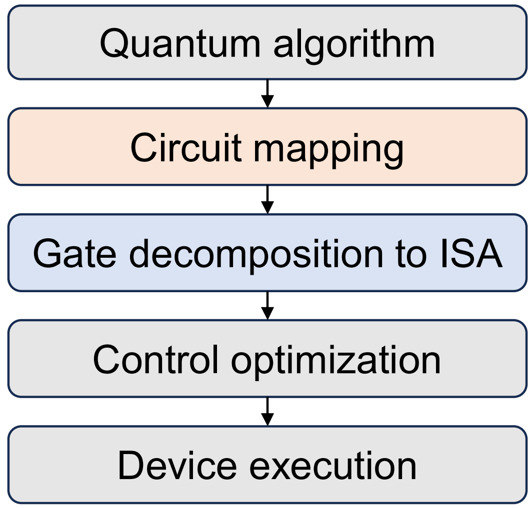
Co-designed architectures for modular superconducting quantum computers

Evan McKinney[†],
M. Hatridge[§], A.K. Jones[†]

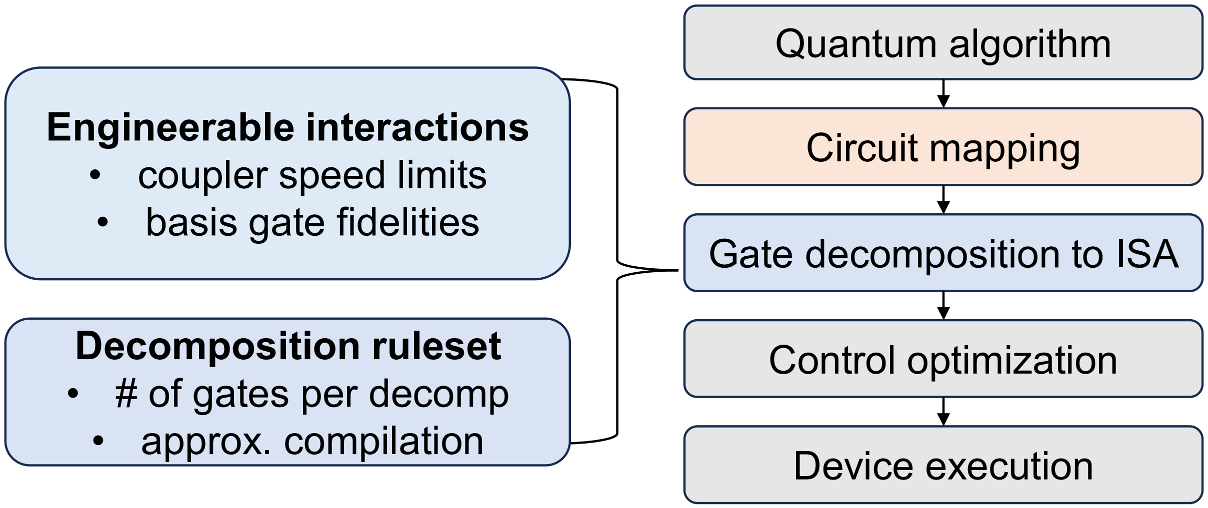
[†]Department of Electrical and Computer Engineering, University of Pittsburgh
[§]Department of Applied Physics, Yale University

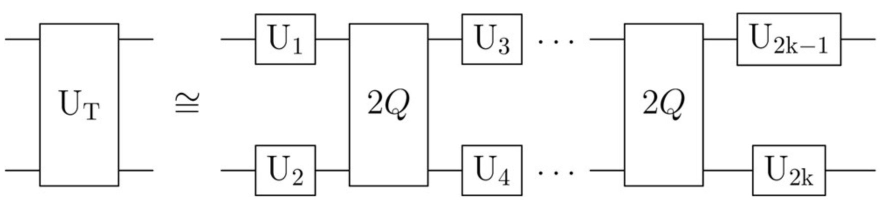
YQI, September 2024

Transpilation as co-design framework



Transpilation as co-design framework





Transpilation as co-design framework

Engineerable interactions
- coupler speed limits
- basis gate fidelities

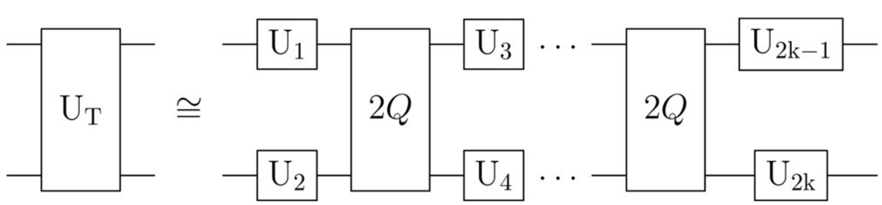
Decomposition ruleset
- # of gates per decomp
- approx. compilation

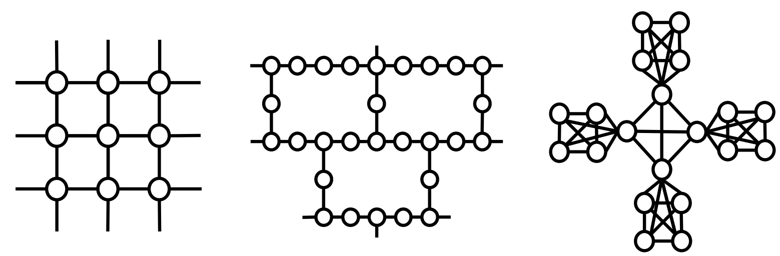
Quantum algorithm
↓
Circuit mapping
↓
Gate decomposition to ISA
↓
Control optimization
↓
Device execution

Device Constraints
- mode frequency bandwidth
- fabrication precision
- hybridization strengths
- coupler speed-limit

Gate fidelity threshold
- # qubits per module
- qubit-coupler connectivity

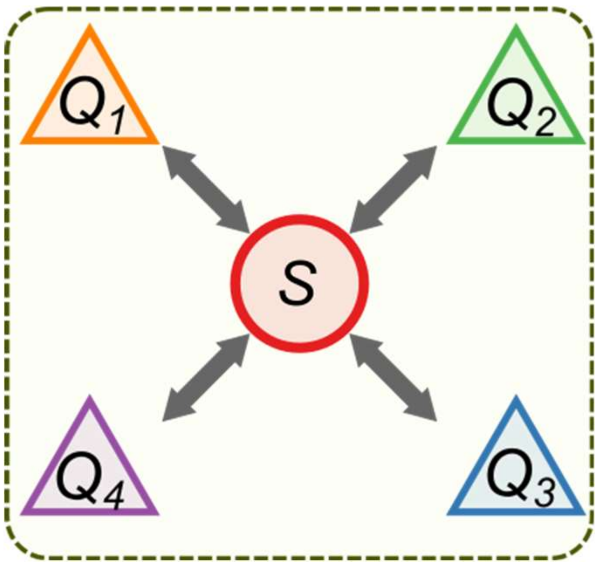
Qubit frequency allocation
- # unique qubit frequencies

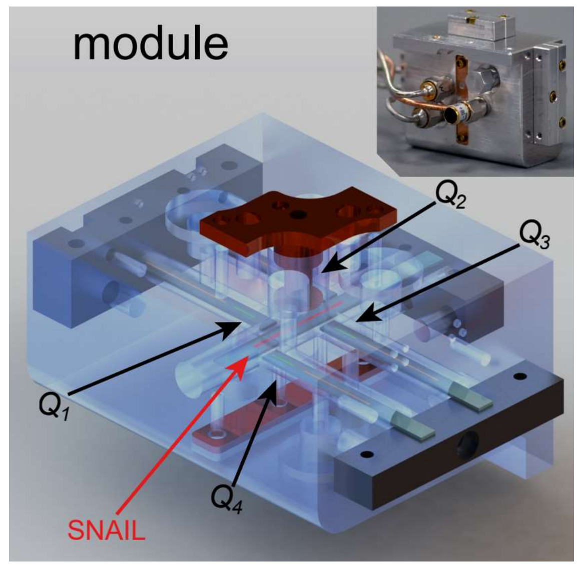


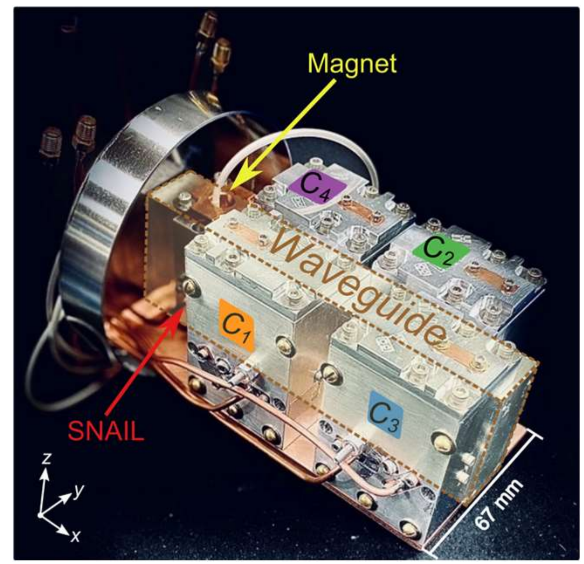


SNAIL module

Four qubit SNAIL-based quantum module

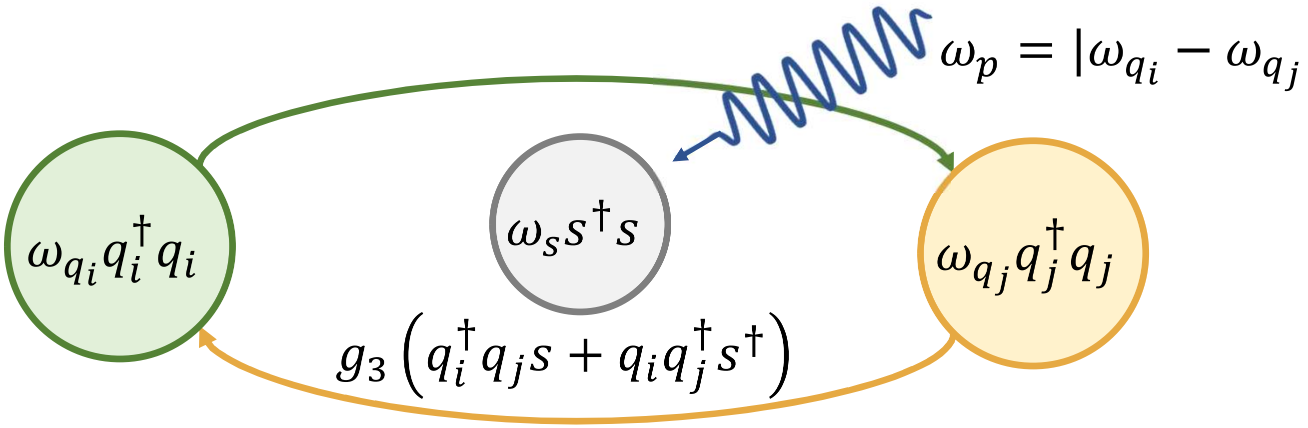






Xia, et al. **arXiv:2306.10162** (2023)
Zhou, et al. **npj Quantum Inf 9**, 54 (2023)

Parametric two qubit gates

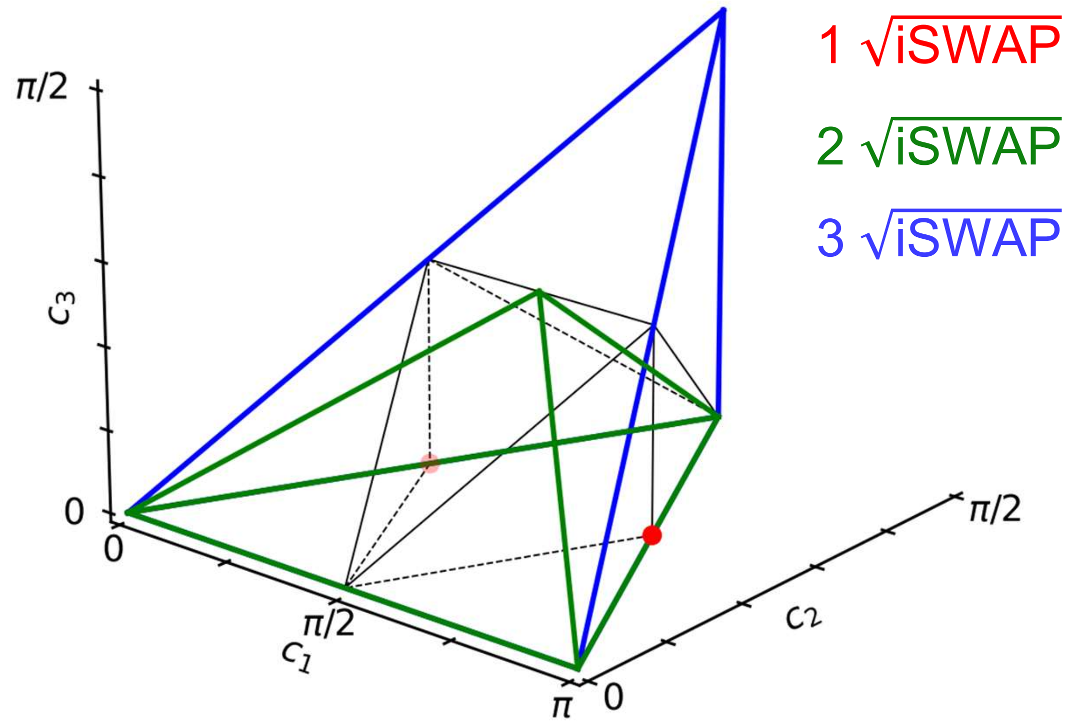


$$\frac{H_{eff}}{\hbar} = g_2^{eff}\left(a^\dagger b + a b^\dagger\right), \qquad where \quad g_2^{eff} = g_3\left(\frac{g}{\Delta}\right)^2 \sqrt{n_p}$$

Clerk, et al. **Reviews of Modern Physics** (2010)
Bergeal, et al. **Nature Physics** (2010)
Frattini, et al. **Applied Physics Letters** (2017)

Efficient instruction sets using $\sqrt{\text{iSWAP}}$

$$\sqrt{\texttt{iSWAP}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{i}{\sqrt{2}} & 0 \\ 0 & \frac{i}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$



