Optimal quantum circuit cuts

(with application to clustered Hamiltonian simulation)

Aram W. Harrow & **Angus Lowe**

Center for Theoretical Physics, MIT

(arxiv: 2403.01018)

Overview

I. **Background & summary of results**

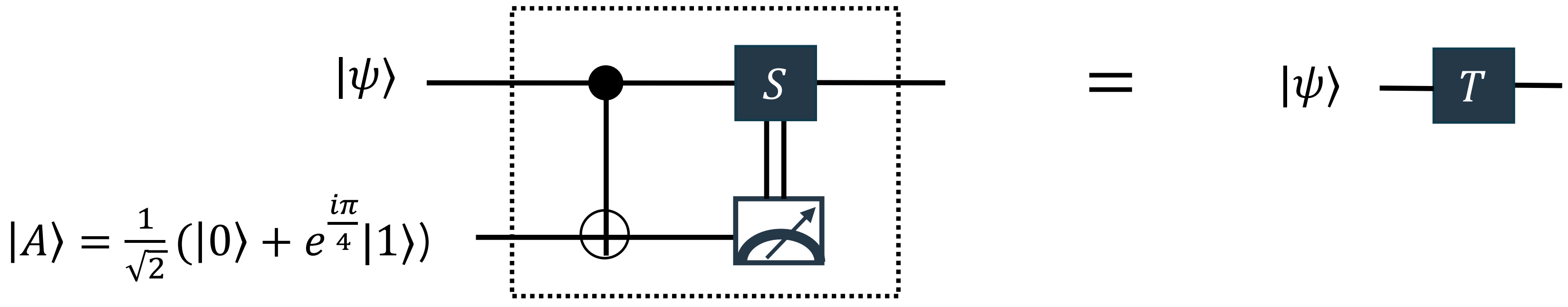
II. Optimal space-like cuts

III. Optimal time-like cuts

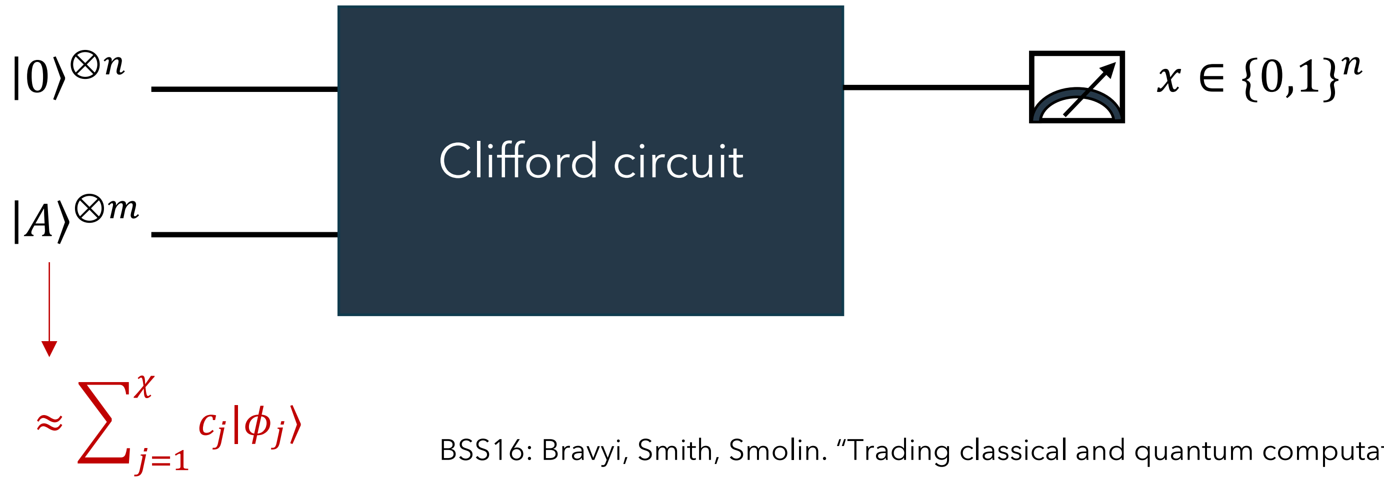
IV. Lower bounds

Quantum information processing consumes quantum **resources**. E.g., **entanglement** & **magic**.

