U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

INTERNATIONAL IBERIAN
NANOTECHNOLOGY
LABORATORY

Circuit knitting
a state-of-the-art review

Group: Quantum and Linear Optical Computation (INL)
PhD Project: Optimizing models of hybrid quantum/classical computation

F. C. R. Peres | 9th of November 2022

Supervisor: Professor Ernesto Galvão
Co-supervisor: Professor João Lopes dos Santos

Outline

1. Introduction
2. Circuit Knitting: Gate Cutting
3. Circuit Knitting: Wire Cutting

What is circuit knitting?

Circuit knitting

Circuit cutting[2-8]

- Wire cutting[3,4,7,8]
- Gate cutting[2,5,6]

Entanglement forging[1]

[1] A. Eddins, M. Motta, T. P. Gujarati, S. Bravyi, A. Mezzacapo,C. Hadfield, and S. Sheldon, “Doubling the size of quantum simulators by entanglement forging,” PRX Quantum **3**, 010309 (2022).

GaCut. [2] S. Bravyi, G. Smith, and J. A. Smolin, “Trading classical and quantum computational resources,” Physical Review X **6**, 021043 (2016).

WiCut. [3] T. Peng, A. W. Harrow, M. Ozols, and X. Wu, “Simulating large quantum circuits on a small quantum computer,” Physical Review Letters **125**, 150504 (2020).

WiCut. [4] W. Tang, T. Tomesh, M. Suchara, J. Larson, and M. Martonosi, “Cutqc: using small quantum computers for large quantum circuit evaluations,” in *Proceedings of the 26th ACM International Conference on Architectural Support for ProgrammingLanguages and Operating Systems* (2021) pp. 473–486.

GaCut. [5] K. Mitarai and K. Fujii, “Constructing a virtual two-qubit gate by sampling single-qubit operations,” New Journal of Physics **23**, 023021 (2021).

GaCut.

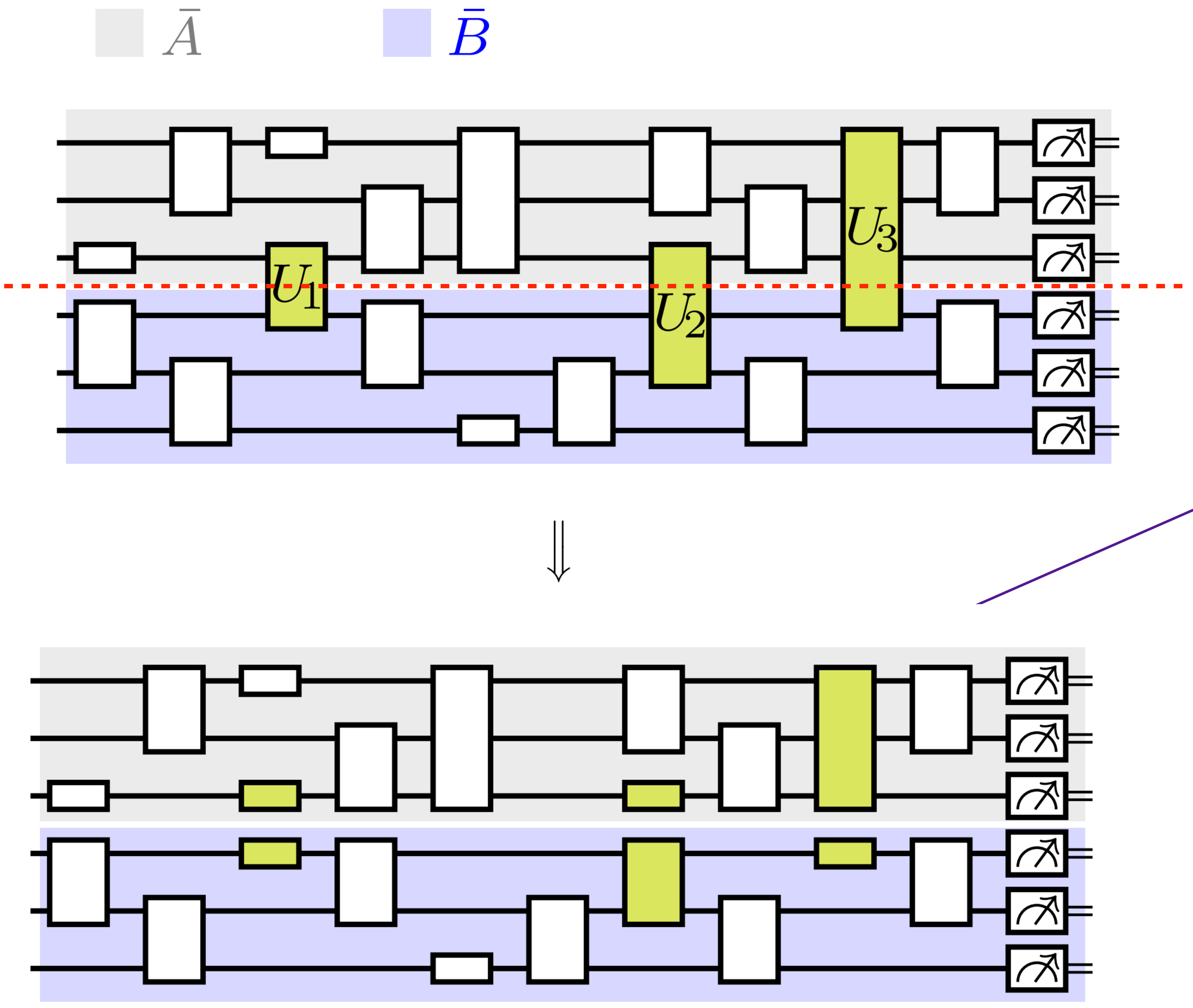
[6] C. Piveteau and D. Sutter, “Circuit knitting with classical communication”, arXiv:2205.00016 (April, 2022)

WiCut.