McKinney, et al. **ISCA** (2023)
Huang, et al. **Physical Review Letters** (2023)
Chen, et al. **arxiv:2312.05652** (2023)

Geometrically representing quantum gates

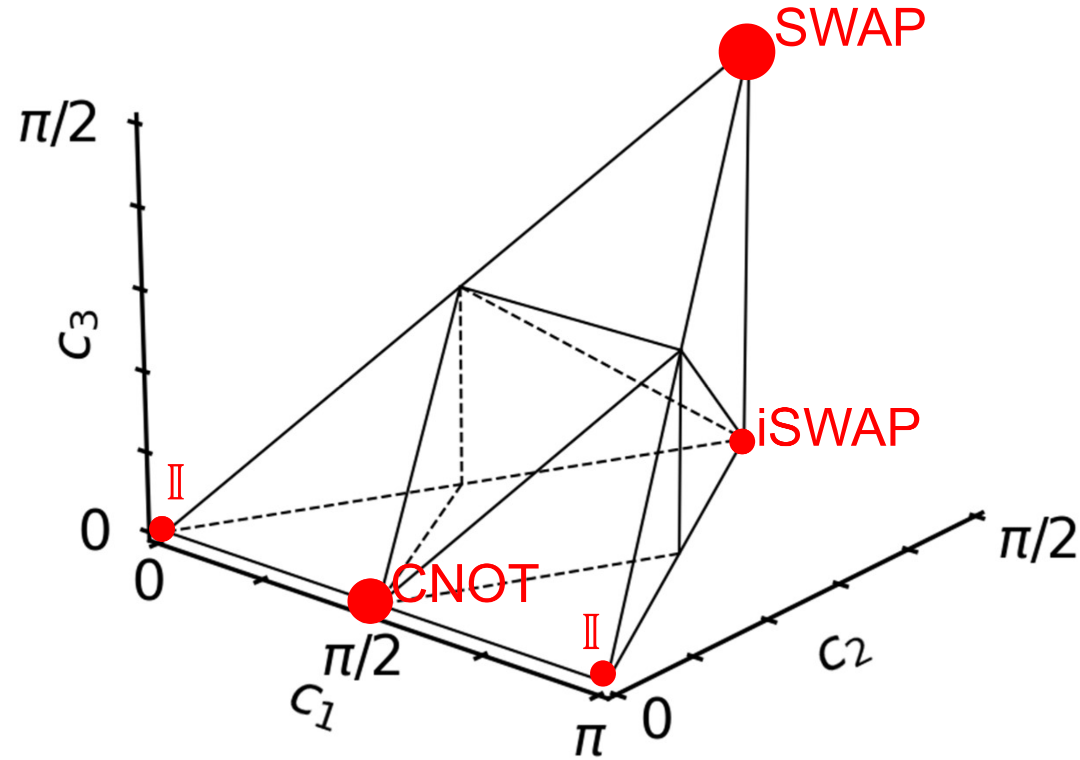
Quantum control

↓

Unitary gate

↓

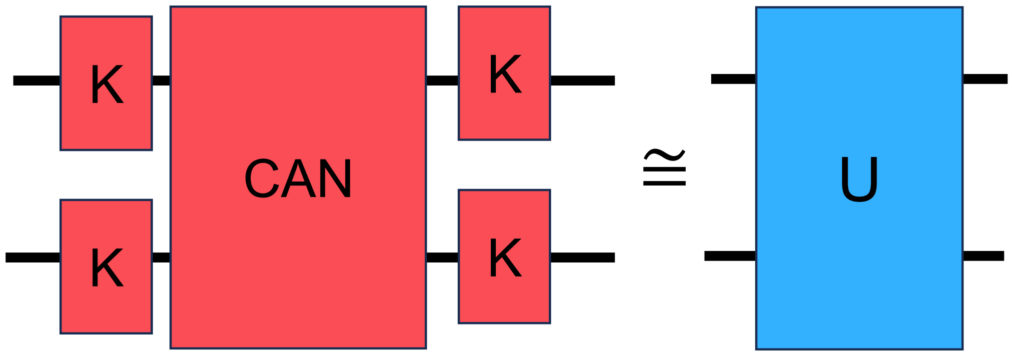
Weyl chamber coordinate



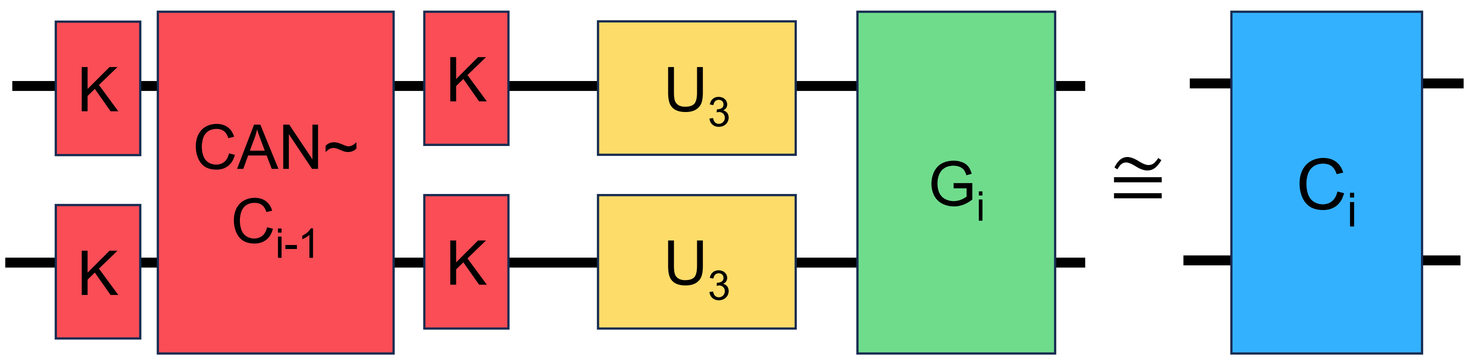
Peterson, et al. **Quantum 4** (2020)
Makhlin, Yuriy. **Quantum Information Processing** (2002)
Watts, et al. **Entropy 15** (2013)
Zhang, et al. **Physical Review A** (2003)

Cartan KAK decomposition

All 2Q gates can be specified by 3 invariants up to local rotations (1Q)



$$U = (K_1 \otimes K_2) e^{-i(k_x \sigma_{XX} + k_y \sigma_{YY} + k_z \sigma_{ZZ})} (K_3 \otimes K_4)$$



Conversion/Gain candidate basis gates

- **Engineerable interactions from 3-wave mixing**

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

- Native gates limited to XX,YY components

$$\hat{H} = \frac{1}{2}[(g_c + g_g)XX + (g_c - g_g)YY]$$

- Cartan trajectories from Makhlin invariants

$$G_1 = (\cos(2g_c t) + \cos(2g_g t))^2/4$$

$$G_2 = \cos(2(g_c - g_g)t) + \cos(2(g_c + g_g)t) + 1$$

Conversion/Gain candidate basis gates

- **Engineerable interactions from 3-wave mixing**

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

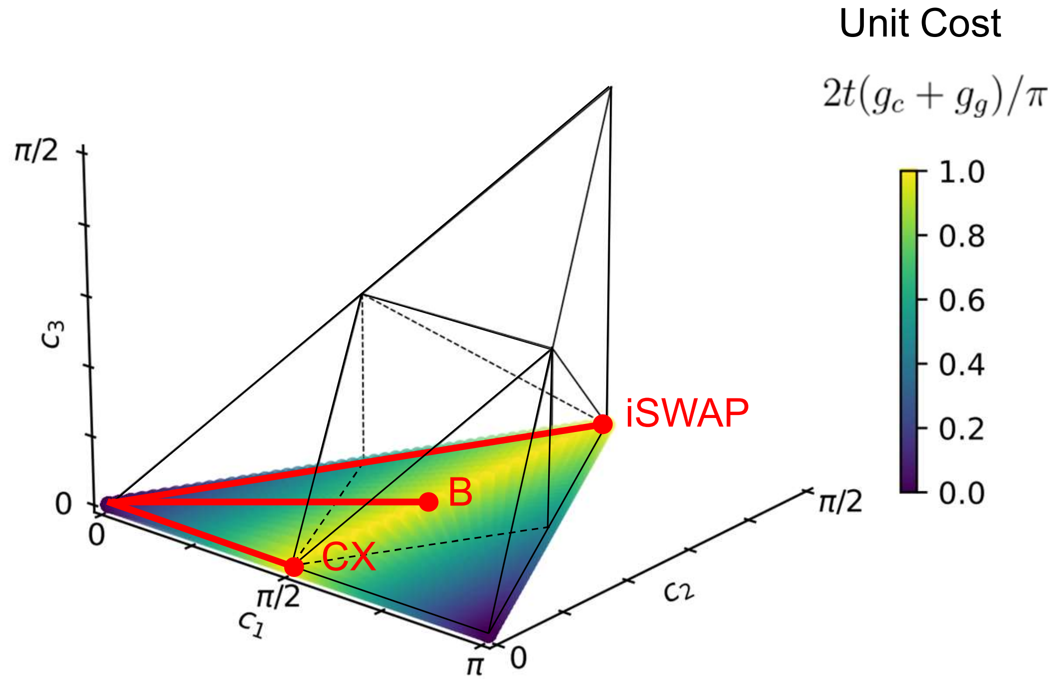
- Native gates limited to XX,YY components

$$\hat{H} = \frac{1}{2}[(g_c + g_g)XX + (g_c - g_g)YY]$$

- Cartan trajectories from Makhlin invariants

$$G_1 = (\cos(2g_c t) + \cos(2g_g t))^2/4$$

$$G_2 = \cos(2(g_c - g_g)t) + \cos(2(g_c + g_g)t) + 1$$



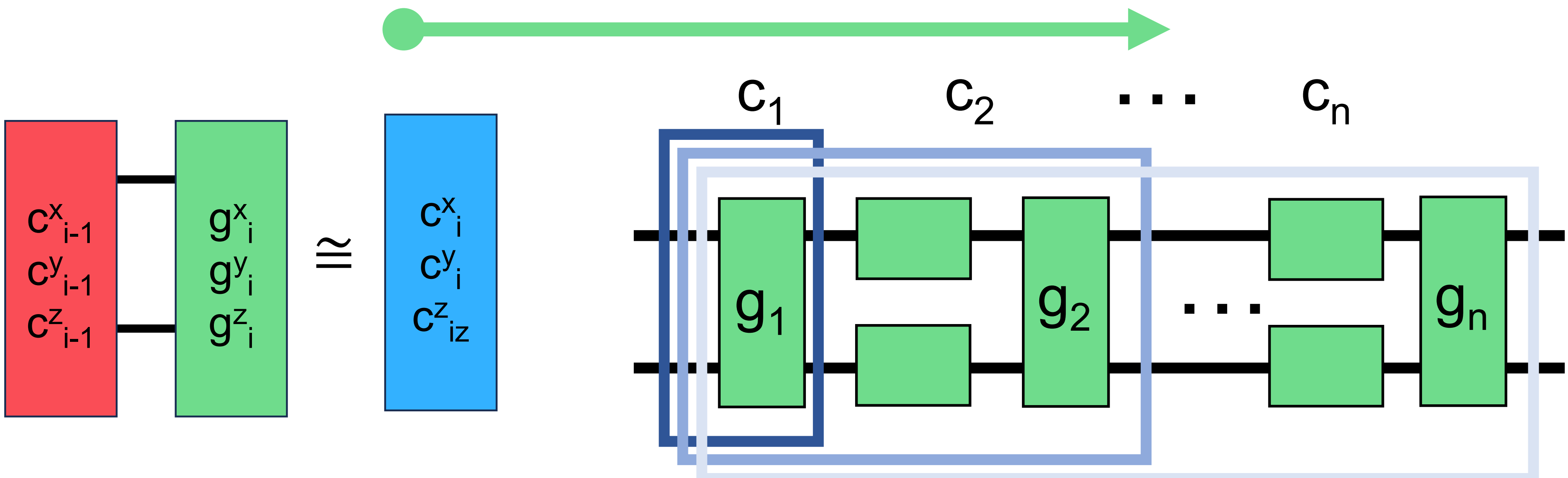
Monodromy polytopes

Satisfying all 72 **quantum Littlewood-Richardson** linear inequalities implies

given a circuit with invariant $\mathbf{c}_{i-1}$,

appending $\mathbf{g}_i$, a gate from the device's instruction set,

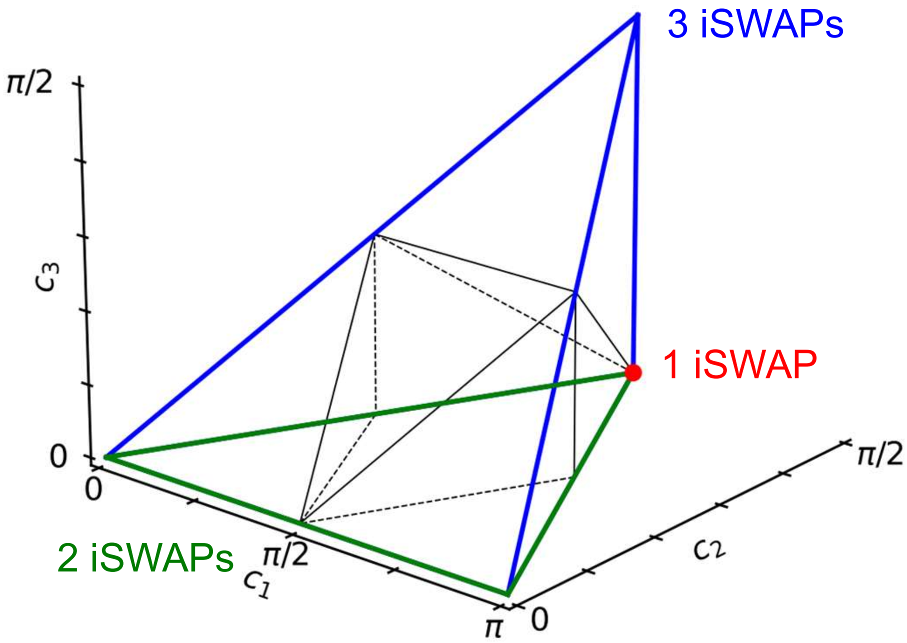
can yield a circuit with invariant $\mathbf{c}_i$.



