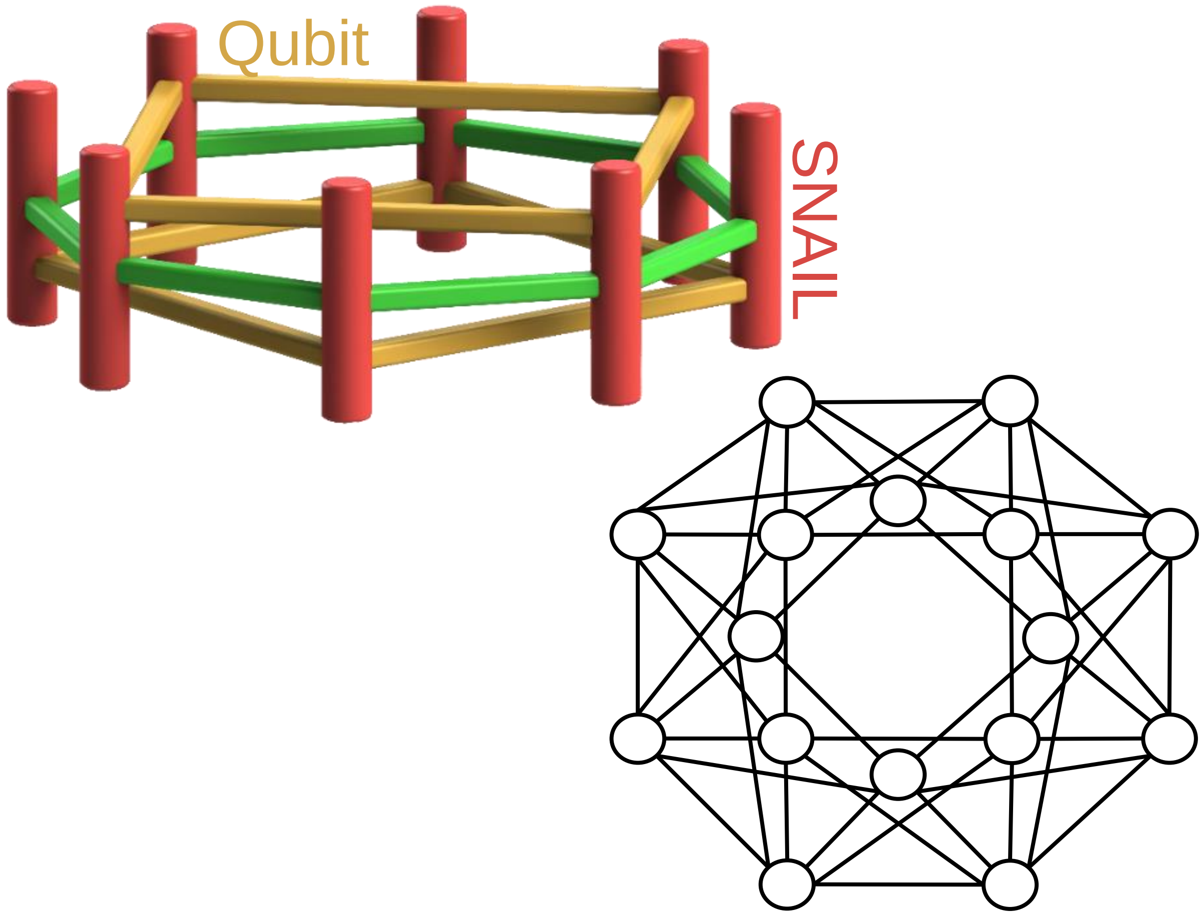
Co-designed SNAIL topologies

Objective: Maintain the low- diameter property of hypercubes without the poor dimensionality scaling.

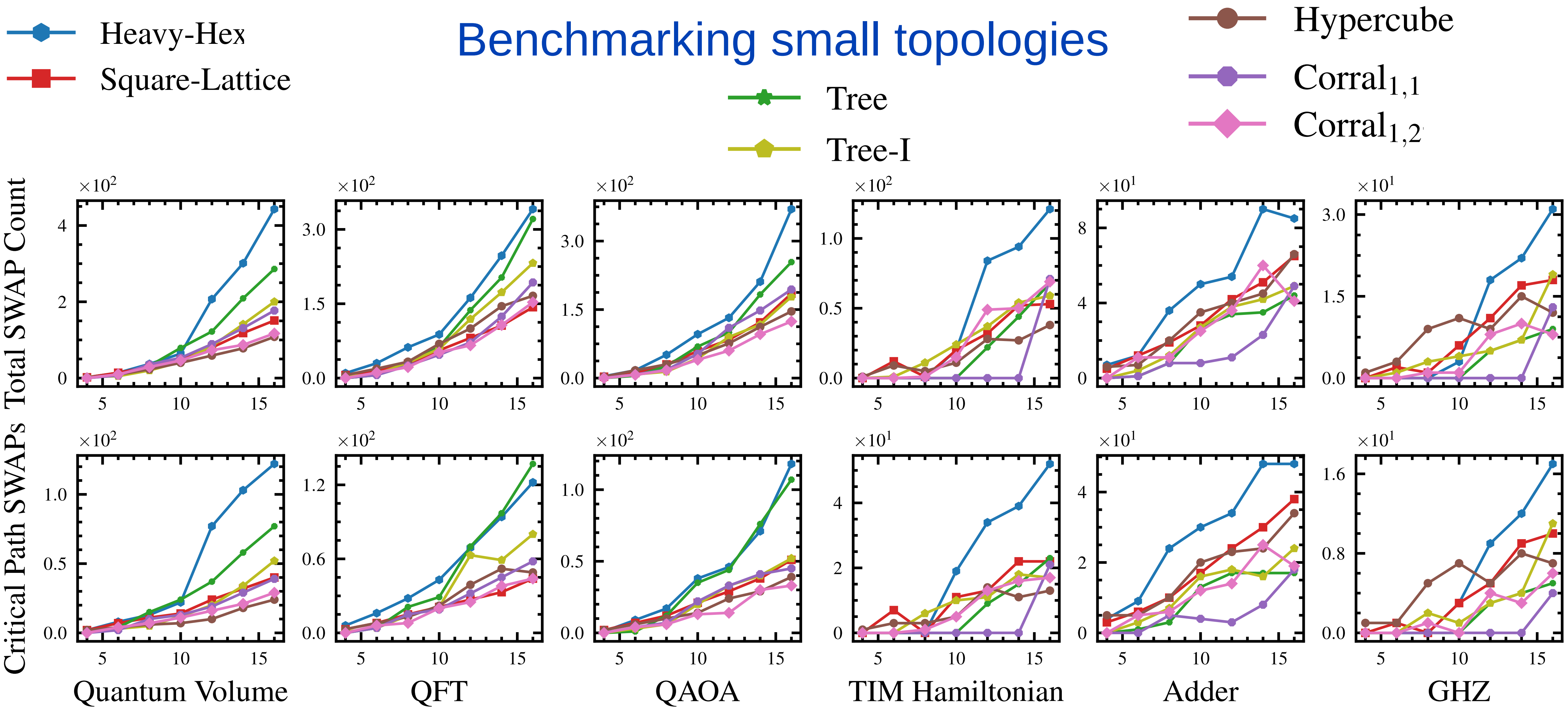
16 Qubit Corral$_{1,1}$



16 Qubit Corral$_{1,2}$

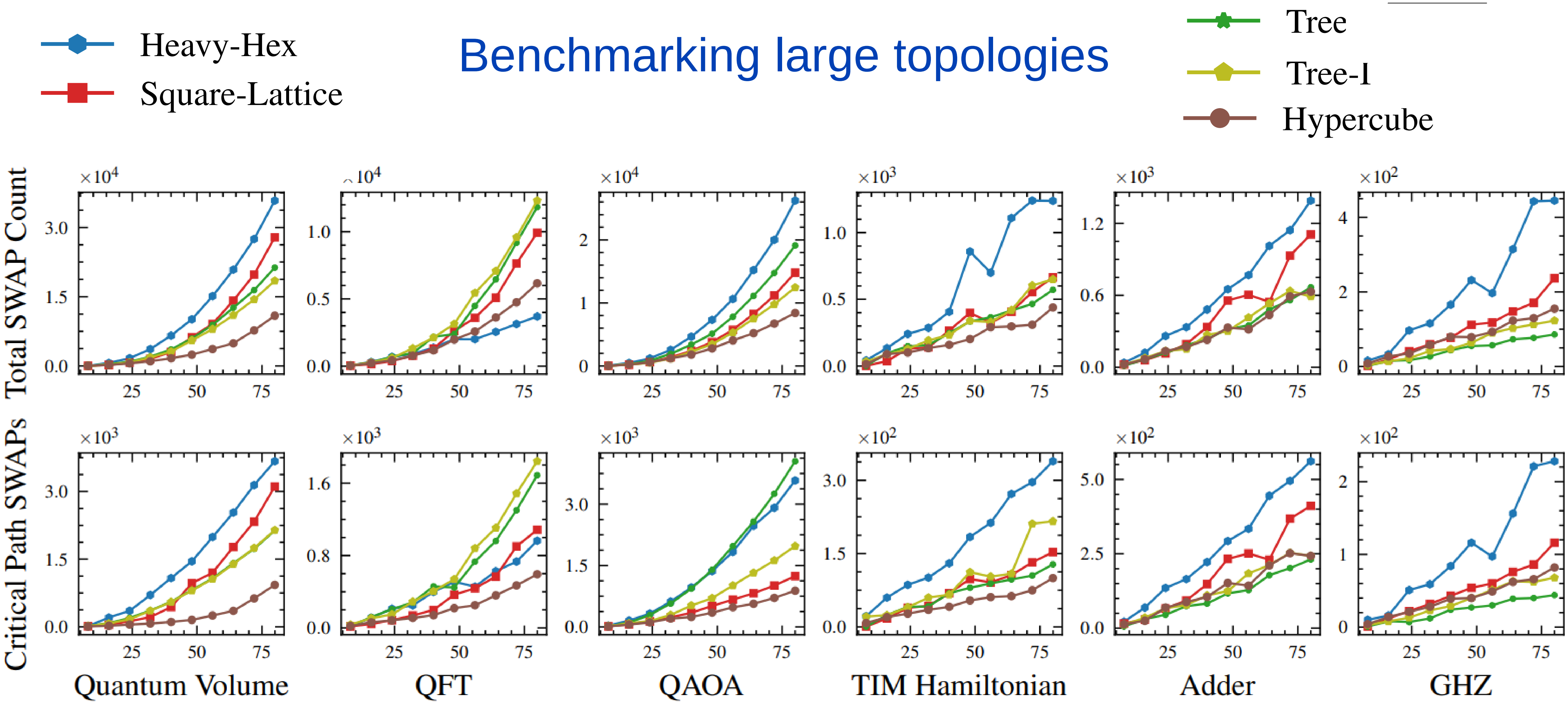


Benchmarking small topologies



- **Even smaller 16Q “neighborhoods” can benefit**
  - Heavy-Hex is 82% slower (<< fidelity) than $Corral_{11}$