Fault-tolerant quantum computation with **magic** states:



Replace **magic** with random Clifford operations [BSS16, Bra+19]:

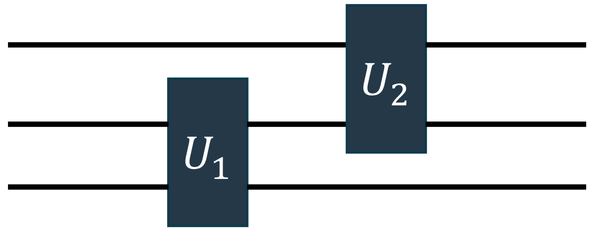


BSS16: Bravyi, Smith, Smolin. "Trading classical and quantum computational resources" PRX 6.2 (2016)

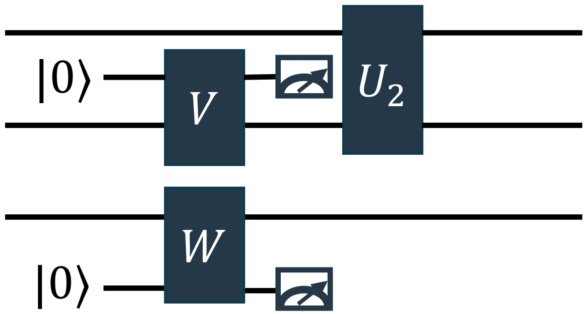
Bra+19: Bravyi et al. "Simulation of quantum circuits by low-rank stabilizer decompositions" Quantum 3 (2019)

Q1: Can we replace **entanglement** with random local operations?

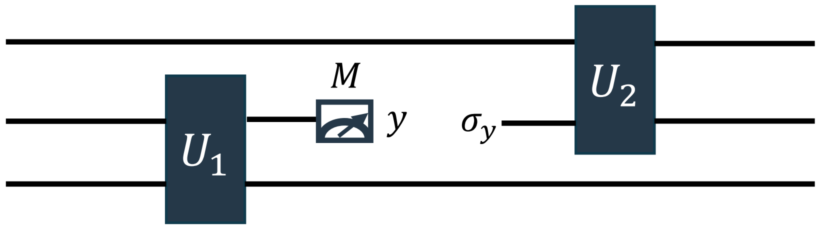
→ If the goal is to compute expectation values, then yes.



(Original circuit)



(Space-like cut)



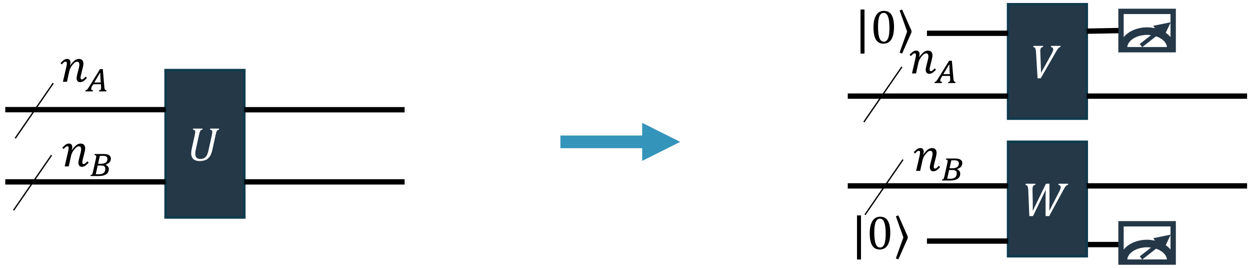
(Time-like cut)

V, W, and M are potentially random.

Q2: At what cost?

Summary of results

For **space-like cuts** the cost is related to the **entangling power**.