Satisfy all $\mathbf{L}_i(\mathbf{c}_{i-1}, \mathbf{g}_i, \mathbf{c}_i)$ such that $\forall i, g_i \in \mathrm{ISA}$

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

Basis coverage volumes



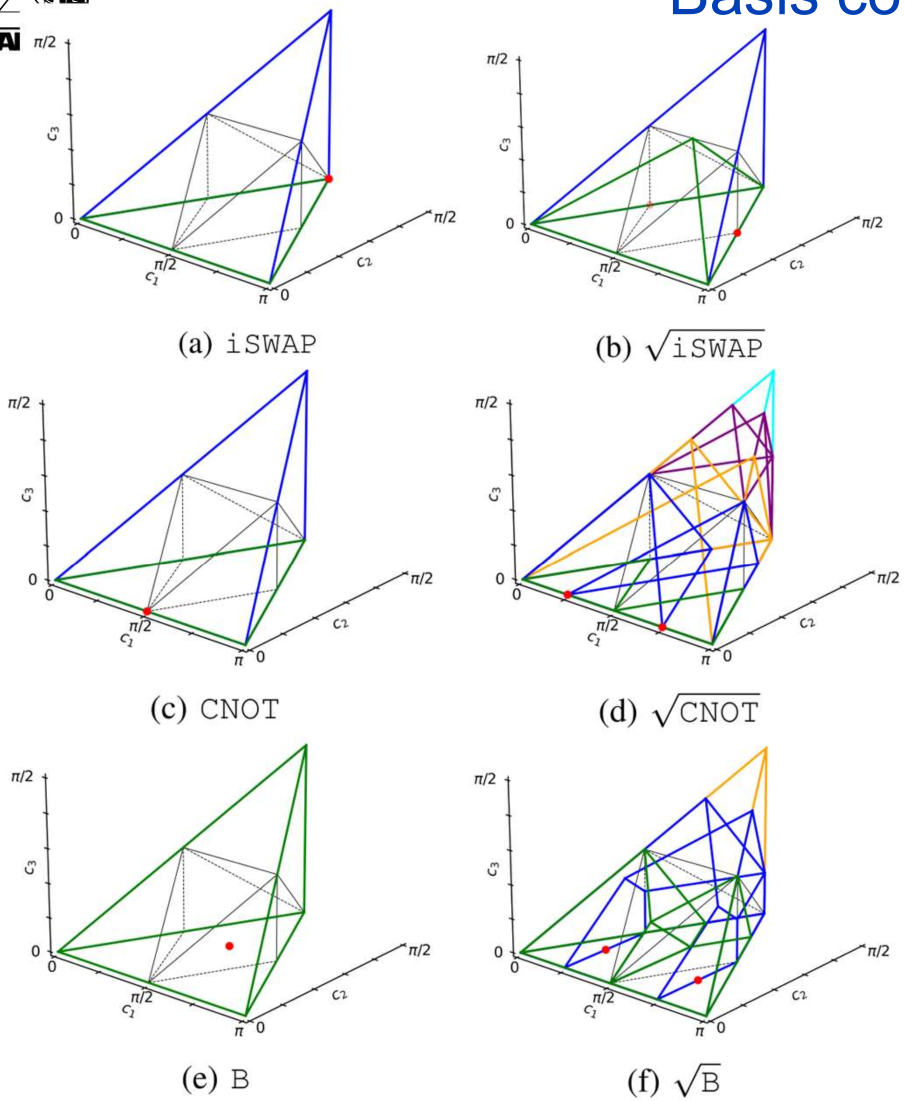
- **Monodromy polytopes** finds minimum gate applications for any 2Q target gate!

Target\Basis	iSWAP
CNOT	2.0
SWAP	3.0
Haar	3.0

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

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

Basis coverage volumes



- **Monodromy polytopes** finds minimum gate applications for any 2Q target gate!

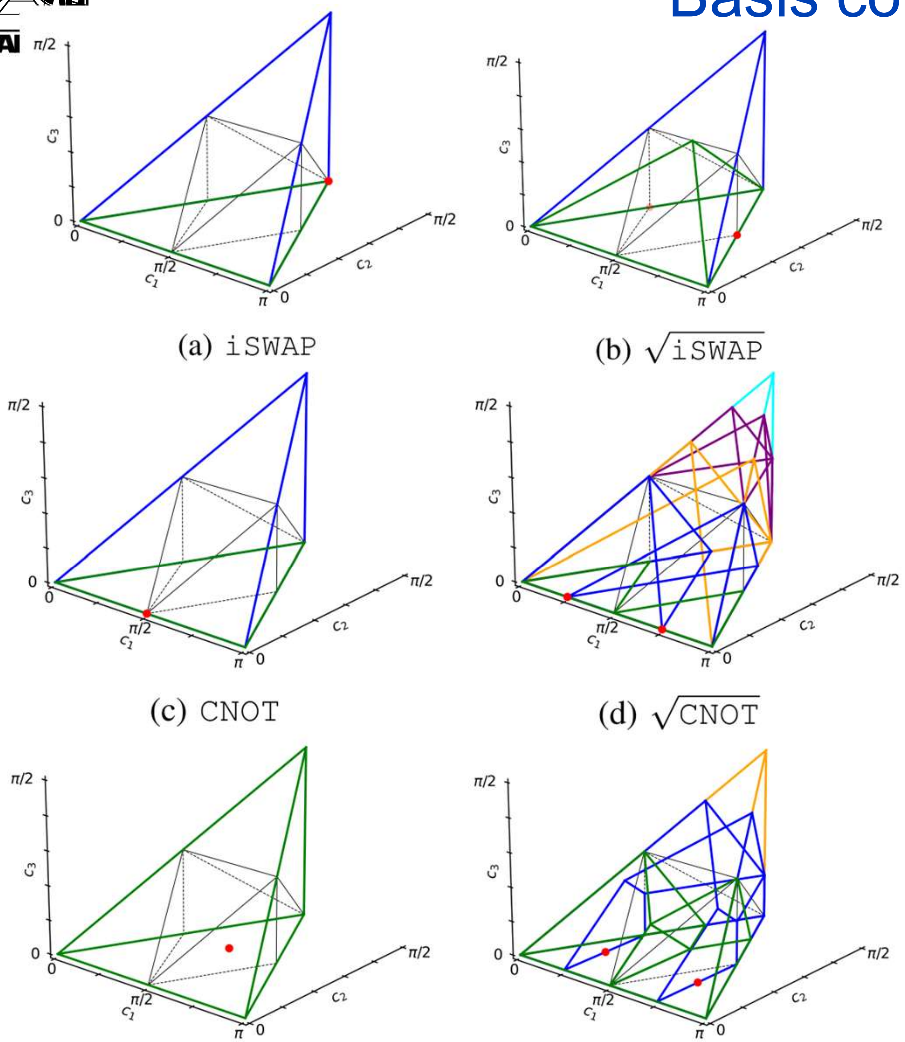
Decomposition *gate count* costs

Target\Basis	iSWAP	$\sqrt{\text{iSWAP}}$	CX	$\sqrt{\text{CX}}$	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)

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

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!

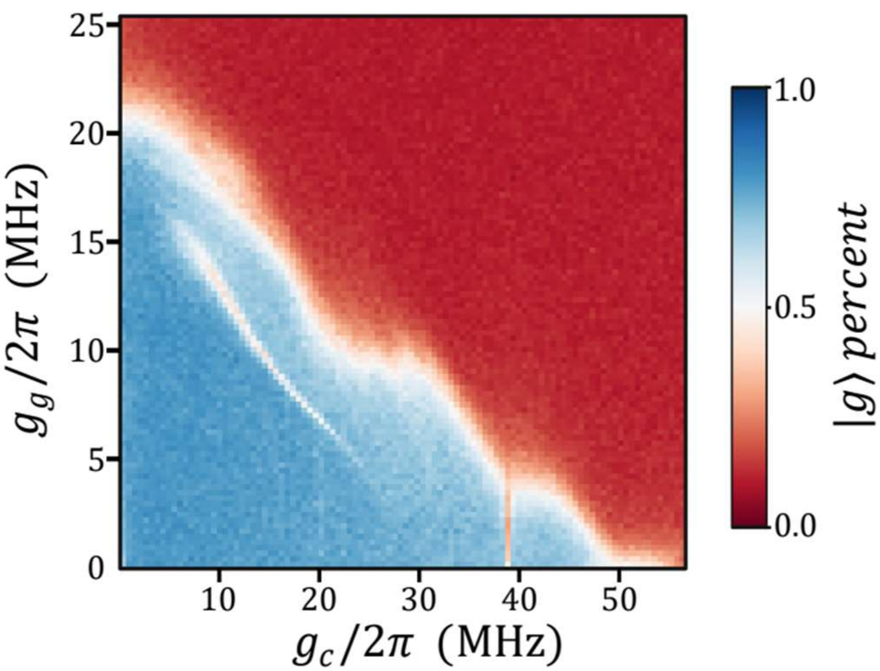
Decomposition *gate count* costs

Target\Basis	iSWAP	$\sqrt{\text{iSWAP}}$	CX	$\sqrt{\text{CX}}$	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)

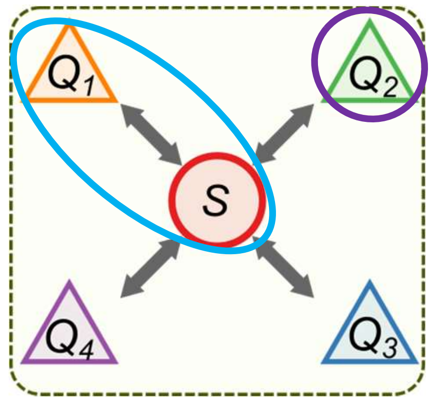
McKinney, et al. **ISCA** (2023)

Hardware speed limits



Limitation of SNAIL when driving both gain and conversion

Module

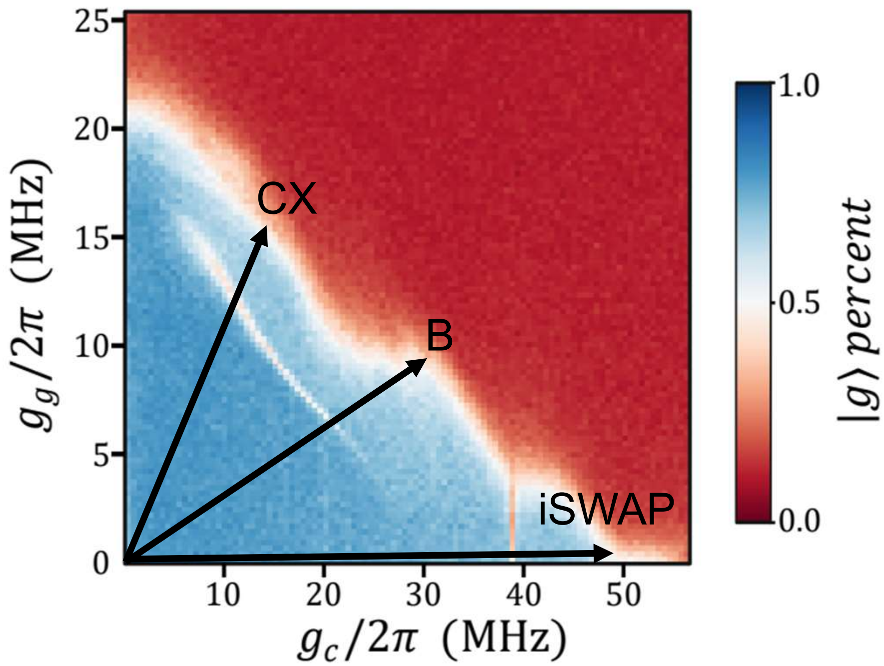


Drives applied between SNAIL and qubit

Measure second qubit to witness SNAIL breakpoint

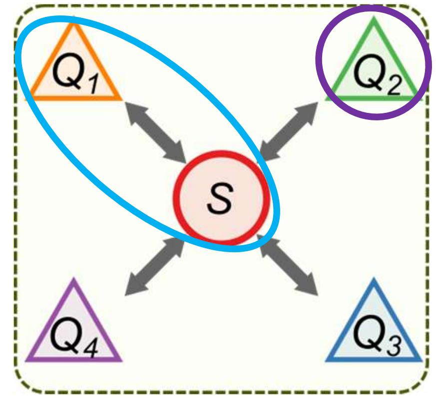
Zhou, et al. **npj Quantum Inf 9**, 54 (2023).

Hardware speed limits



Limitation of SNAIL when driving both gain and conversion

Module

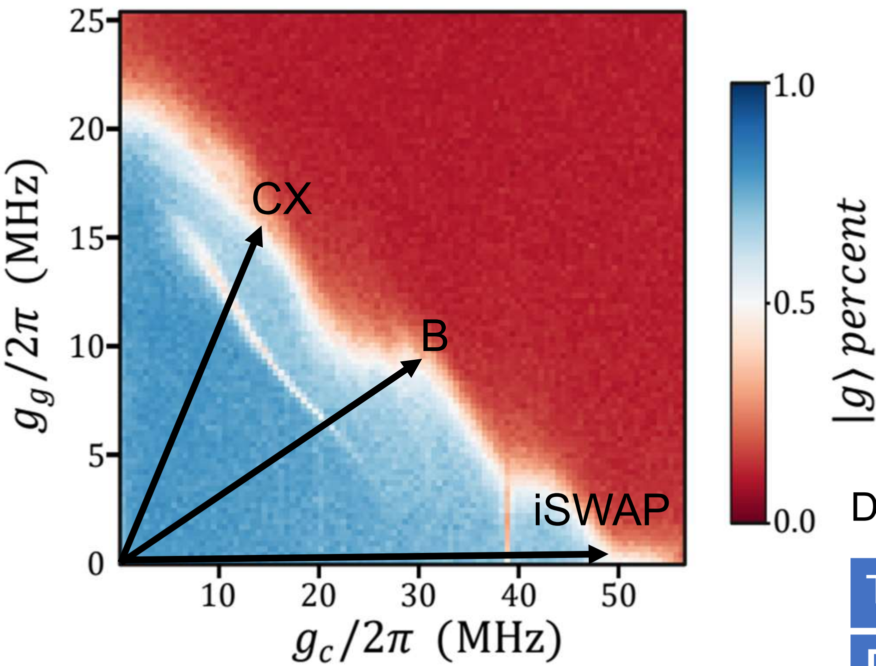


Drives applied between SNAIL and qubit

Measure second qubit to witness SNAIL breakpoint

Zhou, et al. **npj Quantum Inf 9**, 54 (2023).