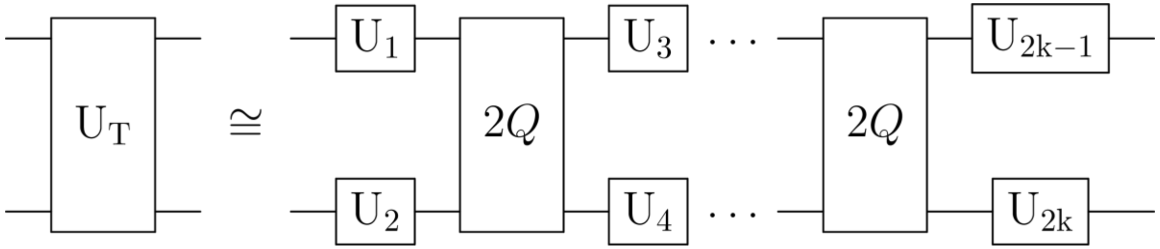
Benchmarking large topologies



- **Sparse topologies require more SWAP gates when scaled**
  - 80Q Heavy-Hex induces 3X critical path SWAPs vs. Hypercubes

Two-qubit basis gates

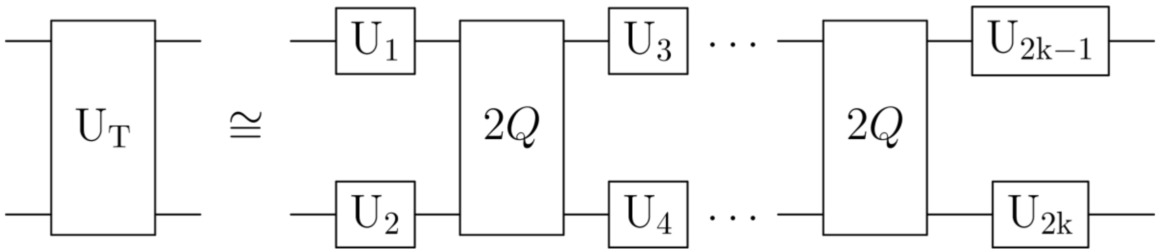
- Decompose all algorithm gates into new basis using repeated applications



- An optimal basis gate *reduces overall duration*
  - Powerful gates need less applications
  - Fidelity limited by decoherence in time

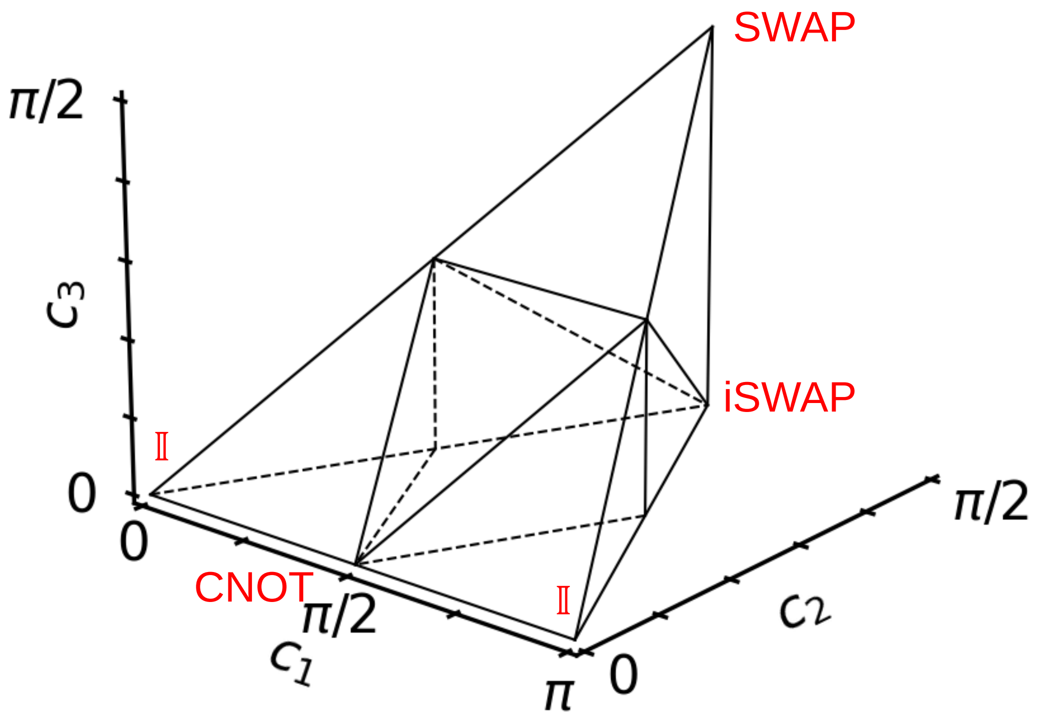
Two-qubit basis gates

- Decompose all algorithm gates into new basis using repeated applications



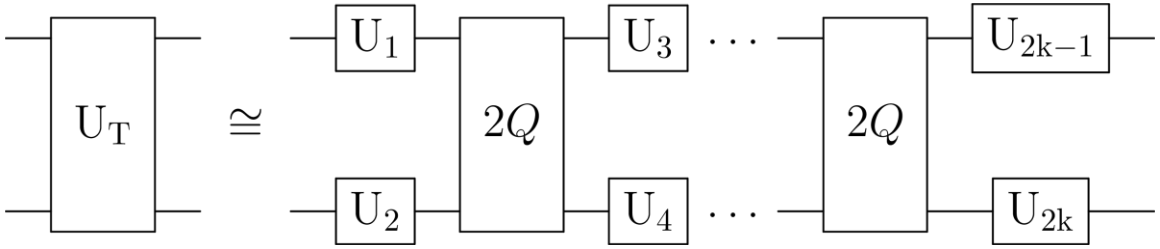
- An optimal basis gate *reduces overall duration*
  - Powerful gates need less applications
  - Fidelity limited by decoherence in time

- Weyl Chamber visualizes the set of all 2Q gates



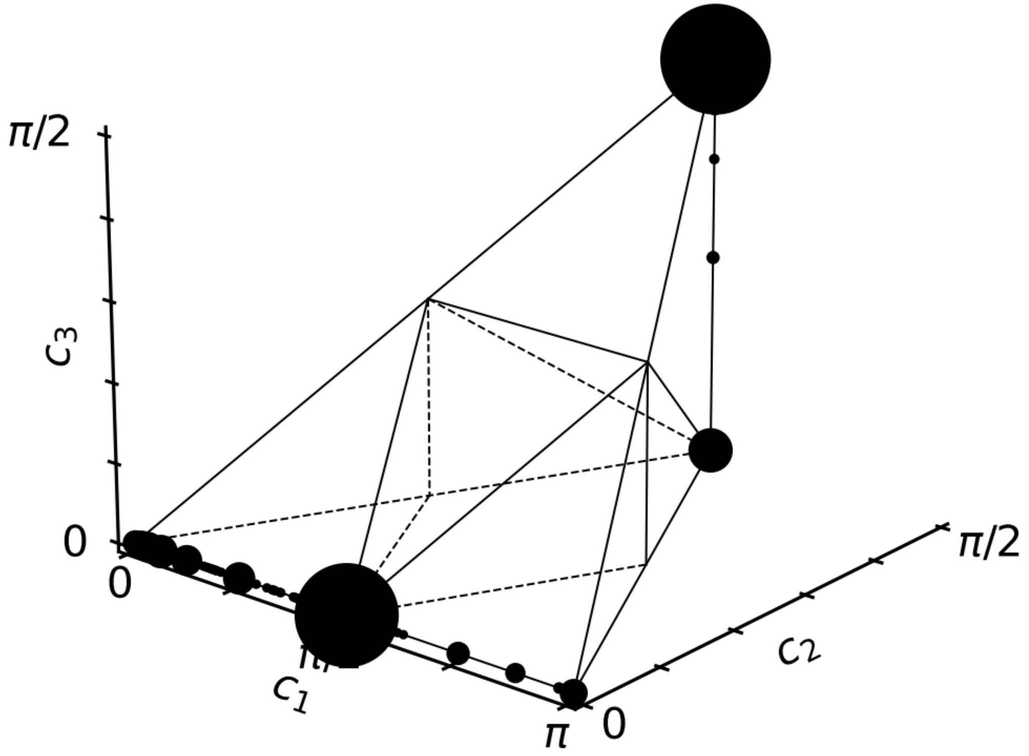
Two-qubit basis gates

- ➢ Decompose all algorithm gates into new basis using repeated applications



- ➢ An optimal basis gate *reduces overall duration*
  - ➢ Powerful gates need less applications
  - ➢ Fidelity limited by decoherence in time