Entanglement measures

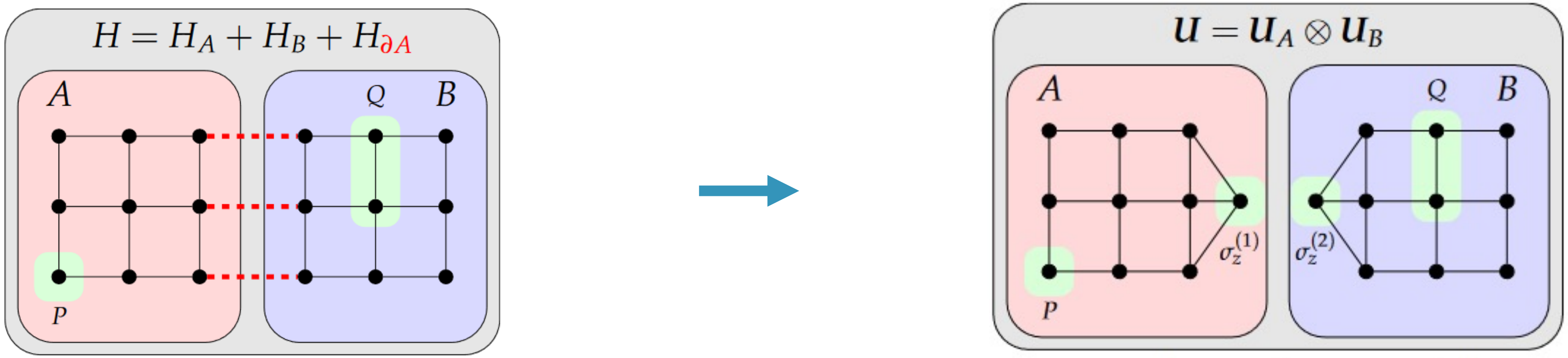
	CC?	Cost	Ancillas	Restrictions
** [BSS16]	No	$2^{O(n_B)}$	1	$O(1)$-sparse
[PS22]	Yes	$1 + 2R(J(\mathcal{U}))$	$n_A + n_B$	Clifford
[This work]	No	$\xi(\mathcal{U})$	2	None

We have $1 + 2R\big(J(\mathcal{U})\big) \leq \xi(\mathcal{U}) \leq 2^{2(n_A+n_B)+1} - 1$. In many cases, $1 + 2R\big(J(\mathcal{U})\big) = \xi(\mathcal{U})$

PS22: Piveteau and Sutter, "Circuit Knitting With Classical Communication," in *IEEE ToIT*, vol. 70, no. 4 (2024)

Summary of results

Space-like cuts can be applied to **remove interactions** in a Hamiltonian simulation.



$\langle \psi_0 | e^{iHt} (P \otimes Q) e^{-iHt} | \psi_0 \rangle$

$\mathbb{E}_{\gamma, U} \, \gamma \langle \psi_0 | U^\dagger \left(P \otimes \sigma_z^{(1)} \otimes Q \otimes \sigma_z^{(2)} \right) U | \psi_0 \rangle$

Space-like cut of the 1st order Trotter formula

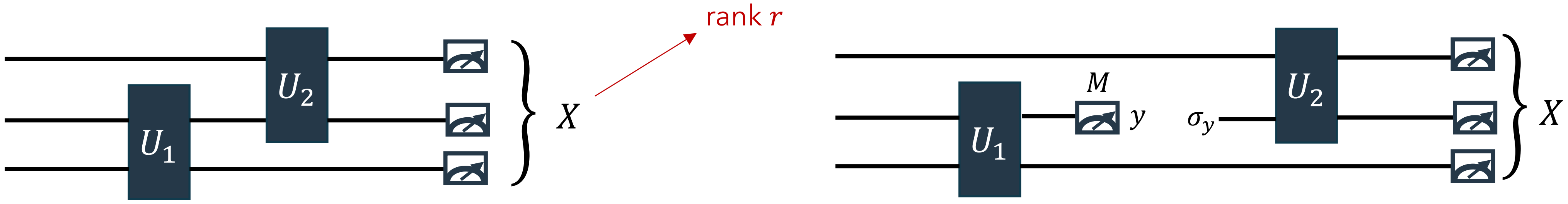
Runtime is on the order of $\exp(8\eta t)$ where $\eta = \sum_{k \in \partial A} \|H_k\|$.

Best prior work achieved $\exp(|\partial A| \eta^{1/p} t^{1+1/p} / \epsilon^{1/p})$ using a p^{th} order formula [Chi+21].

Chi+21: Childs et al. "Theory of Trotter Error with Commutator Scaling". PRX 11.1 (2021)

Summary of results

For **time-like cuts** the cost depends on the **number of wires** replaced.