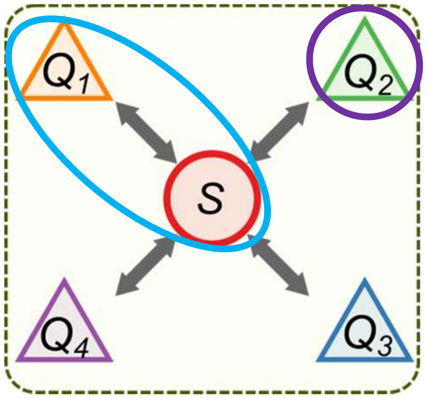
Hardware speed limits



Limitation of SNAIL when driving both gain and conversion

Zhou, et al. **npj Quantum Inf 9**, 54 (2023).

Module



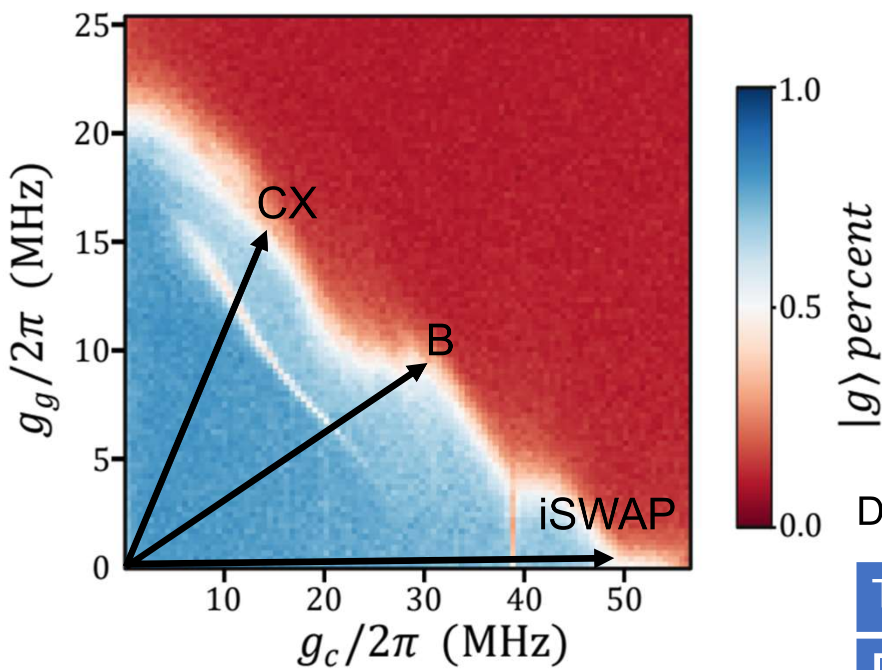
Drives applied between SNAIL and qubit

Measure second qubit to witness SNAIL breakpoint

Decomposition normalized *duration* costs

Target\Basis	iSWAP	$\sqrt{\text{iSWAP}}$	CX	$\sqrt{\text{CX}}$	B	$\sqrt{\text{B}}$
Duration	**1.0**	**0.5**	**1.8**	**0.9**	**1.4**	**0.7**
CNOT	2.0	1.0	1.8	1.8	2.8	1.4
SWAP	3.0	1.5	5.4	5.4	2.8	2.8
Haar	3.0	1.1	5.4	3.2	2.8	2.2

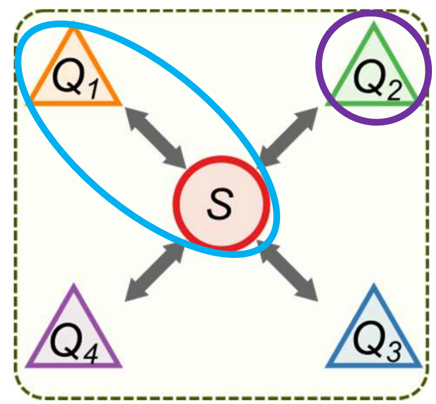
Hardware speed limits



Limitation of SNAIL when driving both gain and conversion

Zhou, et al. **npj Quantum Inf 9**, 54 (2023).

Module



Drives applied between SNAIL and qubit

Measure second qubit to witness SNAIL breakpoint

Decomposition normalized *duration* costs

Target\Basis	iSWAP	$\sqrt{\text{iSWAP}}$	CX	$\sqrt{\text{CX}}$	B	$\sqrt{\text{B}}$
Duration	**1.0**	**0.5**	**1.8**	**0.9**	**1.4**	**0.7**
CNOT	2.0	1.0	1.8	1.8	2.8	1.4
SWAP	3.0	1.5	5.4	5.4	2.8	2.8
Haar	3.0	1.1	5.4	3.2	2.8	2.2

Transpilation as co-design framework

Engineerable interactions
- coupler speed limits
- basis gate fidelities

Decomposition ruleset
- # of gates per decomp
- approx. compilation

Quantum algorithm

Circuit mapping

Gate decomposition to ISA

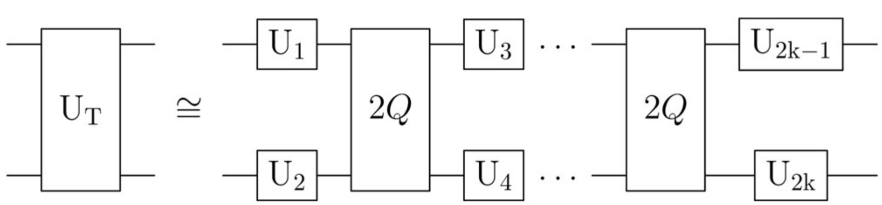
Control optimization

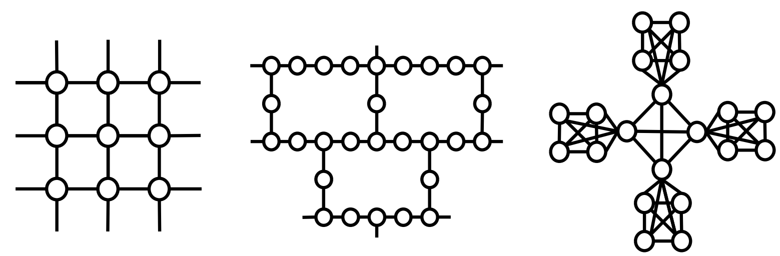
Device execution

Device Constraints
- mode frequency bandwidth
- fabrication precision
- hybridization strengths
- coupler speed-limit

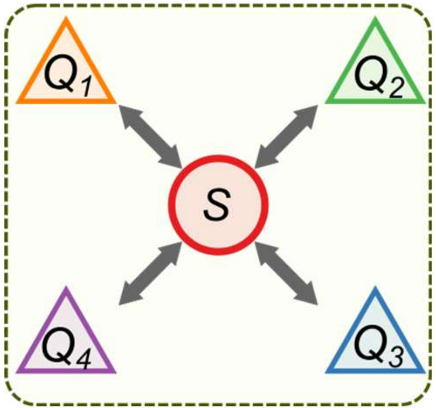
Gate fidelity threshold
- # qubits per module
- qubit-coupler connectivity