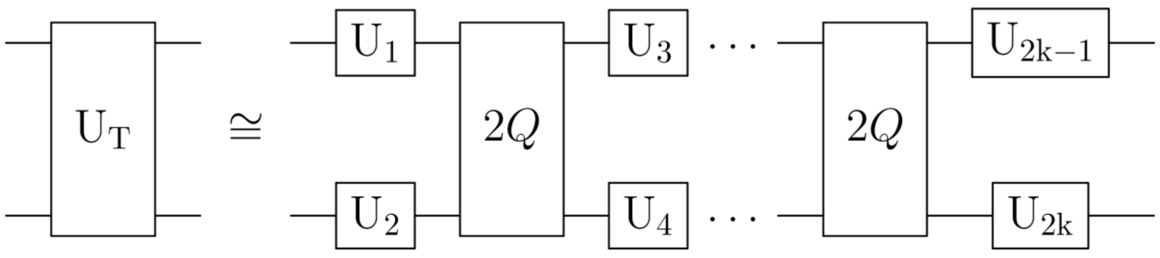
- ➢ Weyl Chamber visualizes the set of all 2Q gates



- ➢ NISQ algorithms dominated by CX and SWAP gates

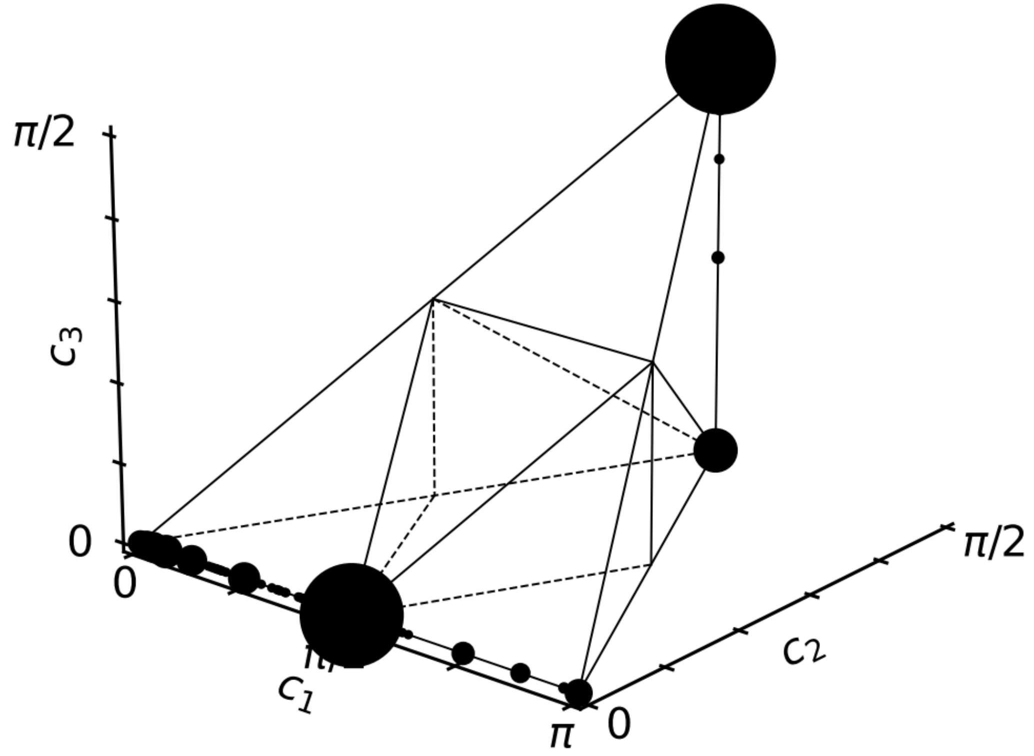
Two-qubit basis gates

- Decompose all algorithm gates into new basis using repeated applications



- An optimal basis gate *reduces overall duration*
  - Powerful gates need less applications
  - Fidelity limited by decoherence in time

- Weyl Chamber visualizes the set of all 2Q gates

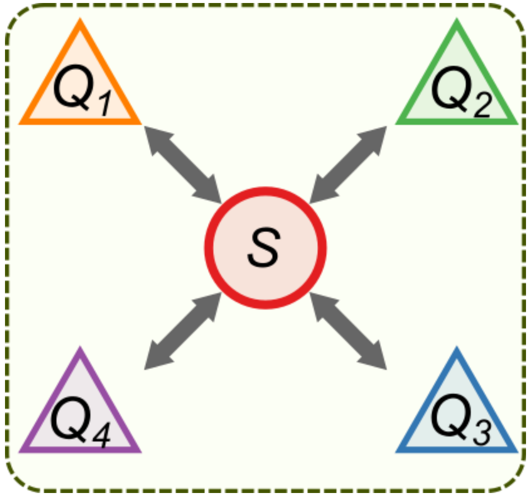


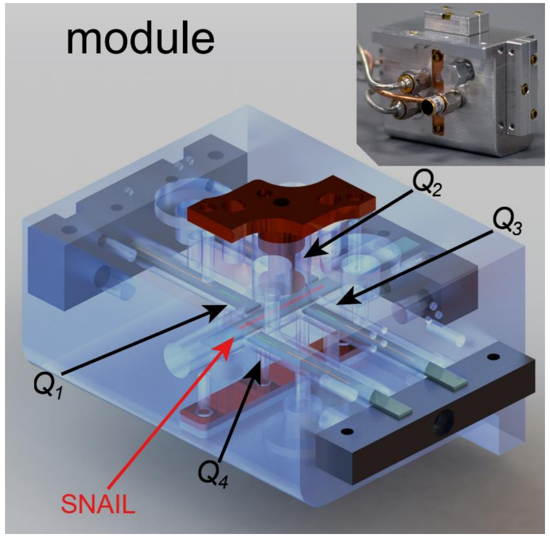
- NISQ algorithms dominated by CX and SWAP gates

- **Goal: Use both decomposition efficiency and hardware latency = overall duration**

Conversion/Gain candidate basis gates

Four qubit SNAIL-based quantum module





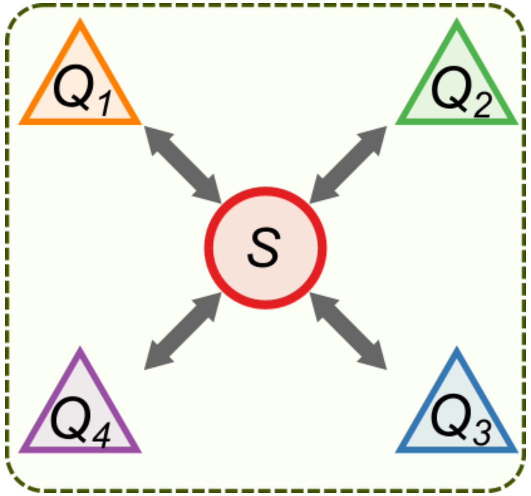
- **Engineerable interactions yields a basis gate design-space**

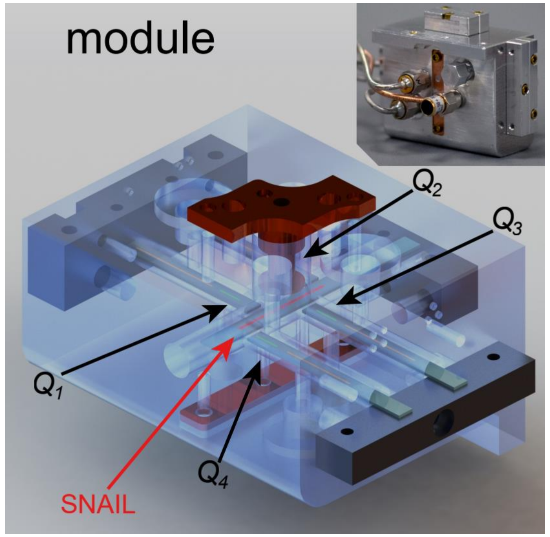
$$\hat{H} = g_c(e^{i\phi_c}a^\dagger b + e^{-i\phi_c}ab^\dagger) + g_g(e^{i\phi_g}ab + e^{-i\phi_g}a^\dagger b^\dagger)$$

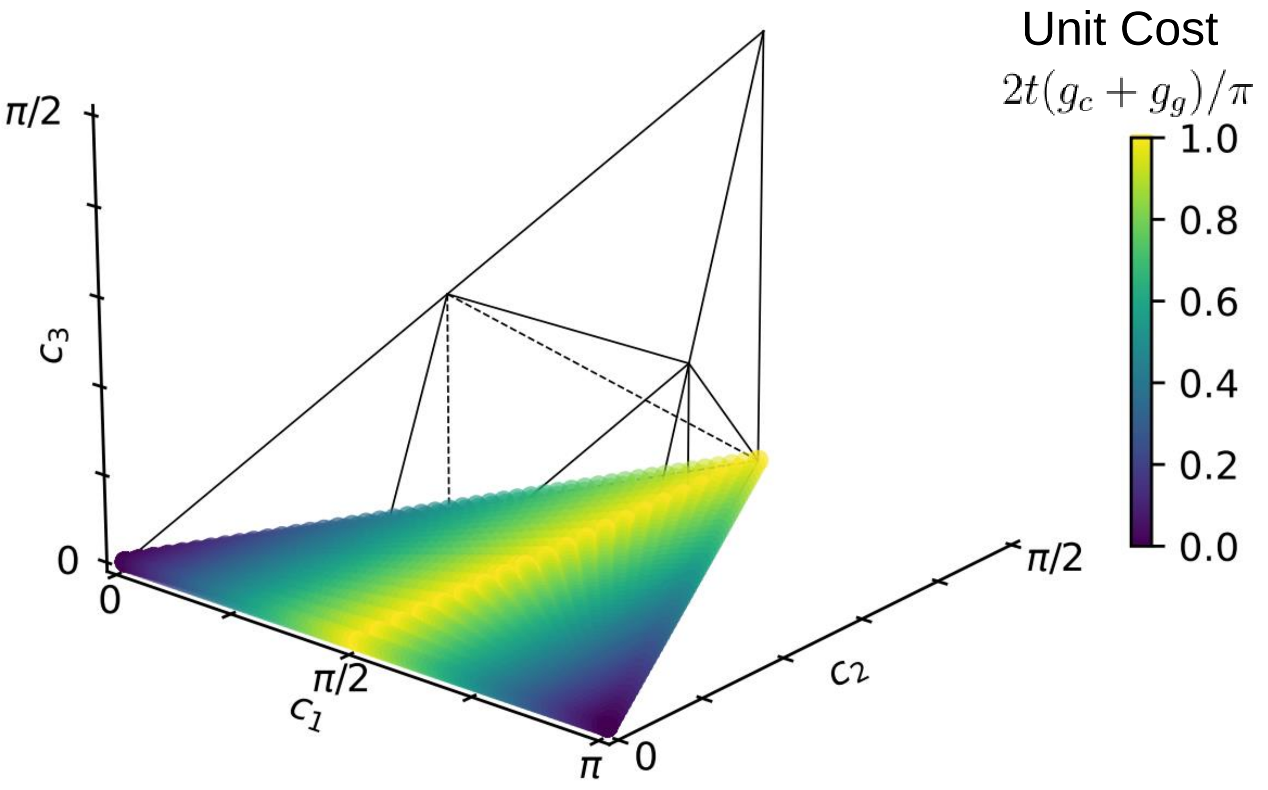
Zhou, et al. ***arXiv**:2109.06848* (2021)