	CC?	Cost	Ancillas
[Pen+19]	No	16^k	None
[BPS23]	Yes	4^k	2k
[This work]	Yes	$2^k r$	None

Lower bound also obtained: $\Theta(2^k)$ samples are necessary and sufficient to estimate output probabilities.

Pen+19: Peng et al. “Simulating Large Quantum Circuits on a Small Quantum Computer”. PRL 125 (2020)

BPS23: Brenner, Piveteau, and Sutter. Optimal wire cutting with classical communication. arxiv:1506.01396

Overview

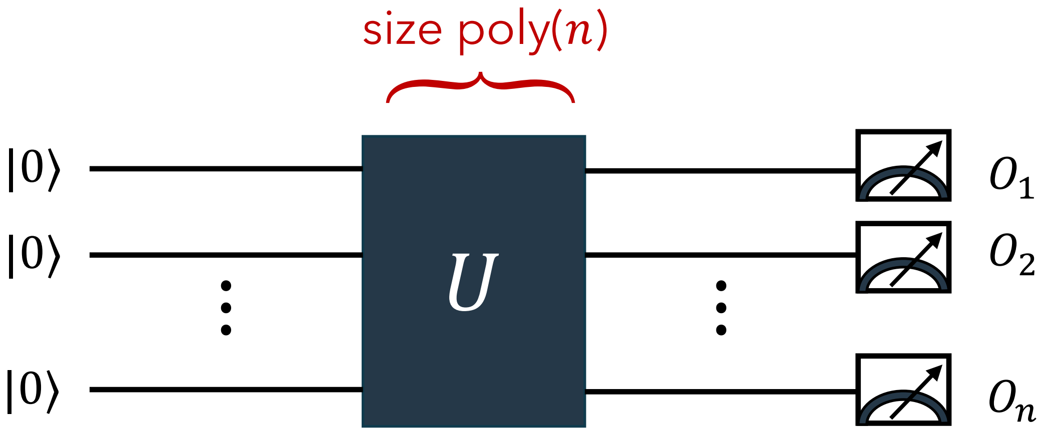
I. Background & summary of results

II. Optimal space-like cuts

III. Optimal time-like cuts

IV. Lower bounds

Computational model



Goal: Compute a good estimate of $\langle O_1 \otimes O_2 \otimes \cdots \otimes O_n \rangle = \langle 0 | U^\dagger (O_1 \otimes O_2 \otimes \cdots \otimes O_n) U | 0 \rangle$.

Standard solution: Run the circuit M times and compute the empirical mean,

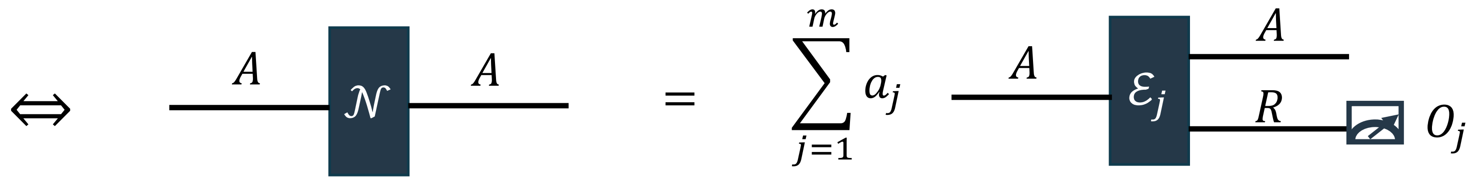
$$\mu = \frac{1}{M}(\mu_1 + \mu_2 + \cdots + \mu_M).$$

$Var[\mu] \sim \frac{1}{M}$

Quasiprobability-based simulation

A **quasiprobability decomposition (QPD)** of a channel $\mathcal{N}_{A\to A}$ has the form