Qubit frequency allocation
- # unique qubit frequencies

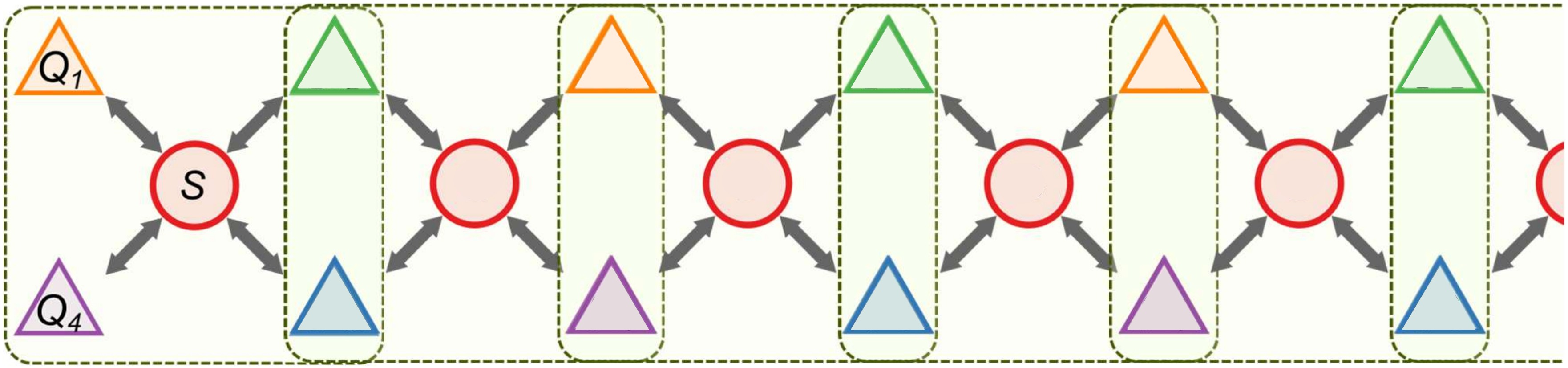


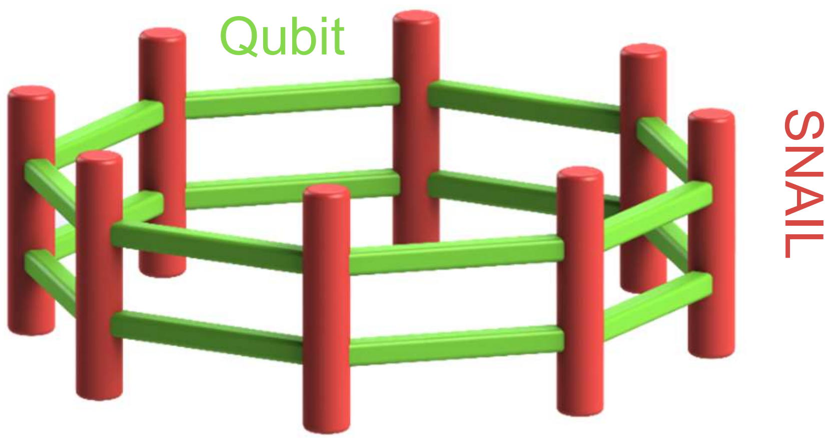


Extending the module

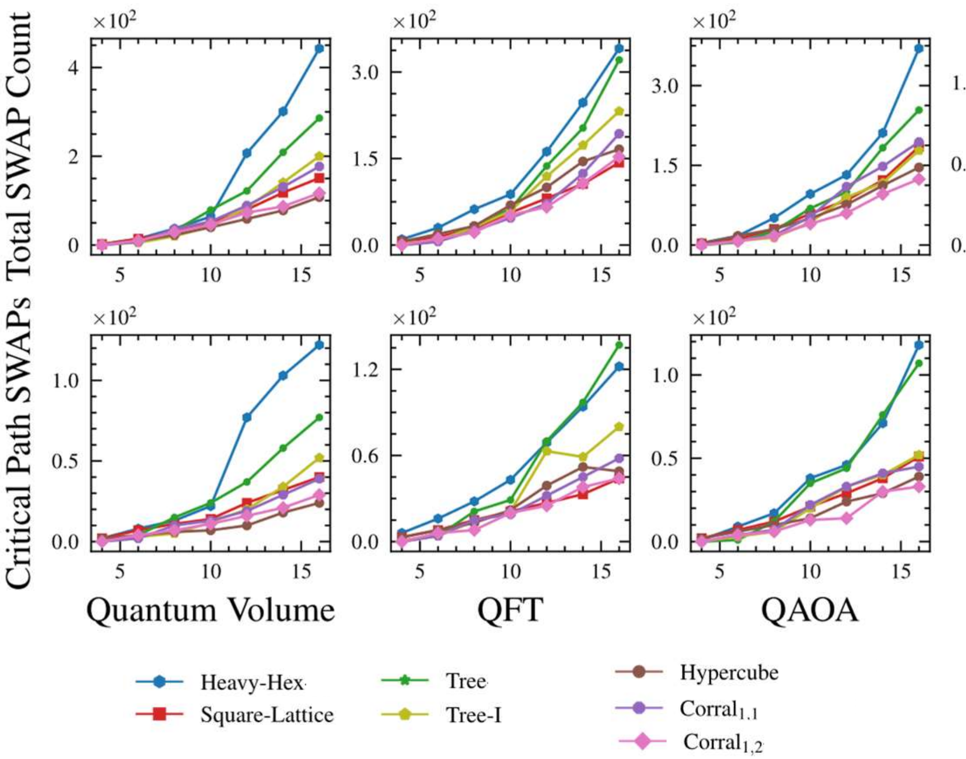


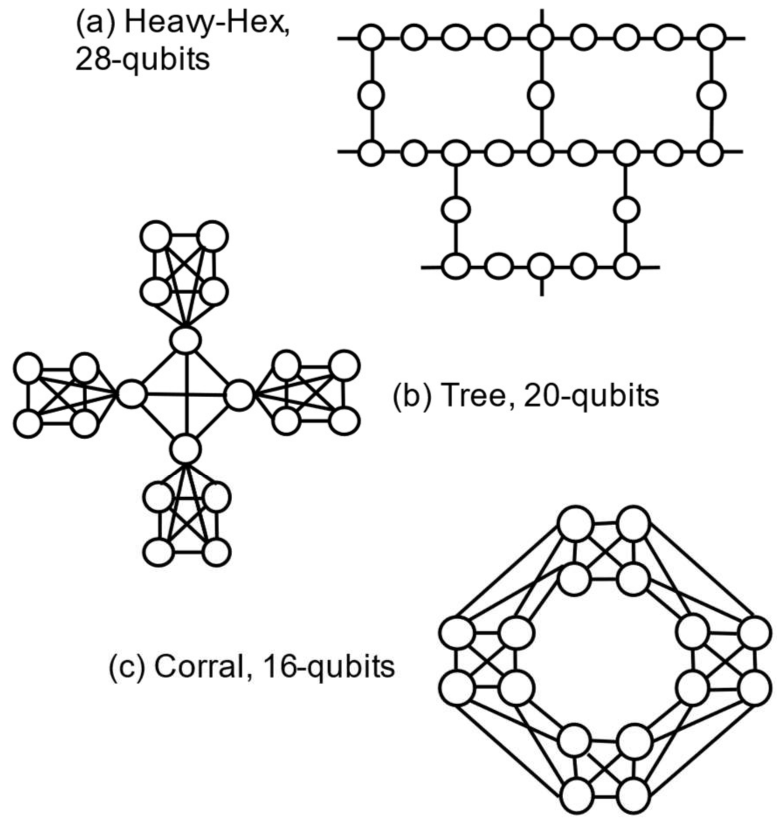
Extending the module





Connectivity reduces circuit costs

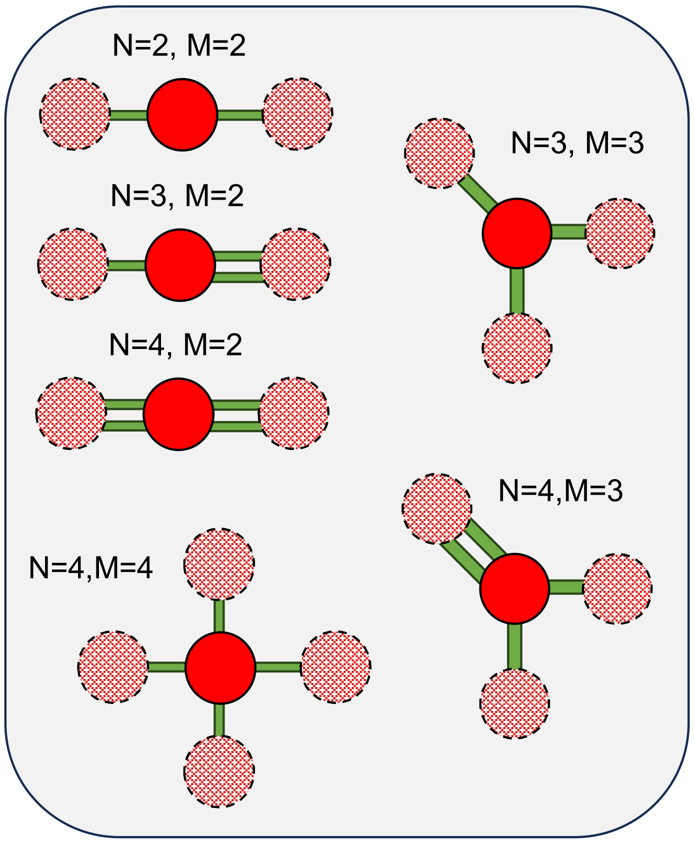




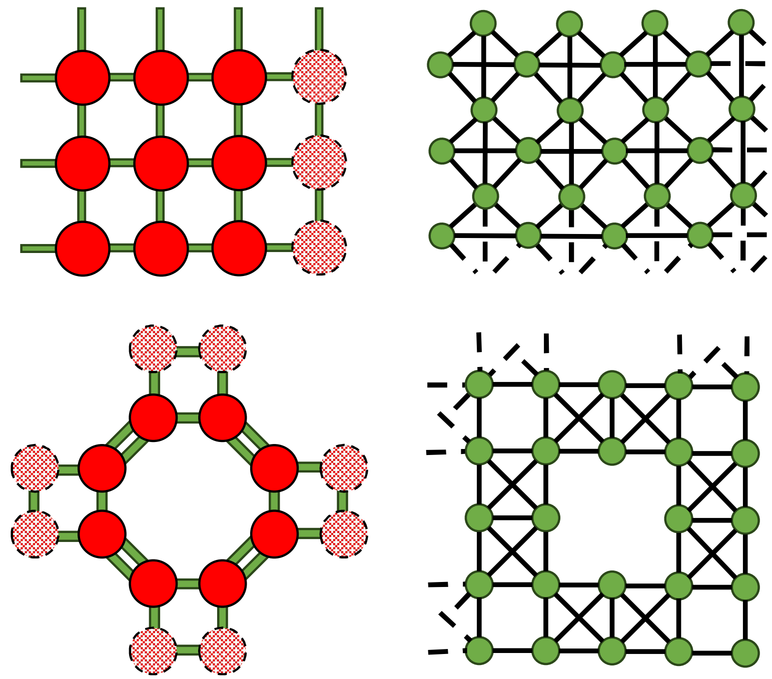
McKinney, et al. **HPCA** (2023)

Alternative designs?

SNAIL and qubit frequencies **must yield unique sets of parametric drives** per neighborhoods

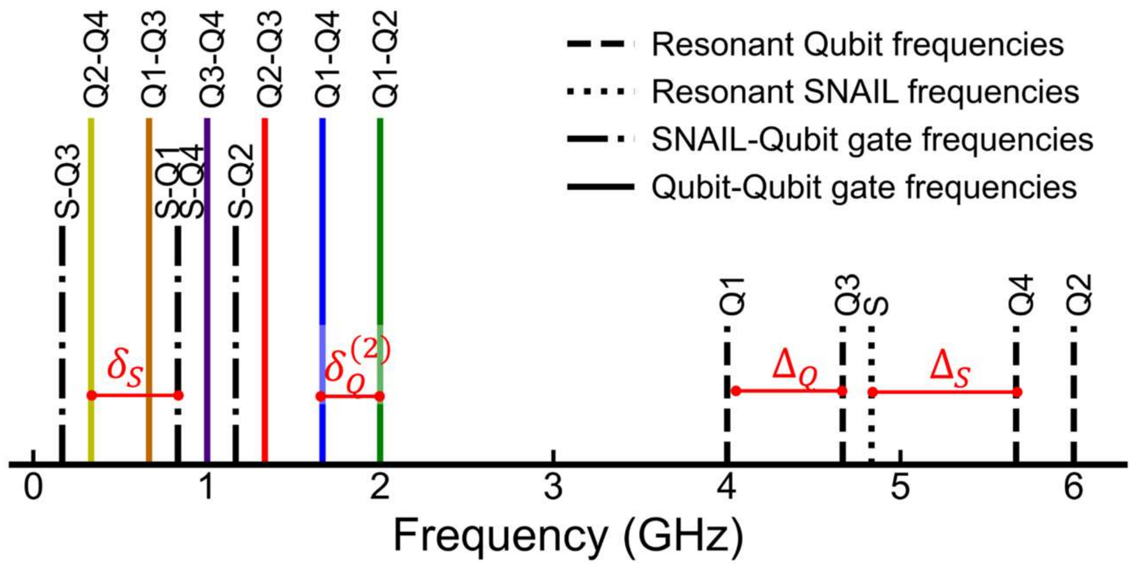


Maximum incident edges per node (N)?
Maximum adjacent vertices (M)?
Maximum vertex variants (k)?



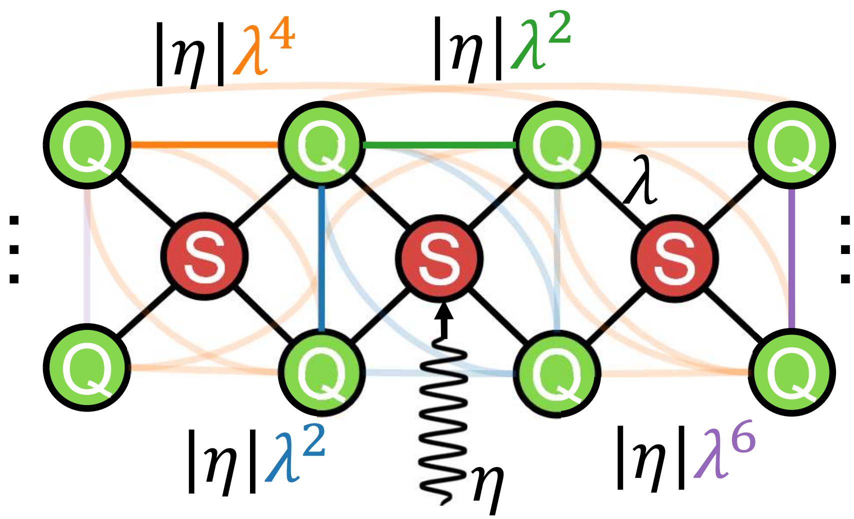
Spectral crowding of terms

$$\tilde{H}_I = g_3 \left(s e^{-i\tilde{\omega}_s t} + \eta e^{-i\omega_p t} + \sum_i \lambda_{si} q_i e^{-i\tilde{\omega}_{q_i} t} + \text{h.c.} \right)^3$$
$$+ \sum_i \frac{\alpha_i}{12} \left(q_i e^{-i\tilde{\omega}_{q_i} t} - \lambda_{si} (s e^{-i\tilde{\omega}_s t} + \eta e^{-i\omega_p t}) + \text{h.c.} \right)^4$$



Spectator error budgeting

- Intra-module terms protected by frequency separation
- Inter-module terms protected by weak SNAIL-SNAIL hybridization



$$\eta = (\epsilon^x + i\epsilon^y)\frac{\omega_s}{(\omega_p^2 - \omega_s^2)}$$

$$H_{conv} = \sum_{i>j} q_i^\dagger q_j \left(6\eta g_3 \lambda^2 e^{-it(\omega_{q_i} - \omega_{q_j} + \omega_p)}\right) + \text{h.c.}$$

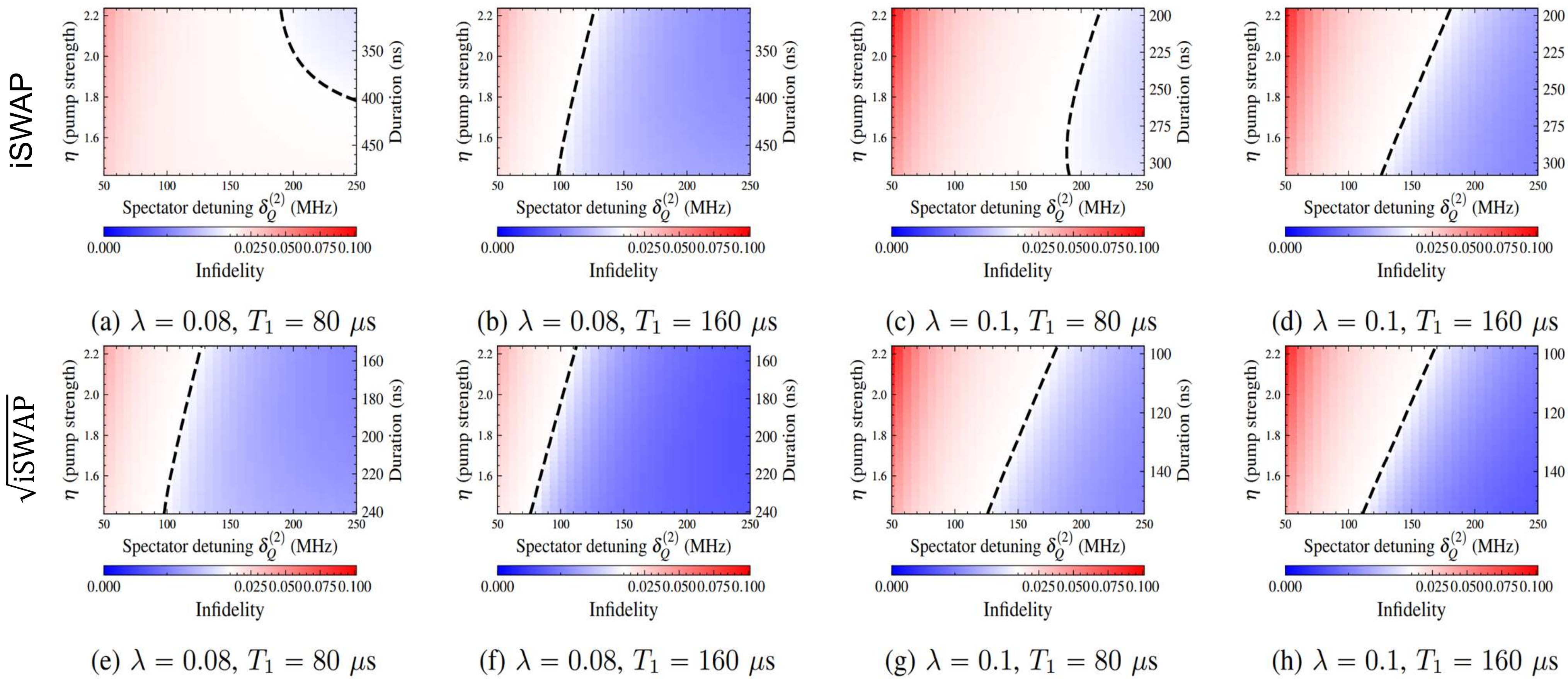
$$H(t) = 6|\eta| g_3 \lambda^2 \left(\mathbf{q_1^\dagger q_2} + \frac{2}{\delta_Q^{(2)} \log 2}\mathbf{q_3^\dagger q_4}\right)$$

When is RWA a good approximation?