Conversion/Gain candidate basis gates

Four qubit SNAIL-based quantum module







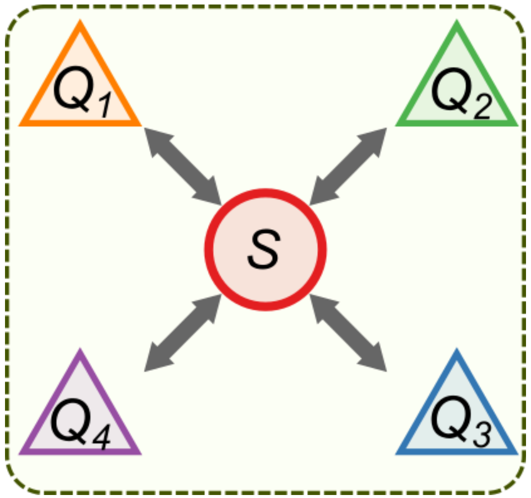
- **Engineerable interactions yields a basis gate design-space**

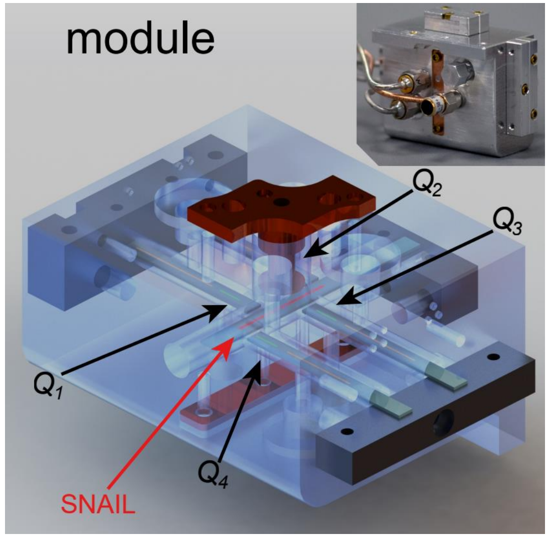
$$\hat{H} = g_c(e^{i\phi_c}a^\dagger b + e^{-i\phi_c}ab^\dagger) + g_g(e^{i\phi_g}ab + e^{-i\phi_g}a^\dagger b^\dagger)$$

Zhou, et al. ***arXiv****:2109.06848* (2021)

Conversion/Gain candidate basis gates

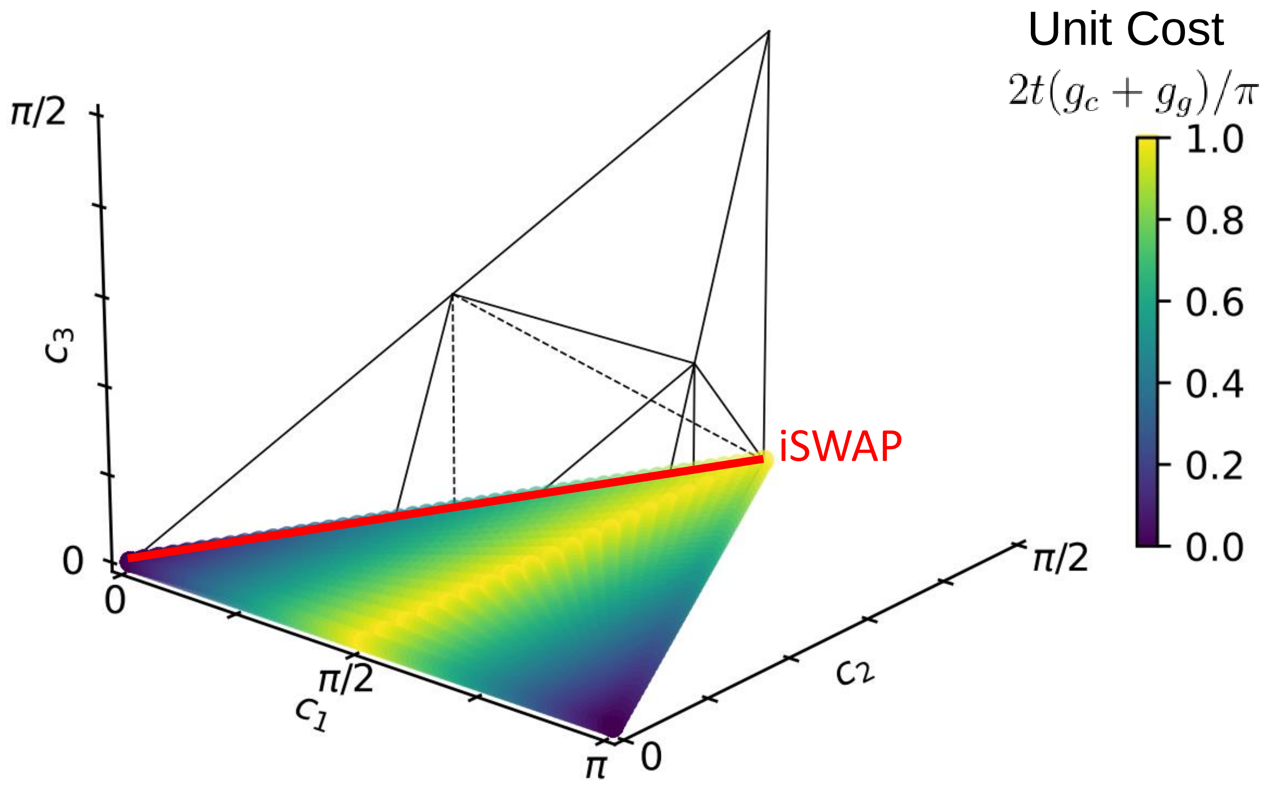
Four qubit SNAIL-based quantum module





- **Engineerable interactions yields a basis gate design-space**

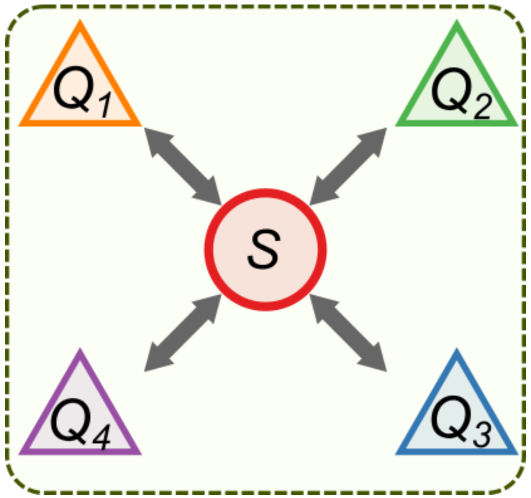
$$\hat{H} = g_c(e^{i\phi_c}a^\dagger b + e^{-i\phi_c}ab^\dagger) + g_g(e^{i\phi_g}ab + e^{-i\phi_g}a^\dagger b^\dagger)$$

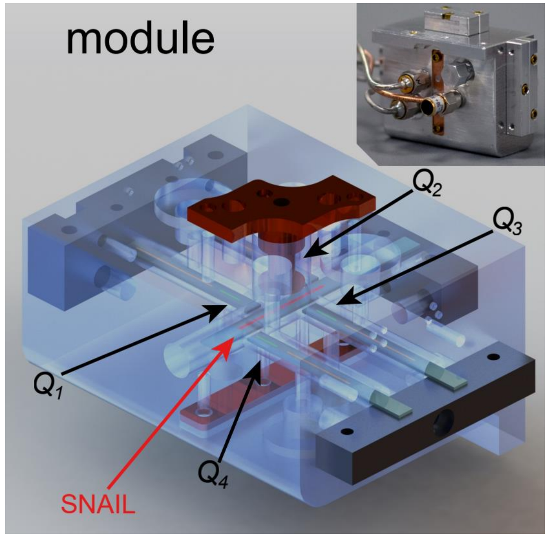


Zhou, et al. ***arXiv****:2109.06848* (2021)