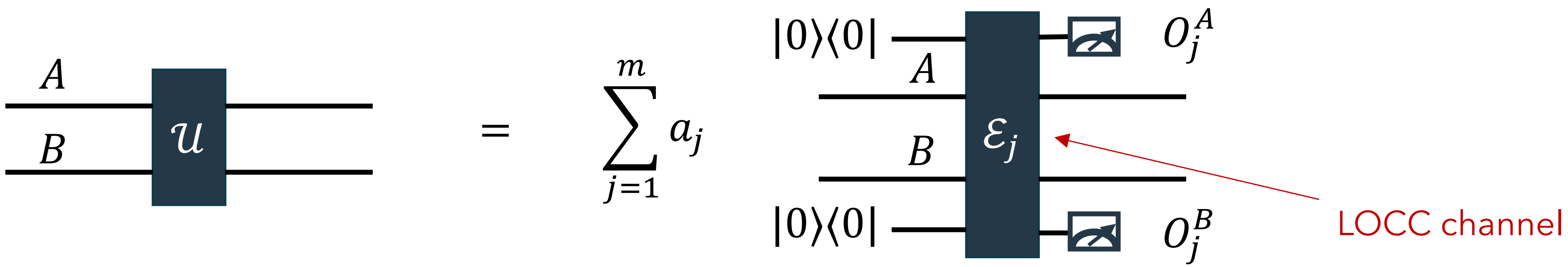
$$\mathcal{N} = \sum_{j=1}^{m} a_j \mathcal{T}_j \circ \mathcal{E}_j = \|a\|_1 \sum_{j=1}^{m} \frac{|a_j|}{\|a\|_1} sign(a_j) \mathcal{T}_j \circ \mathcal{E}_j$$



The 1-norm of the QPD is defined as $\|a\|_1 = \sum_{j=1}^{m} |a_j|$.

Fact: Let X be a Hermitian observable on A. Measuring the observable $\|a\|_1 sign(a_j)(O_j \otimes X)$ on the ensemble of states $\left\{\left(|a_j|/\|a\|_1, \mathcal{E}_j(\rho)\right)\right\}_j$ yields an unbiased estimator of $Tr(X\mathcal{N}(\rho))$ for any ρ.

A **space-like cut is** a QPD of the form



$$\Leftrightarrow \qquad \mathcal{U}_{AB\to AB} = \sum_{j=1}^{m} a_j \left(id_{AB\to AB} \otimes \mathcal{T}_j\right) \circ \mathcal{E}_j$$

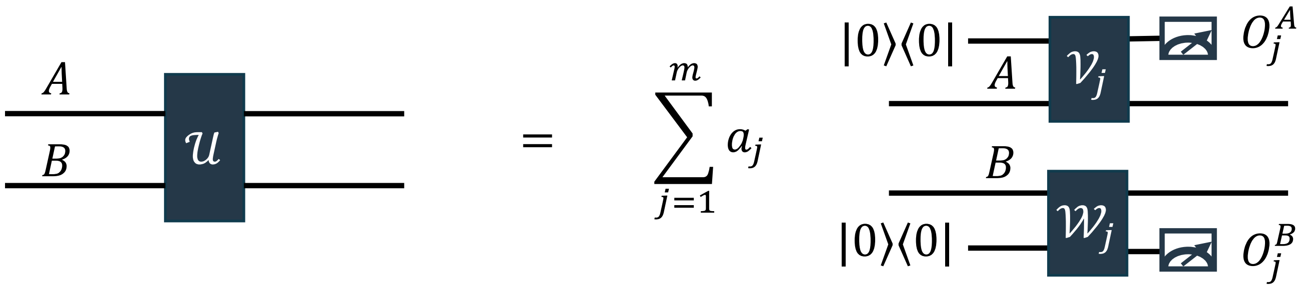
$$\mathcal{T}_j(X) := Tr\left(\left(O_j{}^A \otimes O_j^B\right) X\right)$$

Theorem [PS22]: The minimal 1-norm of a space-like cut of $\mathcal{U}_{AB}: \rho \mapsto U\rho U^\dagger$ is at least $1 + 2R(J(\mathcal{U}))$.

If in addition $\mathcal{E}_j = \mathcal{V}_j \otimes \mathcal{W}_j$ for every $j \in [m]$ we call this a **local space-like cut**.

