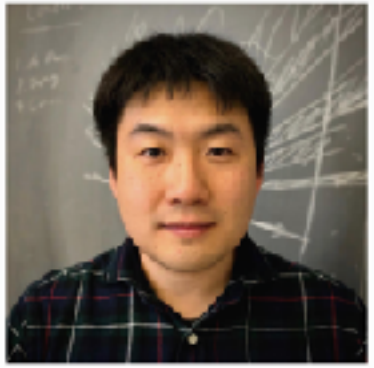
Understanding Linear Cross-Entropy Benchmark through a Statistical Physics Model

Xun Gao
Harvard





Mathematical Picture Seminar
May 11, 2021

Outline

- Quantum Advantage based on Linear Cross-Entropy Benchmark (XEB)
- Overview of our results (a new spoofing algorithm) and their implication
- Understanding our spoofing algorithm by mapping XEB to a diffusion-reaction model If time permitted
- Conclusion and outlook

- Quantum Advantage based on Linear Cross-Entropy Benchmark (XEB)
- Overview of our results (a new spoofing algorithm) and their implication
- Understanding our spoofing algorithm by mapping XEB to a diffusion-reaction model
- Conclusion and outlook

Quantum Advantage (a.k.a. quantum supremacy)

Google:
Sycamore
superconductor

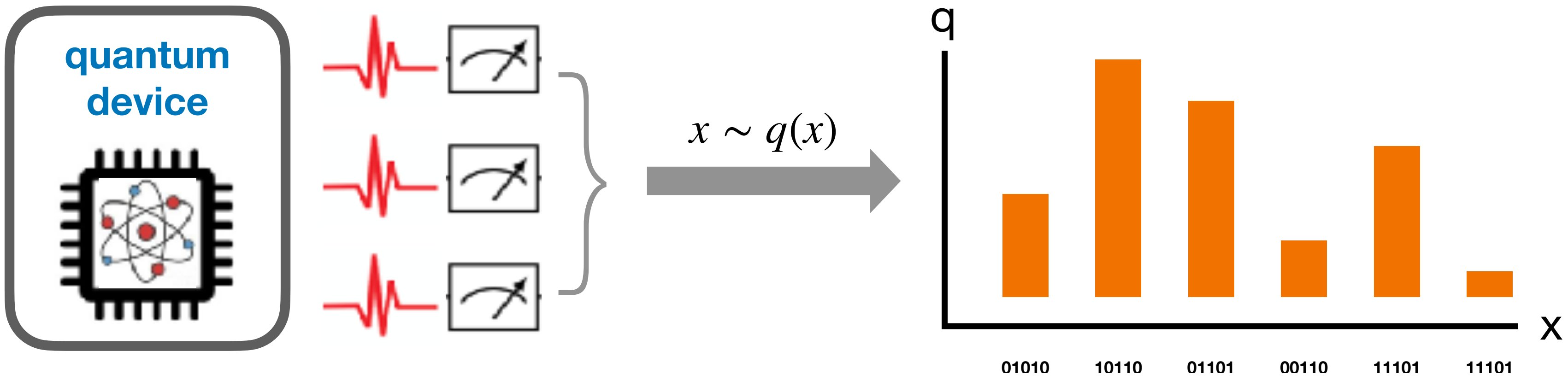
USTC:
Jiuzhang
photonics

quantum advantage: experimental demonstration of the computational power of a quantum device far beyond that of any existing classical devices.

- **constitute a milestone of quantum technology** (high-precision control of complex quantum systems)
- **challenge Extended Church-Turing Thesis** (the foundation of computational complexity theory: the concept of efficiency is expected to be independent of the physical realization of computers)

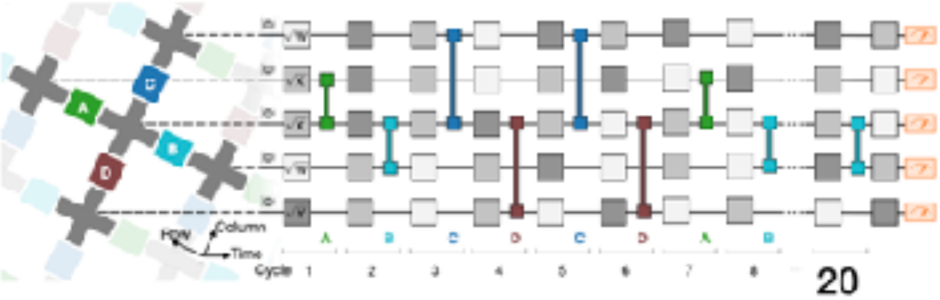
Quantum Advantage based on Random Unitary Circuits (RUC)

Sampling based protocol: (1) suitable for NISQ device (in contrast to implementing Shor's algorithm);
(2) with complexity-theoretic support



is believed hard to simulate classically with complexity-theoretic evidence

A.Bouland, et al. Nature Physics (2019)



Google's Sycamore chip: RUC on a 2D array

But how to benchmark?

Linear Cross-Entropy Benchmark (XEB)

$$\text{XEB}_U = 2^N \sum_x \underbrace{p_U(x)}_{\text{ideal circuit}} \overbrace{q(x)}^{\text{experiment}} - 1 = \mathbb{E}_{\underbrace{x \sim q}_{\text{experiment}}}[2^N \underbrace{p_U(x)}_{\text{ideal circuit}} - 1]$$

- p ≈ q, XEB ≈ 1 for chaotic quantum circuits
- p and q uncorrelated, XEB = 0
- non-vanishing XEB —> p and q have non-trivial correlation

Linearization of KL-divergence $\langle \alpha \log \left(p_U(x)\right) - \beta \rangle_{x\sim q}$

More general cross-entropy $\langle f\left(p_U(x)\right) \rangle_{x\sim q}$

Google. Nature (2019)

among several candidates, ***linear*** cross-entropy has the ***smallest*** statistical fluctuation when estimated empirically

More motivations: two purposes of using XEB

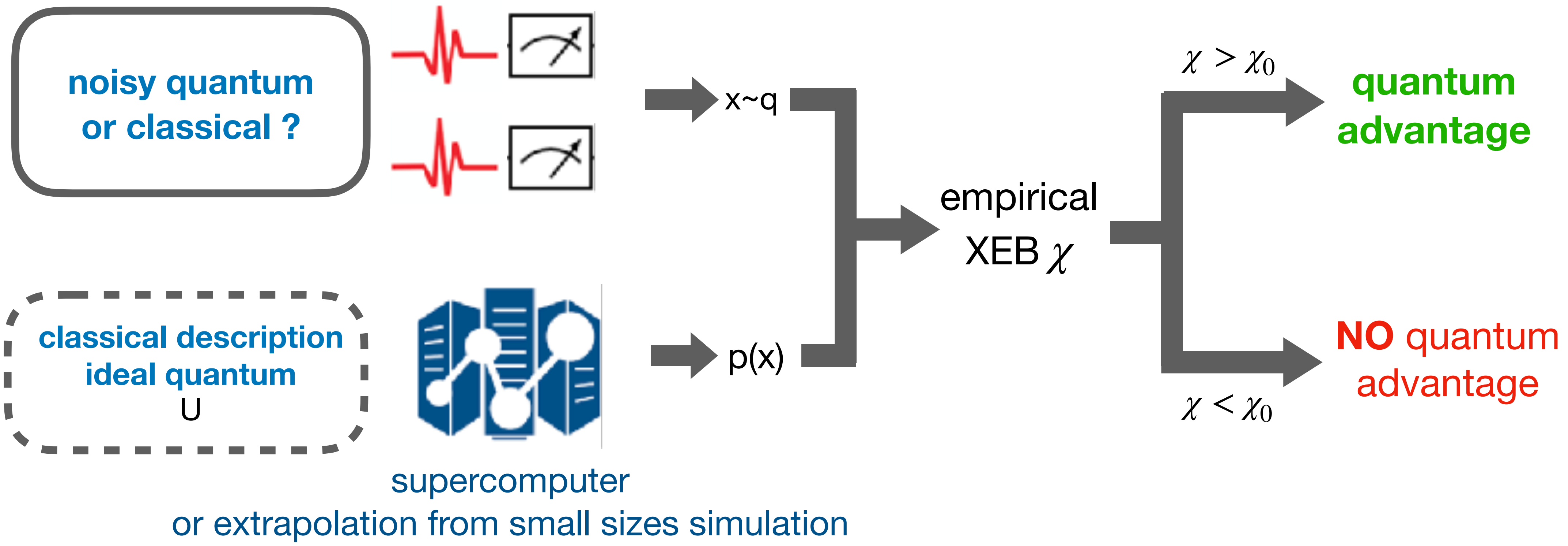
- **proxy for fidelity**
  - linear cross-entropy is similar to fidelity (numerics and analytics under strong assumptions)

Google. Nature (2019)
Google. Nature Physics (2018)

- **certification of quantum advantage**
  - high fidelity is hard to achieve classically —> its proxy should also be hard
  - direct complexity-theoretic evidence (achieving high XEB is classically hard)

Scott Aaronson and Lijie Chen. CCC (2017)
Scott Aaronson and Sam Gunn. arXiv:1910.12085

XEB as certification of quantum advantage



computational complexity version of
Bell’s inequality

Classical Challenger?

53 qubits in 200s:

7,000,000 samples
XEB>0.002

Google's claim: 10,000 yr for a classical computer to achieve XEB>0.002 with 7 million distinct bitstrings

seems challenged by:

heuristic algorithms for tensor network contraction — Gray, J. & Kourtis, S. Hyper-optimized tensor network contraction. *arXiv preprint arXiv:2002.01935* (2020).

improves the above, 20 days in Summit supercomputer — Huang, C. *et al.* Classical simulation of quantum supremacy circuits. *arXiv preprint arXiv:2005.06787* (2020).

improves it further, 5 days using 60 GPUs — Pan, F. & Zhang, P. Simulating the sycamore quantum supremacy circuits. *arXiv preprint arXiv:2103.02074* (2021).

highly correlated bitstrings thus possible to distinguish from experiment (but we don't know how); XEB not enough?

Classical Challenger?

53 qubits in 200s:

7,000,000 samples
XEB>0.002

70 qubits in hours:

100 million samples
XEB>0.0005

Scott Aaronson's blog:


Shtetl-Optimized

Another axe swung at the Sycamore

I guess my main thoughts for now are:

1. Once you knew about this particular attack, you could evade it and get back to where we were before by switching to a more sophisticated verification test — namely, one where you not only computed a Linear XEB score for the observed samples, you *also* made sure that the samples didn't share too many bits in common. (Strangely, though, the paper never mentions this point.)
2. The other response, of course, would just be to redo random circuit sampling with a slightly bigger quantum computer, like the ~70-qubit devices that Google, IBM, and others are now building!

these classical algorithms will fail due to the intrinsic complexity (exp resources)
They are not scalable!

the end of the story?

Classical Challenger?

53 qubits in 200s:

7,000,000 samples
XEB>0.002

70 qubits in hours:

100 million samples
XEB>0.0005

Scott Aaronson's blog:

Another axe swung at the Sycamore

I guess my main thoughts for now are:

1. Once you knew about this particular attack, you could evade it and get back to where we were before by switching to a more sophisticated verification test — namely, one where you not only computed a Linear XEB score for the observed samples, you *also* made sure that the samples didn't share too many bits in common. (Strangely, though, the paper never mentions this point.)
2. The other response, of course, would just be to redo random circuit sampling with a slightly bigger quantum computer, like the ~70-qubit devices that Google, IBM, and others are now building!

these classical algorithms will fail due to the intrinsic complexity (exp resources)
They are not scalable!

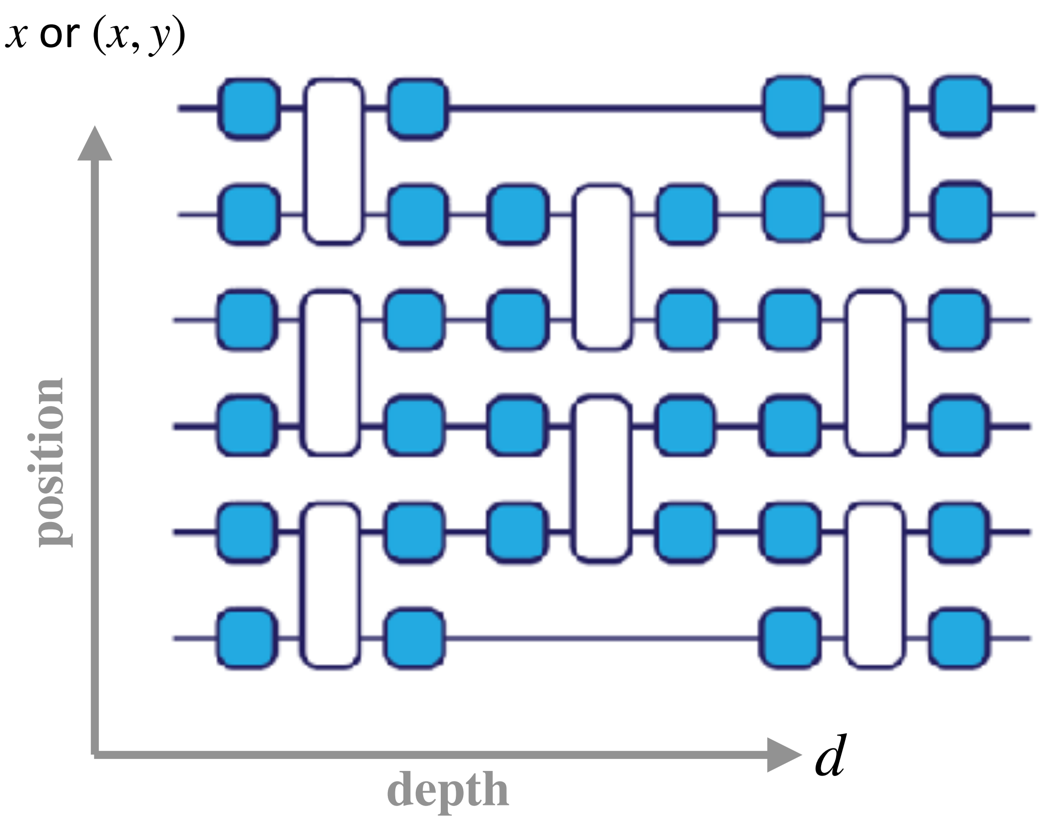
the end of the story?

No, we have a different type of spoofing algorithm

- Quantum Advantage based on Linear Cross-Entropy Benchmark (XEB)
- Overview of our results (a new spoofing algorithm) and their implication
- Understanding our spoofing algorithm by mapping XEB to a diffusion-reaction model
- Conclusion and outlook

- Overview of our results and their implications
  - Random Unitary Circuit (RUC)
  - Description of our spoofing algorithm
  - Key results of our algorithm: (1) scalability; (2) XEB vs. fidelity
  - Implication on the two usages of XEB

Random Unitary Circuit (RUC)

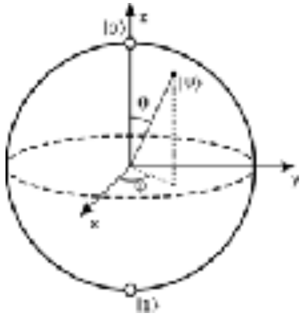


could be arranged on arbitrary architectures (1D, 2D, Sycamore architecture, etc.)

random 2-qubit gate ensemble

1-qubit Haar independent

U



$|\psi\rangle = U|0\rangle$
uniform on Bloch sphere

arbitrary 2-qubit gate

either fixed or from an ensemble

Control-Z (CZ)

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

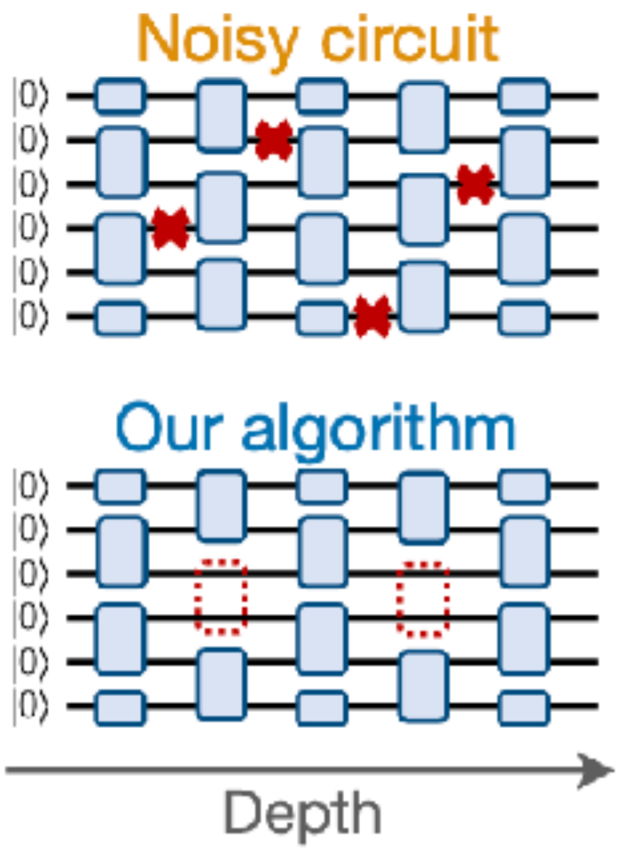
2-qubit Haar

fSim
Google's gate

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -i & 0 \\ 0 & -i & 0 & 0 \\ 0 & 0 & 0 & e^{i\pi/6} \end{pmatrix}$$

Is the XEB really hard to spoof?