Conversion/Gain candidate basis gates

Four qubit SNAIL-based quantum module

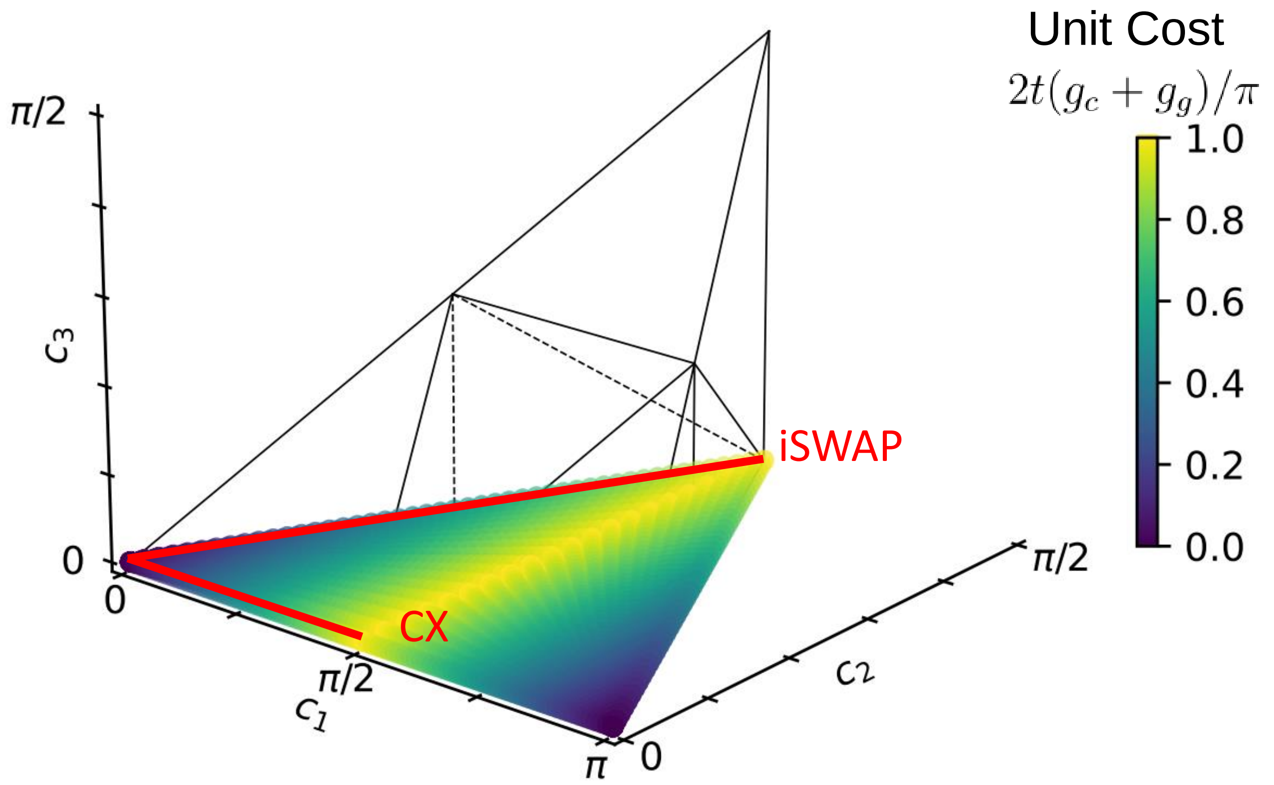




- **Engineerable interactions yields a basis gate design-space**

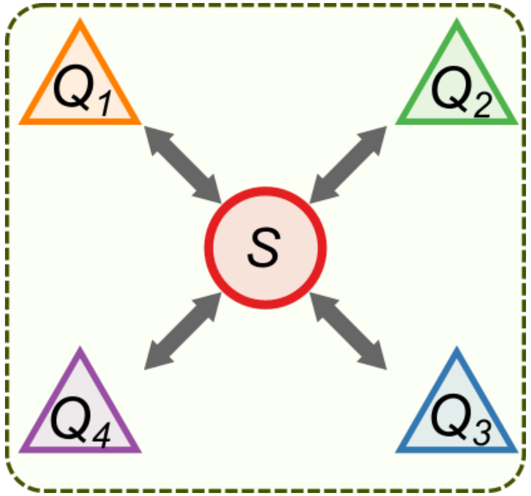
$$\hat{H} = g_c(e^{i\phi_c}a^\dagger b + e^{-i\phi_c}ab^\dagger) + g_g(e^{i\phi_g}ab + e^{-i\phi_g}a^\dagger b^\dagger)$$

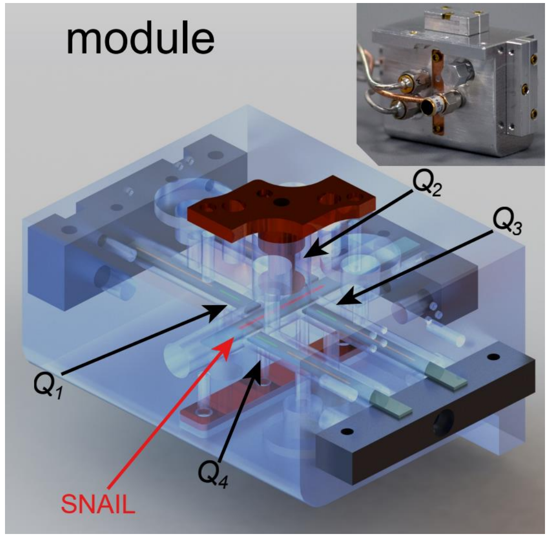
Zhou, et al. ***arXiv****:2109.06848* (2021)



Conversion/Gain candidate basis gates

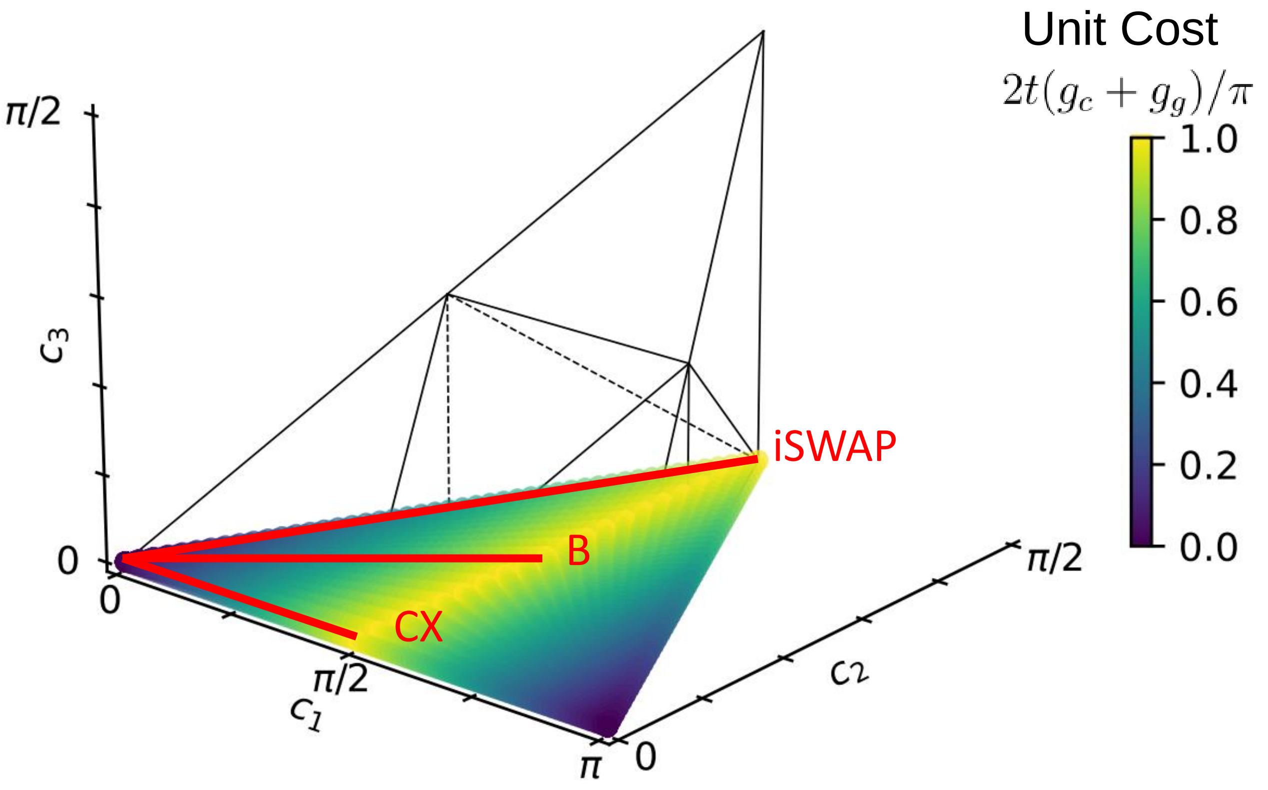
Four qubit SNAIL-based quantum module





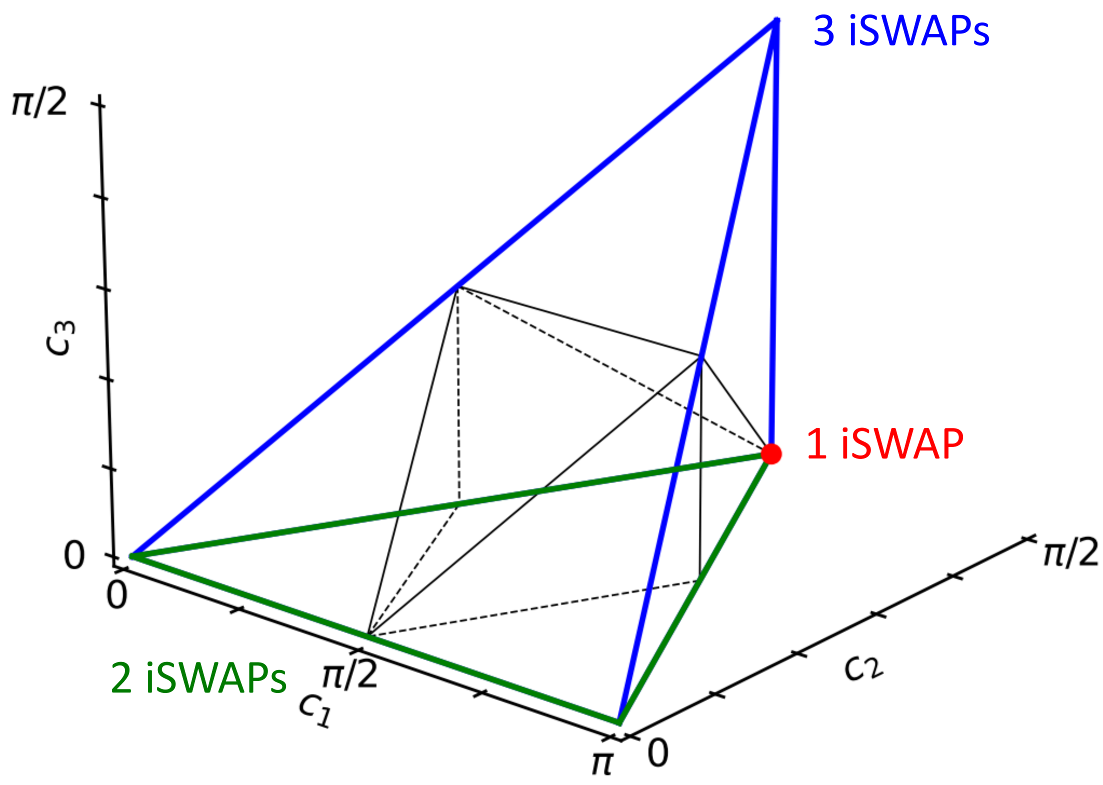
- **Engineerable interactions yields a basis gate design-space**

$$\hat{H} = g_c(e^{i\phi_c}a^\dagger b + e^{-i\phi_c}ab^\dagger) + g_g(e^{i\phi_g}ab + e^{-i\phi_g}a^\dagger b^\dagger)$$