PS22: Piveteau and Sutter, "Circuit Knitting With Classical Communication," in *IEEE ToIT*, vol. 70, no. 4 (2024)

Theorem [PS'22]: The minimal 1-norm of a space-like cut of $\mathcal{U}_{AB}: \rho \mapsto U\rho U^\dagger$ is at least $1 + 2R(J(\mathcal{U}))$.



$$\mathcal{U} = \sum_{j=1}^{m} a_j \,(\ldots)$$

(Local space-like cut)

Theorem: Let $U = \sum_\alpha c_\alpha V_\alpha \otimes W_\alpha$ be a decomposition of U into local unitaries. The **double Hadamard test** is a local space-like cut of $\mathcal{U}: \rho \mapsto U\rho U^\dagger$ with two ancilla qubits and 1-norm $\|a\|_1 = 2\|c\|_1^2 - \|c\|_2^2$. Furthermore, whenever this decomposition is an **operator Schmidt decomposition**,

$$\|a\|_1 = 2\|c\|_1^2 - 1 = 1 + 2R(J(\mathcal{U})).$$

We define the **product extent** $\xi(U)$ as the minimum of $2\|c\|_1^2 - \|c\|_2^2$ over all $U = \sum_\alpha c_\alpha V_\alpha \otimes W_\alpha$.

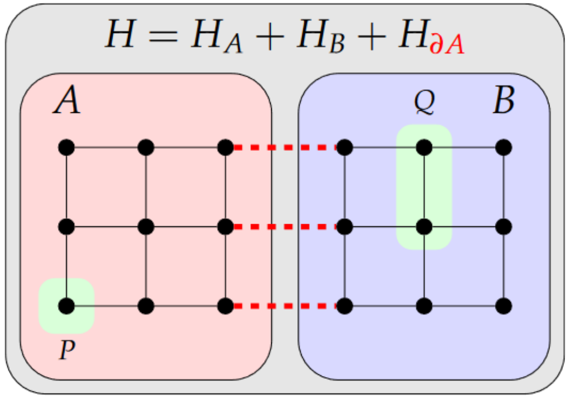
The product extent

We define the **product extent** $\xi(U)$ as the minimum of $2\|c\|_1^2 - \|c\|_2^2$ over all $U = \sum_\alpha c_\alpha V_\alpha \otimes W_\alpha$.

The product extent is an entanglement measure satisfying:

1. Faithfulness: $\xi(U) = 1$ iff U is a product of local unitaries.
2. Local unitary invariance: $\xi((V_A \otimes V_B)U(W_A \otimes W_B)) = \xi(U)$.
3. Submultiplicativity: $\xi(UV) \leq \xi(U)\xi(V)$.

Clustered Hamiltonian simulation



$$\langle \psi_0 | e^{iHt} (P \otimes Q) e^{-iHt} | \psi_0 \rangle$$

Consider the 1st order Trotter formula:

$$U = \left(\prod_{j\in A} e^{-iH_j t/r} \prod_{k\in B} e^{-iH_k t/r} \prod_{\ell\in\partial A} e^{-iH_\ell t/r}\right)^r.$$

$$\xi(U) \le \left(\prod_{j\in A} \xi(e^{-iH_j t/r}) \prod_{k\in B} \xi(e^{-iH_k t/r}) \prod_{\ell\in\partial A} \xi(e^{-iH_\ell t/r})\right)^r$$