[7] A. Lowe, M. Medvidovíc, A. Hayes, L. J. O’Riordan, “Fast quantum circuit cutting with randomized measurements”, arXiv:2207.14734v1 (July, 2022)

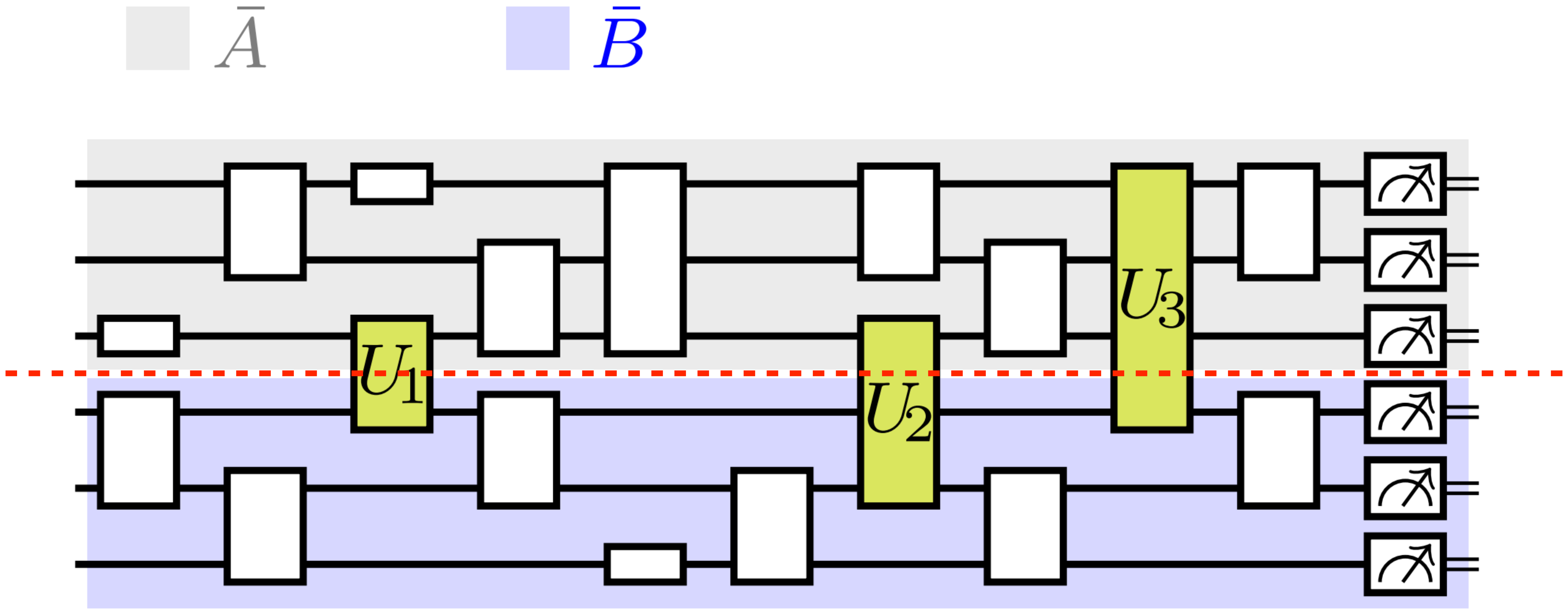
WiCut.

[8] C. Ying *et al* “Experimental Simulation of Larger Quantum Circuits with Fewer Superconducting Qubits”, arXiv:2207.14142v1 (July, 2022)



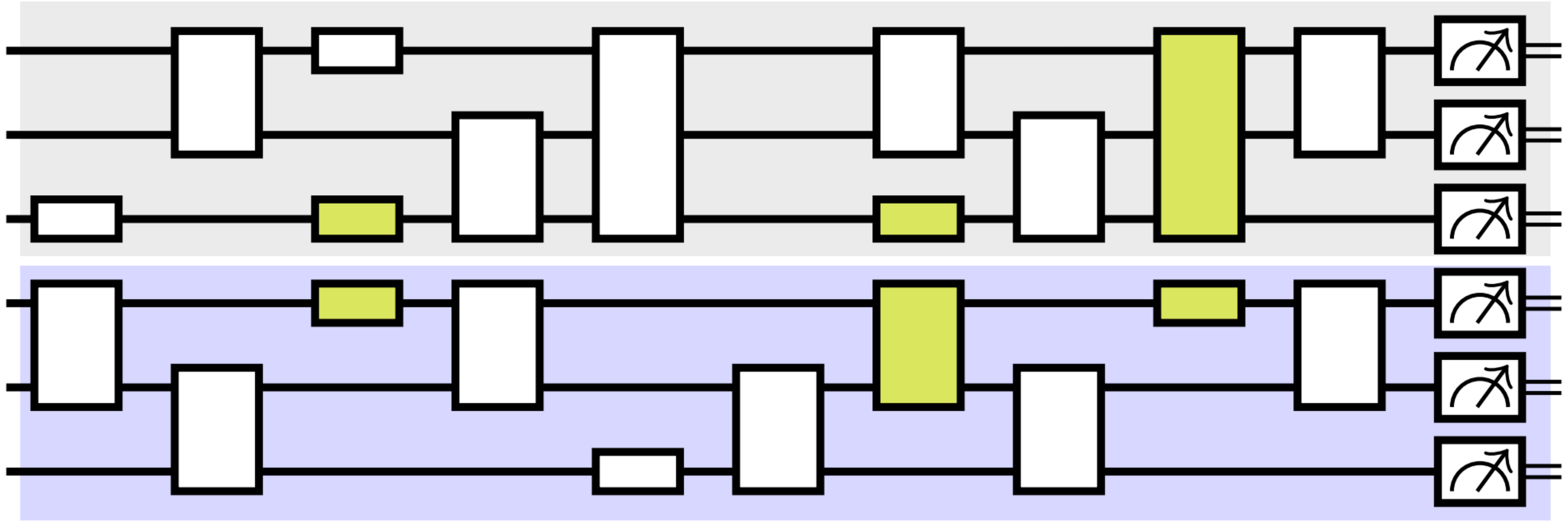
Quasi-probability simulation of non-local gates

The expectation value of the measurement outcomes can be estimated with the two smaller quantum computers.



The cost of this technique is a sampling overhead that scales exponentially in the number of nonlocal gates involved.

$\Downarrow$



Does classical communication help?

3 settings: (1) LO; (2) $\mathrm{LO}\overrightarrow{\mathrm{CC}}$; (3) LOCC.