Zhou, et al. ***arXiv***:*2109.06848* (2021)

Basis coverage volumes

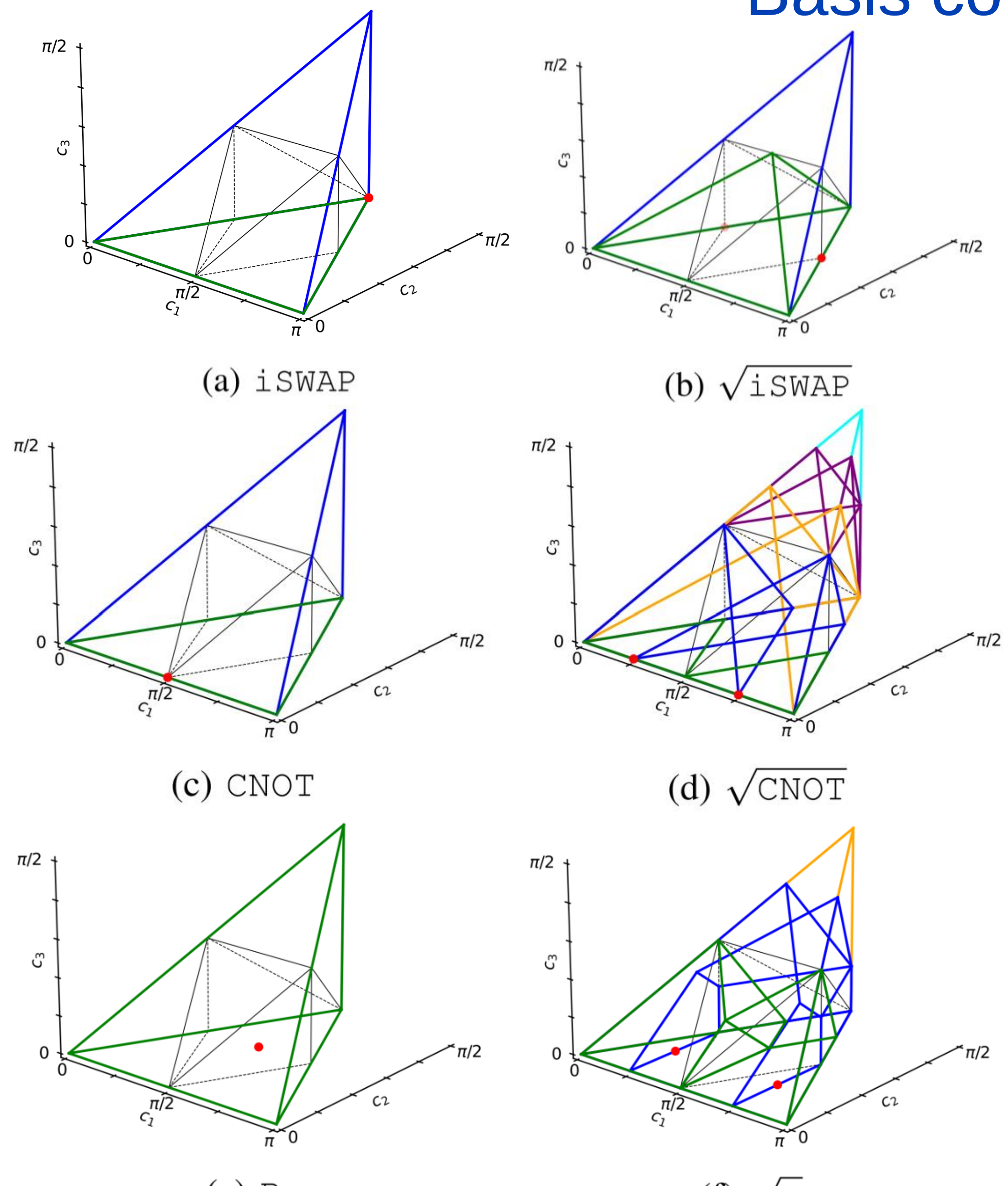


- **Monodromy polytopes** finds minimum gate applications for any 2Q target gate
- A single gate locally equivalent to itself
- SWAP is the most expensive target

Basis	iSWAP√
CNOT	2.0
SWAP	3.0
Haar	3.0

Peterson, et al. ***Quantum* 4** (2020): 247

Basis coverage volumes



(a) iSWAP (b) $\sqrt{\text{iSWAP}}$ (c) CNOT (d) $\sqrt{\text{CNOT}}$ (e) B (f) $\sqrt{\text{B}}$

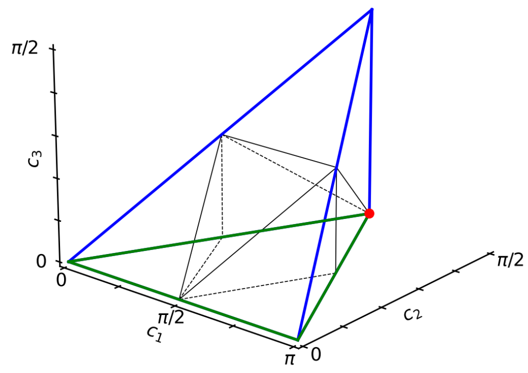
- **Monodromy polytopes** finds minimum gate applications for any 2Q target gate
- A single gate locally equivalent to itself
- SWAP is the most expensive target

Decomposition *gate count* costs

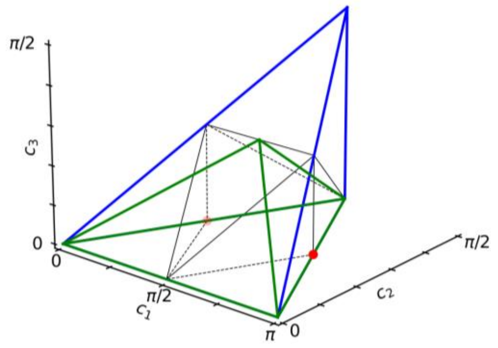
Basis	iSWAP	$\sqrt{\text{iSWAP}}$	CNOT	$\sqrt{\text{CNOT}}$	B	$\sqrt{\text{B}}$
CNOT	2.0	2.0	1.0	2.0	2.0	2.0
SWAP	3.0	3.0	3.0	6.0	2.0	4.0
Haar	3.0	2.2	3.0	3.5	2.0	3.1

Peterson, et al. ***Quantum* 4** (2020): 247

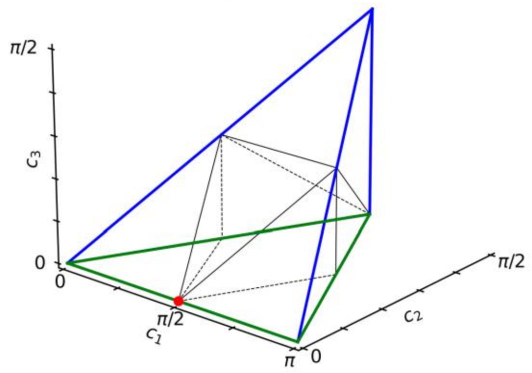
Basis coverage volumes



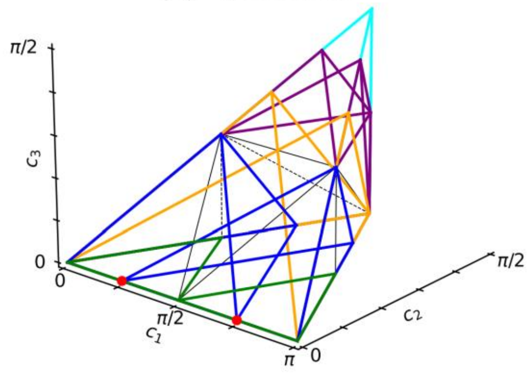
(a) iSWAP



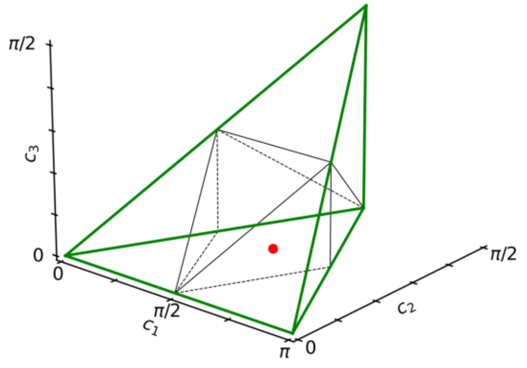
(b) $\sqrt{\text{iSWAP}}$



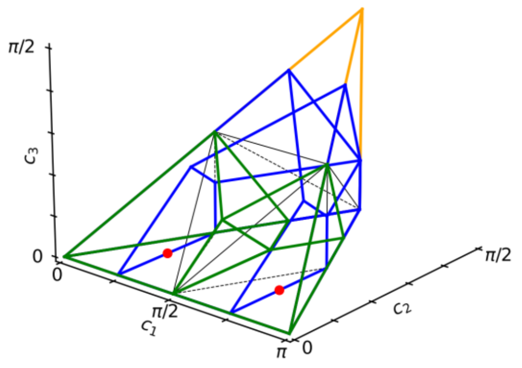
(c) CNOT



(d) $\sqrt{\text{CNOT}}$



(e) B



(f) $\sqrt{\text{B}}$

- **Monodromy polytopes** finds minimum gate applications for any 2Q target gate
- A single gate locally equivalent to itself
- SWAP is the most expensive target