A simple example of spoofing algorithms



Motivation of the spoofing algorithm:

- **in noisy circuit**: noises truncate entanglement or correlations effectively among subsystems
- **in our algorithm**: introducing "effective noise": gate defects by omitting some gates removing correlations explicitly —> subsystems decouple —> much easier to simulate

not to simulate the noisy circuit

but $XEB_{\text{our algorithm}}$ and $XEB_{\text{noisy circuit}}$ compete

Performance of our algorithm

- **Complexity-theoretic: 1D Haar random**, **linear-time classical algorithm** outperforms any noisy circuit (no matter how small the noise is) for large enough number of qubits in XEB test

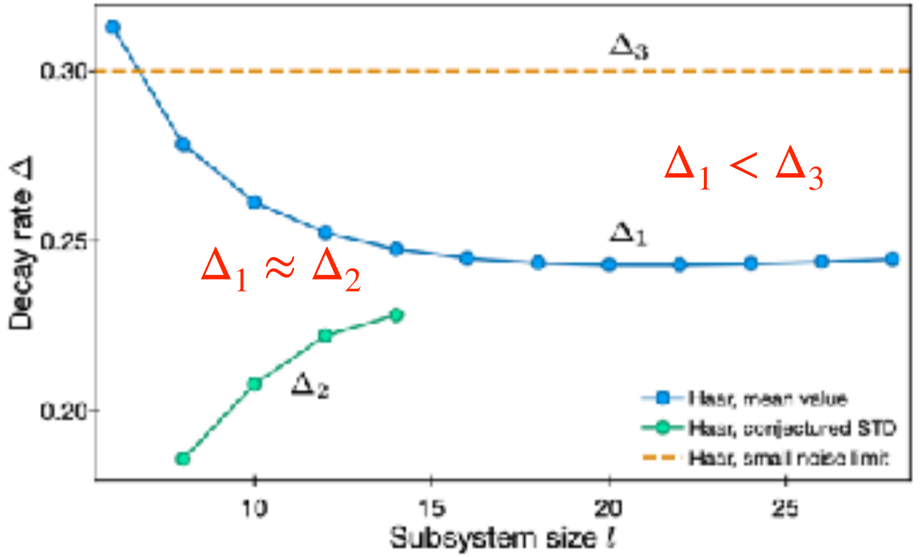
Our algorithm

mean value of XEB $\sim e^{-\Delta_1 d}$

STD of XEB over U $\sim e^{-\Delta_2 d}$

up to a simple generalization

XEB for noisy circuit $\sim e^{-\Delta_3 d}$ when $\epsilon \to 0, N\epsilon > 1$

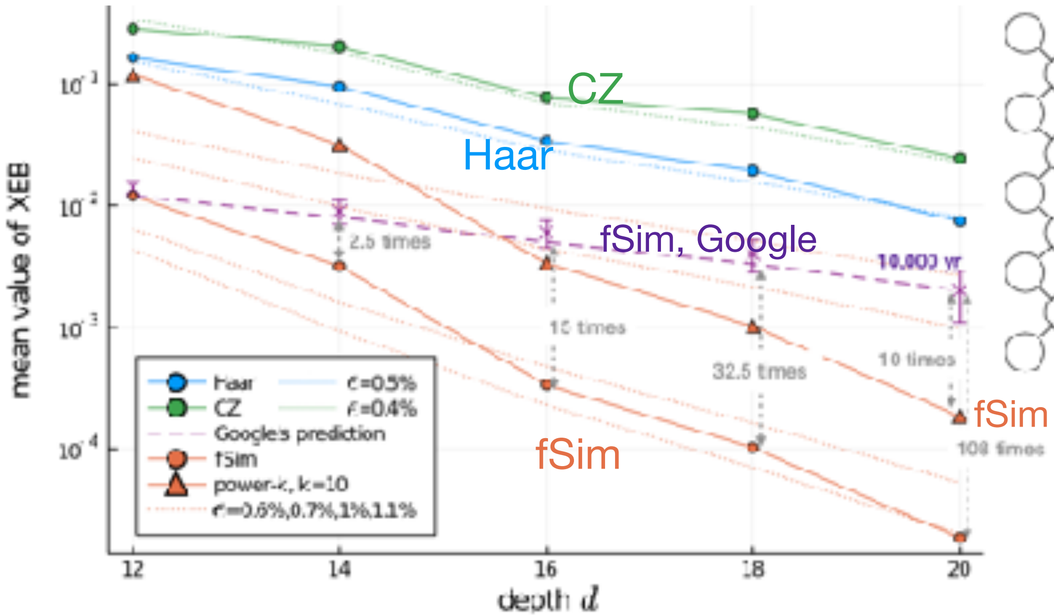


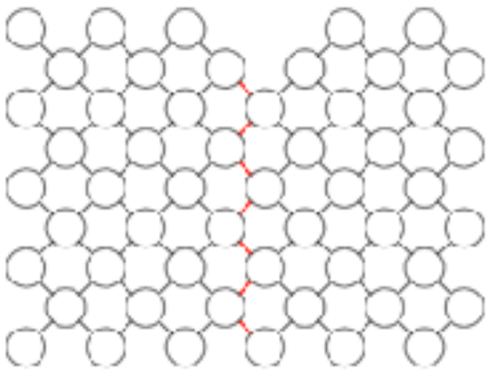
There is a nice interpretation using Ising model

decay rate of our algorithm (**constant subsystem size**)
< decay rate of any noisy circuit in large N case

Performance of our algorithm

- **Current experiment**: Sycamore architecture with 53 qubits & 20 depth
  - super-fast classical algorithm (**<10s** using only **1 GPU**); XEB comparable to Google's (which is ~ 0.002)
  - in our algorithm, XEB$\approx$ 0.007 for Haar random, XEB$\approx$ 0.024 for CZ
  - for fSim (Google's gate), XEB ~ 10% of Google's in average enhanced by power-k method





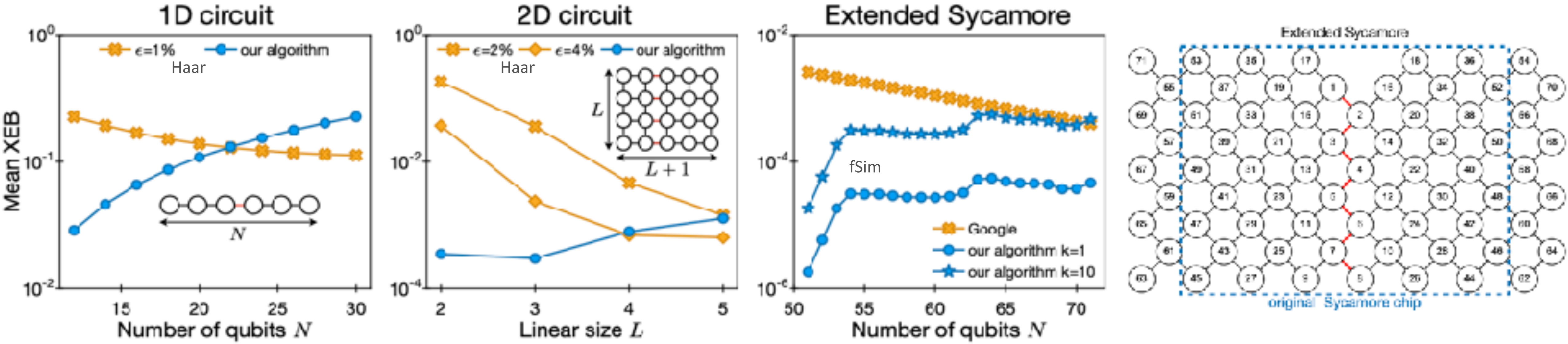
amplifying XEB by a heuristic post-processing: power-k method

$$p_x^{(k)} = \frac{p_x^k}{\sum_y p_y^k} = \frac{p_{x_L}^k}{\sum_{y_L} p_{y_L}^k} \cdot \frac{p_{x_R}^k}{\sum_{y_R} p_{y_R}^k}$$

$E_U[\text{XEB}_k] \approx k \cdot E_U[\text{XEB}_1]$ up to $k \approx 10$ supported by analytics for a toy model and numerics for Sycamore circuit

Performance of our algorithm

- **Scaling to more qubits**: XEB in noisy circuit **decreases**, in our algorithm usually **increases**
  - extending Sycamore to 71 qubits: XEB ~ 10% of Google's & > Google's in average enhanced by power-k



our algorithm has better scaling for achieving high XEB
will eventually outperform noisy circuit

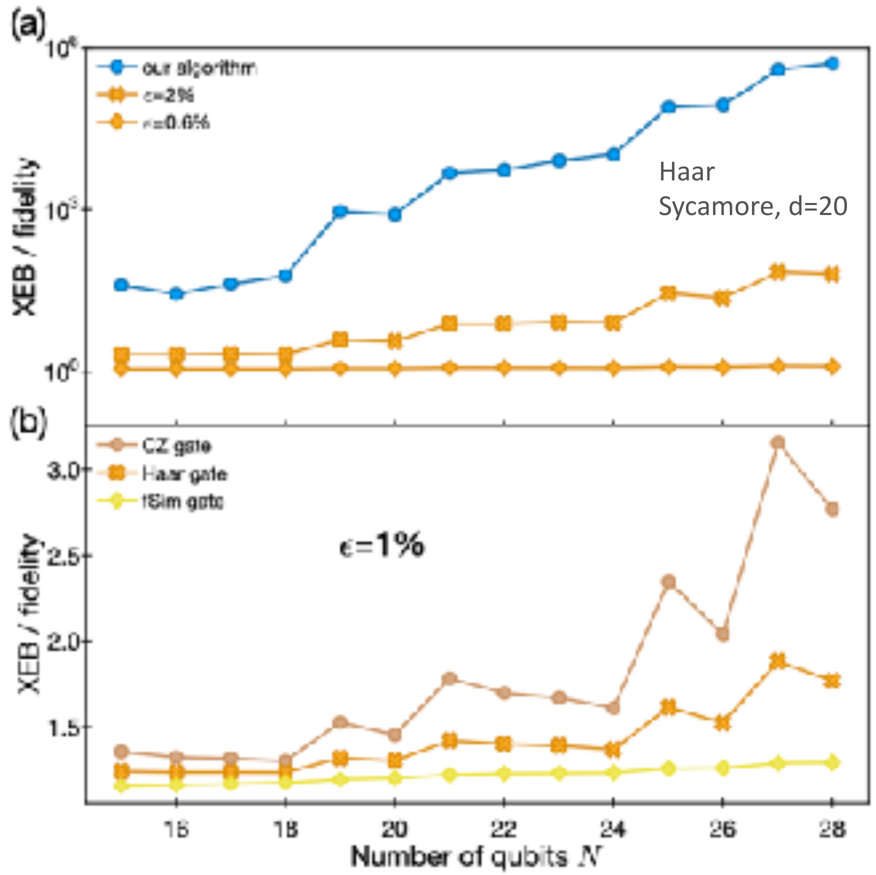
Implication 1

- In 1D Haar random ensemble, linear-time classical algorithm better than any large noisy circuit
- 53 qubits Sycamore architecture, super-fast classical algorithm (<10s, 1GPU), comparable to Google's
- Our classical algorithm has better scaling with system size than noisy circuit

These challenge XEB as a **scalable** measure to **directly certify** quantum advantage

Our algorithm is just an example;
it's pretty simple and there is a lot of room to improve

XEB vs. fidelity in our algorithm



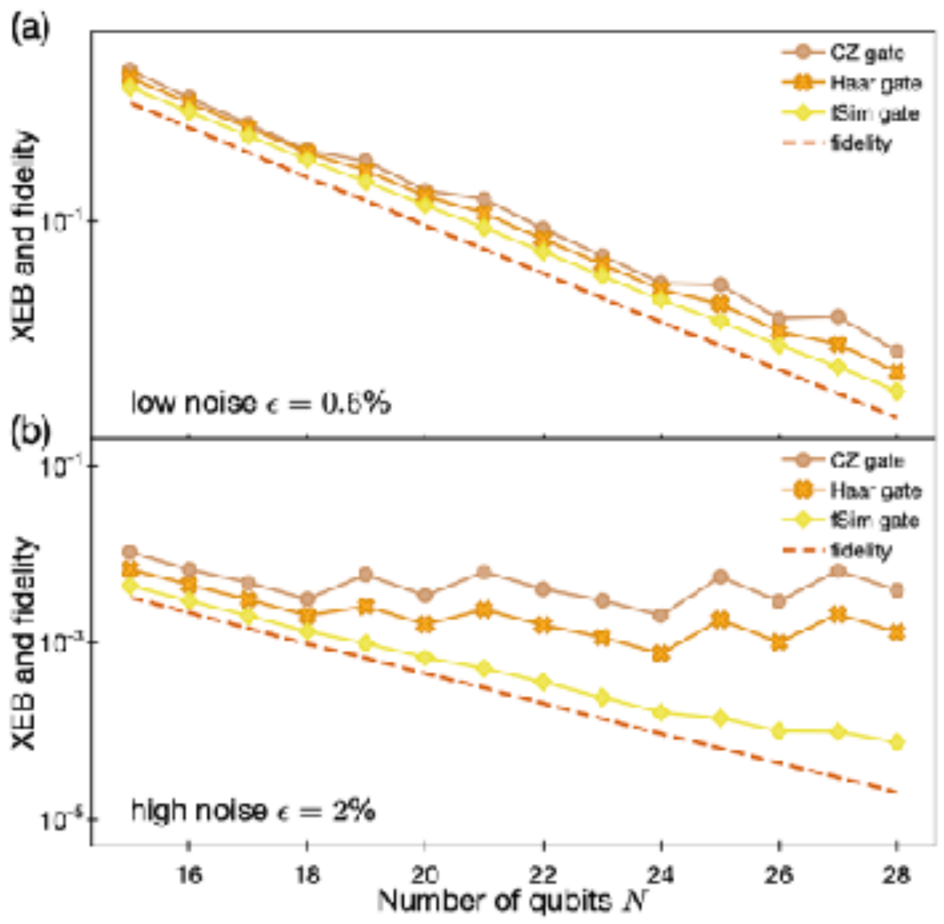
In our algorithm
XEB highly overestimates fidelity

In noisy circuit (i.i.d. 1-qubit noise)
larger noise, more overestimation

In noisy circuit (i.i.d. 1-qubit noise)
The overestimation is **minimized by fSim**

XEB could highly overestimates fidelity
not a good proxy for fidelity in general

Implication 2



assuming independent weak noise

fSim is optimal for XEB≈fidelity

A good choice by Google!

XEB as proxy for fidelity is not un-conditioned
the conditions should be checked experimentally

violation of these conditions are
like the loopholes in Bell's test

Discussion

other measure for quantum advantage? the original quantum supremacy protocols are based on total variation distance (TVD)

$$\frac{1}{2}\sum_x \left| p_{\text{ideal}}(x) - q_{\text{experiment}}(x) \right|$$

e.g., BosonSampling, Ising dynamics (IQP), theory for RUC, etc.

hard to measure; but high fidelity implies this quantity not small

An indirect usage of XEB to certify quantum advantage?

- XEB estimates fidelity only for noisy circuit (need to make sure, e.g., weak independent noise)
- not as a score for a classical challenger to achieve (the score is e.g., TVD)
- but only supports that experiment achieves other score like TVD

based on XEB as a proxy for fidelity

Conclusion

XEB is not a good measure, for both estimating fidelity and certifying quantum advantage, if without further assumptions on the experiment

checking loopholes or new benchmark is required

questions?

Summary

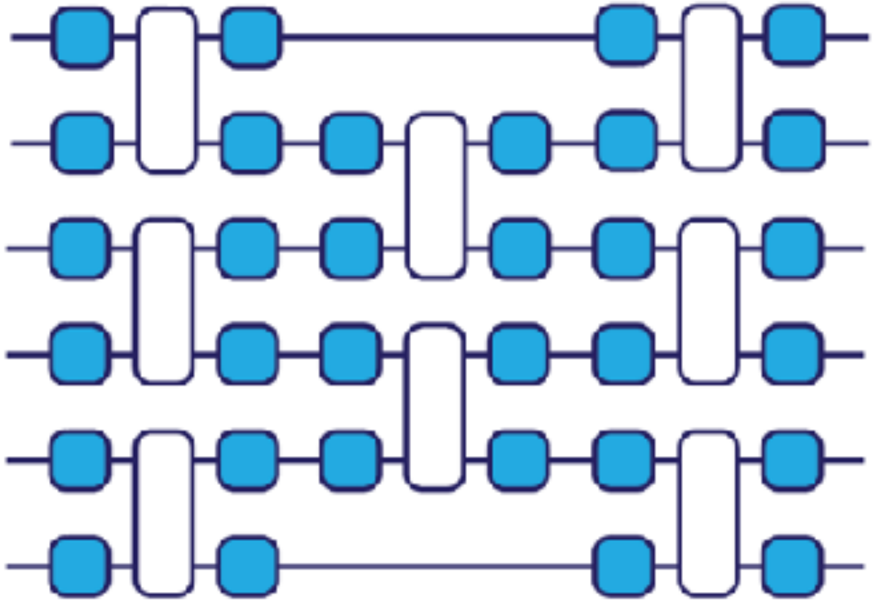
- Our classical algorithm has better scaling with system size than noisy circuit in terms of XEB.