U: non-local unitary gate

$\mathscr{U}$: the corresponding unitary channel

Quasi-probability decomposition:

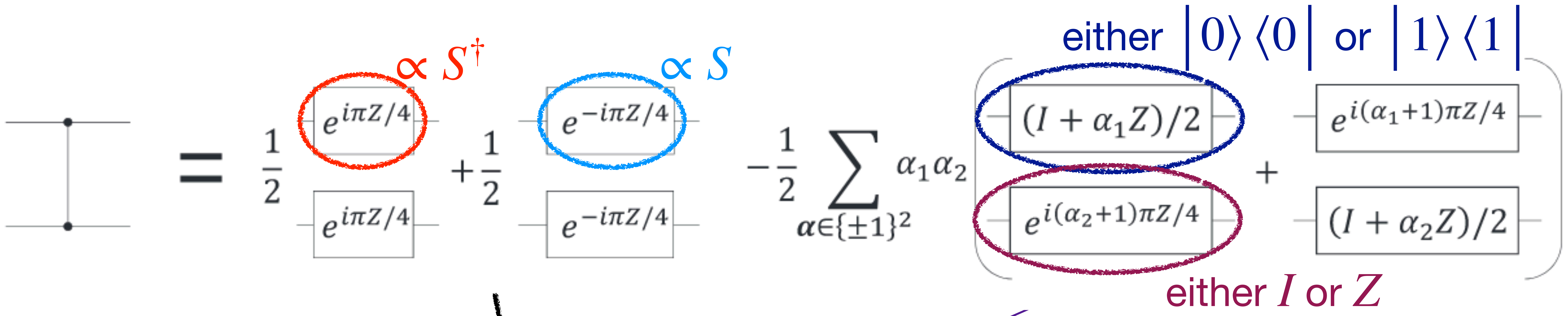
$$\boxed{\mathscr{U} = \sum_i a_i \mathscr{F}_i, \quad \mathscr{F}_i \in \{\mathrm{LO}, \mathrm{LO}\overrightarrow{\mathrm{CC}}, \mathrm{LOCC}\}.}$$

Monte Carlo simulation procedure:

$$\mathcal{U} = \sum_i a_i \mathcal{F}_i$$

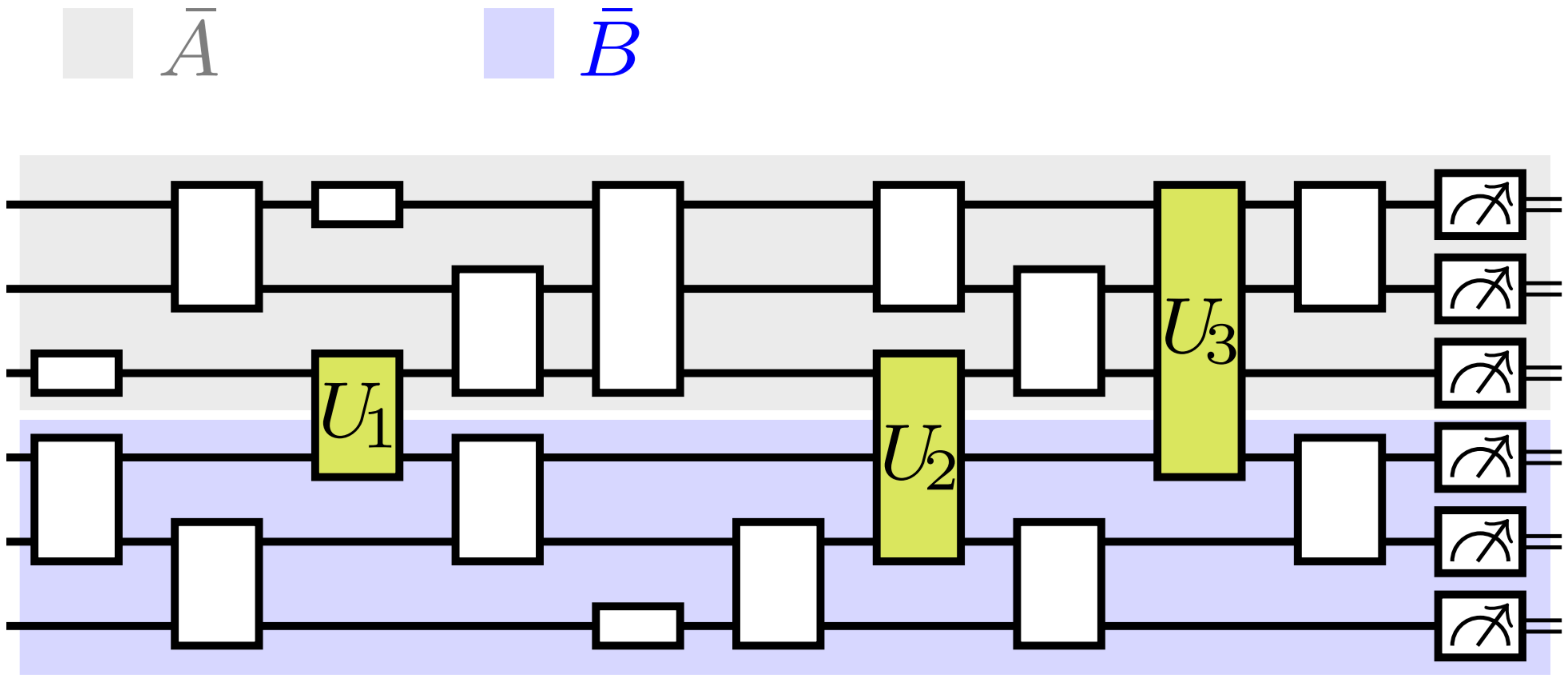
$$p_i = \frac{|a_i|}{\sum_i |a_i|}$$

$\kappa = \sum_i |a_i|$ determines the sampling overhead.