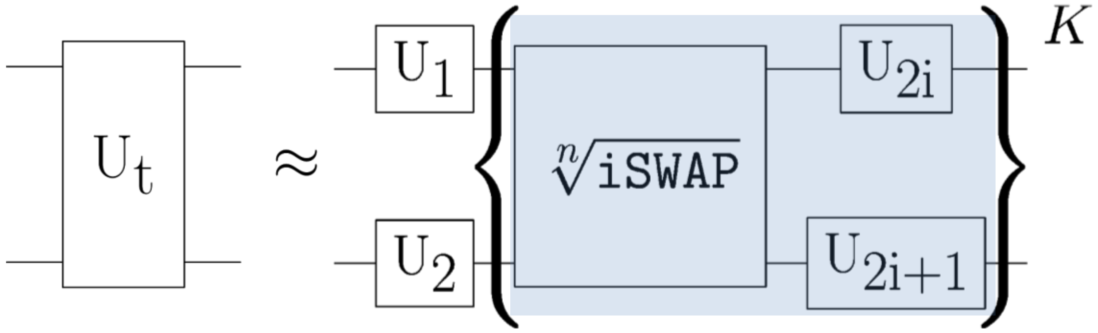
Decomposition *gate count* costs

Basis	iSWAP	$\sqrt{\text{iSWAP}}$	CNOT	$\sqrt{\text{CNOT}}$	B	$\sqrt{\text{B}}$
CNOT	2.0	2.0	1.0	2.0	2.0	2.0
SWAP	3.0	3.0	3.0	6.0	2.0	4.0
Haar	3.0	2.2	3.0	3.5	2.0	3.1

Peterson, et al. ***Quantum* 4** (2020): 247

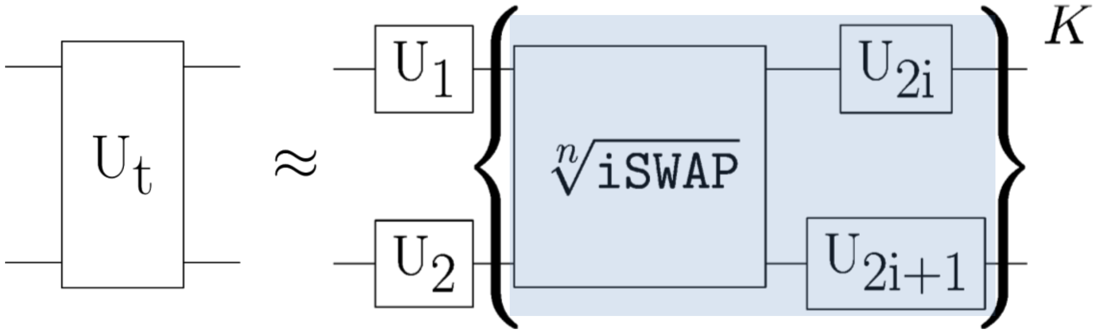
Approximate decomposition into continuous iSWAP

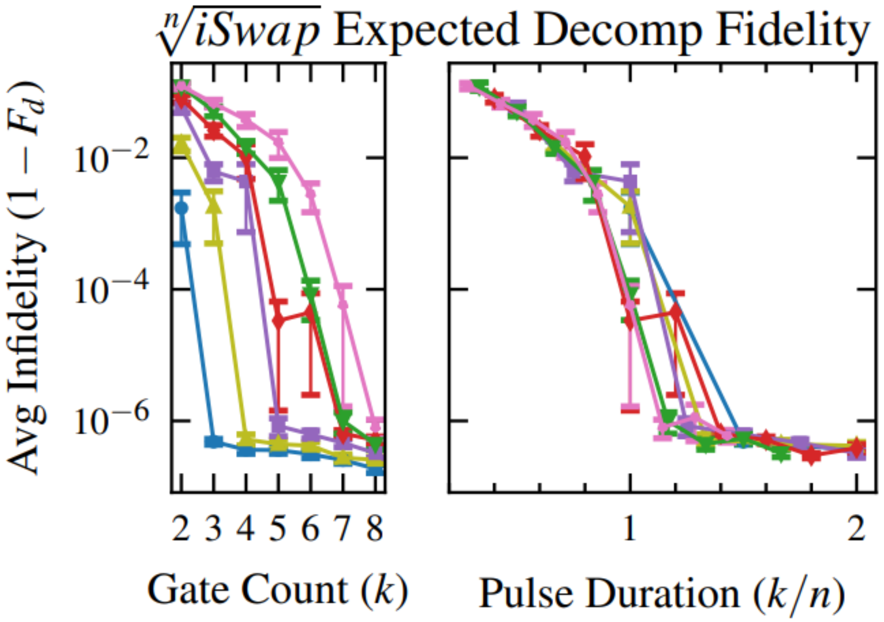
Repeats *K* times | i ➜ 1 to *K*



Approximate decomposition into continuous iSWAP

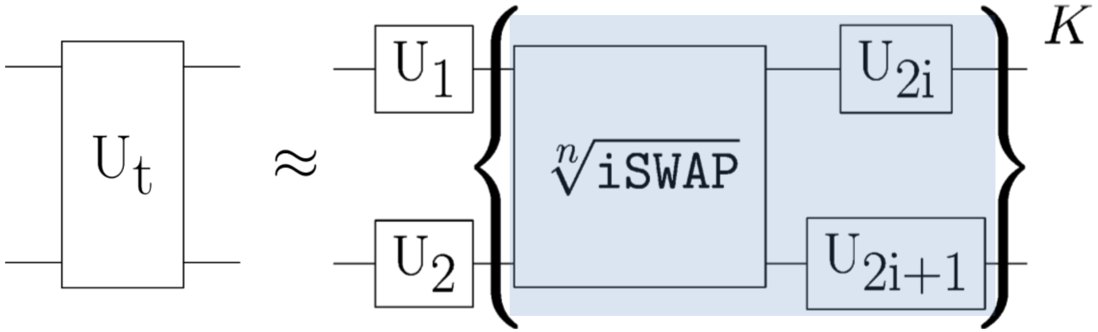
Repeats K times | i ➜ 1 to K

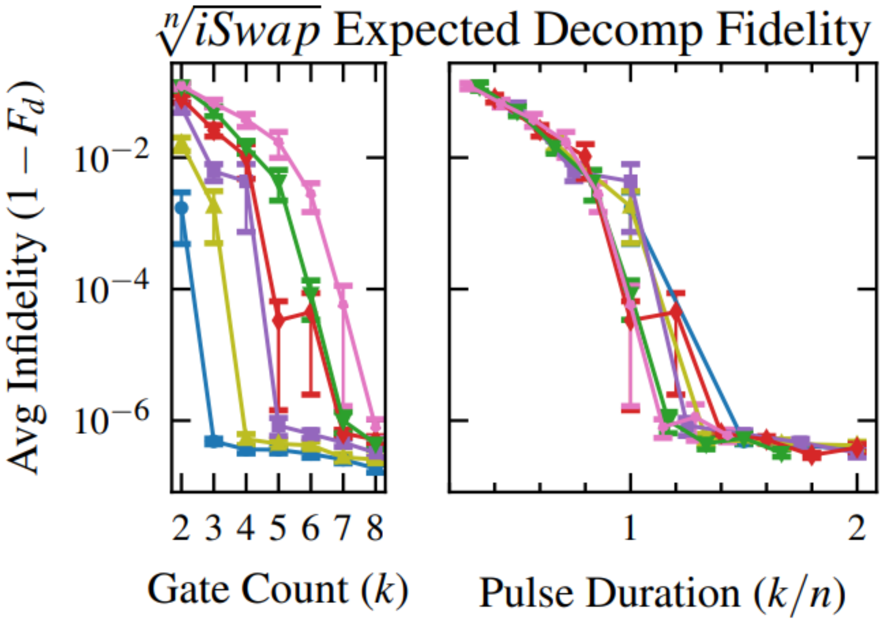




Approximate decomposition into continuous iSWAP

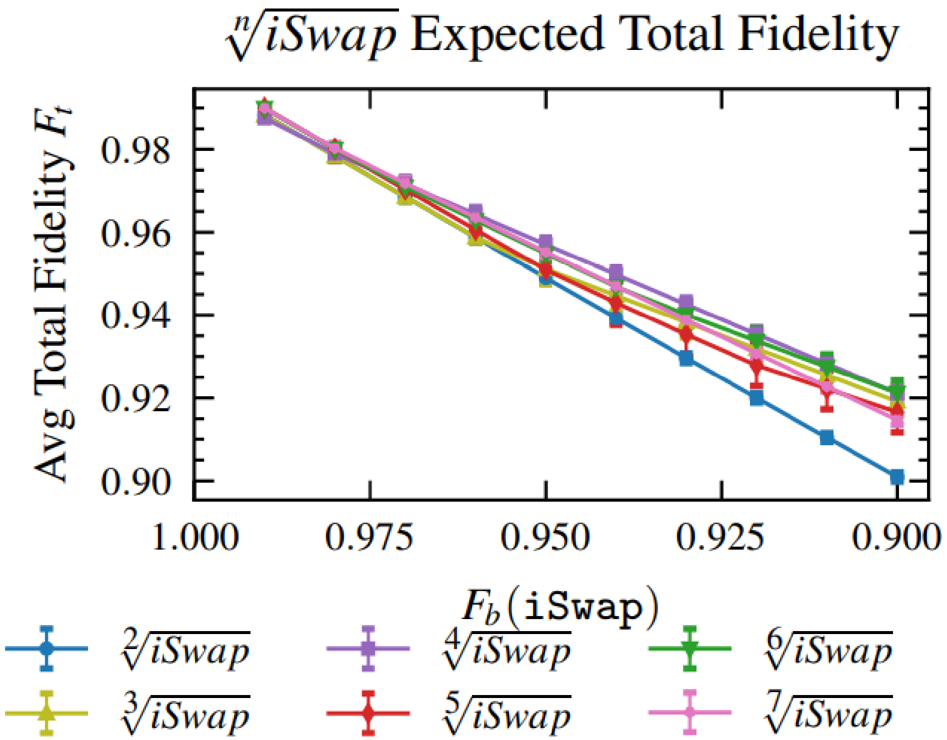
Repeats K times | i ➔ 1 to K





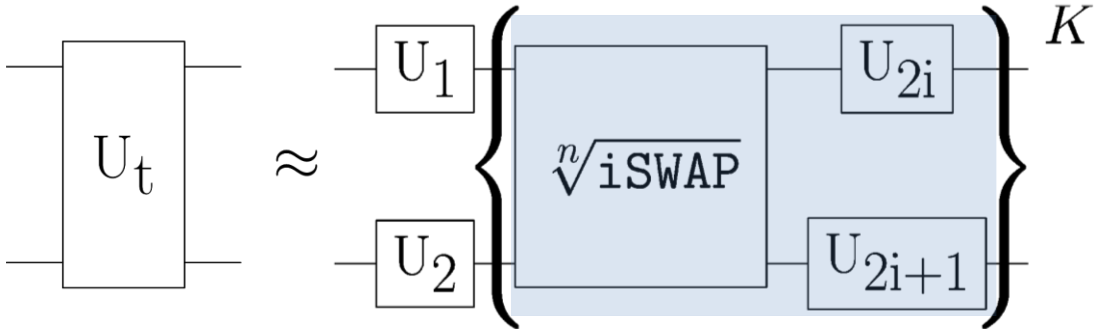
- ➢ Coherence-limited fidelity as a function of gate duration