- XEB overestimates fidelity in general but sometimes the overestimation is small.

These could be understood by a statistical physics model

- Quantum Advantage based on Linear Cross-Entropy Benchmark (XEB)
- Overview of our results (a new spoofing algorithm) and their implication
- Understanding our spoofing algorithm by mapping XEB to a diffusion-reaction model
- Conclusion and outlook

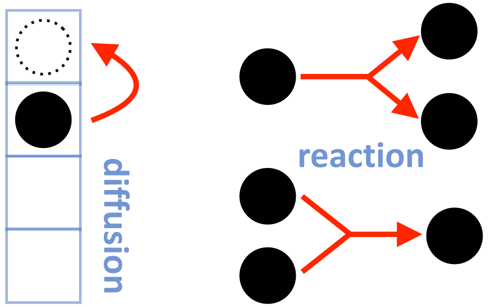
Mapping RUC to Statistical Physics Model



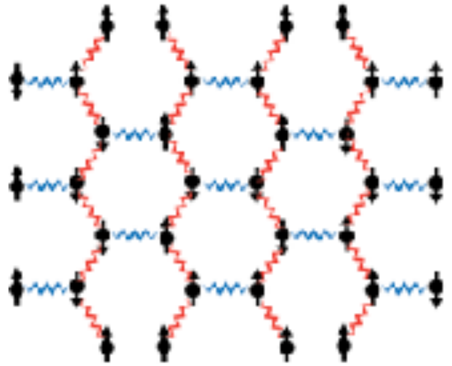
Analyzing individual quantum circuit/dynamics is extremely difficult

averaging over unitaries

Diffusion-reaction model



Ising spin model



Quantum spin chain

Nicholas Hunter-Jones. arXiv:1905.12053
T.Zhou and A.Nahum. PRB(2019)
Y.Bao, S.Choi, and E.Altman. PRB(2020)
C.Jian, et al. PRB(2020)
J.Napp, et al. arXiv:2001.00021
Y.Bao, S.Choi, and E.Altman. arXiv:2102.09164

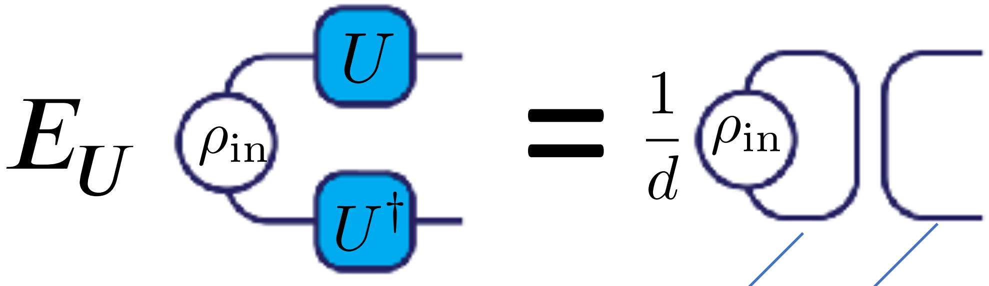
Analyzing the **behavior averaged over an ensemble** is possible

- Mapping XEB to a diffusion-reaction model
  - How does the classical degree of freedom emerge from taking average
  - Mapping XEB and fidelity to a diffusion-reaction model
  - Explaining when XEB $\approx$ (or $\not\approx$) fidelity
  - Understanding the scaling feature of our algorithm

- Mapping XEB to a diffusion-reaction model
  - How does the classical degree of freedom emerge from taking average
  - Mapping XEB and fidelity to a diffusion-reaction model
  - Explaining when XEB $\approx$ (or $\not\approx$) fidelity
  - Understanding the scaling feature of our algorithm

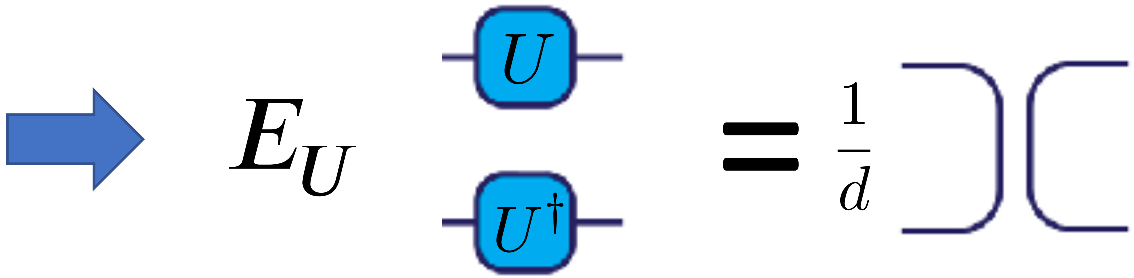
How does it work?
Emergent degree of freedom by randomization

- 1 copy (unitary 1-design) property



$$E_U[U\rho_{\text{in}}U^\dagger] = (\text{tr}\rho_{\text{in}})\frac{I}{d}$$

via Choi-Jamiolkovski isomorphism (channel-state duality) $\sum_{ij} \rho_{ij}|i\rangle\langle j| \to \sum_{ij} \rho_{ij}|i\rangle|j\rangle$

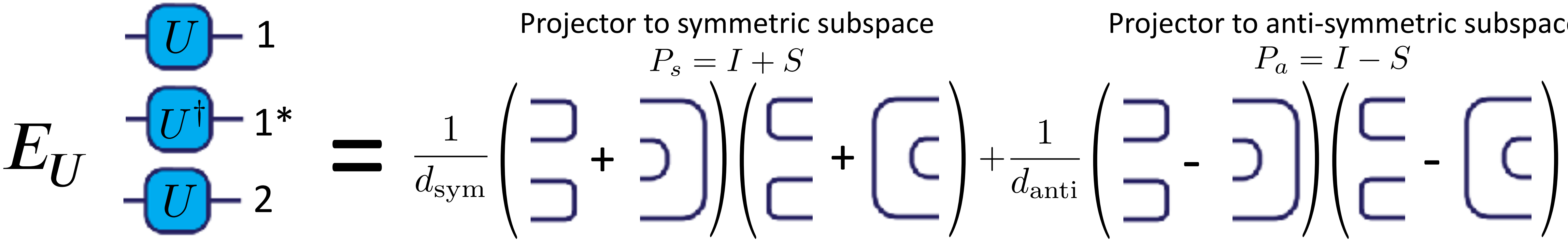


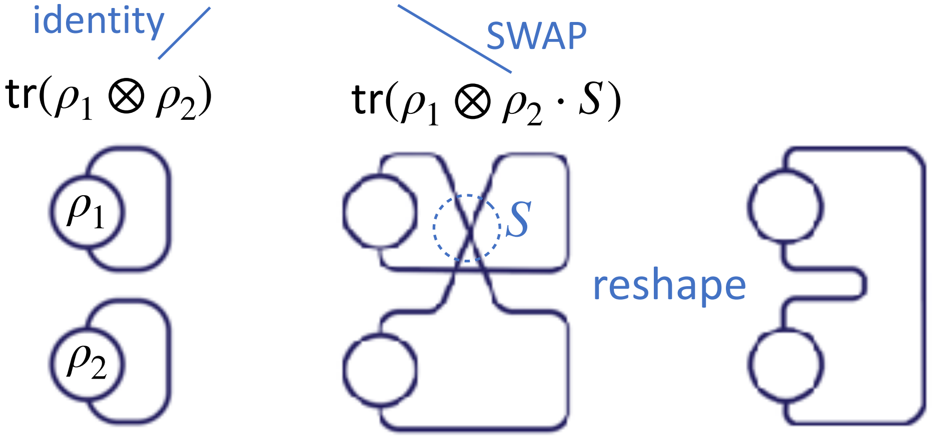
projector of $\frac{1}{\sqrt{d}}\sum_i |ii\rangle$

Delete all information except for the normalization condition (total probability conservation or tr$\rho = 1$)

How does it work?
Emergent degree of freedom by randomization

- 2 copies (unitary 2-design) property (related to Schur-Weyl duality)





Weingarten formula

$$E_U[U^{\otimes 2}\rho_{\mathrm{in}}U^{\dagger\otimes 2}] = \frac{1}{d_s}P_s\mathrm{tr}(P_s\rho_{in}) + \frac{1}{d_a}P_a\mathrm{tr}(P_a\rho_{in})$$

where $\rho_{in} = \rho_1 \otimes \rho_2$

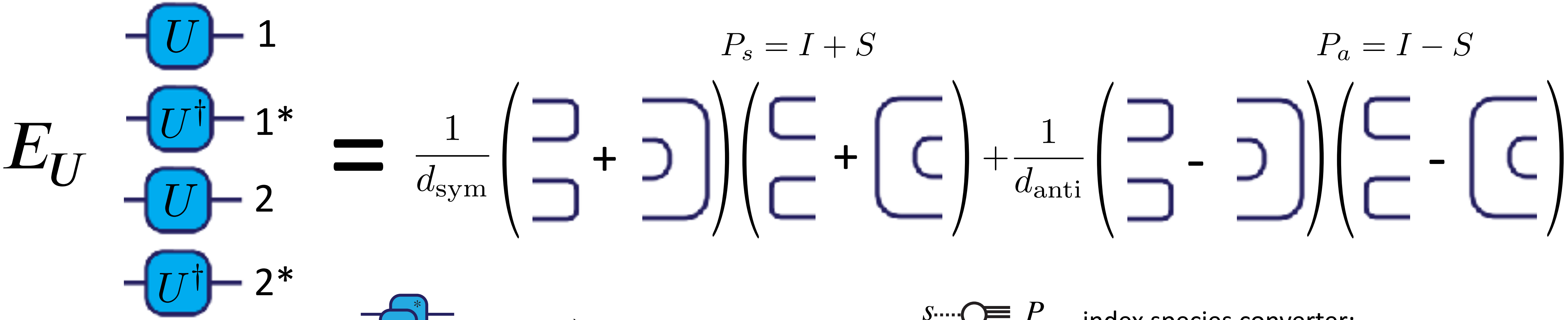
ideal noisy

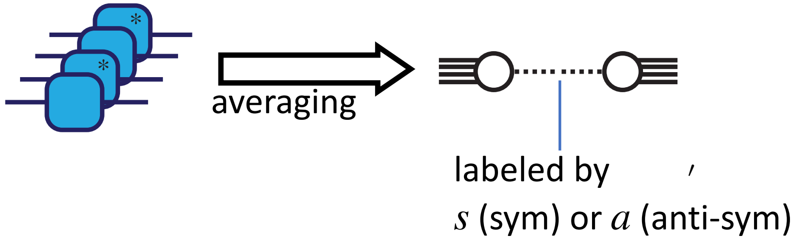
Delete **all information except** for (1) the normalization condition (total probability conservation)

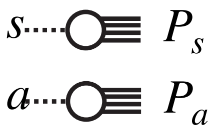
And (2) symmetry wrt swap (either symmetric or anti-sym).

How does it work?
Emergent degree of freedom by randomization

- 2 copies (unitary 2-design) property







index species converter:
effective degree of freedom <—> projectors

Simple basis change !

effective degree of freedom:

I or S — Ising spin model

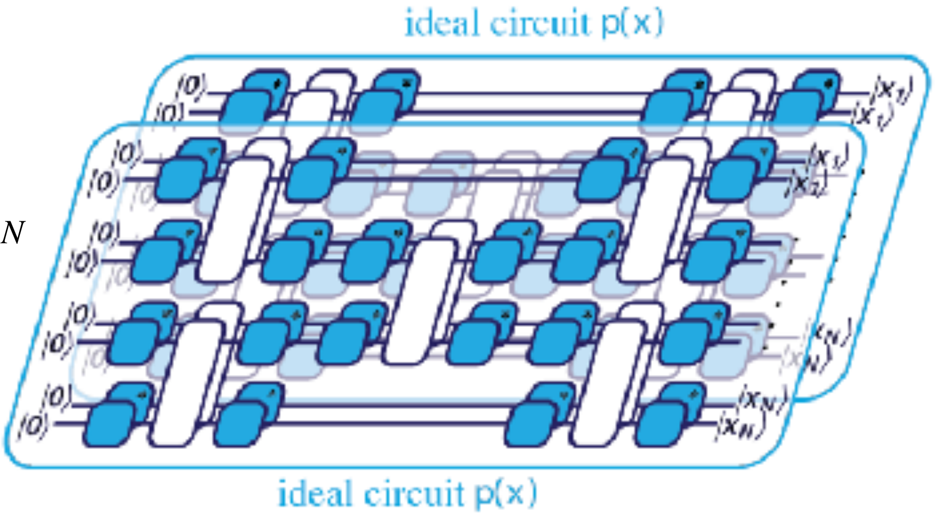
P_s or P_a

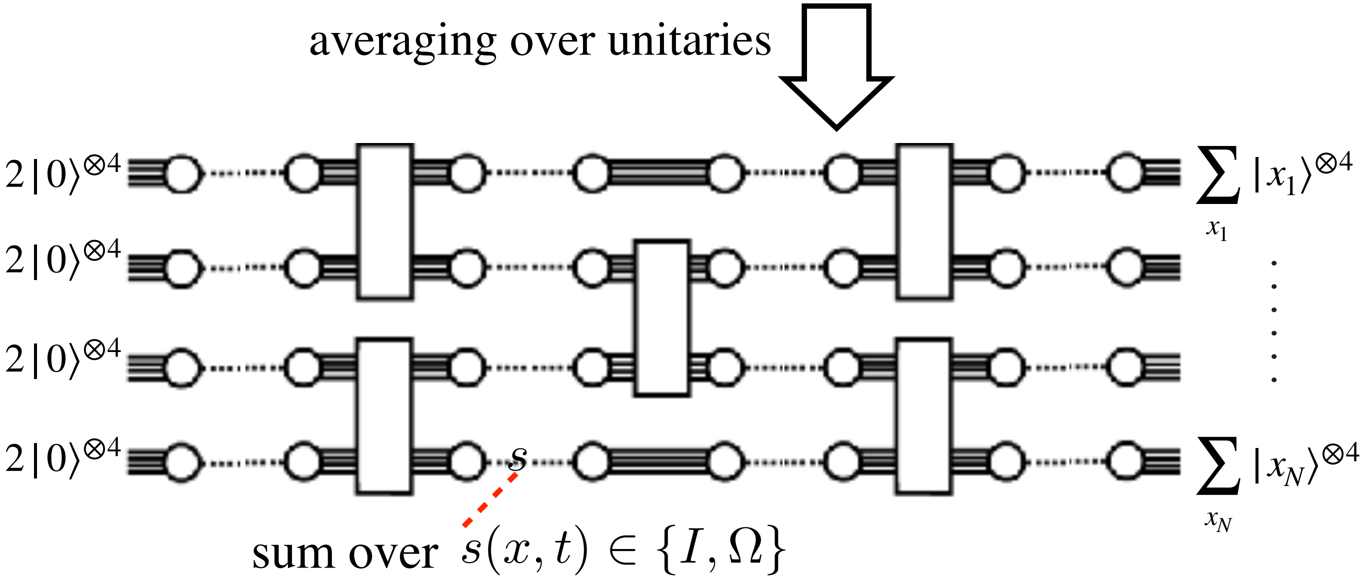
I or $\Omega_{\propto 2S - I}$ — diffusion-reaction model

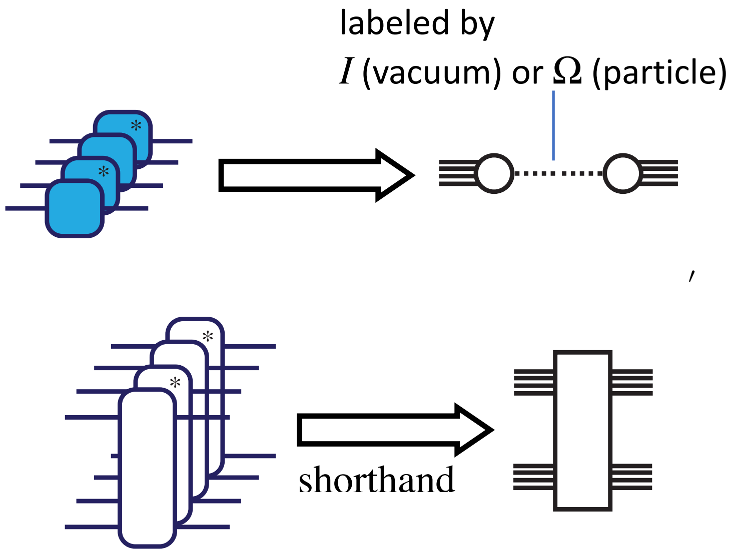
- Mapping XEB to a diffusion-reaction model
  - How does the classical degree of freedom emerge from taking average
  - Mapping XEB and fidelity to a diffusion-reaction model
  - Explaining when XEB $\approx$ (or $\not\approx$) fidelity
  - Understanding the scaling feature of our algorithm