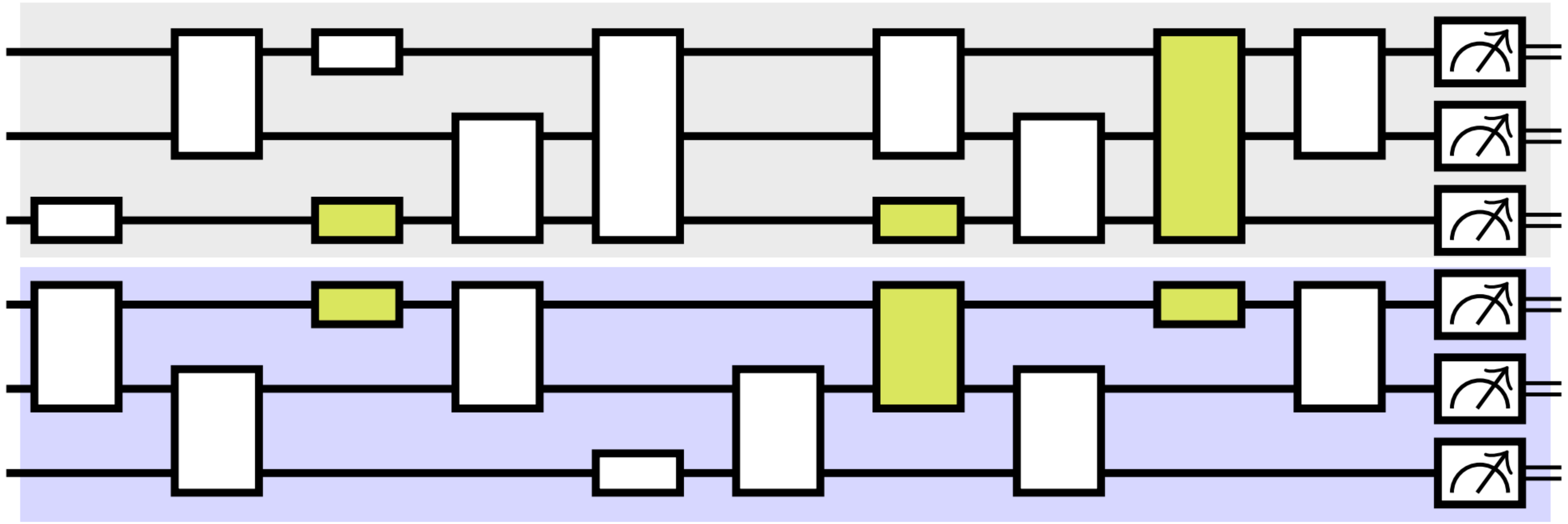


$\kappa = 3 \longrightarrow$ **Optimal decomposition in LO!**

[5] K. Mitarai and K. Fujii, “Constructing a virtual two-qubit gate by sampling single-qubit operations”, New Journal of Physics **23**, 023021 (2021).



$\Downarrow$



Based on the previous decomposition, we have:
$\mathcal{S} = \kappa^2 = 3^{2n} = 9^n$.

The smallest achievable value of κ for the gate U in a setting $S \in \{\mathrm{LO}, \mathrm{LO}\overrightarrow{\mathrm{CC}}, \mathrm{LOCC}\}$ is denoted: $\gamma_S(U)$.

The following is necessarily true:

$$\boxed{\gamma_{LOCC}(U) \leq \gamma_{LO\overrightarrow{CC}}(U) \leq \gamma_{LO}(U)}$$

For a single copy of the gate U, the sampling overhead we incur in is:

$$\mathcal{S} = \gamma_S(U)^2 \,.$$

For a single copy of the CNOT gate, it can be shown that:

$$\gamma_{\mathrm{LOCC}}(\mathrm{CNOT}) = \gamma_{\mathrm{LO}\overrightarrow{\mathrm{CC}}}(\mathrm{CNOT}) = \gamma_{\mathrm{LO}}(\mathrm{CNOT}) = 3.$$

Sampling overhead: $\mathcal{S} = 9$, if we cut only a single CNOT gate.

Suppose we cut n copies of the CNOT gate. Then, the sampling overhead for the LO scenario is:

$$\mathcal{S}_{\mathrm{LO}} = 9^n \, .$$

However, for the other two settings better decompositions can be found so that:

$$\mathcal{S}_{\mathrm{LO}\,\overrightarrow{\mathrm{CC}}} = \mathcal{O}\left(8^n\right)$$

$$\mathcal{S}_{\mathrm{LOCC}} = \mathcal{O}\left(4^n\right) .$$

Why? How?

For n copies of the entangling gate U we can write:

$$\gamma_{\mathrm{LO}}\left(U^{\otimes n}\right) = \gamma_{\mathrm{LO}}(U)^n = 3^n\,.$$

For the CNOT gate

But, for the two other settings:

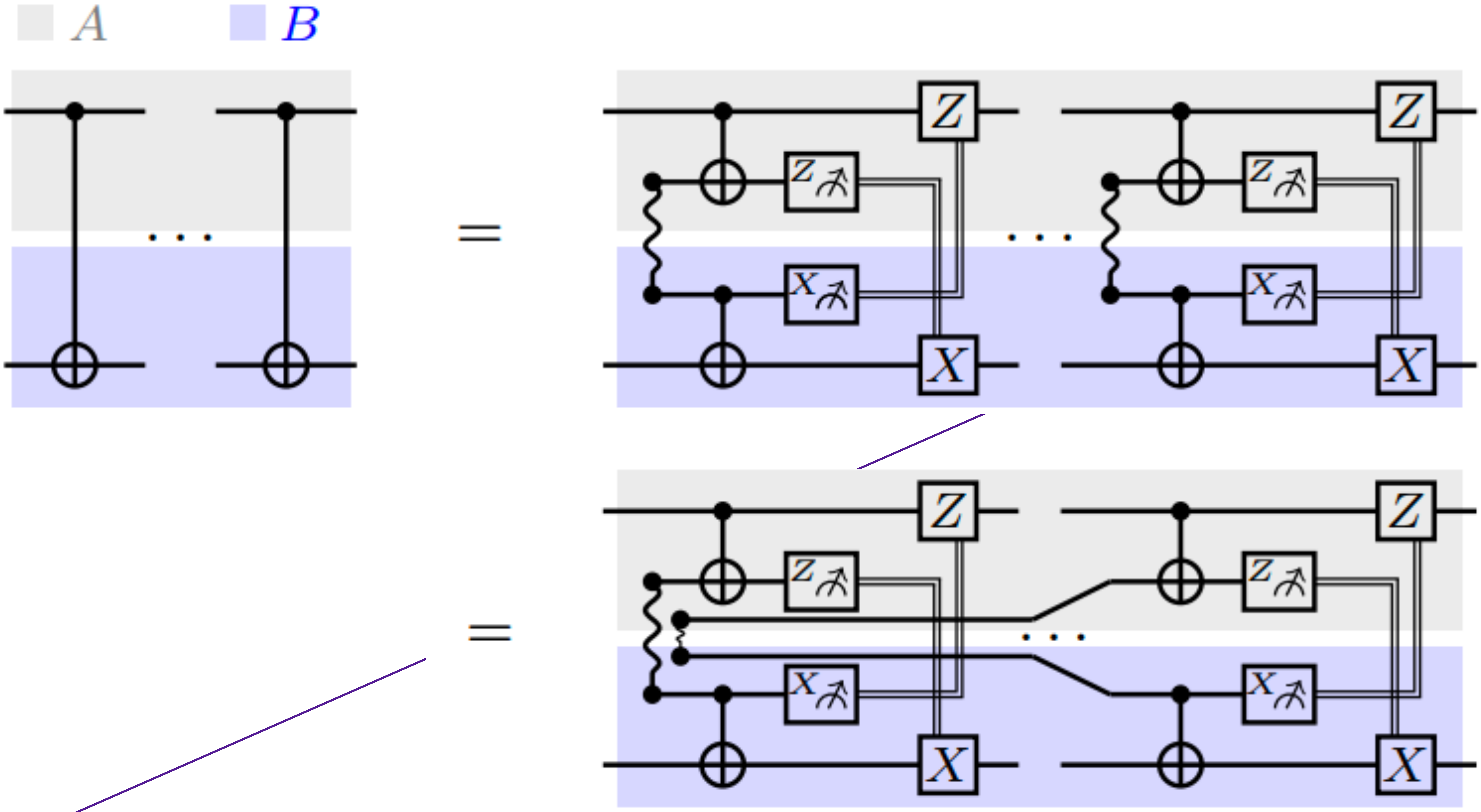
$$\gamma_{S\setminus\{\mathrm{LO}\}}\left(U^{\otimes n}\right) < \gamma_{S\setminus\{\mathrm{LO}\}}(U)^n = \gamma_{\mathrm{LO}}(U)^n\,.$$