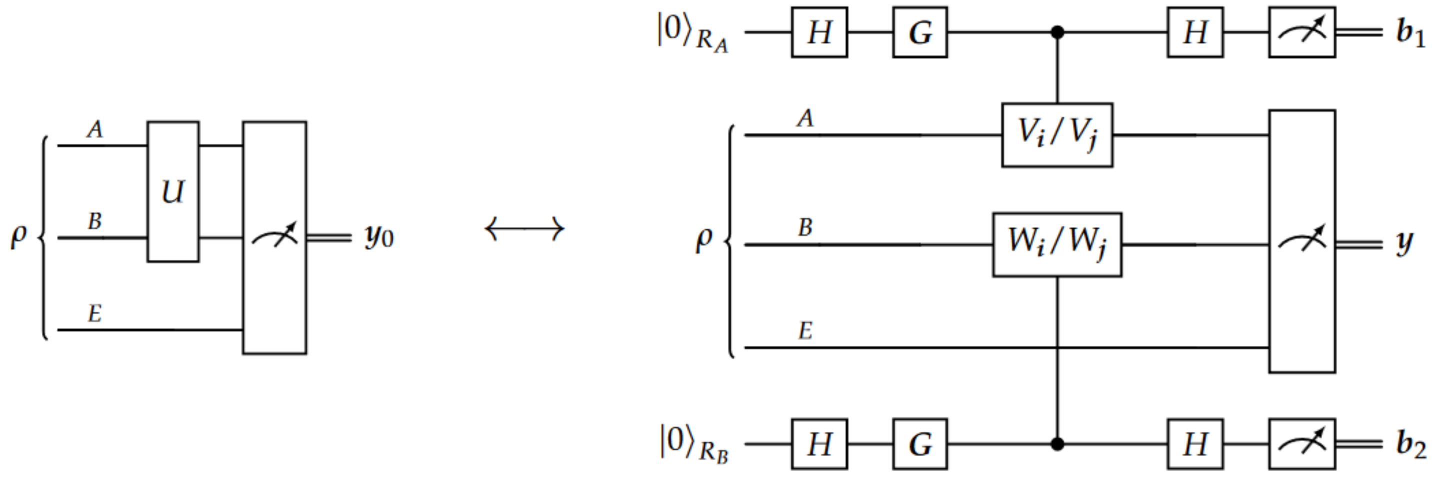
$= 1$

$$= \left(\prod_{\ell\in\partial A} [1 + 2R(J(e^{-iH_\ell t/r}))]\right)^r$$

$$= \left(\prod_{\ell\in\partial A} [1 + |4\cos(\|H_\ell\| t/r)\sin(\|H_\ell\| t/r)|]\right)^r \le \exp\left\{4\left(\sum_{\ell\in\partial A} \|H_\ell\|\right) t\right\}$$

The double Hadamard test

Let $U = \sum_\alpha c_\alpha V_\alpha \otimes W_\alpha$. We can explain the simulation procedure without referring to QPDs.



$i, j, g \sim p(i, j, g)$

Lemma: $\xi(U)(-1)^{b_1+b_2+g}y$ is an unbiased estimator of $\mathbb{E}\, y_0$.

→ If $Var[y_0] = \sigma^2$ then the variance of this estimator is $\xi(U)^2\sigma^2$.

Overview

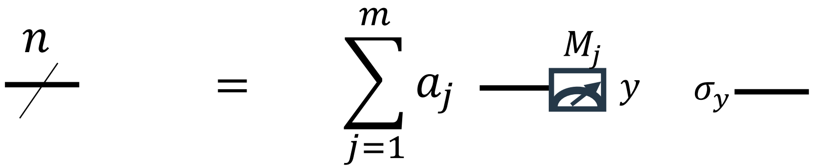
I. Background & summary of results

II. Optimal space-like cuts

III. Optimal time-like cuts

IV. Lower bounds

Time-like cuts



$$\Leftrightarrow \qquad id = \sum_{j=1}^{m} a_j \mathcal{M}_j$$

Measure-and-prepare channels

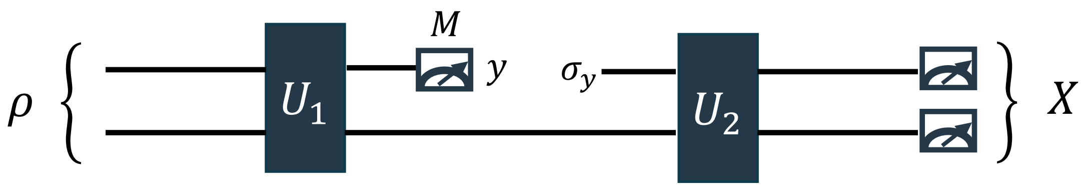
Theorem [Yua+21, This work]: The optimal time-like cut has 1-norm $2^{n+1} - 1$. Moreover, the measure-and-prepare channels can each be implemented using $O(n^2)$ diagonal gates.

$$id = 2^n \mathcal{M}_0 - (2^n - 1)\mathcal{M}_1$$

How well does this work for a concrete estimation task?

Yuan et al. "Universal and operational benchmarking of quantum memories". *npj QI* 7, 108 (2021)

Task: Given fixed U_1, U_2 along with N copies of some unknown state ρ, estimate $Tr(XU_2U_1\rho U_1^\dagger U_2^\dagger)$ using circuits of the form



If we only care about rank-1 observables, $O(2^n)$ copies suffice.

Theorem: For rank-1 observables, $\Omega(2^n)$ copies are necessary.