Error budget as a separation design constraint

$$F_{\mathrm{avg}}(U,V) = \frac{16\,\mathrm{Tr}[UV^\dagger] + 1}{17}$$

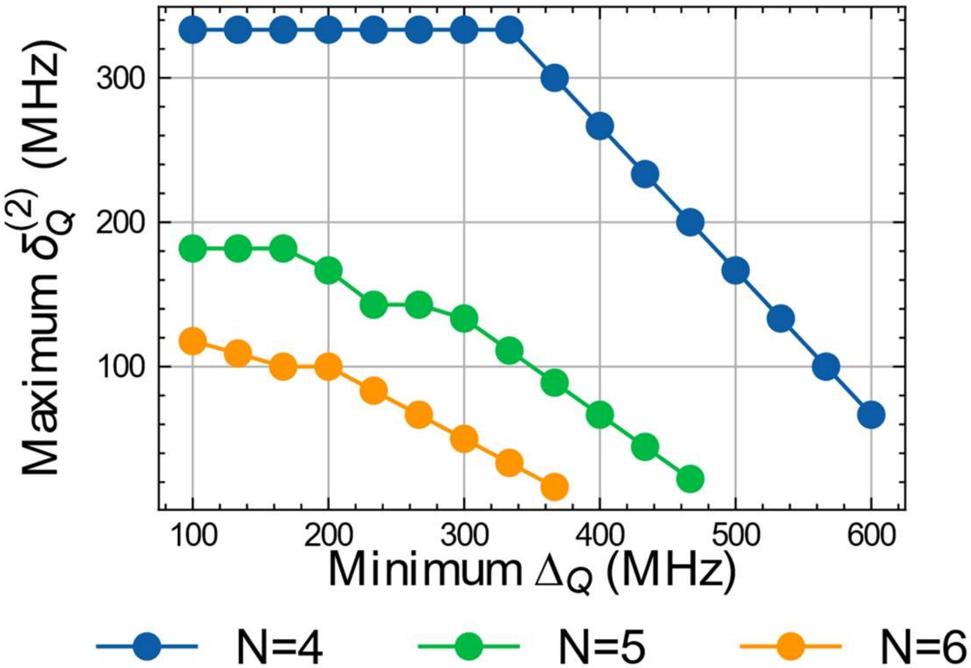
Threshold at F > .99



(a) $\lambda = 0.08,\ T_1 = 80\ \mu s$ (b) $\lambda = 0.08,\ T_1 = 160\ \mu s$ (c) $\lambda = 0.1,\ T_1 = 80\ \mu s$ (d) $\lambda = 0.1,\ T_1 = 160\ \mu s$

(e) $\lambda = 0.08,\ T_1 = 80\ \mu s$ (f) $\lambda = 0.08,\ T_1 = 160\ \mu s$ (g) $\lambda = 0.1,\ T_1 = 80\ \mu s$ (h) $\lambda = 0.1,\ T_1 = 160\ \mu s$

Satisfying constraints using linear programming



➢ Binary search:
maximum conversion separation (y-axis)
given minimum qubit separation (x-axis)

Next steps:

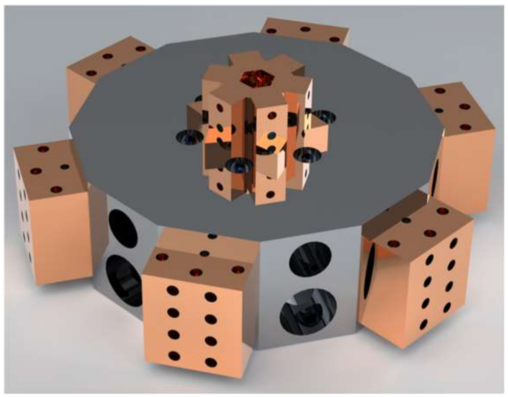
- Constraints for SNAIL-qubit conversion
- Multi-qubit gate + speed limits
- Scheduling pulses for double SNAILS

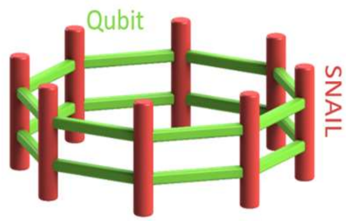
➢ Qubit bandwidth [4,6] GHz

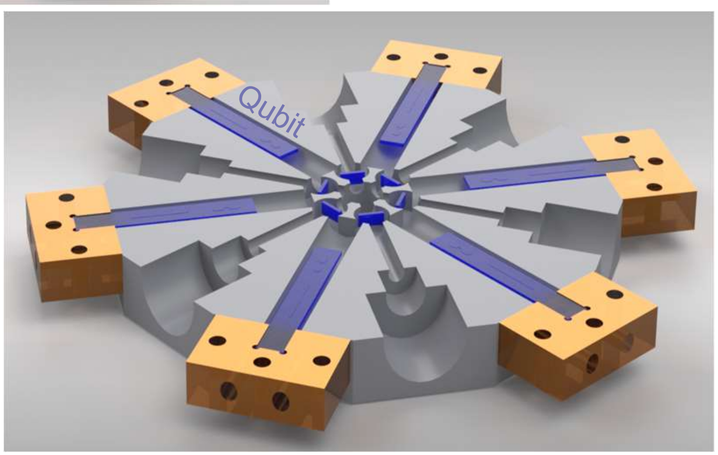
McKinney, et al. **arXiv:2409.18262** (2024)

Physical realization of the module (and potential problems?)

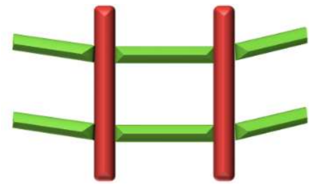
Corral

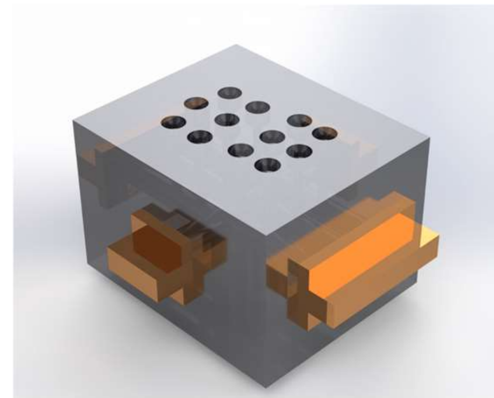


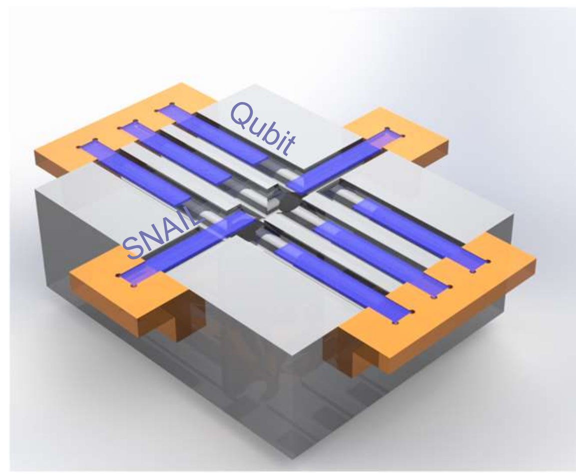




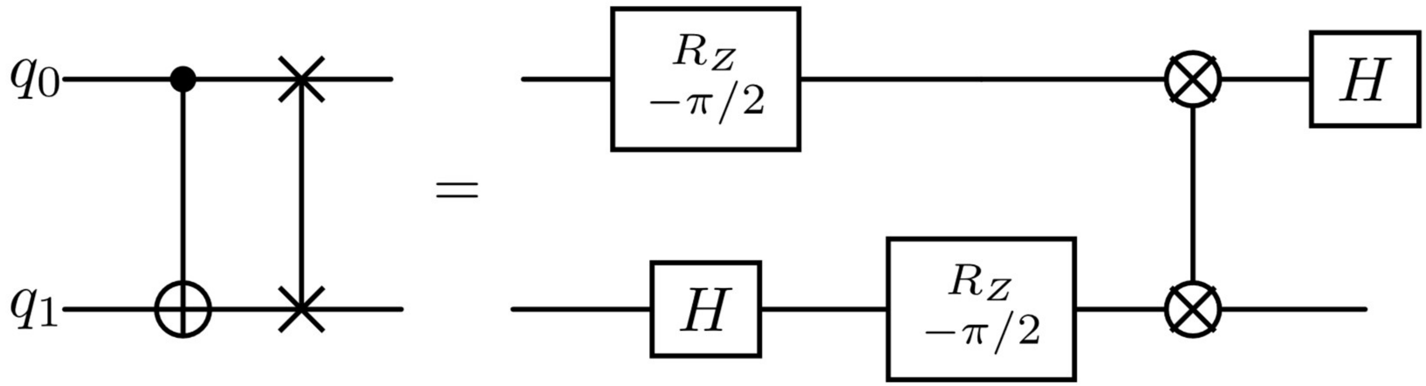
Fence-line

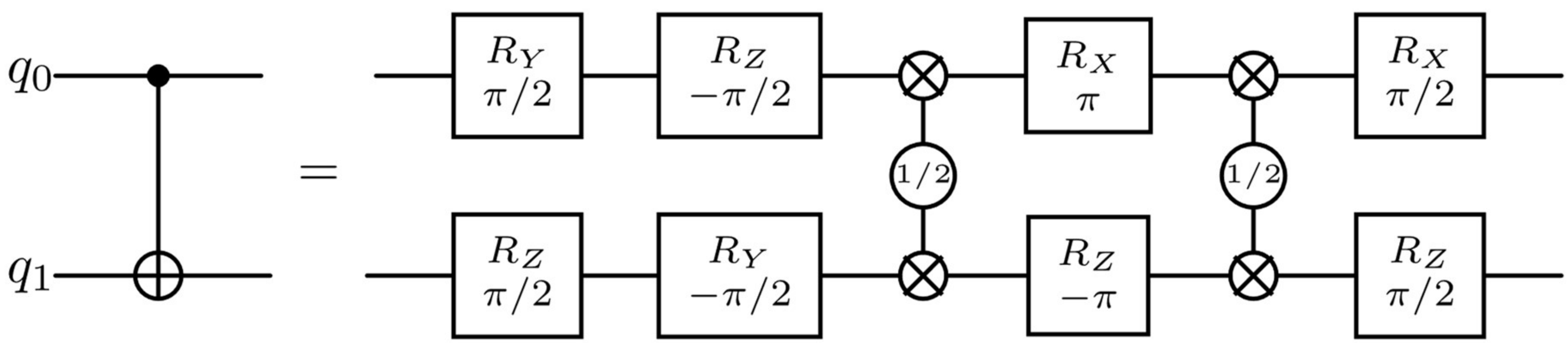






Decomposition identities



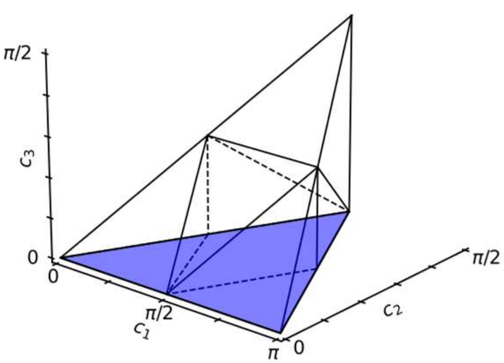


Schuch, et al. **Physical Review A 67.3** (2003)
Huang, et al. **Physical Review Letters** (2023)
McKinney, et al. **HPCA** (2024)

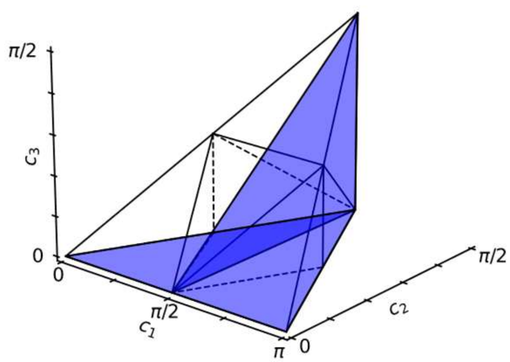
Mirror-inclusive coverage sets

- Compute effective coverage volumes using monodromy polytopes

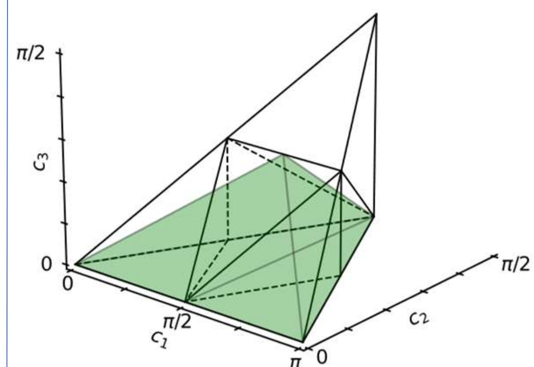
CNOT



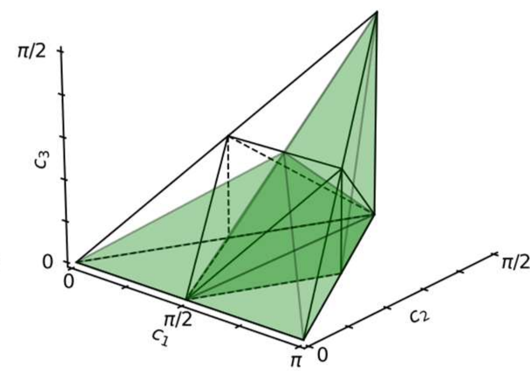
CNOT+Mirrors



$\sqrt{\text{iSWAP}}$



$\sqrt{\text{iSWAP}}$+Mirrors



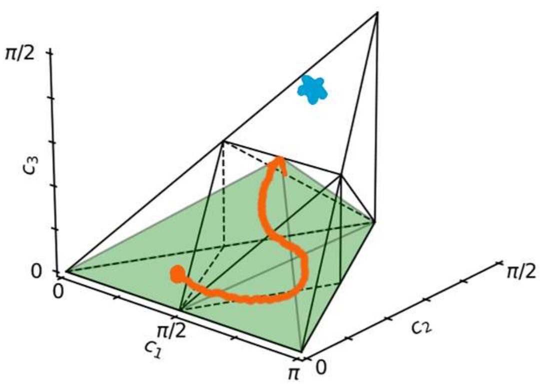
Peterson, et al. **Quantum 4** (2020)
Peterson, et al. **Quantum 6** (2022)
McKinney, et al. **HPCA** (2024)

$k = 2$

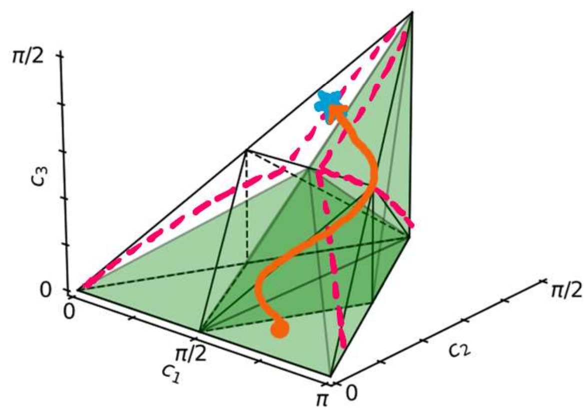
Monte Carlo Haar scores

- Intuition: Approximate decomp threshold defines an acceptable inflated polytope volume.