XEB

$$\mathrm{XEB} + 1 = 2^N \sum_x p(x)^2 = \sum_x 2^N$$

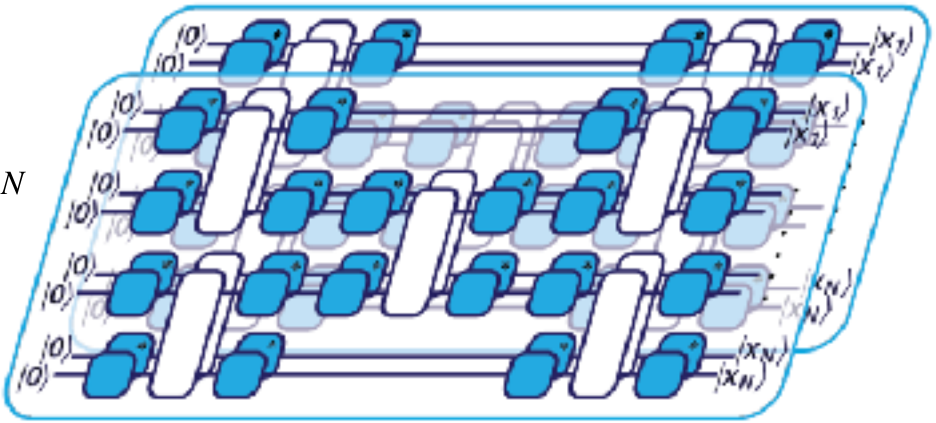


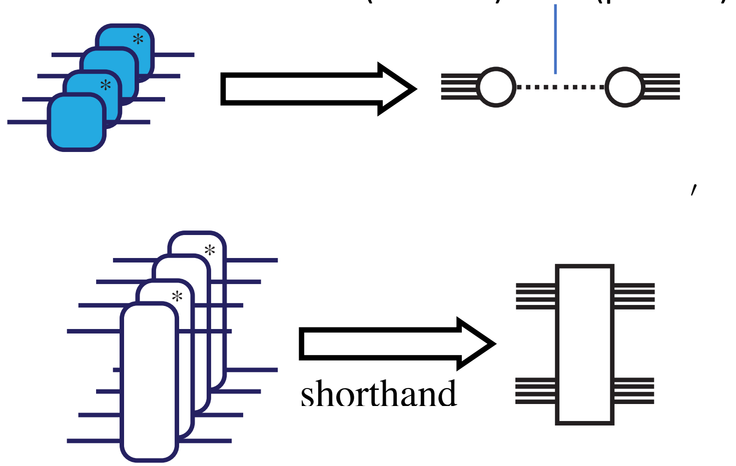




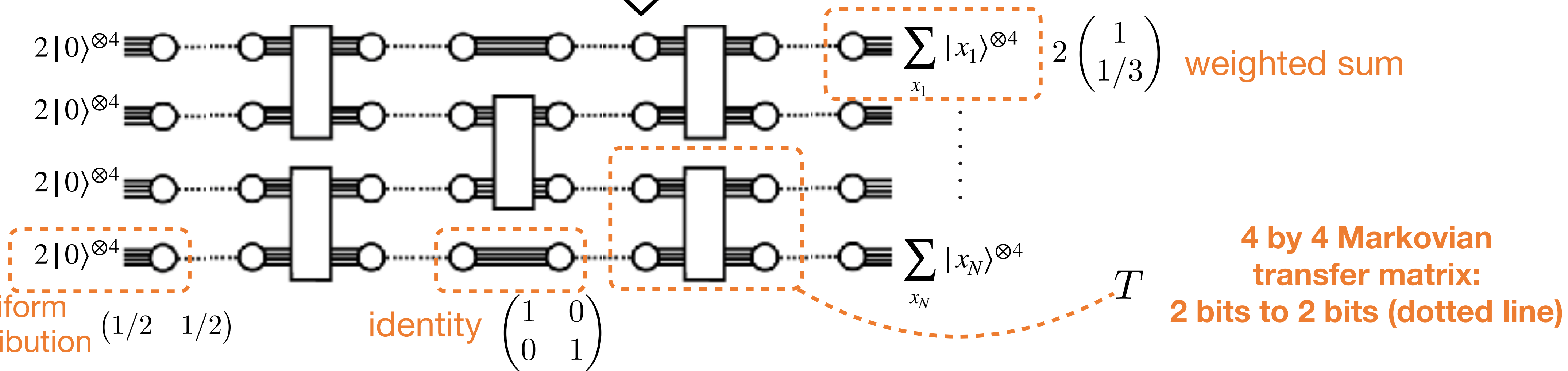
XEB

$$\text{XEB} + 1 = 2^N \sum_x p(x)^2 = \sum_x 2^N$$

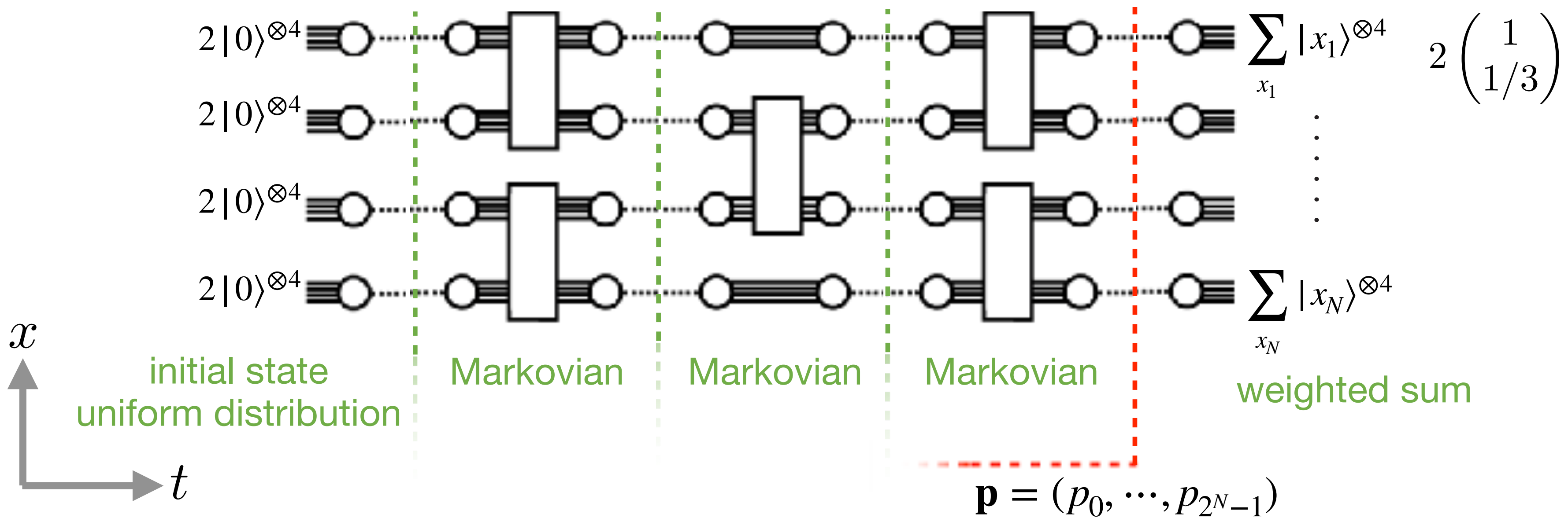




averaging over unitaries



Diffusion-Reaction Model



$$\mathbf{p} \cdot 2^N \begin{pmatrix}1\\1/3\end{pmatrix}^{\otimes N} = \text{path integral of } \mathbb{E}_{\text{histories}}\left[2^N \cdot \frac{1}{3^{\#\Omega} \text{ in the last layer}}\right]$$

the weight 1/3: XEB measured in Z basis, but $\Omega \sim X, Y, Z$
among 3 Paulis, only information in Z can be detected

factor 2 for normalization (will see it later)

Diffusion-Reaction Model

$$T_{\mathrm{Haar}} = \begin{array}{c} \\ II \\ I\Omega \\ \Omega I \\ \Omega\Omega \end{array} \begin{array}{c} \begin{array}{cccc} II & I\Omega & \Omega I & \Omega\Omega \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/5 & 1/5 & 3/5 \\ 0 & 1/5 & 1/5 & 3/5 \\ 0 & 1/5 & 1/5 & 3/5 \end{pmatrix} \end{array}$$

I : vacuum Ω : particle

diffusion process

$$I\Omega \xrightarrow{1/5} \Omega I$$
$$\Omega I \xrightarrow{1/5} I\Omega,$$

random walk

more entanglement "productivity"
—> larger diffusion rate

random walk speed: fSim > Haar > CZ

$$T_{\mathrm{CZ}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 2/3 \\ 0 & 0 & 1/3 & 2/3 \\ 0 & 2/9 & 2/9 & 5/9 \end{pmatrix}$$

$$T_{\mathrm{fSim}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 2/3-\sqrt{3}/6 & 1/3+\sqrt{3}/6 \\ 0 & 2/3-\sqrt{3}/6 & 0 & 1/3+\sqrt{3}/6 \\ 0 & 1/9+\sqrt{3}/18 & 1/9+\sqrt{3}/18 & 7/9-\sqrt{3}/9 \end{pmatrix}$$

reaction process

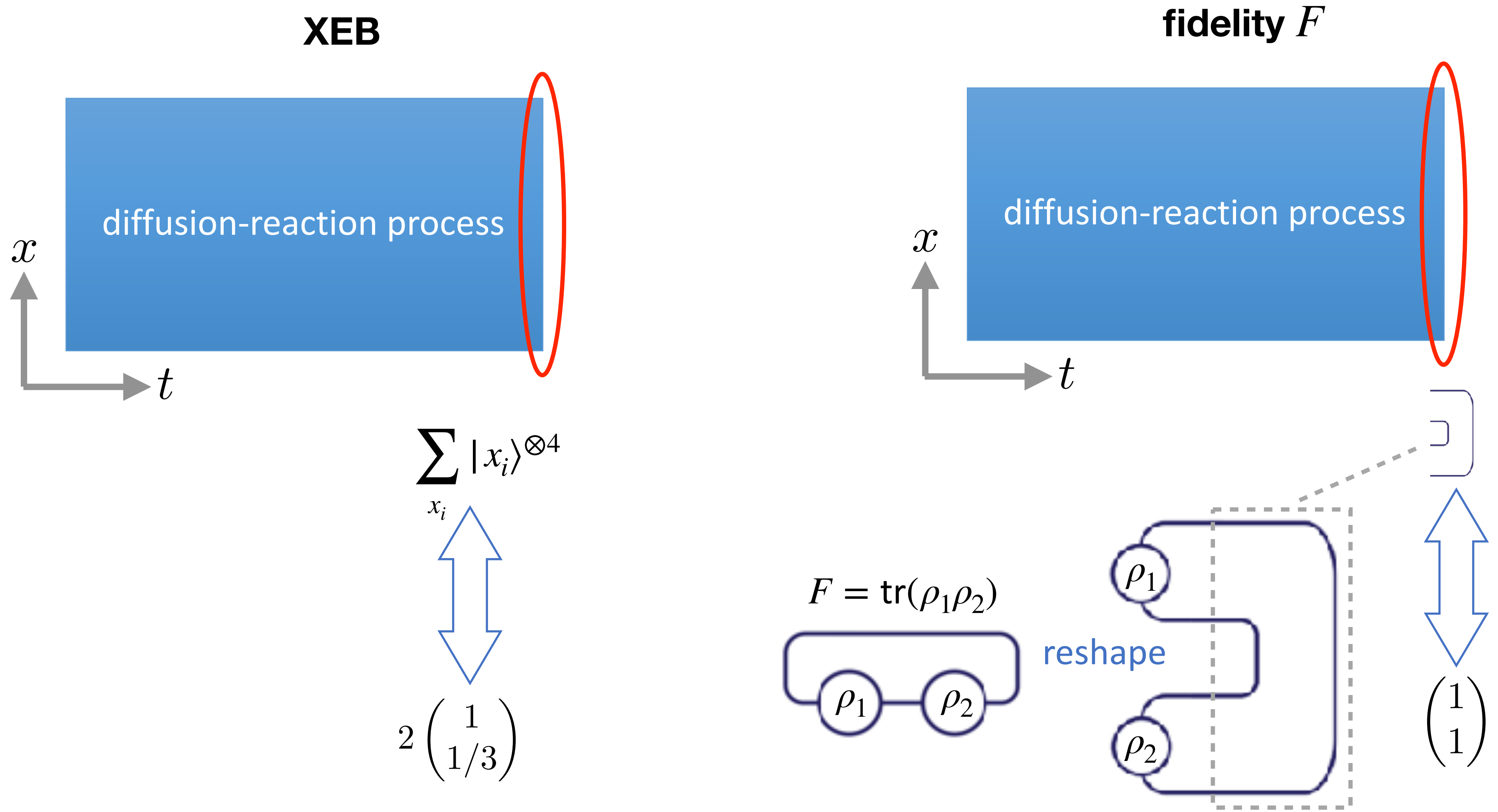
$$I\Omega, \Omega I \xrightarrow{3/5} \Omega\Omega \quad \text{duplication}$$
$$\Omega\Omega \xrightarrow{1/5} I\Omega, \Omega I. \quad \text{recombination}$$

reaction rate: $R = \dfrac{p(\Omega \to \Omega\Omega)}{p(\Omega\Omega \to \Omega)} = 3$

intuitively
Ω represents 3 species
X, Y, Z

reaction rate =3, gate set independent

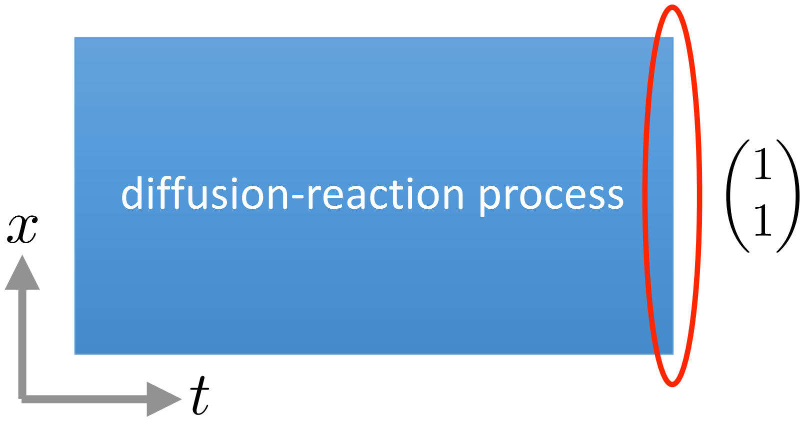
XEB vs. fidelity



XEB vs. fidelity

XEB

fidelity F



XEB = $2^N \mathbf{p} \cdot \begin{pmatrix}1\\1/3\end{pmatrix}^{\otimes N} - 1$ =

path integral of $\mathbb{E}_{\text{histories}}\left[2^N \cdot \frac{1}{3^{\#\Omega \text{ in the last layer}}}\right] - 1$

weighted average of particle populations
(up to normalization)

F = $\mathbf{p} \cdot \begin{pmatrix}1\\1\end{pmatrix}^{\otimes N}$ =

path integral of $\mathbb{E}_{\text{histories}}\left[1\right] = 1$

total sum of particle populations

they only differ in the boundary condition of the last layer !