The γ-factor is strictly submultiplicative for many unitaries

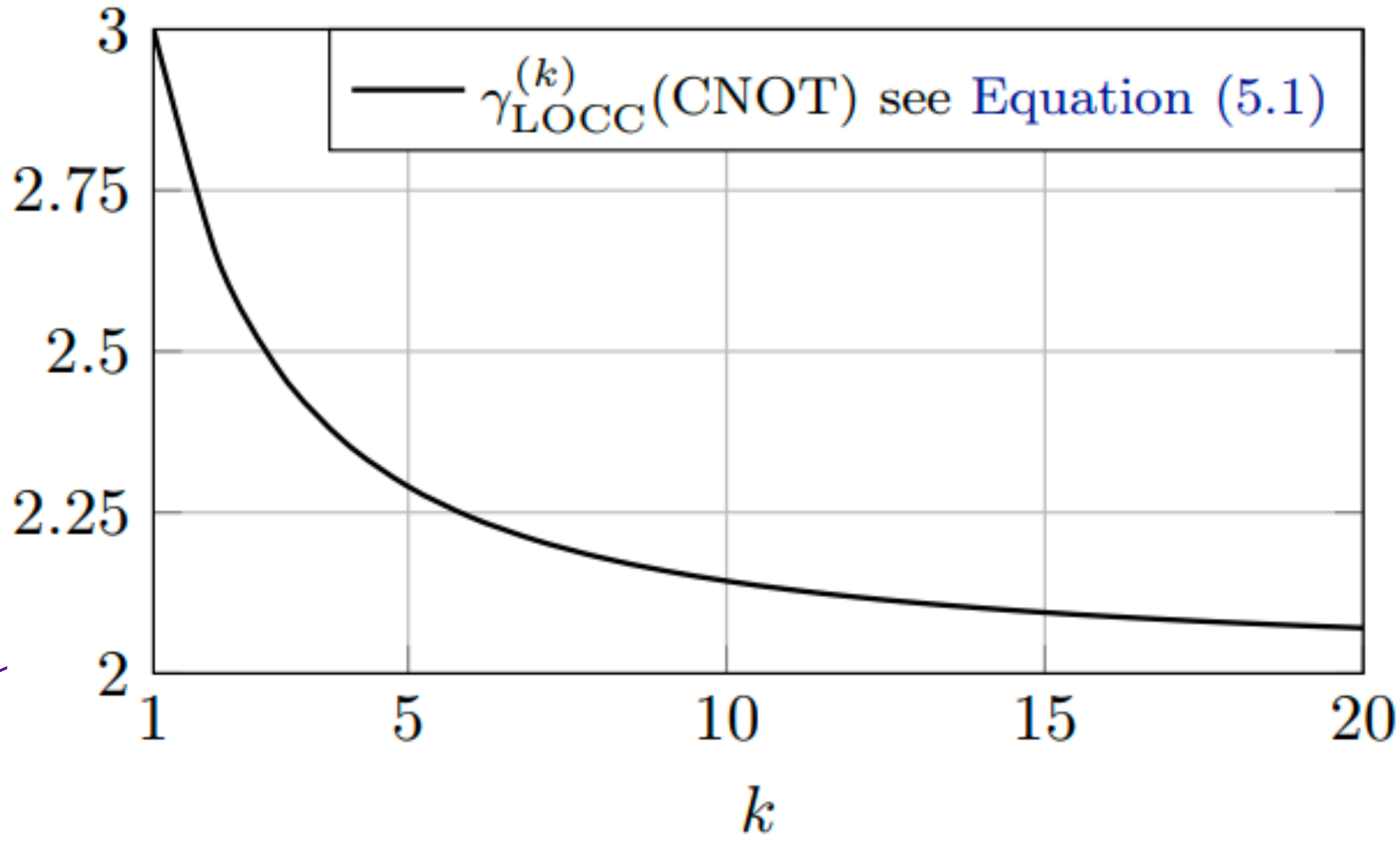
For the CNOT:

$$\gamma_{\mathrm{LOCC}}\left(\mathrm{CNOT}^{\otimes n}\right) = 2^{n+1} - 1 \Rightarrow \mathcal{S}_{\mathrm{LOCC}} = O(4^n)\,.$$

"The main ingredient that enables us to use the submultiplicativity of the γ-factor under the tensor product is the ***ability to realize a gate with a*** LOCC ***protocol that consumes a preexisting entangled state****."*


A
B
=
Z
X
=

Tradeoff between the entanglement factory size k and the effective sampling overhead for the CNOT gate