Overview

I. Background & summary of results

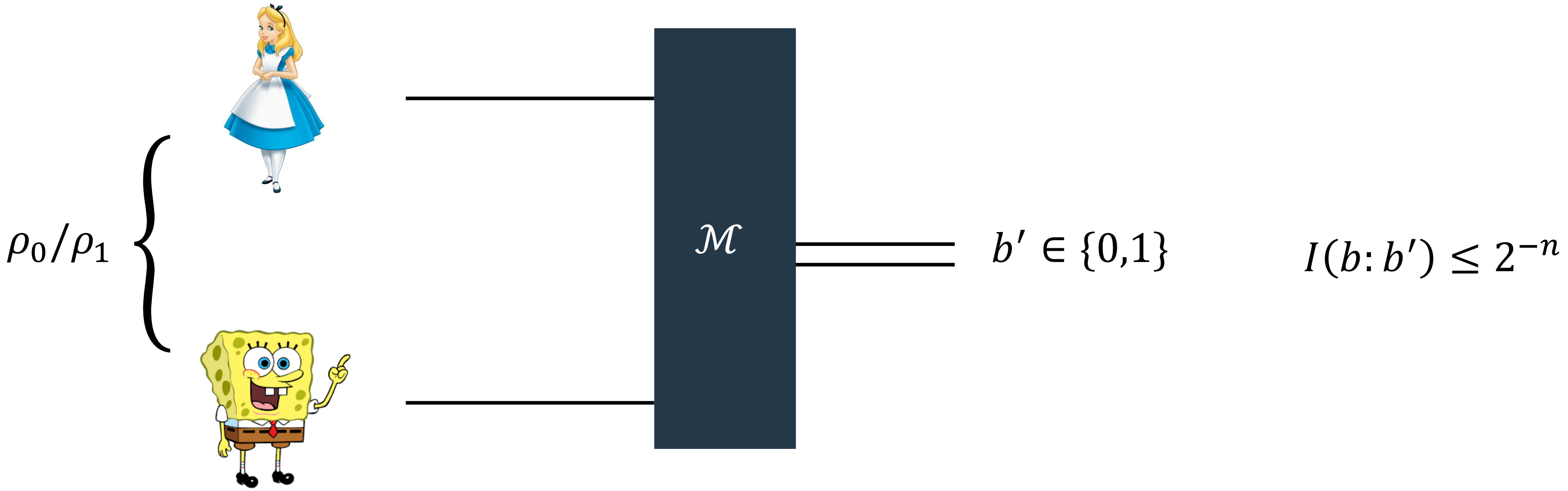
II. Optimal space-like cuts

III. Optimal time-like cuts

IV. Lower bounds

Lower bounds

Use **quantum data hiding** [DLT'02]: there exists a pair of orthogonal states $(\rho_b : b \in \{0,1\})$ on AB such that b is inaccessible to Alice and Bob if they only make LOCC measurements.



Time-like cut that works too well $\Rightarrow I(b' : b) > 2^{-n} \Rightarrow$ contradiction!

Open questions

- Can we find an example of a U such that $\xi(U) \neq 1 + 2R(J(U))$?
- What are some lower bounds in the space-like case?
- Can we extend the clustered Hamiltonian simulation algorithm to compute correlation functions in thermal states?
- Is there a classically-hard family of n-qubit circuits that can be simulated in time poly(n) using just polylog(n)-qubit circuits?

- Can a quantum computer outperform classical computers at some useful task? → Central motivating question
  - Now (NISQ VQEs), Far future (Fault-tolerance) https://arxiv.org/pdf/2203.17181.pdf has a graph cartoon
  - Interesting to look at "soon" regime (hundreds of logical qubits). Quantum simulation?
    - Highlight Quera paper as example of status quo
- Maintaining long-range entanglement is a significant barrier to large-scale quantum computation. How do we mitigate this?
  - Error mitigation: try our best with noise, mitigate the errors
  - Circuit cutting idea: we don't do the entangling operation at all.
  - Can we apply circuit cutting to practical problems?
- Philosophical question: quantum Shannon theory has been (somewhat) successful in establishing a tradeoff between classical randomness and entanglement, for communication. Is there a similar tradeoff between these resources for computation? Circuit cutting is one answer.