Javadi, Ali. **APS March Meeting** (2023)



Exact: Fail



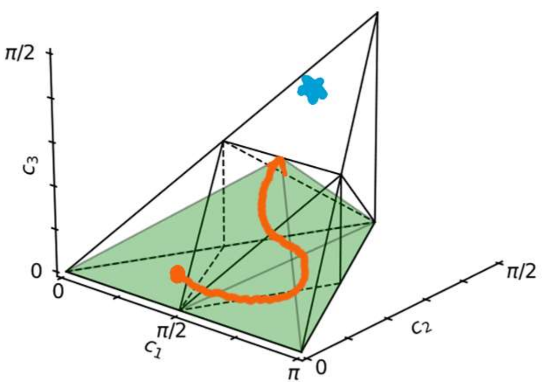
Approx. + Mirrors: Success

Lao, et al. **ISCA** (2021)
McKinney, et al. **HPCA** (2024)

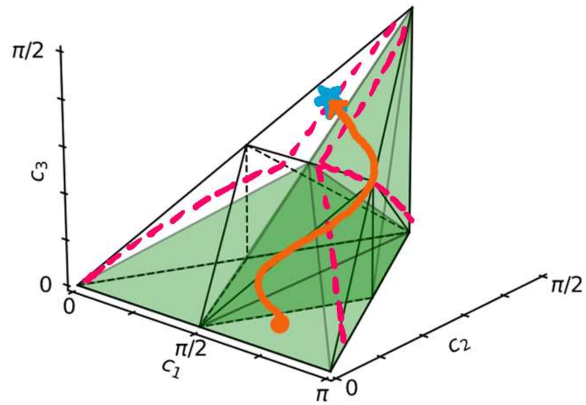
Monte Carlo Haar scores

- Intuition: Approximate decomp threshold defines an acceptable inflated polytope volume.

Javadi, Ali. **APS March Meeting** (2023)



Exact: Fail

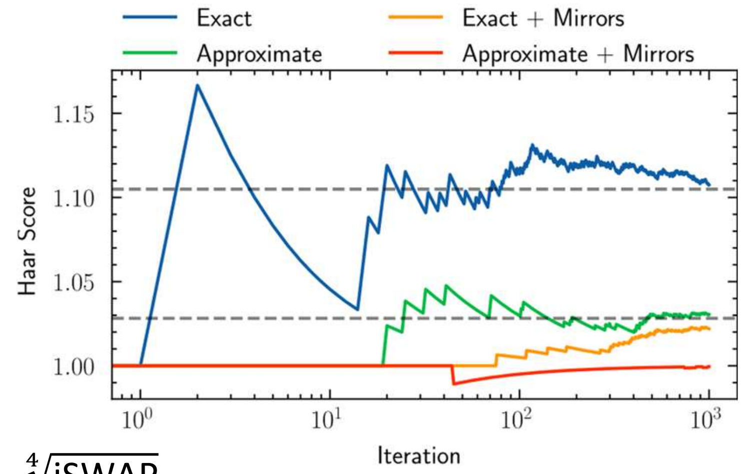


Approx. + Mirrors: Success

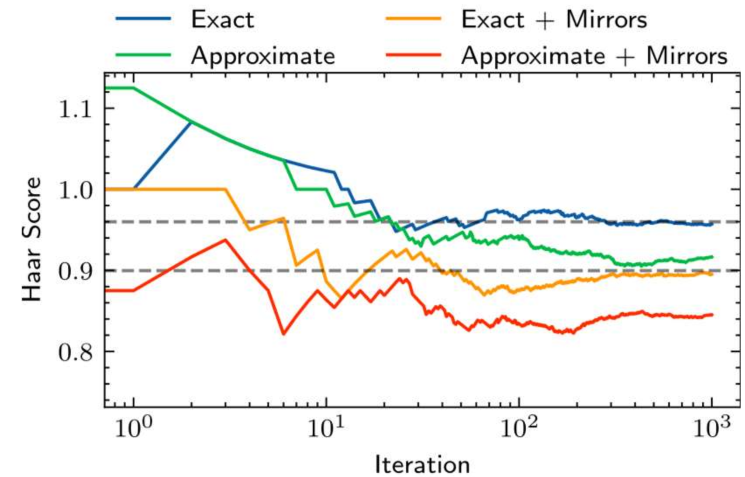
- $\sqrt{\text{iSWAP}}$ with approximate decomp + mirrors has an 8.8% relative decrease in total infidelity

Lao, et al. **ISCA** (2021)
McKinney, et al. **HPCA** (2024)

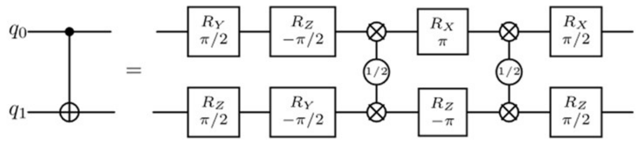
$\sqrt{\text{iSWAP}}$

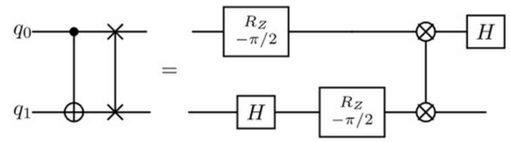


$\sqrt[4]{\text{iSWAP}}$



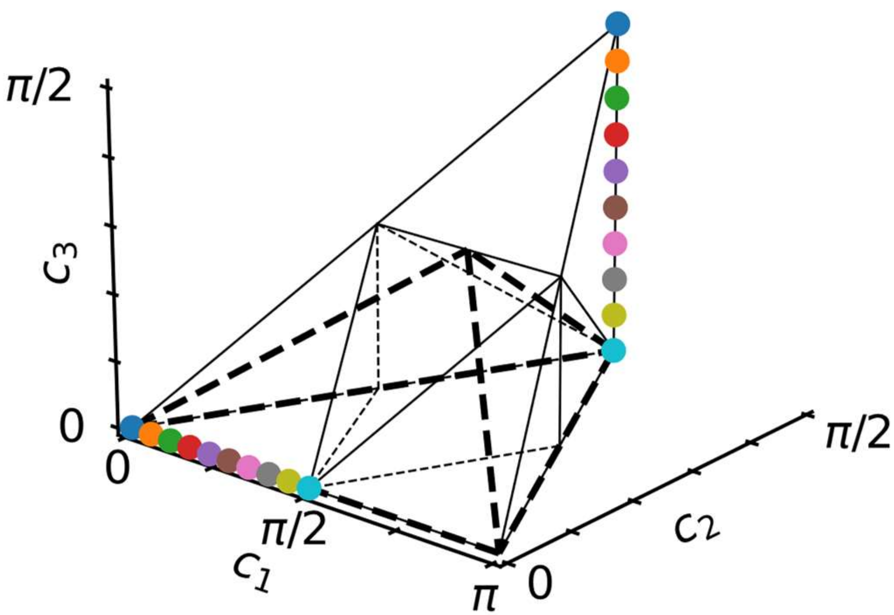
Why does this work?





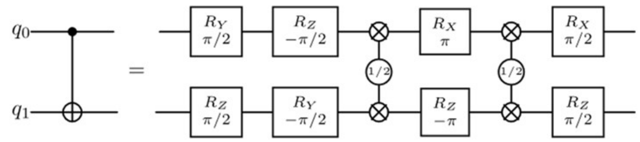
- CPHASE gates mirror to pSWAP gates

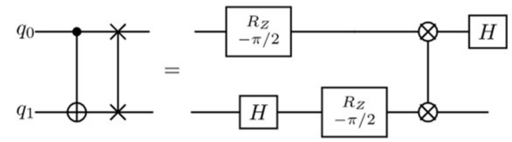
$$(a', b', c') = \begin{cases} (\frac{\pi}{4} + c, \frac{\pi}{4} - b, \frac{\pi}{4} - a) & \text{if } a \leq \frac{\pi}{4} \\ (\frac{\pi}{4} - c, \frac{\pi}{4} - b, a - \frac{\pi}{4}) & \text{else} \end{cases}$$



Peterson, et al. **Quantum 6** (2022)

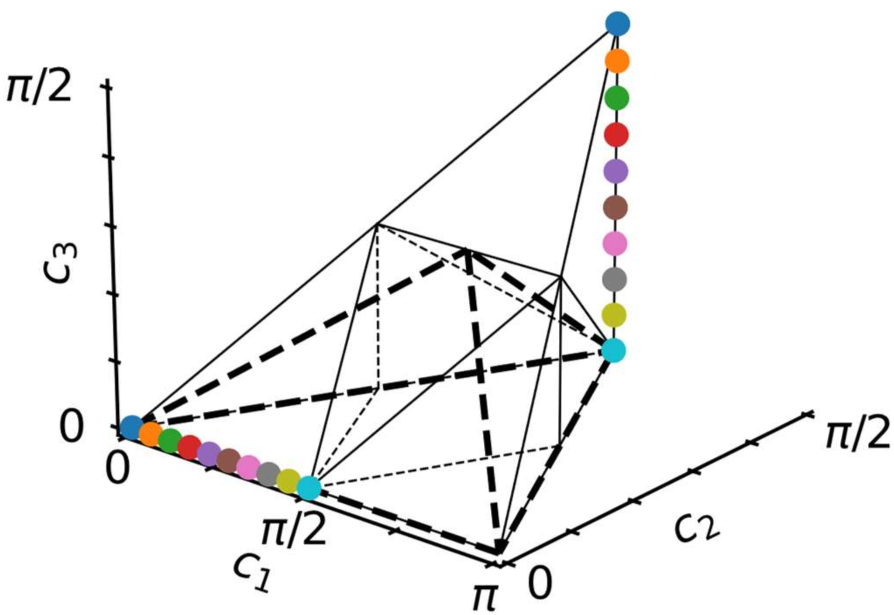
Why does this work?

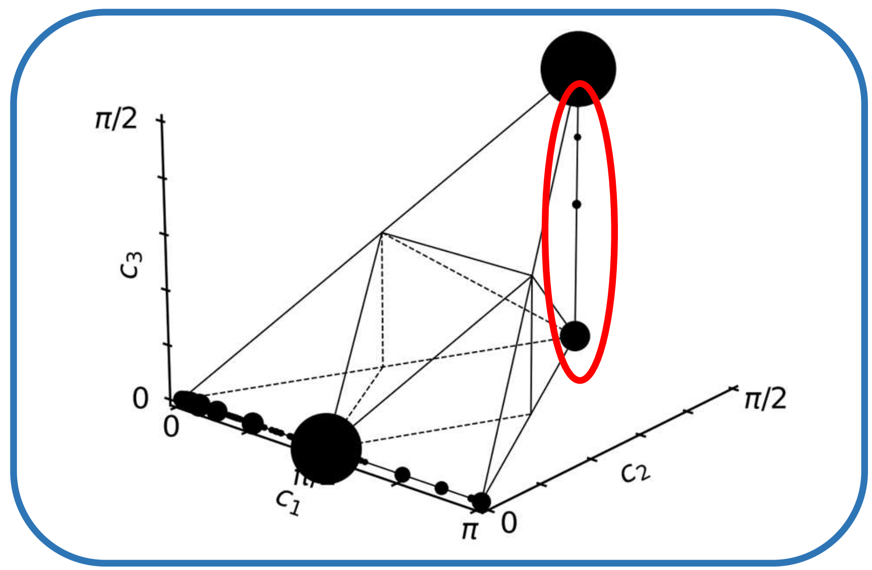




- CPHASE gates mirror to pSWAP gates

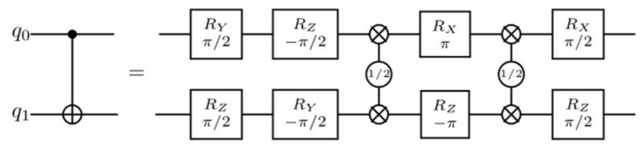
$$(a', b', c') = \begin{cases} (\frac{\pi}{4} + c, \frac{\pi}{4} - b, \frac{\pi}{4} - a) & \text{if } a \leq \frac{\pi}{4} \\ (\frac{\pi}{4} - c, \frac{\pi}{4} - b, a - \frac{\pi}{4}) & \text{else} \end{cases}$$

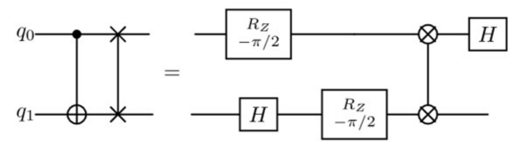




Peterson, et al. **Quantum 6** (2022)

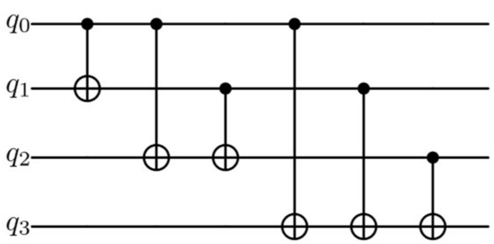
Using mirrors for data movement



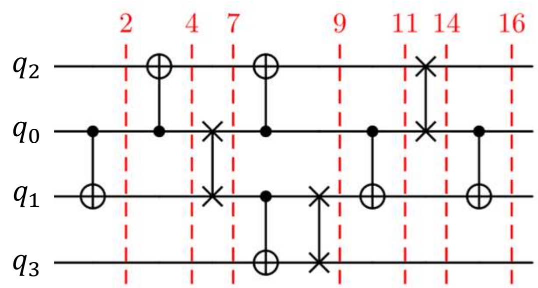


- Intuition: For every CX, decide whether output qubit ordering is (q0, q1) or (q1, q0) based on whether it makes the qubits closer to their next qubit pair

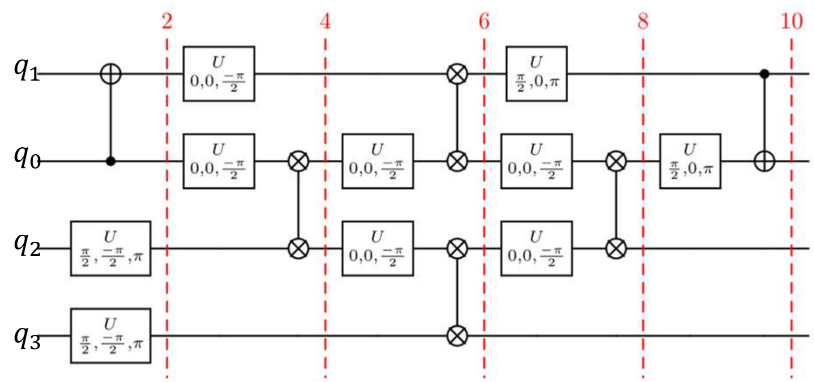
Goal: Full entanglement on a line topology