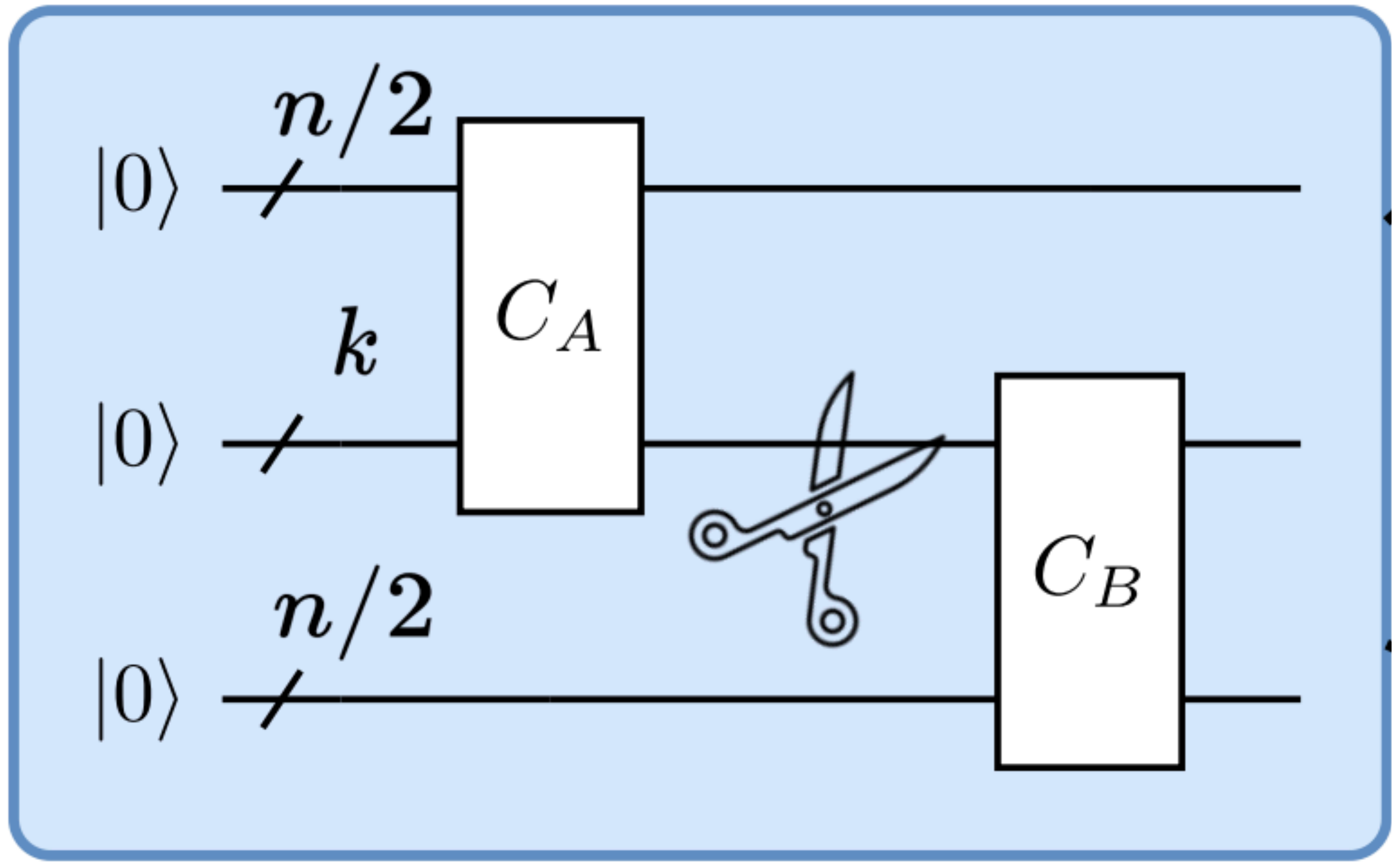


	Clifford gates		non-Clifford gates	
	two-qubit gates	n-qubit gates for $n \geq 3$	gates defined in Theorem 4.3 e.g. $\mathrm{CR}_Z(\theta)$ & $\mathrm{R}_{ZZ}(\theta)$	others
CC helps for **single** instances	✗ Corollary 4.4		✗ Theorem 4.3 & Corollary 4.4	
CC helps for **multiple** instances	✓ Theorem 5.1	✓ Theorem 5.1	(✓) Section 5.2 for $\mathrm{CR}_Z(\theta)$ gate	

Open questions:

1. Identifying useful applications where this framework fits well.

2. Generalizing the presented technique to non-Clifford gates is more complicated and not well understood.



- Randomly inserting measure-and-prepare channels to express the output state of a large circuit as a separable state across distinct devices.

- Sample overhead: $\tilde{O}\left(4^k/\epsilon^2\right)$.

	Type	Overhead	Communication
Thm. 1 in [10]	Gate cutting	$2^{O(nD)}$	None
Thm. 1 in [12]	Wire cutting	16^k	None
Thm. 5 in [14]	Gate cutting	$\min\{9^{kD}, 16^k\}$	None
Eq. (5.1) in [15]	Gate cutting	4^{kD}	LOCC (shared entanglement)
Thm. II.2	Wire cutting	4^k	Classical communication

"(…) *our method likely outperforms all other proposed circuit cutting methods for simulating quantum computation* (…)"

- Similarly, to what we have seen in the previous work, the authors exploit a decomposition:

$$\mathrm{id} = \sum_i a_i \Phi_i \,.$$

- The ℓ_1-norm of this decomposition scales with the dimension of the subspace upon which the channel acts.

- This "*results in our method outperforming the state-of-the-art for a natural problem*".

- The scheme considers two different measure-and-prepare channels.

$$\Psi_0(X) = \mathbb{E}_U \left[\sum_{j=1}^{d} \langle j | U^\dagger X U | j \rangle U | j \rangle \langle j | U^\dagger \right];$$