- Mapping XEB to a diffusion-reaction model
    - How does the classical degree of freedom emerge from taking average
    - Mapping XEB and fidelity to a diffusion-reaction model
    - Explaining when XEB $\approx$ (or $\not\approx$) fidelity
    - Understanding the scaling feature of our algorithm

Why XEB —> fidelity=1 for ideal circuit?

stationary distribution of the last layer

in the absence of error & long time limit / deep circuit

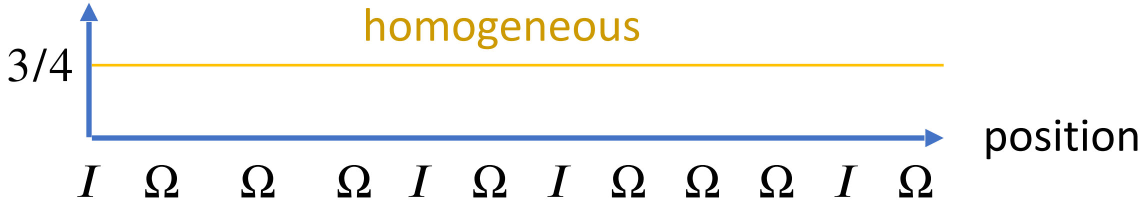
particle population equilibrates: $\mathbf{p} = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \end{pmatrix}^{\otimes N}$

entangling gate independent: fully determined by reaction rate

up to a small correction to eliminate the -1 in the def of XEB
Self-consistent Eq. (Or solving eigenstate with eigenvalue 1)

intuition: in chaotic system, uniform mixture of I, X, Y, Z
I —> vacuum (1 species); X, Y, Z —> Ω (3 species)

density of Ω
at the last layer



time to reach equilibrium: depends on diffusion rate

mixing time $\sim$ scrambling time

Why XEB —> fidelity=1 for ideal circuit?

weighted average of particle populations

$$\begin{pmatrix}\frac{1}{4} & \frac{3}{4}\end{pmatrix} \xrightarrow[\text{re-weight}]{2\begin{pmatrix}1\\1/3\end{pmatrix}} \frac{1}{4}\cdot 2+\frac{3}{4}\cdot\frac{2}{3}=1$$

more weights on I
(2 on I, 2/3 on Ω)

XEB $\approx 1^N = 1$

total sum of particle populations

$$\begin{pmatrix}\frac{1}{4} & \frac{3}{4}\end{pmatrix} \xrightarrow{\begin{pmatrix}1\\1\end{pmatrix}} \frac{1}{4}\cdot 1+\frac{3}{4}\cdot 1=1$$

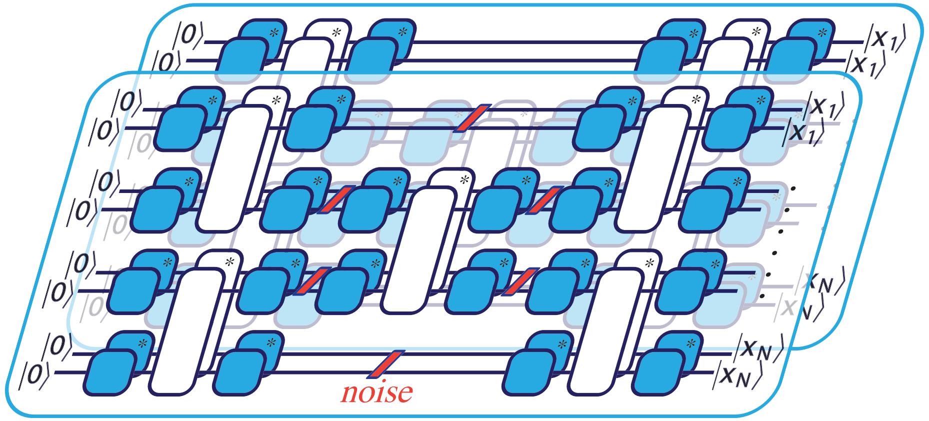
equal weights on I and Ω

$F = 1^N = 1$

Effect of noise

$$\mathrm{XEB}+1 = 2^N \sum_x p(x)q(x) = \sum_{x_1\cdots x_N}$$

ideal circuit p(x)



noisy circuit q(x)

noise (any type)

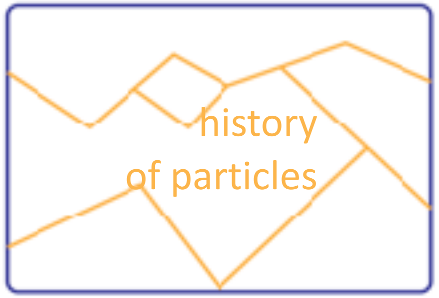
$$\begin{pmatrix} 1 & 0 \\ 0 & 1-\epsilon \end{pmatrix}$$

particle loss

path integral in the bulk:

$$\sum_{\text{histories}}$$

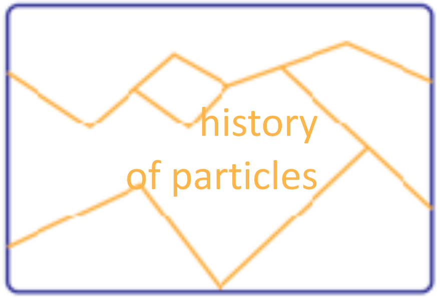
weight of this history



Ideal circuit

$$\Rightarrow \sum_{\text{histories}} (1-\epsilon)^{\text{total }\#\Omega\text{ during this history}}$$

weight of this history



Noisy circuit

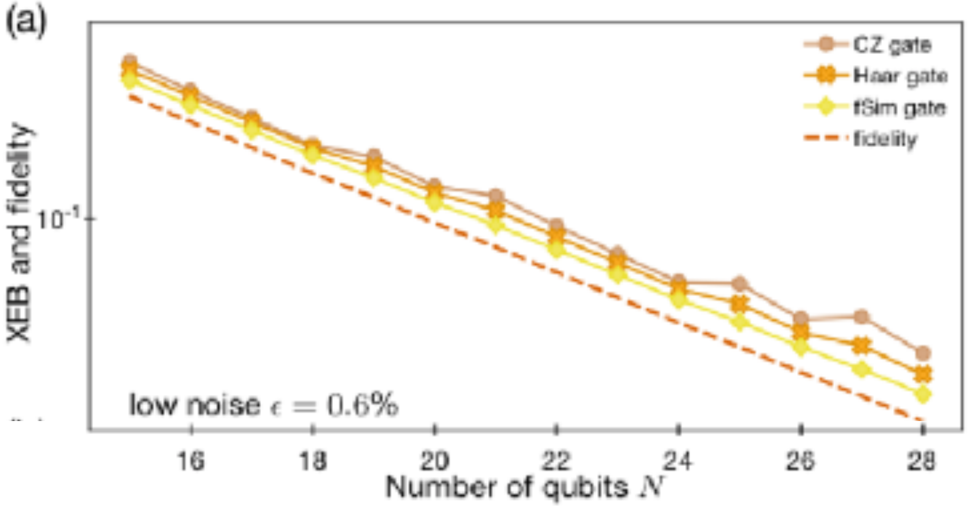
particle loss leads to decline of XEB and fidelity

XEB vs. fidelity in noisy circuit (small noise)

perturbative regime $\mathbf{p} \to \alpha \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \end{pmatrix}^{\otimes N}$ (“equilibration time” < “loss time” : weak noise regime)

final states and the histories approximately decouple
both I and Ω experienced the same amount of loss α

Just re-scale the distribution —> XEB $\approx F \approx \alpha^N$

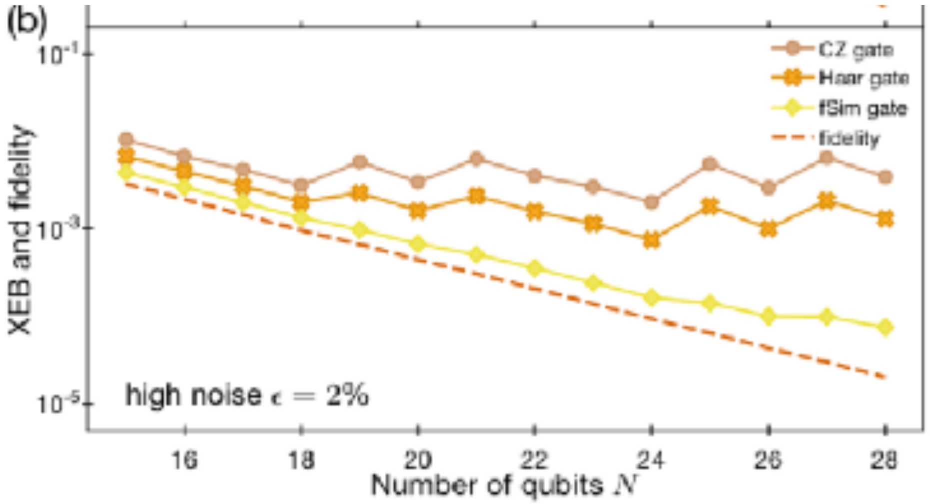


XEB vs. fidelity in noisy circuit (large noise)

non-perturbative regime $\mathbf{p} \rightarrow \begin{pmatrix} \frac{\alpha}{4} & \frac{3\beta}{4} \end{pmatrix}^{\otimes N}$, with $\alpha > \beta$

final states and the histories are correlated
I in the last layer —> more likely I in the previous layers
—> suffer less from particle loss —> $\alpha > \beta$

in XEB, state I (which suffer less from loss) has more weight —> XEB $> F$



The discrepancy is minimized by fSim gate ensemble

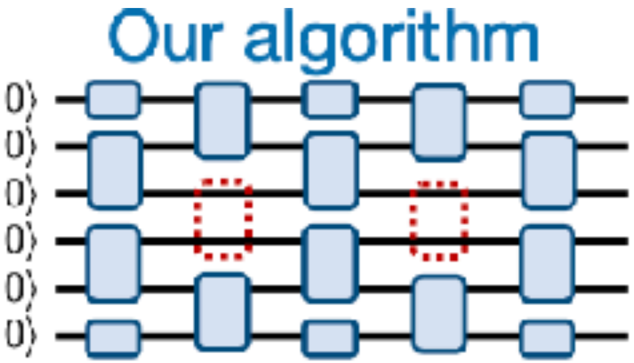
fSim produces more entanglement —>
diffusion rate is large —>
faster equilibration time

- Mapping XEB to a diffusion-reaction model
  - How does the classical degree of freedom emerge from taking average
  - Mapping XEB and fidelity to a diffusion-reaction model
  - Explaining when XEB $\approx$ (or $\not\approx$) fidelity
  - **Understanding the scaling feature of our algorithm**

Our classical algorithm

deviation from the distribution $\begin{pmatrix} \frac{1}{4} & \frac{3}{4} \end{pmatrix}^{\otimes N}$ —> discrepancy between XEB and fidelity

provides a strategy to spoof XEB: disturbing this distribution dramatically



skipping some gates —> 1 copy of $U \otimes U^*$ remained
—> unitary 1-design —> all information is deleted
—> strong depolarizing noise, $\epsilon = 1$

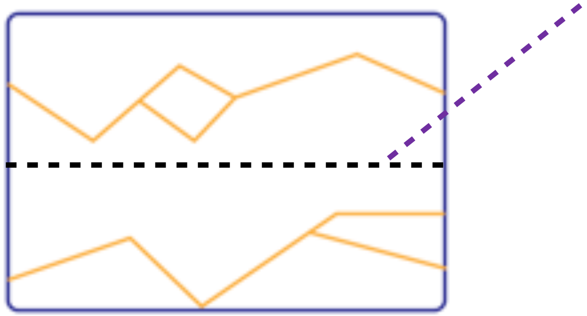
violating homogeneity of the distribution

Our classical algorithm

path integral in the bulk:

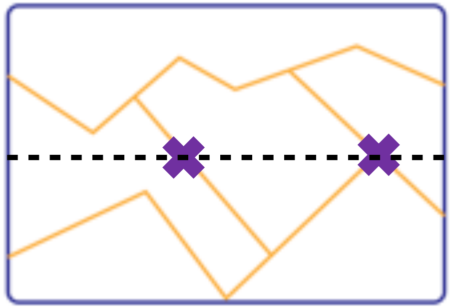
Ideal circuit

$$\sum_{\text{histories without loss}}$$ weight of this history

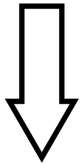


history no loss

$$+ \sum_{\text{histories with loss}}$$ weight of this history

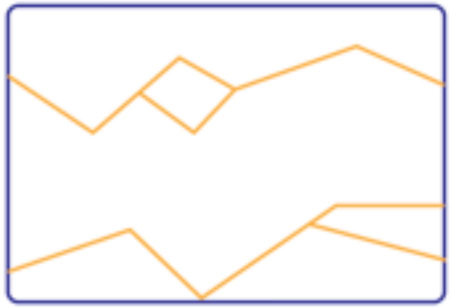


history with loss



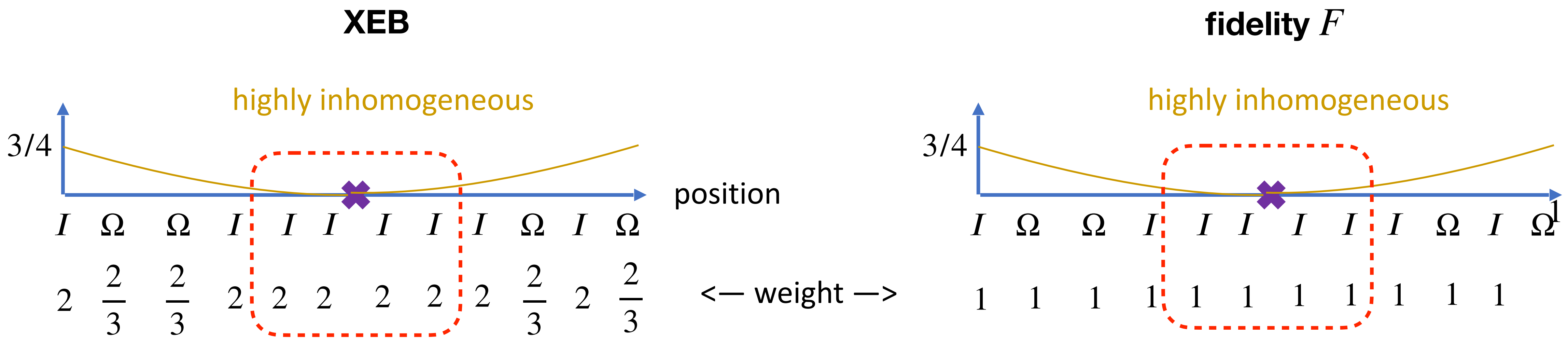
Our algorithm

$$\sum_{\text{histories without loss}}$$



$$\begin{pmatrix} 1 & 0 \\ 0 & 1-\epsilon \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$ projector to I

Our classical algorithm



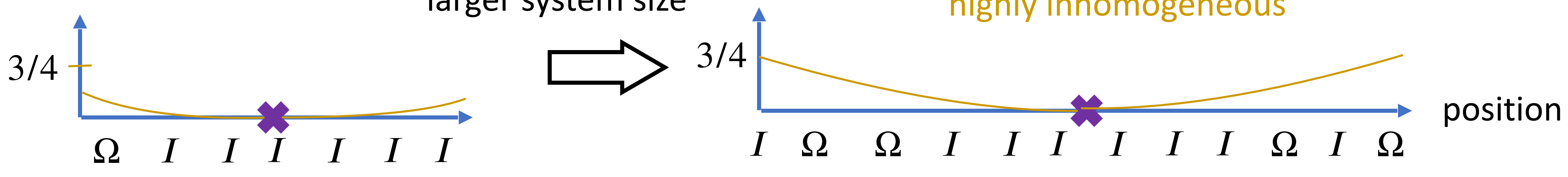
XEB has a larger weight for I & there are a lot of I —>

XEB highly overestimates fidelity in our algorithm

Our classical algorithm

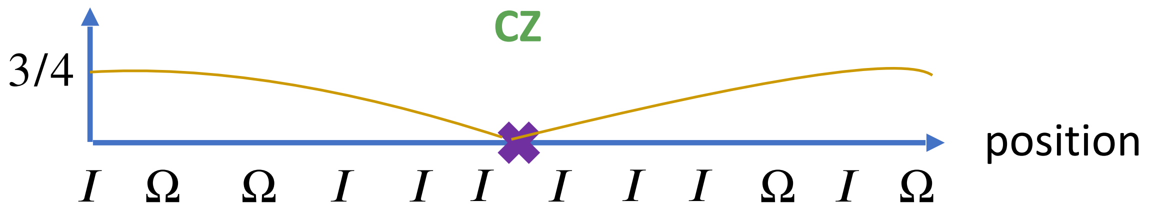
density of Ω
at the last layer

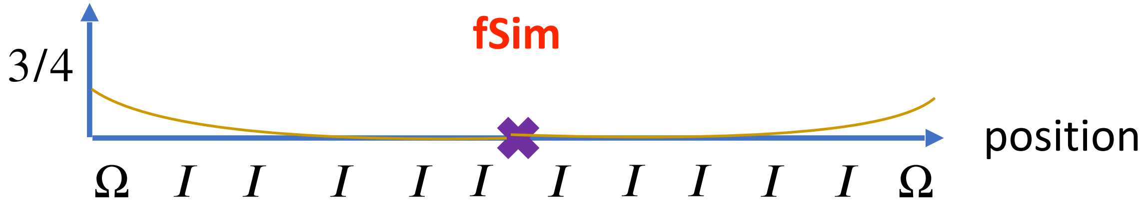
larger system size



more rooms for particles to escape from loss —>

in our algorithm, more qubits, larger XEB in general (in contrast to noisy circuit)





faster diffusion —> easier to be lost —> less XEB

XEB: CZ > Haar > fSim

- Quantum Advantage based on Linear Cross-Entropy Benchmark (XEB)
- Overview of our results (a new spoofing algorithm) and their implication
- Understanding our spoofing algorithm by mapping XEB to a diffusion-reaction model
- Conclusion and outlook

Conclusion

1. XEB is not a good measure, for both estimating fidelity and certifying quantum advantage, if without further assumptions on the experiment