Qiskit

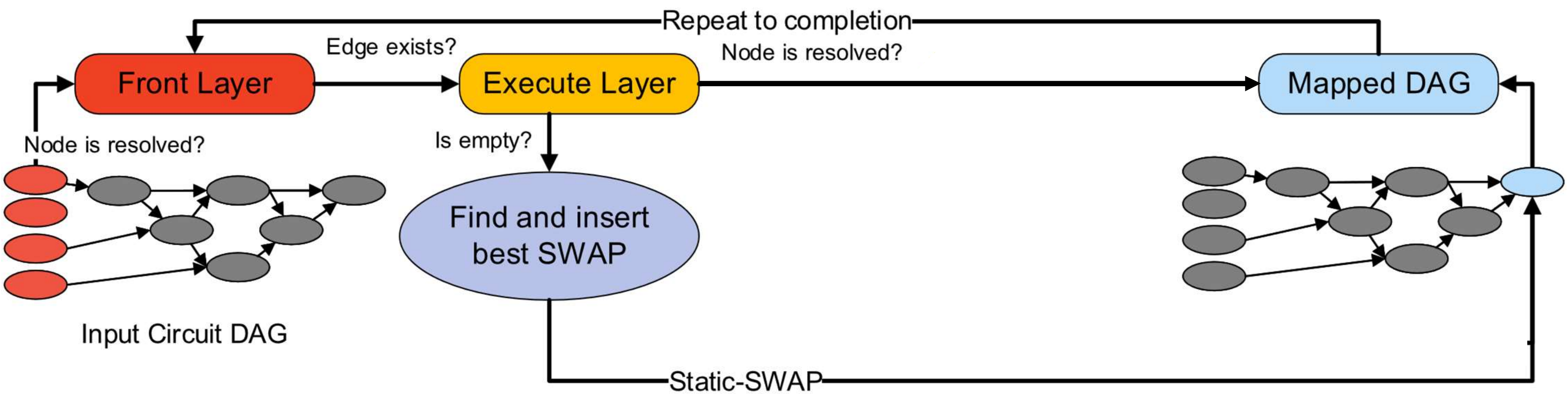


MIRAGE



McKinney, et al. **HPCA** (2024)

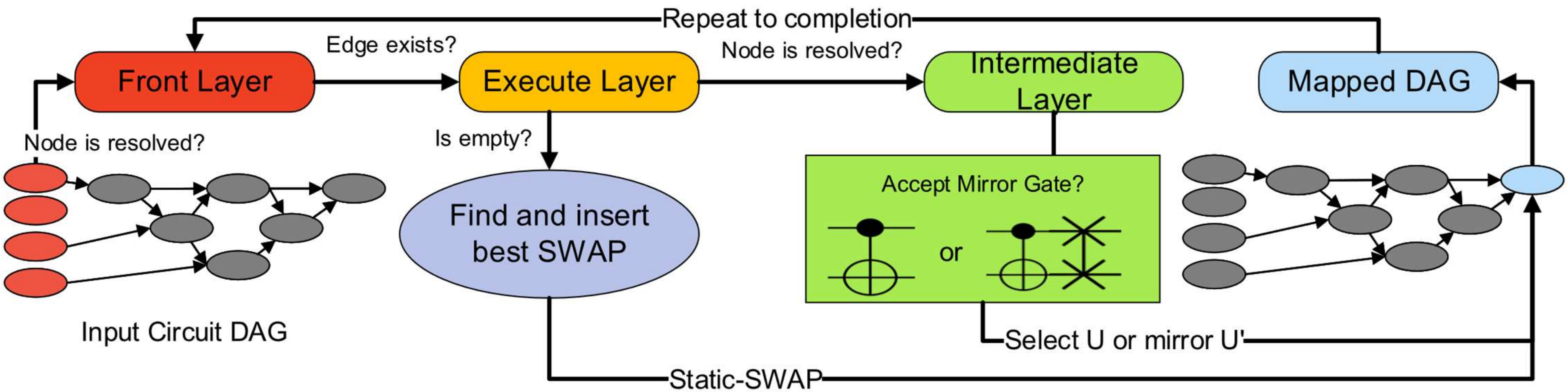
Mirage flow



- Simple yet powerful modification to SABRE:
  - Each gate must pass through an Intermediate Layer
  - Considers if substituting the *mirror* would reduce topological distance cost

Li, et al. **ASPLOS** (2019)

Mirage flow

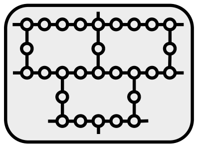


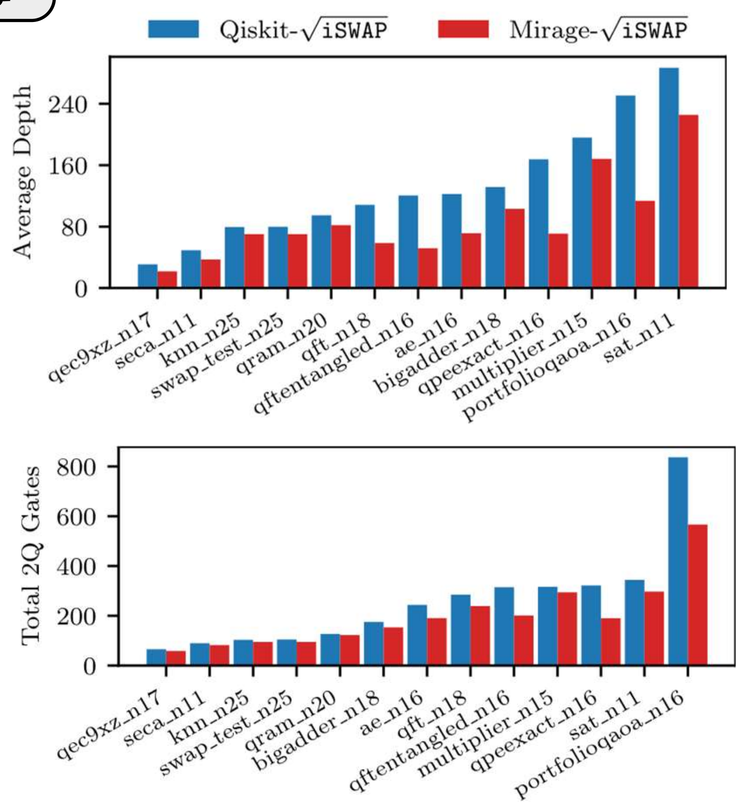
- Simple yet powerful modification to SABRE:
  - Each gate must pass through an Intermediate Layer
  - Considers if substituting the *mirror* would reduce topological distance cost

Li, et al. **ASPLOS** (2019)

Circuit depth reduction

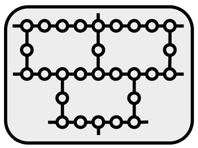
- For the Heavy-Hex topology
  - Average depth decrease of **31.19%**
  - Average total gate decrease of **16.97%**

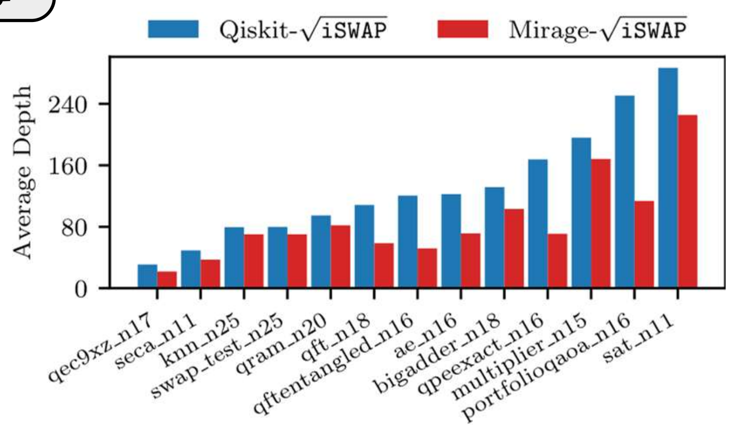


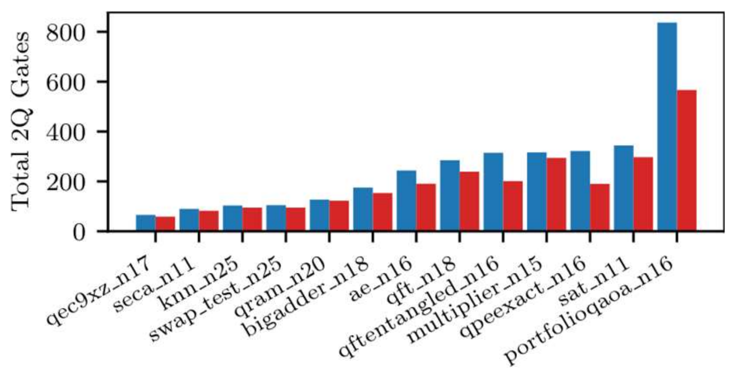


Circuit depth reduction

- For the Heavy-Hex topology
  - Average depth decrease of **31.19%**
  - Average total gate decrease of **16.97%**







- Qiskit Transpiler Plugin

Usage

```
from qiskit.transpiler import CouplingMap
coupling_map = CouplingMap.from_grid(6, 6)
```

1. Use as a Qiskit-Plugin

Integrate MIRAGE into your existing transpilation pipeline:

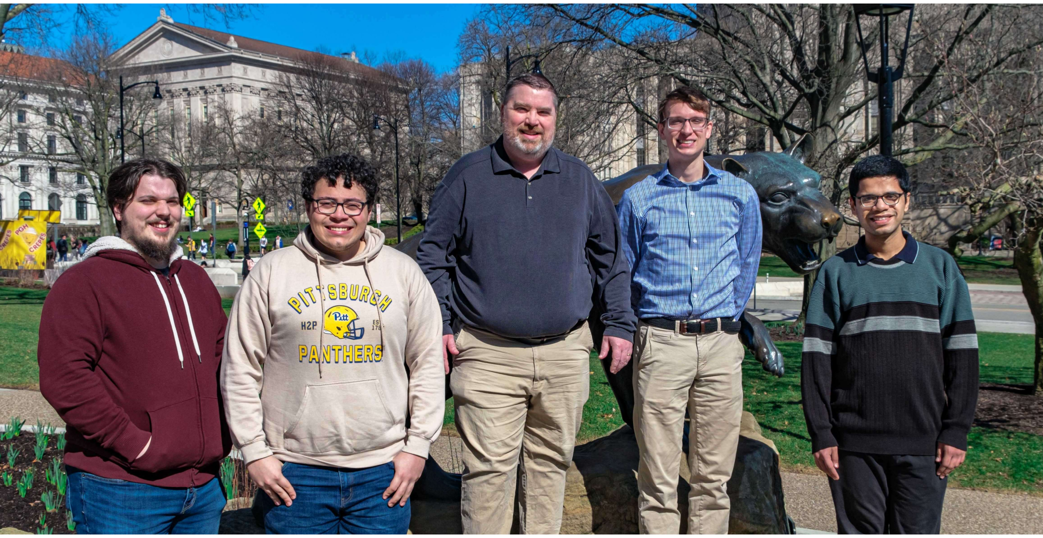
```
from qikist import transpile
mirage_qc = transpile(
            qc, # input circuit
            optimization_level = 3, # default: Qiskit's highest level
            coupling_map=coupling_map,
            basis_gates= ["u", "xx_plus_yy", "id"],
            routing_method="mirage",
            layout_method="sabre_layout_v2",
)
```

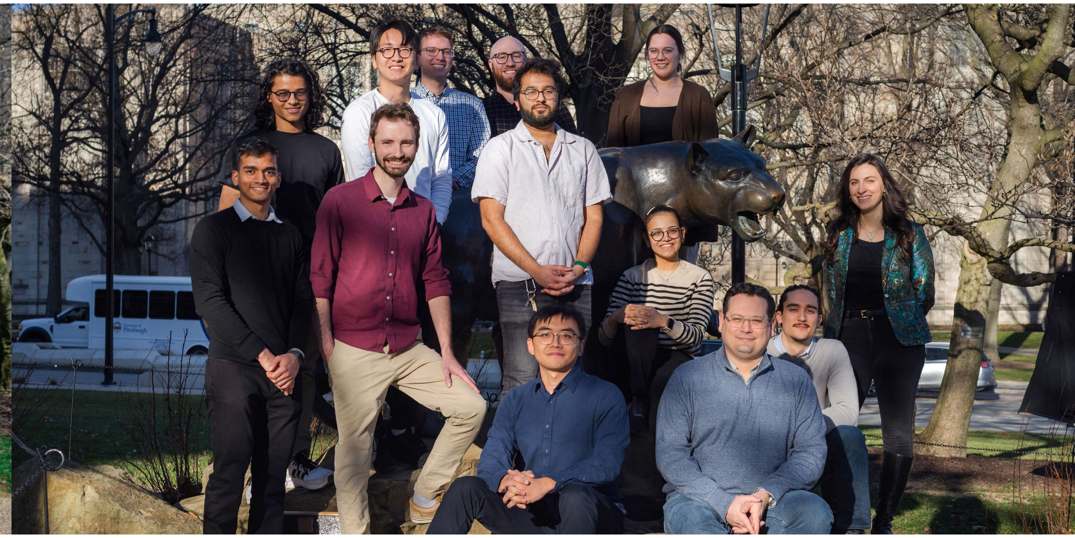
- Software optimizations:
  - Depth post-selection criteria
  - Variable mirror acceptance thresholds
  - Fast block consolidate w/ coord caching

https://github.com/Pitt-JonesLab/mirror-gates

Conclusions

- Evaluate $\sqrt{\text{iSWAP}}$ as a choice basis gate for optimized quantum ISAs
- Qubit frequency allocation over hardware-aware gate errors for SNAIL Corral design
- Significant circuit optimization with MIRAGE, **reducing depth by ~30%**





evm9.dev