Basis:
$\{ | j \rangle \}_{j=1}^{d}$

Random unitary operator
which forms a unitary 2-design

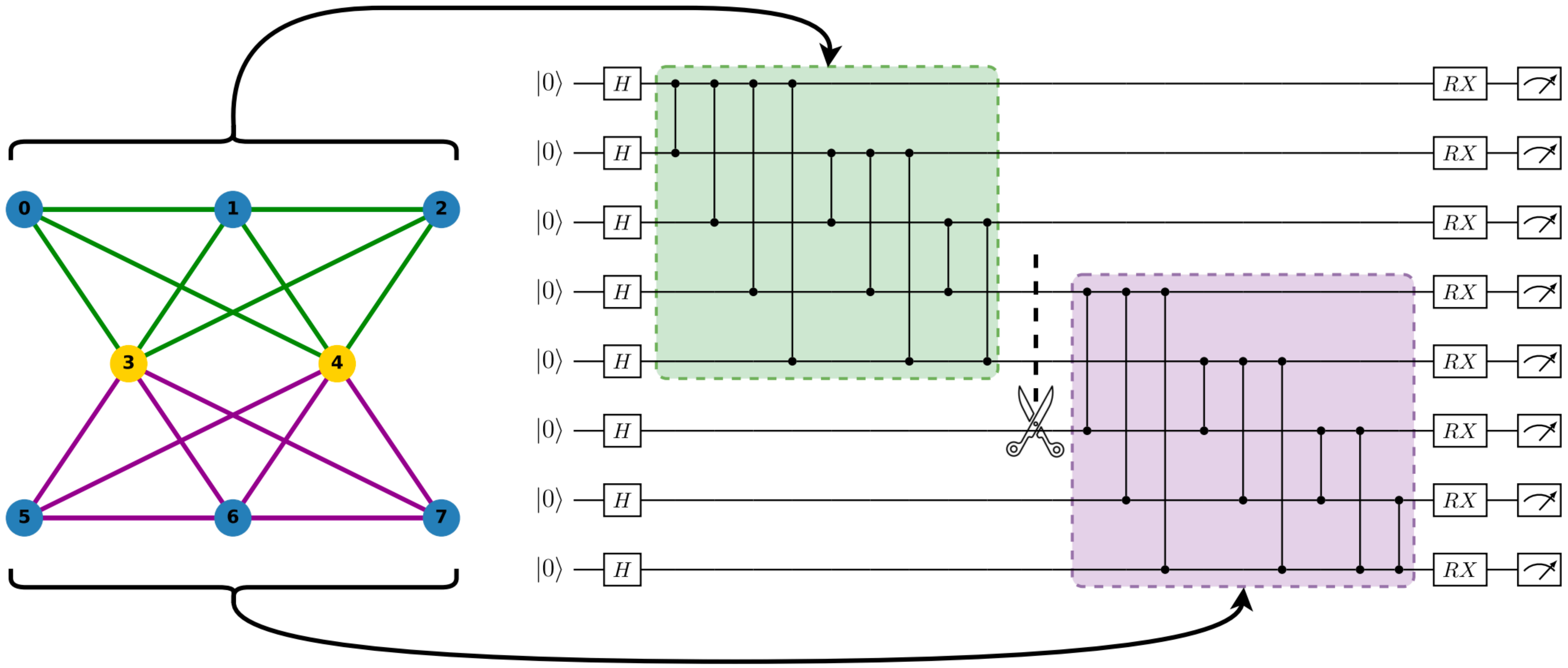
Random POVM
$\{ U | j \rangle \langle j | U^\dagger \}_{j=1}^{d}$

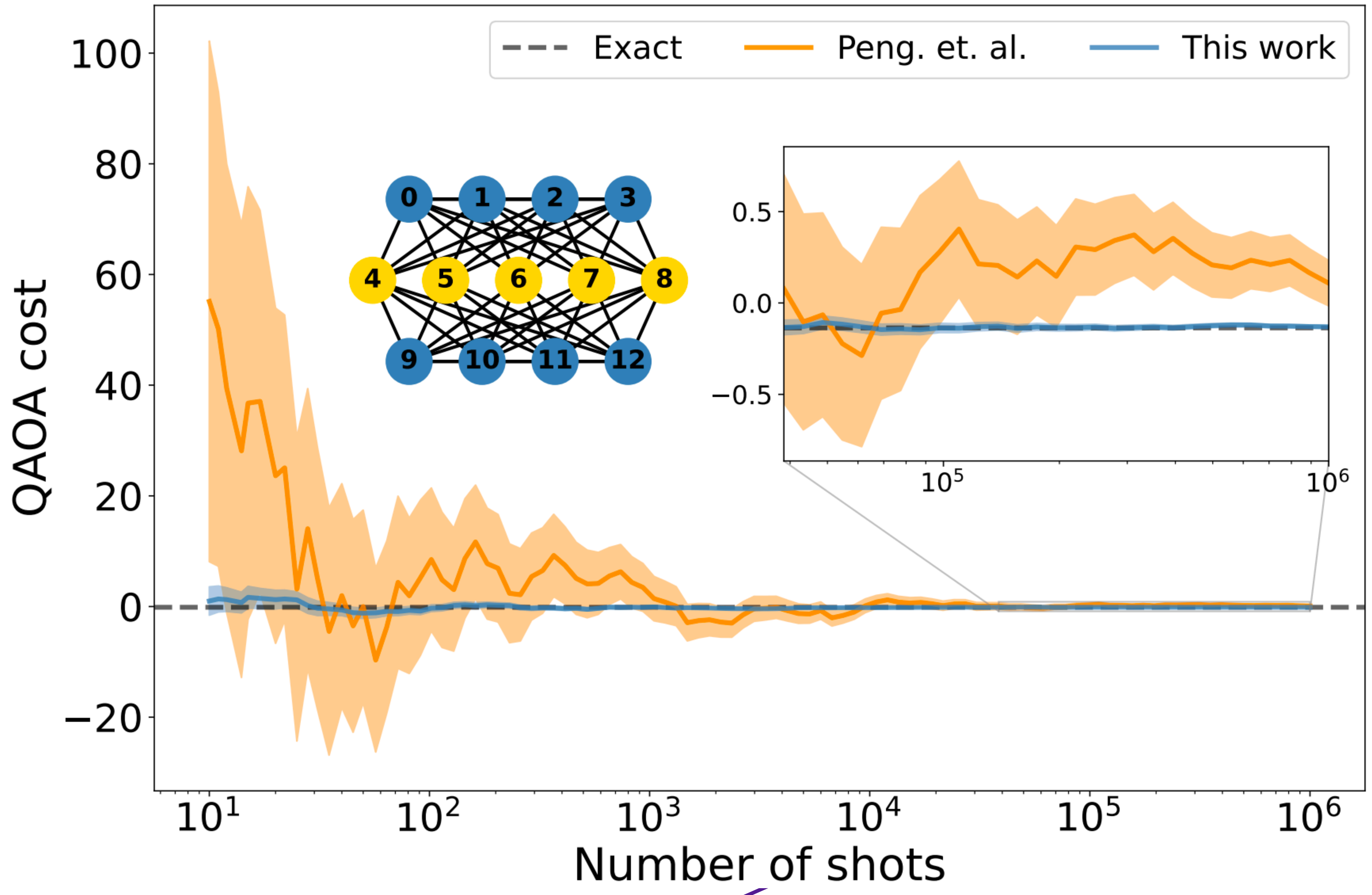
$$\Psi_1(X) = \mathrm{Tr}\,(X)\,\frac{\mathbf{1}}{d}\,.$$ (completely depolarizing channel)

Lemma II.1. *Let d be a positive integer and $\Psi_0\,,\Psi_1$ be the channels defined above. Define the Bernoulli random variable $z \in \{0,1\}$ to be equal to 1 with probability $d/(2d+1)$. It holds that:*