2. The mapping to statistical physics models is an interesting approach to understand RUC

   e.g., the explanation when XEB $\approx$ fidelity

Outlook

1. Extending our algorithm: other post-processing, adding entanglement

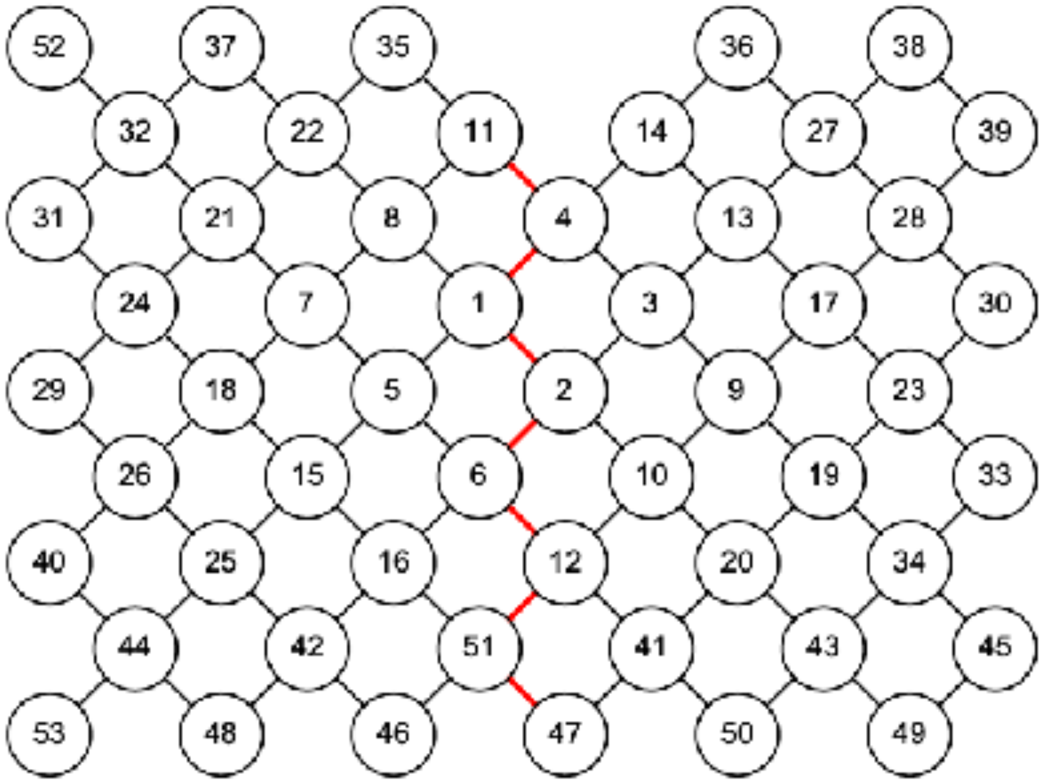
2. Other applications of the diffusion-reaction model: generalizing XEB and learning properties of quantum circuit

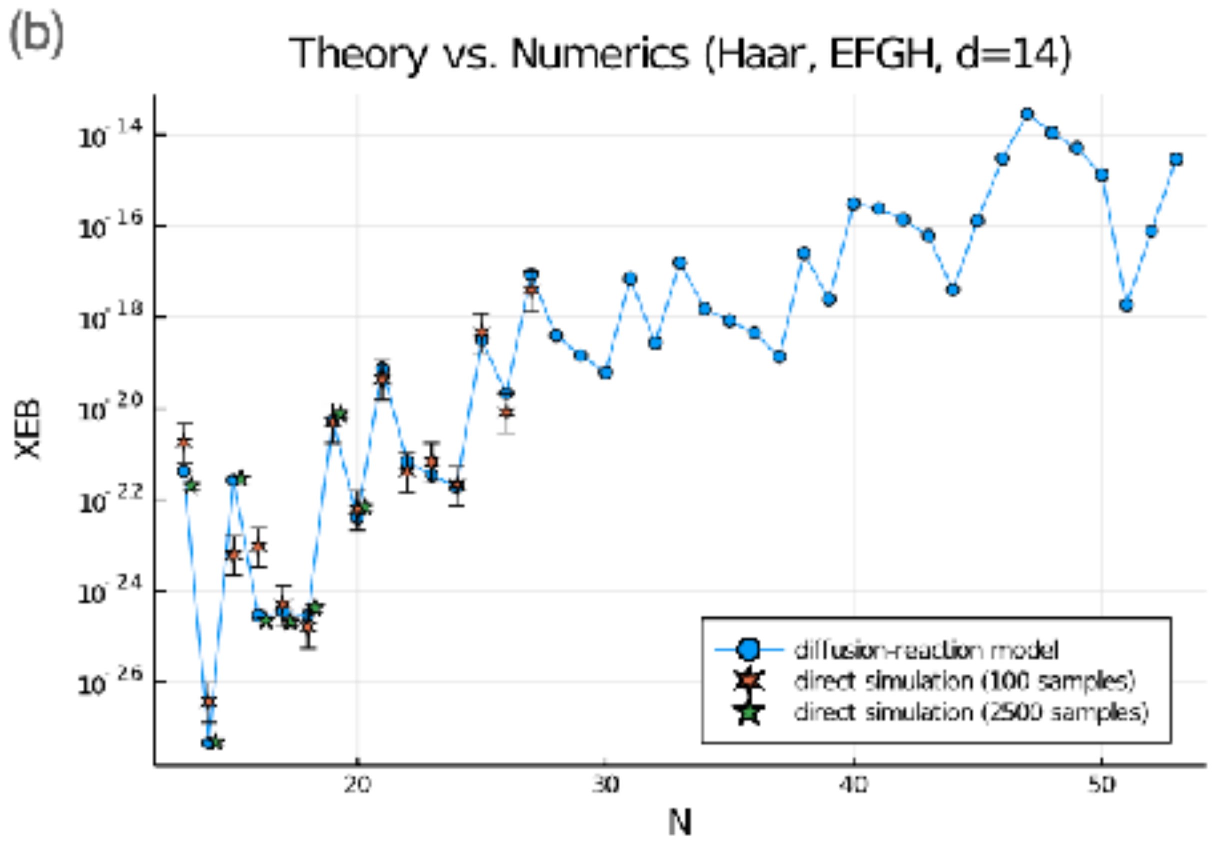
3. Other benchmarks of quantum advantage

Thank you for your attention!

Backup

Google's ordering





We can even explain the valleys and peaks in the curve using diffusion-reaction model

Diffusion-Reaction Model

$$T_{\text{Haar}} = \begin{array}{c} \\ II \\ I\Omega \\ \Omega I \\ \Omega\Omega \end{array} \begin{array}{c} \begin{array}{cccc} II & I\Omega & \Omega I & \Omega\Omega \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/5 & 1/5 & 3/5 \\ 0 & 1/5 & 1/5 & 3/5 \\ 0 & 1/5 & 1/5 & 3/5 \end{pmatrix} \end{array}$$

vacuum: $II \xrightarrow{1} II$

stay: $I\Omega \xrightarrow{1/5} I\Omega, \Omega I \xrightarrow{1/5} \Omega I$

move: $I\Omega \xrightarrow{1/5} \Omega I, \Omega I \xrightarrow{1/5} I\Omega, \Omega\Omega \xrightarrow{3/5} \Omega\Omega$

duplication: $I\Omega, \Omega I \xrightarrow{3/5} \Omega\Omega$

recombination: $\Omega\Omega \xrightarrow{1/5} I\Omega, \Omega I.$

diffusion {stay, move}

reaction {duplication, recombination}

diffusion-reaction model

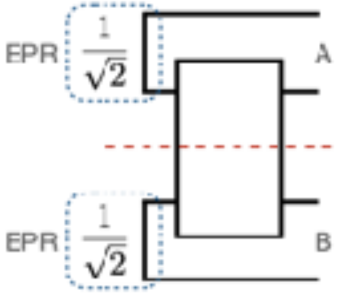
Diffusion-Reaction Model

$$T_{\text{Haar}} = \begin{array}{c} \\ II \\ I\Omega \\ \Omega I \\ \Omega\Omega \end{array} \begin{array}{c} \begin{array}{cccc} II & I\Omega & \Omega I & \Omega\Omega \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \boxed{1/5} & 1/5 & 3/5 \\ 0 & 1/5 & 1/5 & 3/5 \\ 0 & 1/5 & 1/5 & 3/5 \end{pmatrix} \end{array}$$

$$T_{\text{fSim}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \boxed{0} & 2/3 - \sqrt{3}/6 & 1/3 + \sqrt{3}/6 \\ 0 & 2/3 - \sqrt{3}/6 & 0 & 1/3 + \sqrt{3}/6 \\ 0 & 1/9 + \sqrt{3}/18 & 1/9 + \sqrt{3}/18 & 7/9 - \sqrt{3}/9 \end{pmatrix}$$

$$T_{\text{CZ}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \boxed{1/3} & 0 & 2/3 \\ 0 & 0 & 1/3 & 2/3 \\ 0 & 2/9 & 2/9 & 5/9 \end{pmatrix}$$

$\boxed{p_{\text{stay}}}$



diffusion

speed of random walk (diffusion rate)

$$D = 1 - p_{\text{stay}} = \frac{4}{3}(1 - \text{tr}\rho_B^2)$$

more entanglement (smaller purity) <—> faster random walk

$$D_{\text{CZ}} < D_{\text{Haar}} < D_{\text{fSim}}$$

Diffusion-Reaction Model

$$T_{\text{Haar}} = \begin{array}{c} \\ II \\ I\Omega \\ \Omega I \\ \Omega\Omega \end{array} \begin{array}{c} \begin{array}{cccc} II & I\Omega & \Omega I & \Omega\Omega \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/5 & 1/5 & 3/5 \\ 0 & 1/5 & 1/5 & 3/5 \\ 0 & 1/5 & 1/5 & 3/5 \end{pmatrix} \end{array}$$

$$T_{\text{fSim}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 2/3 - \sqrt{3}/6 & 1/3 + \sqrt{3}/6 \\ 0 & 2/3 - \sqrt{3}/6 & 0 & 1/3 + \sqrt{3}/6 \\ 0 & 1/9 + \sqrt{3}/18 & 1/9 + \sqrt{3}/18 & 7/9 - \sqrt{3}/9 \end{pmatrix}$$

$$T_{\text{CZ}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 2/3 \\ 0 & 0 & 1/3 & 2/3 \\ 0 & 2/9 & 2/9 & 5/9 \end{pmatrix}$$

reaction

reaction rate: $R = \dfrac{p(\Omega \to \Omega\Omega)}{p(\Omega\Omega \to \Omega)} = 3$

fully determined by the symmetry in $U \otimes U^* \otimes U \otimes U^*$

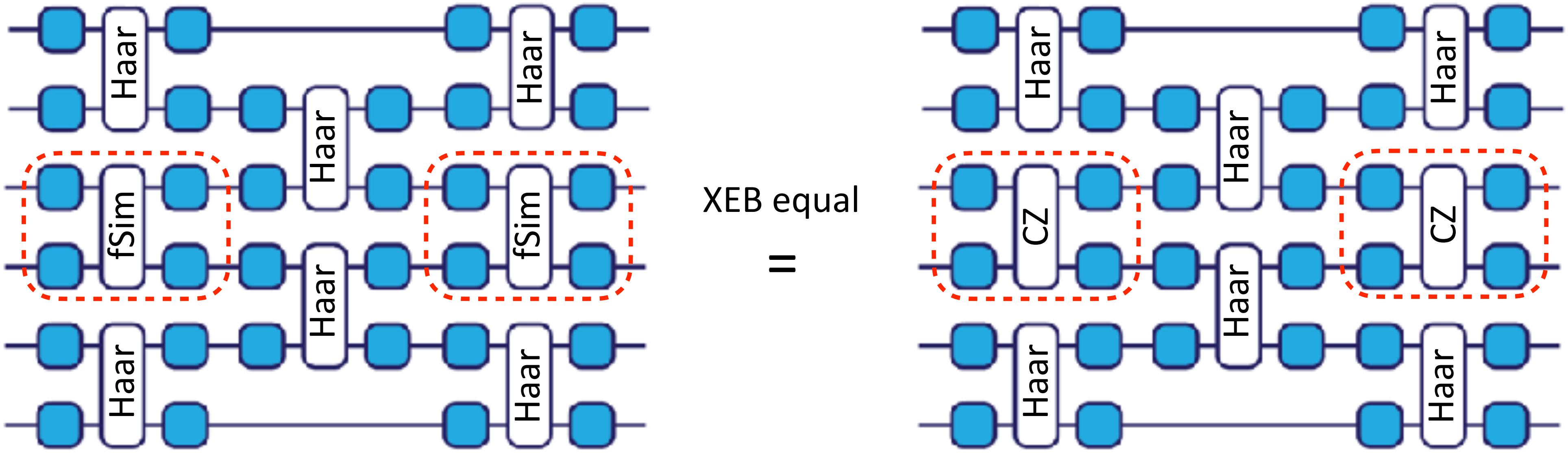
gate set independent, a universal constant

$$R_{\text{CZ}} = R_{\text{Haar}} = R_{\text{fSim}}$$

Our classical algorithm: different gate set

faster diffusion —> less XEB

trivial result? fSim has larger bond dimensions (more entangled) —> **No**
more dangerous to skip because of more "entanglement" being removed?

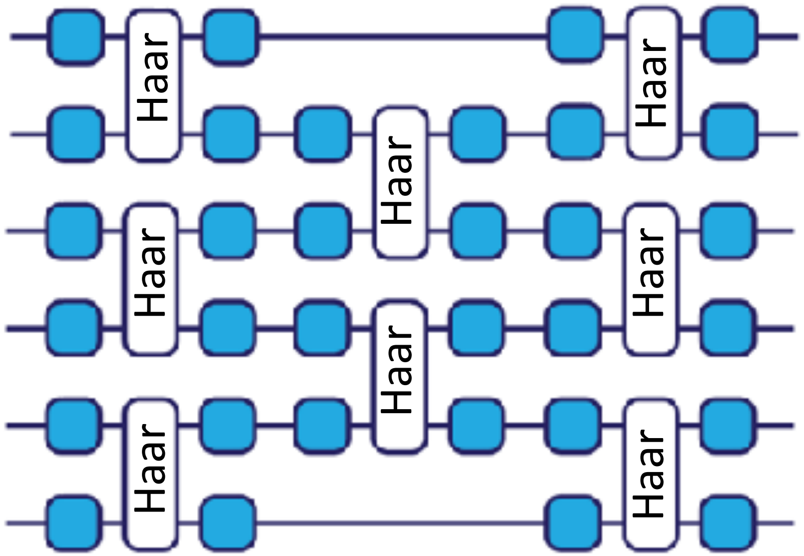


XEB determined by the "entanglement productivity" of the remaining gates instead of those being skipped

Or equivalently, total "entanglement productivity" is mainly determined by the (globally) scrambling property

Mapping XEB to Ising model

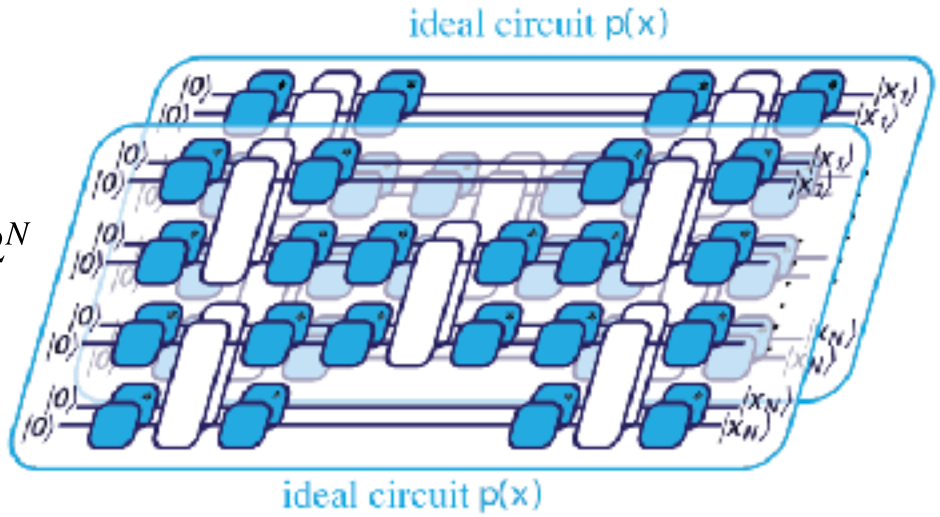
for more quantitative analysis of asymptotic behavior for complexity-theoretic purpose



1+1D Haar —> 2D Ising model

XEB

$$\text{XEB} + 1 = 2^N \sum_x p(x)^2 = \sum_x 2^N$$



XEB of ideal circuit

$$\frac{1}{4}|I\rangle\langle I| + \frac{3}{4}|\Omega\rangle\langle\Omega|$$

in Choi-Jamiolkovski representation
operation from 4 qubits to 4 qubits
only 2 dim subspace

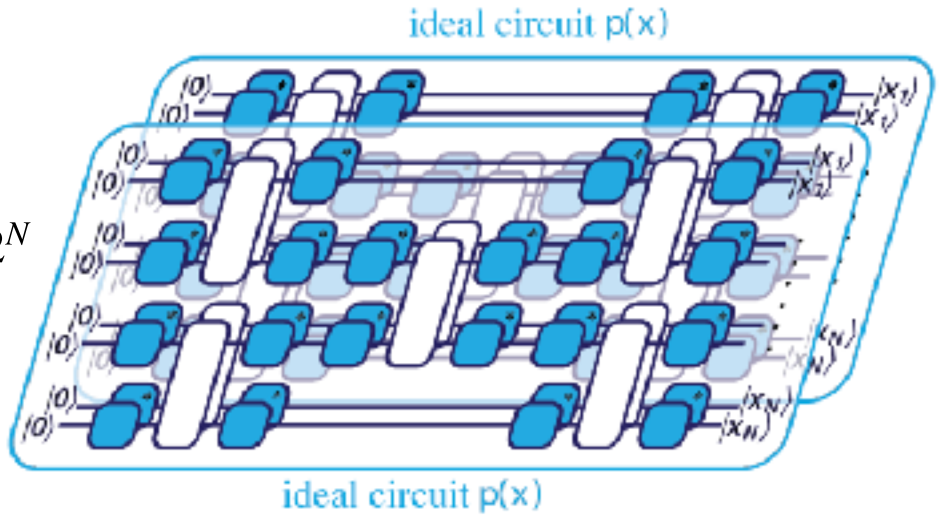
$|I\rangle =$  $= (|00\rangle + |11\rangle)^{\otimes 2}$