$$F(\sqrt[n]{\texttt{iSWAP}}) = 1 - \frac{1 - F(\texttt{iSWAP})}{n}$$



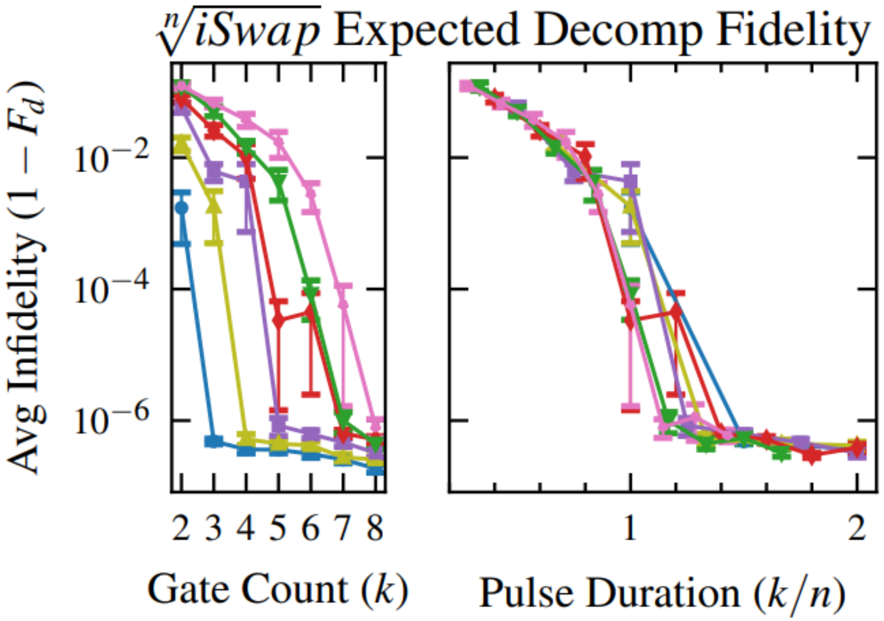
Approximate decomposition into continuous iSWAP

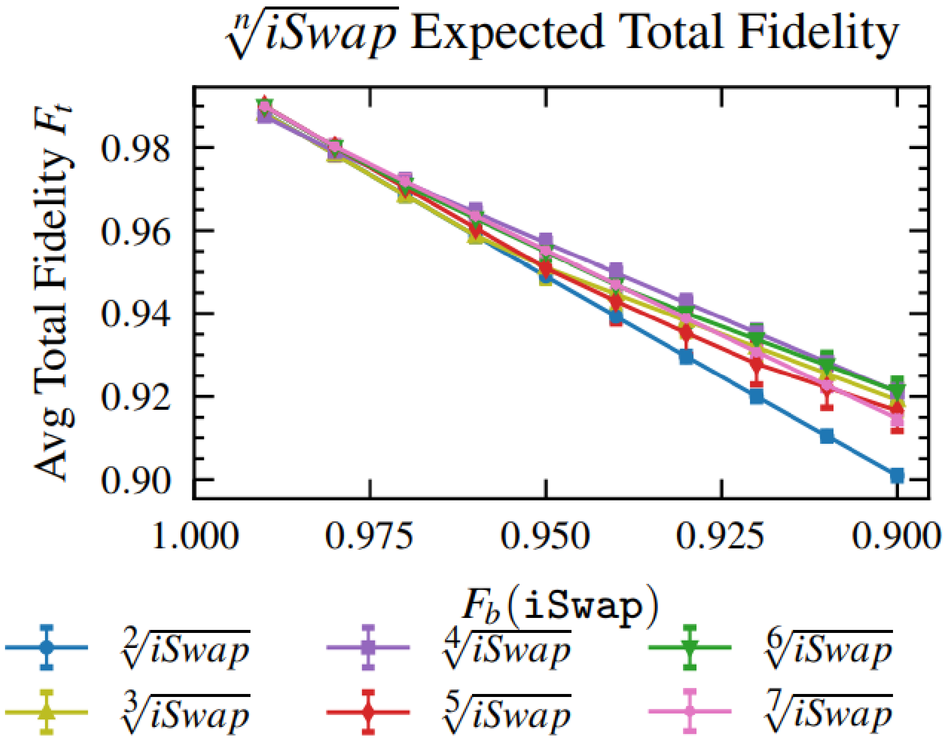
Repeats K times | i ➜ 1 to K



➢ Coherence-limited fidelity as a function of gate duration

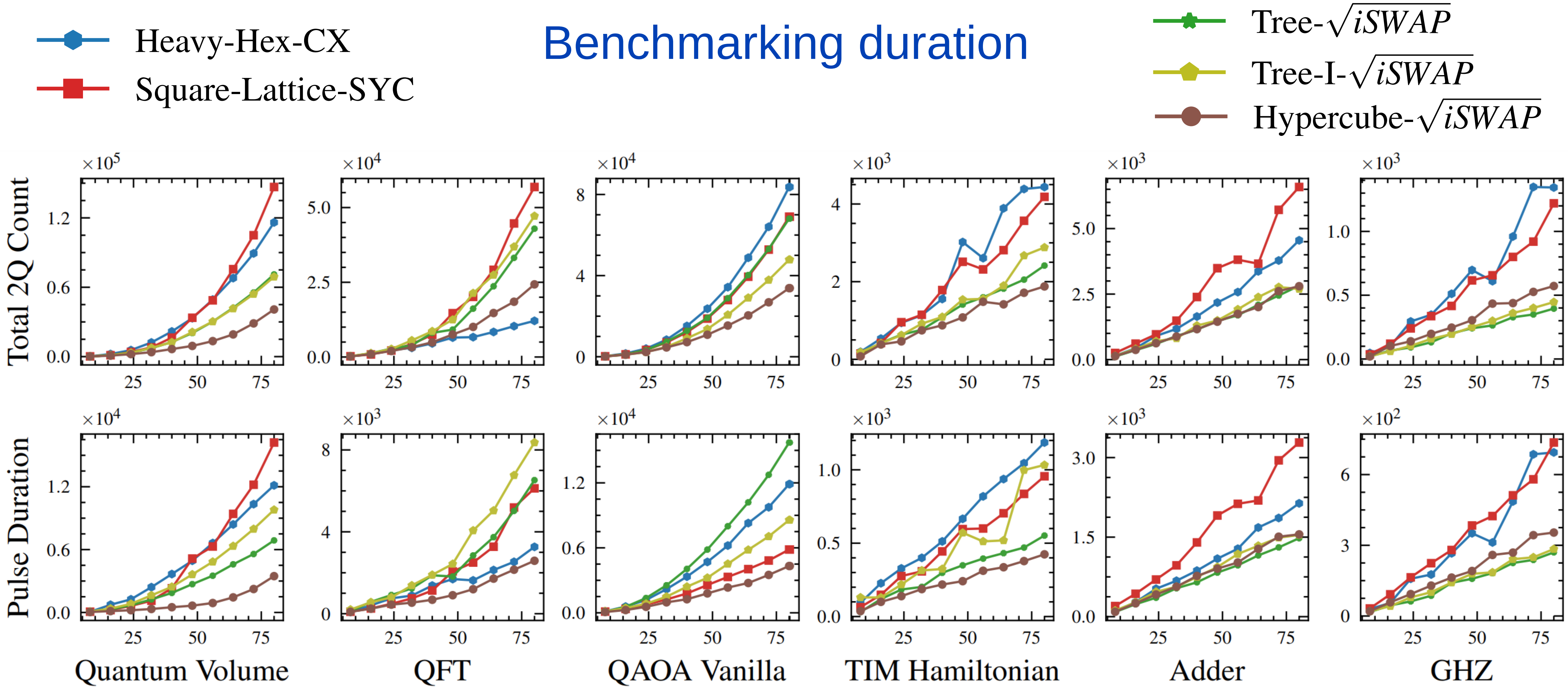
$$F(\sqrt[n]{\text{iSWAP}}) = 1 - \frac{1 - F(\text{iSWAP})}{n}$$





$\sqrt[2]{\text{iSwap}}$ decreases infidelity by 51%, $\sqrt[4]{\text{iSwap}}$ by 58% vs iSwap

Benchmarking duration



- **Considering the impact from both topology and basis gate**
  80Q Heavy-Hex 54% slower (<< fidelity) vs. Hypercubes

Connectivity Topologies + Native Hardware Gates

Conclusion

1. SNAIL coupling provides both powerful topologies and basis gates
2. Corral increases parallelism for near-term quantum applications
3. Continuous iSwap gate set shortens overall duration.

McKinney, et al. **HPCA** (2023).

evm9.dev