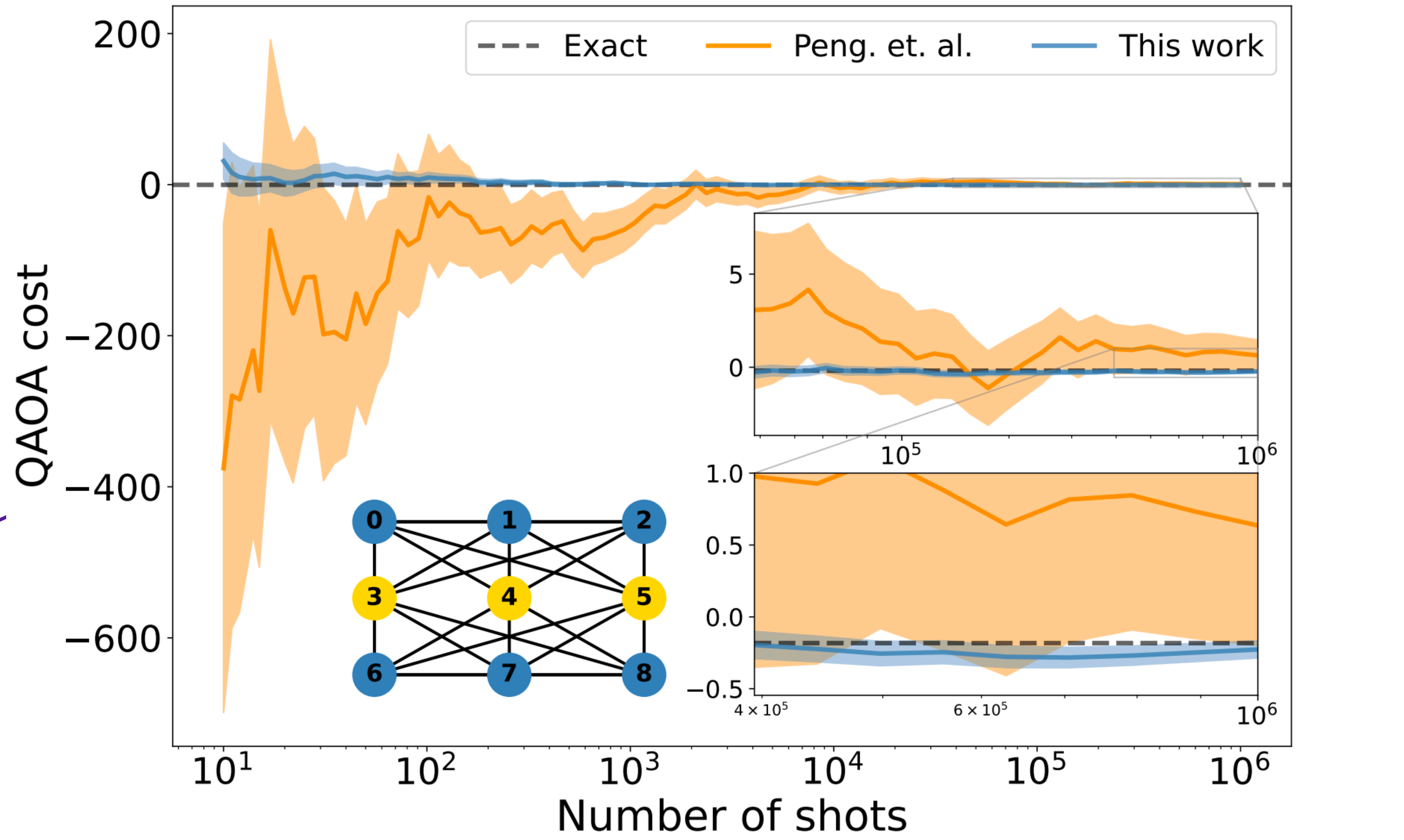
$$\mathrm{id} = (2d+1)\mathbb{E}_z\left[(-1)^z\Psi_z\right].$$

|0⟩
|0⟩
|0⟩
n/2
k
n/2
C_A
C_B
z = 0
z = 1
U†
y
|y⟩
U
x
|y'⟩


0
1
2
3
4
5
6
7
|0⟩
H
RX



Exact
Peng. et. al.
This work
QAOA cost
Number of shots
0 1 2 3 4 5 6 7 8 9 10 11 12



Exact
Peng. et. al.
This work
QAOA cost
Number of shots
0 1 2 3 4 5 6 7 8

Important nuances and final comments:

1. This method requires classical communication between circuit fragments to coordinate measurement outcomes and state preparation.

2. This paper gives an algorithm based on randomized measure-and-prepare channels as well as a way to **sample the cut circuit**.

3. Difficulties may arise when employing quantum hardware instead of simulators. However, these numerical experiments are a testament to the practicality of large-scale circuit cutting workflows.

Thank you for your attention!

$$
\begin{aligned}
\left|\Phi^+\right\rangle\left\langle\Phi^+\right| = {} & \frac{1}{2}\left|0\right\rangle\left\langle 0\right| \otimes \left|0\right\rangle\left\langle 0\right| + \frac{1}{2}\left|1\right\rangle\left\langle 1\right| \otimes \left|1\right\rangle\left\langle 1\right| + \\
& + \frac{1}{4}\left| + \right\rangle\left\langle + \right| \otimes \left| + \right\rangle\left\langle + \right| + \frac{1}{4}\left| - \right\rangle\left\langle - \right| \otimes \left| - \right\rangle\left\langle - \right| + \\
& + \frac{1}{4}\left|+i\right\rangle\left\langle +i\right| \otimes \left|-i\right\rangle\left\langle -i\right| + \frac{1}{4}\left|-i\right\rangle\left\langle -i\right| \otimes \left|+i\right\rangle\left\langle +i\right| - \\
& - \frac{1}{4}\left|+i\right\rangle\left\langle +i\right| \otimes \left|+i\right\rangle\left\langle +i\right| - \frac{1}{4}\left|-i\right\rangle\left\langle -i\right| \otimes \left|-i\right\rangle\left\langle -i\right| - \\
& - \frac{1}{4}\left| + \right\rangle\left\langle + \right| \otimes \left| - \right\rangle\left\langle - \right| - \frac{1}{4}\left| - \right\rangle\left\langle - \right| \otimes \left| + \right\rangle\left\langle + \right| .
\end{aligned}
$$

$$\gamma\left(\left|\Phi^{+}\right\rangle\left\langle\Phi^{+}\right|\right) = 2 \times \frac{1}{2} + 8 \times \frac{1}{4} = 1 + 2 = 3.$$