$|\Omega\rangle = \frac{1}{3}\sum_{\sigma=X,Y,Z}$  $= \frac{1}{3}\sum_{\sigma=X,Y,Z}[\sigma \otimes I(|00\rangle + |11\rangle)]^{\otimes 2}$

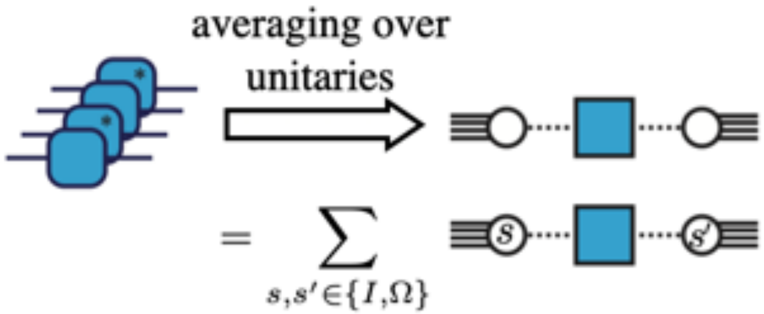
two labels are enough

XEB

$$\text{XEB} + 1 = 2^N \sum_x p(x)^2 = \sum_x 2^N$$

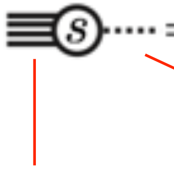


XEB of ideal circuit



 $= \frac{1}{4}\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$

I : vacuum $\quad\Omega$: particle

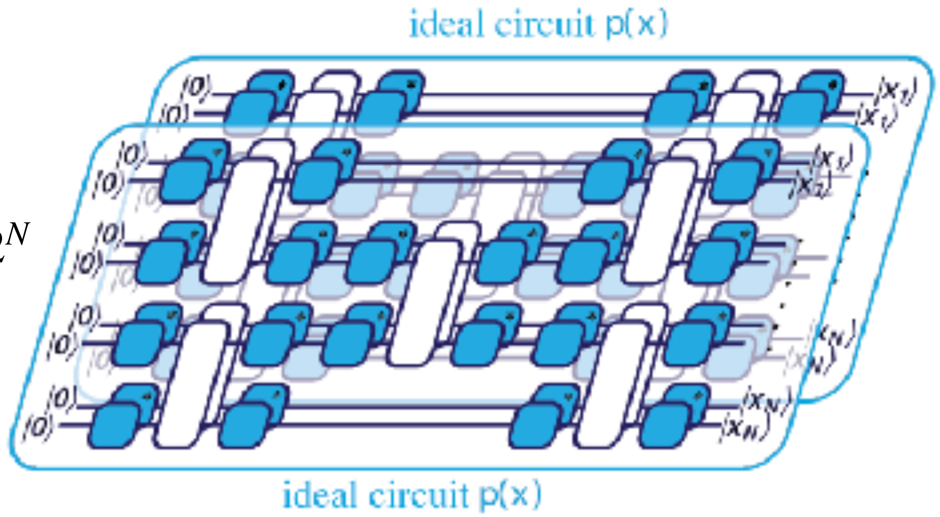
 $= \begin{cases} |I\rangle, & \text{if } s = I \\ |\Omega\rangle, & \text{if } s = \Omega \end{cases}$

label **a subspace of 4 qubits (dim=2)** by **vacuum/particle**

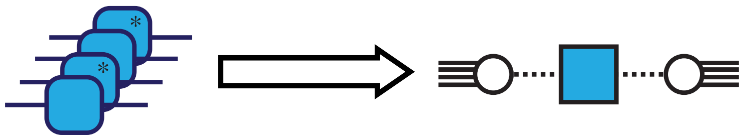
index species converter

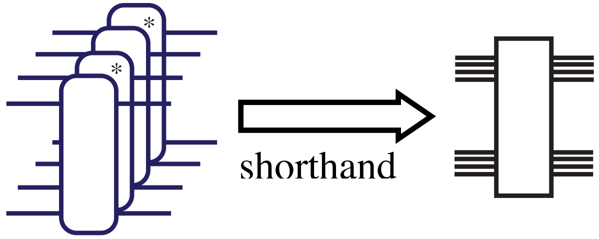
XEB

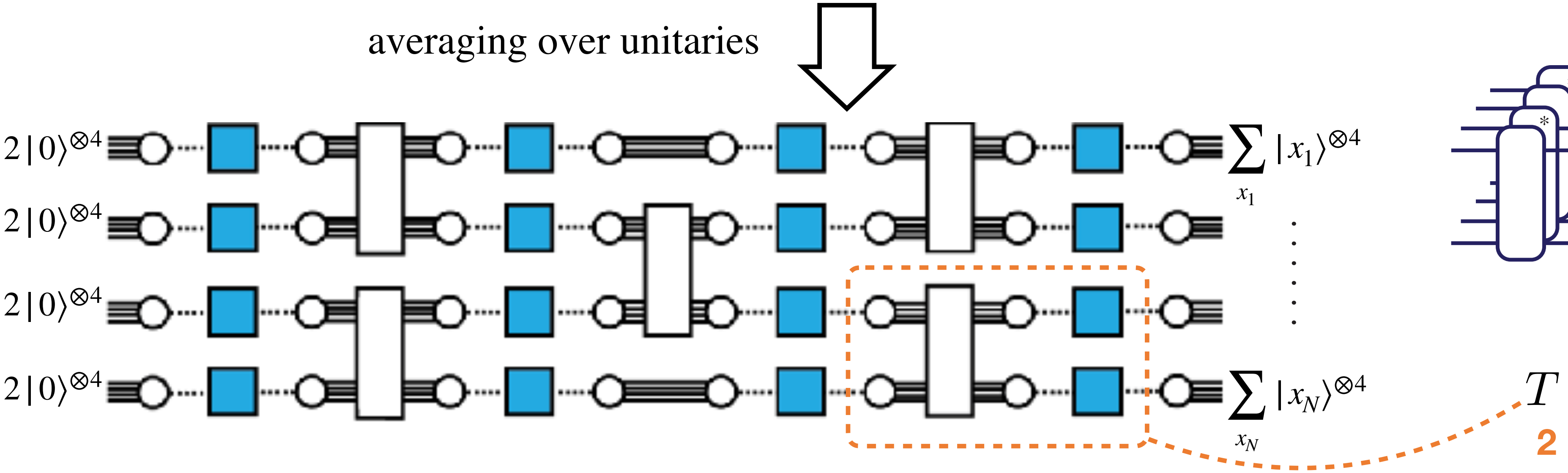
$$\text{XEB} + 1 = 2^N \sum_x p(x)^2 = \sum_x 2^N$$



Remove black box

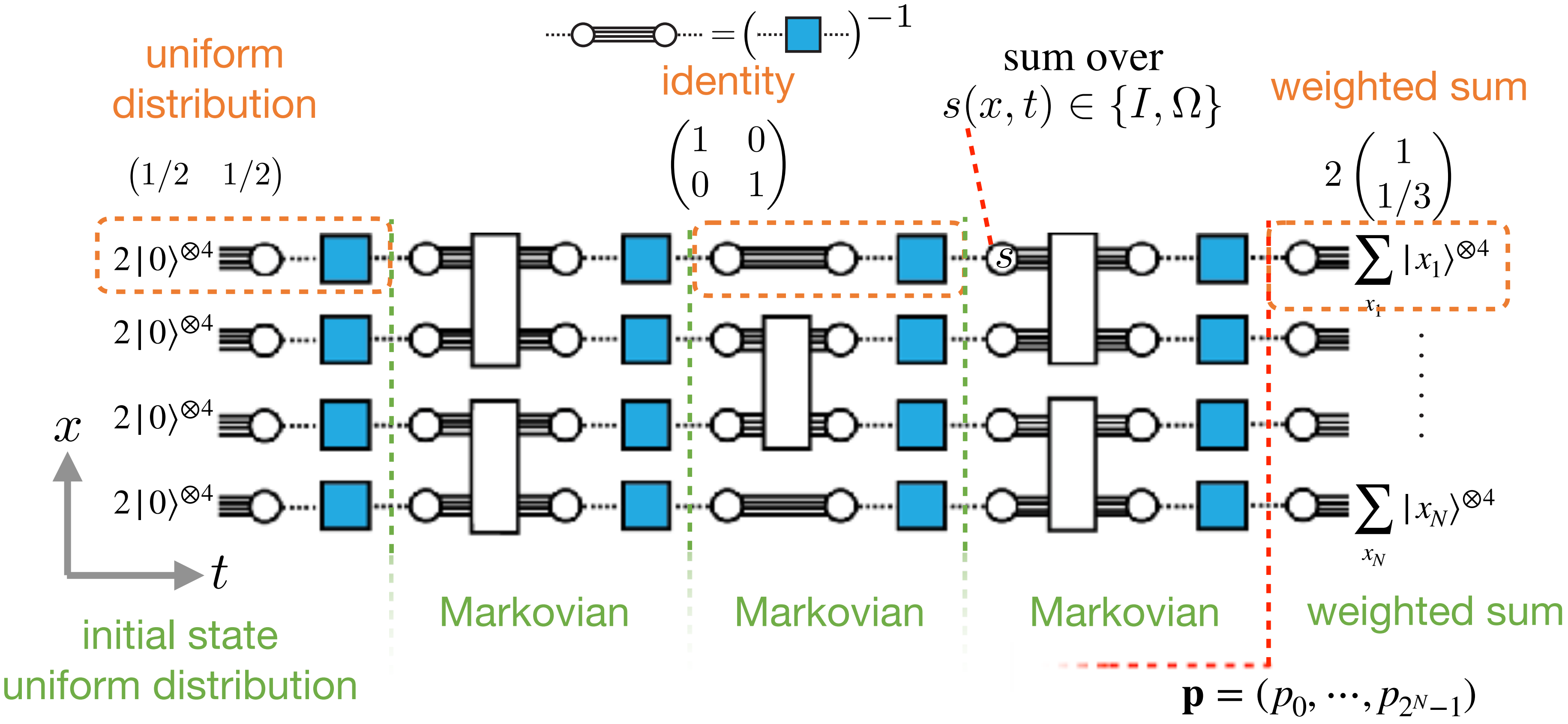






4 by 4 Markovian
transfer matrix:
2 bits to 2 bits (dotted line)

Diffusion-Reaction Model



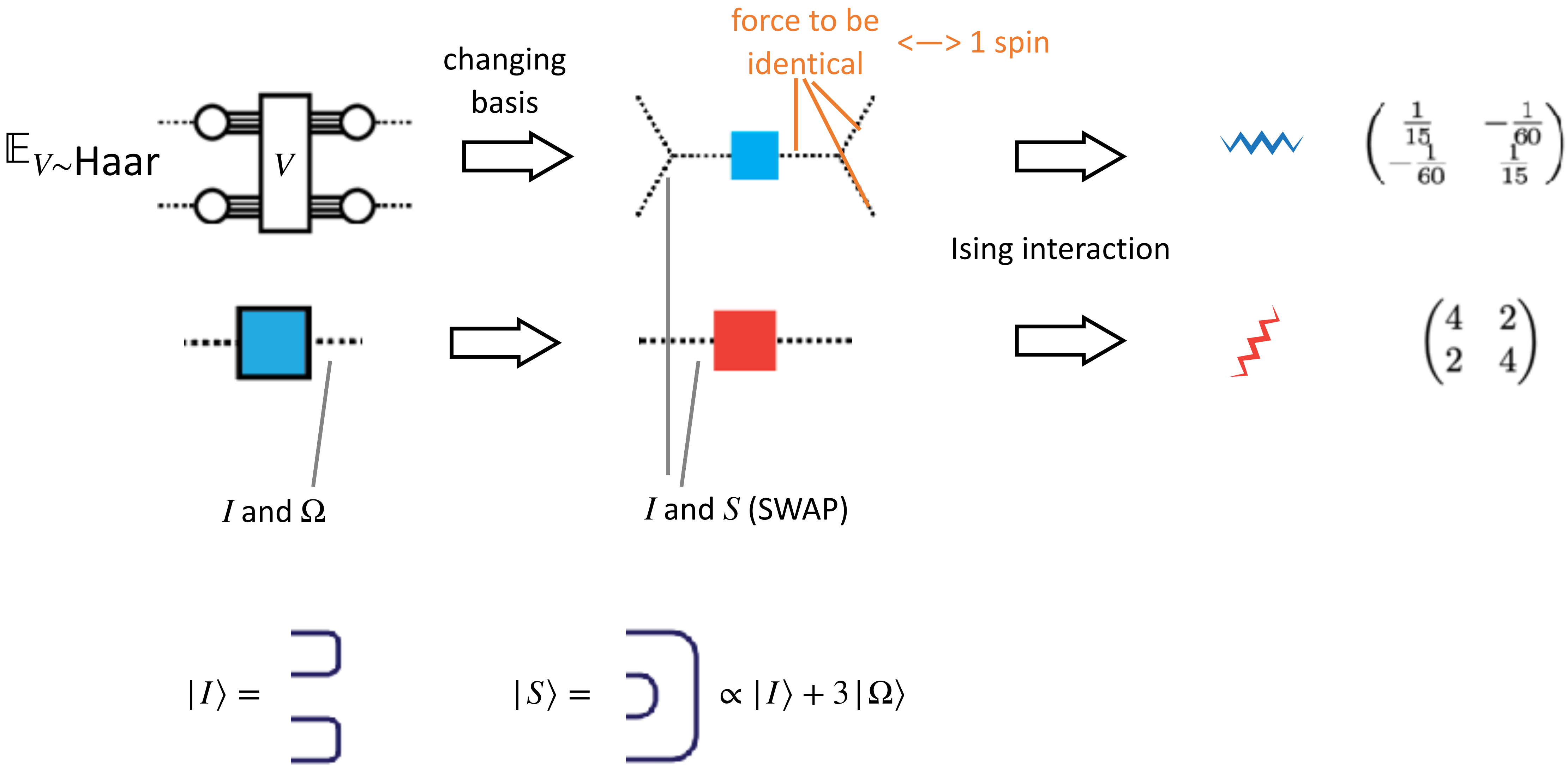
$\mathbf{p} \cdot 2^N \begin{pmatrix}1\\ 1/3\end{pmatrix}^{\otimes N}$ = path integral of $\mathbb{E}_{\text{histories}}\left[2^N \cdot \frac{1}{3^{\#\Omega} \text{ in the last layer}}\right]$

the weight 1/3: XEB measured in Z basis

among 3 Paulis, only information in Z can be detected

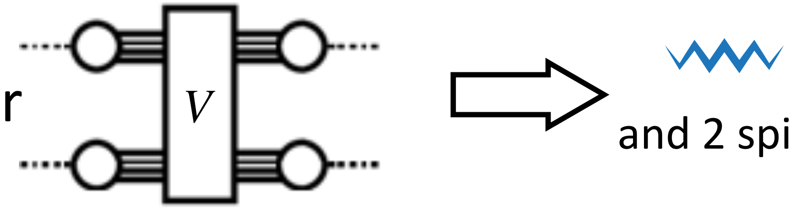
factor 2 for normalization (will see it later)

Mapping XEB to Ising model

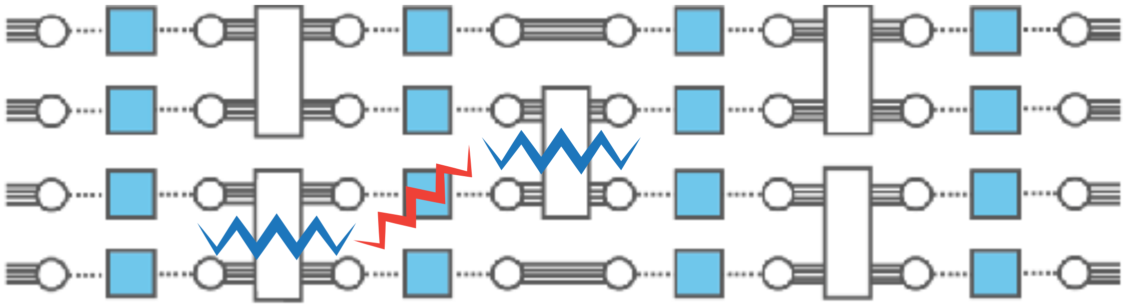


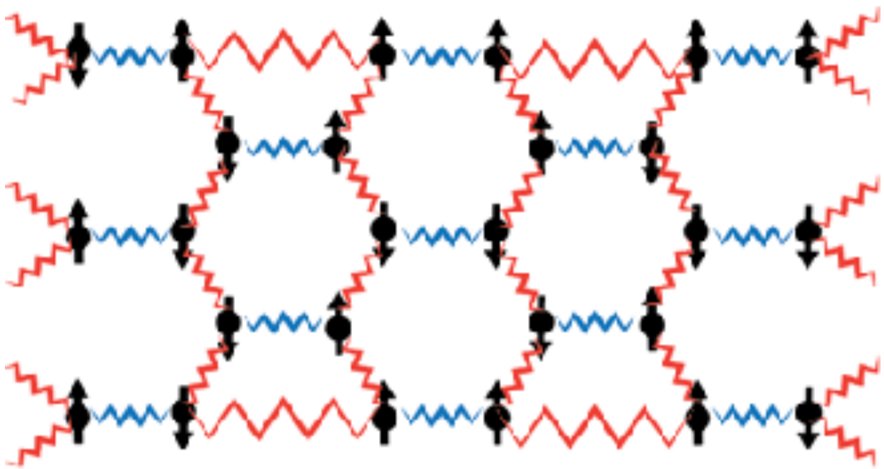
Mapping XEB to Ising model

$\mathbb{E}_{V\sim\text{Haar}}$



and 2 spins



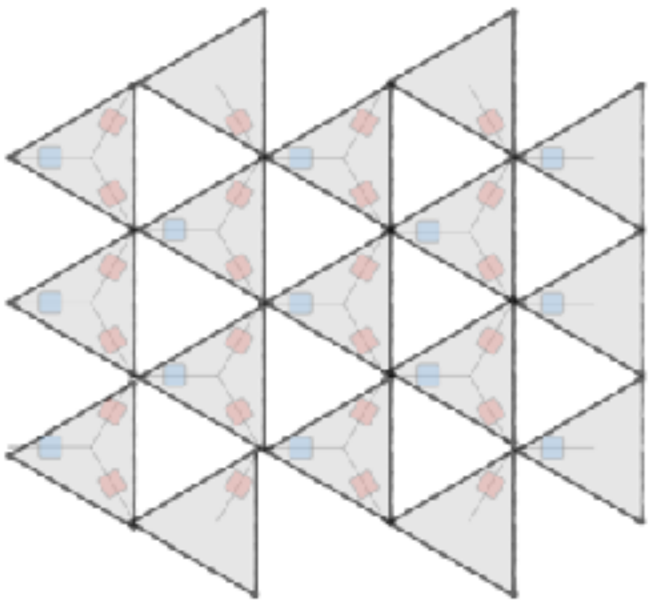


Left and Right boundaries:
equal weight sum of spin up and down

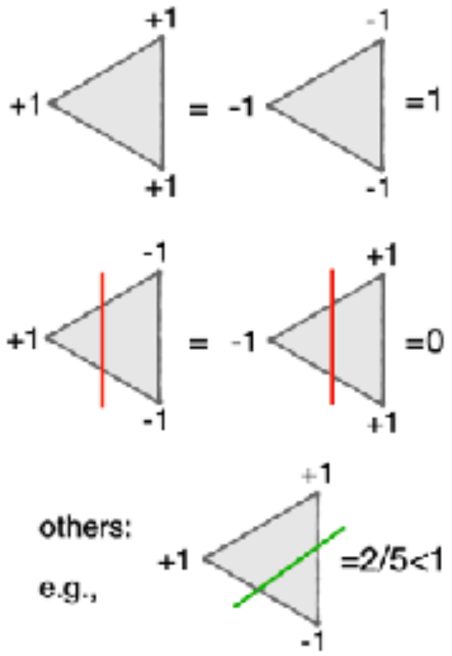
XEB+1=partition function
For deal circuit, global $\mathbb{Z}_2$ Ising symmetry

The Ising symmetry reflects the permutation symmetry in $V \otimes V^* \otimes V \otimes V^*$

Mapping XEB to Ising model

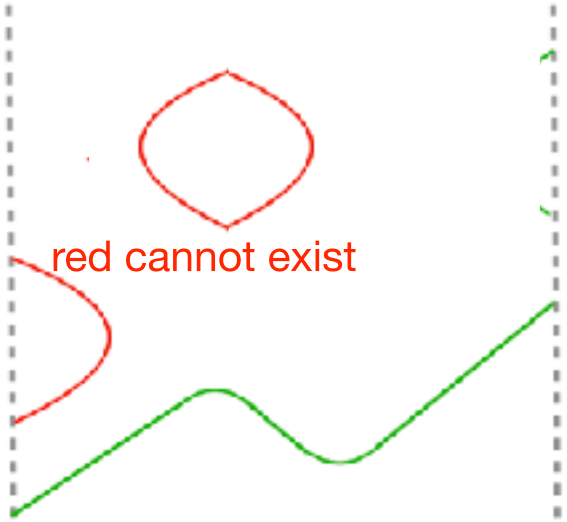


Integrate the spin in the middle of each triangle (star-triangle relation)



red is forbidden

green is allowed



only green domain wall can exist
green domain wall must be **very long**

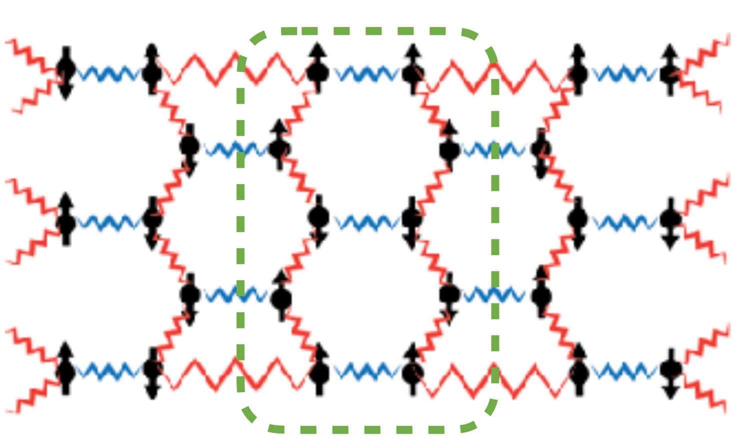
Nicholas Hunter-Jones. arXiv:1905.12053

B.Barak, C.Chou, and X.Gao. arXiv:2005.02421

2 degenerate ground states stable
ferromagnetic phase

Mapping XEB to Ising model

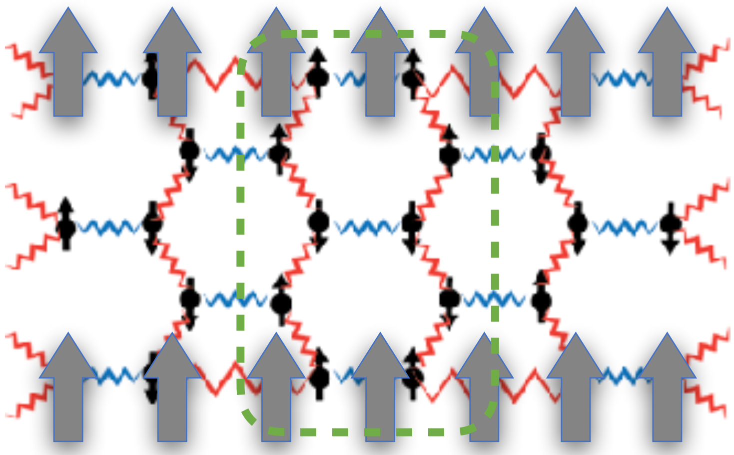
ideal circuit



transfer matrix T

no magnetic field
$\mathbb{Z}_2$ Ising symmetry

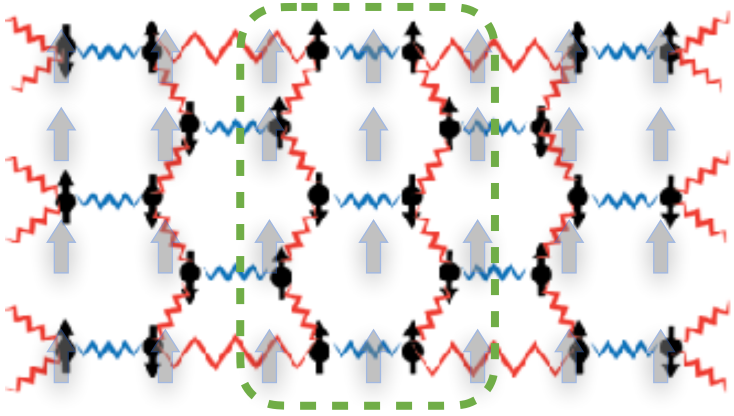
our algorithm



transfer matrix T

strong boundary magnetic field
violates $\mathbb{Z}_2$ Ising symmetry

noisy circuit



transfer matrix T

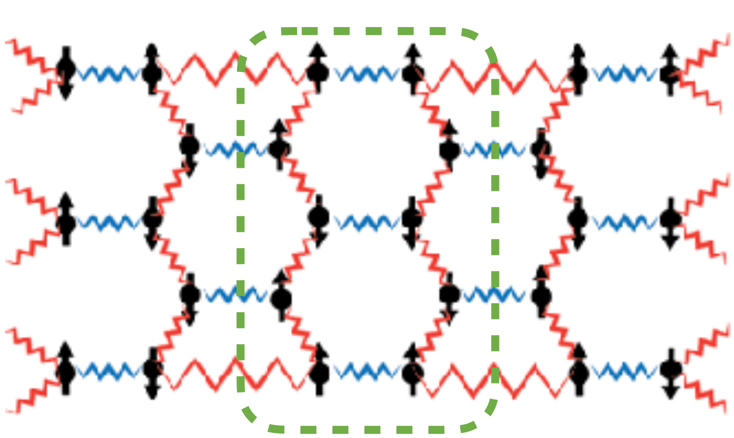
weak bulk magnetic field
violates $\mathbb{Z}_2$ Ising symmetry

eigenvalues of T: $\lambda_0 = 1$, $\lambda_1 = 1 - \Delta$ (spectral gap), $\lambda_3 \geq 0$, $\cdots\cdots$

XEB=$O\left(e^{-\Delta d}\right)$

Mapping XEB to Ising model

ideal circuit



transfer matrix T

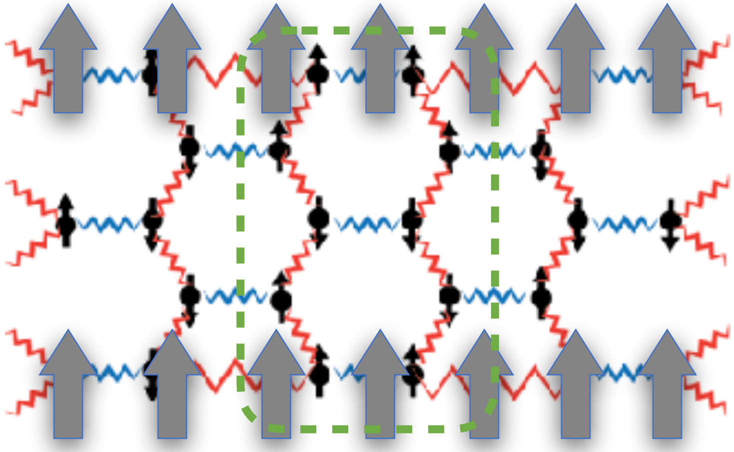
no magnetic field

$\mathbb{Z}_2$ Ising symmetry

$\Delta = 0$

2-fold degeneracy

our algorithm



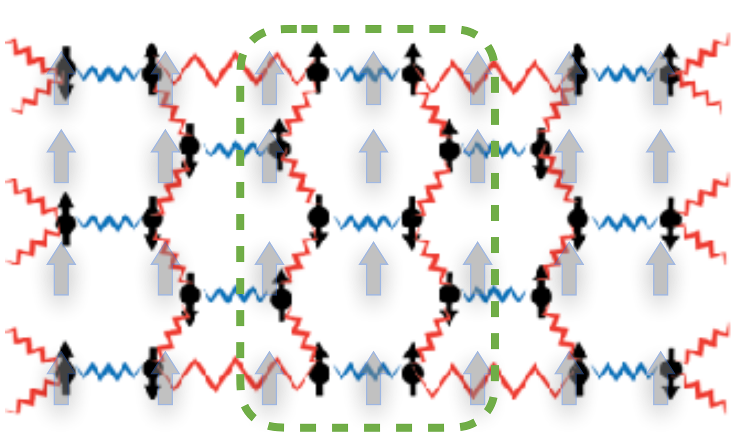
transfer matrix T

strong boundary magnetic field

violates $\mathbb{Z}_2$ Ising symmetry

Δ_1

noisy circuit



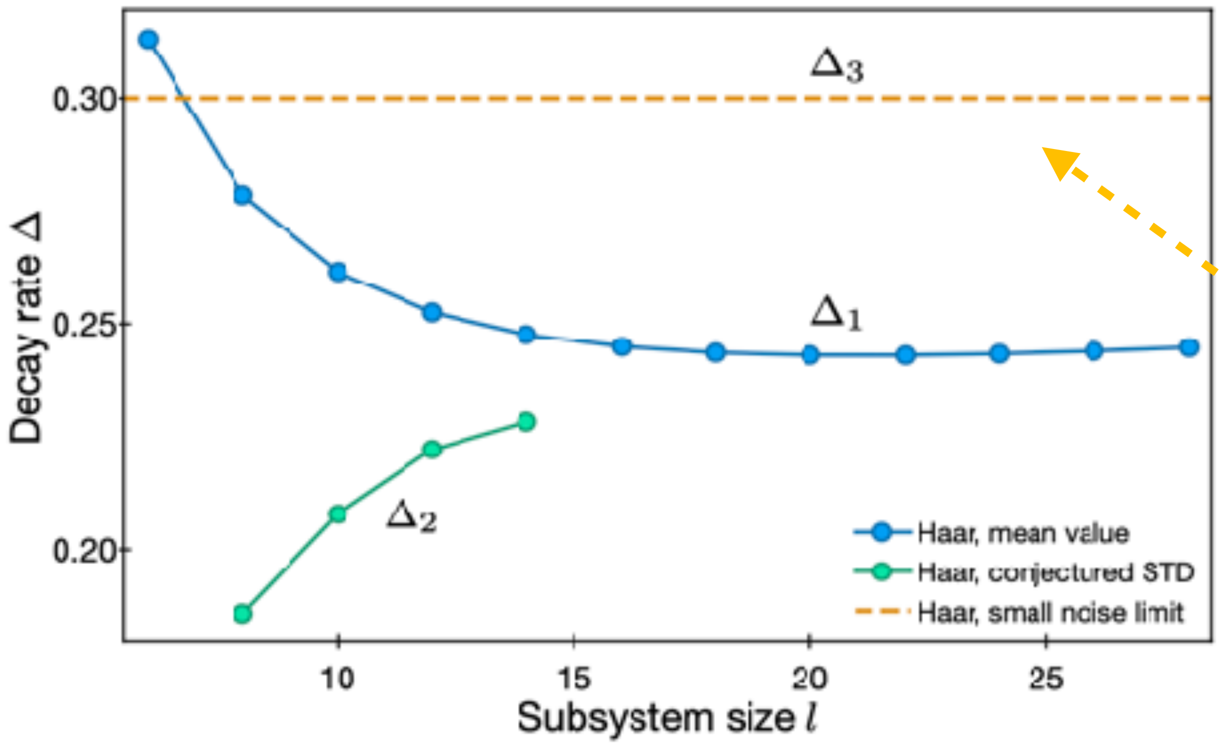
transfer matrix T

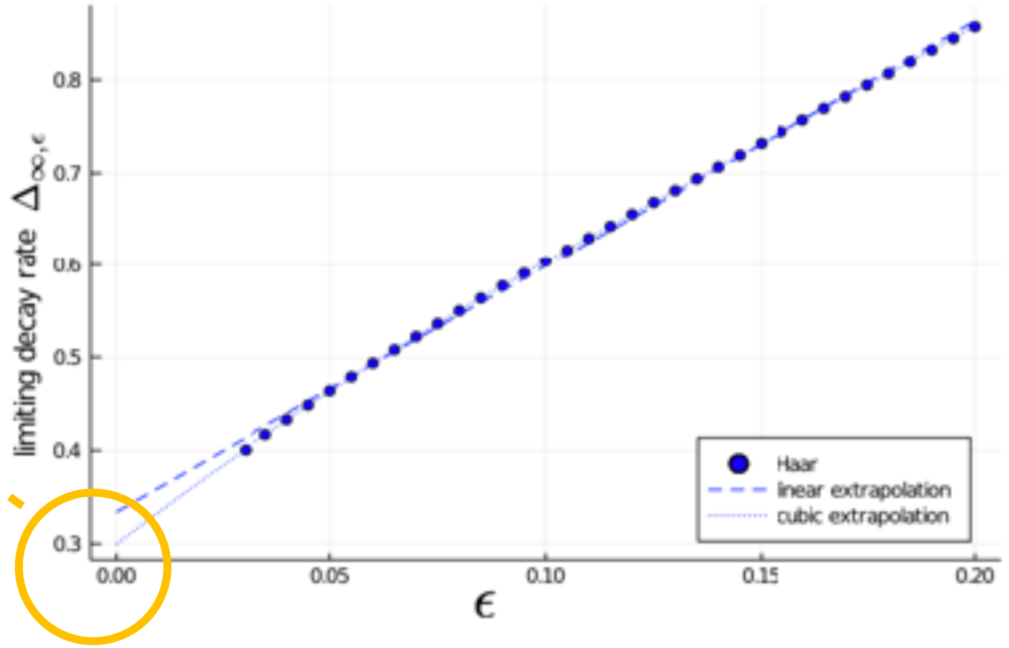
weak bulk magnetic field

violates $\mathbb{Z}_2$ Ising symmetry

$$\Delta_3 = \lim_{\epsilon \to 0} \lim_{N \to \infty} \Delta_{N,\epsilon}$$

Mapping XEB to Ising model





Δ_3 non-vanishing even if $\epsilon \to 0$ in large N case, spontaneous magnetization

decay rate of our algorithm (constant subsystem size)
< decay rate of any noisy circuit in large N case

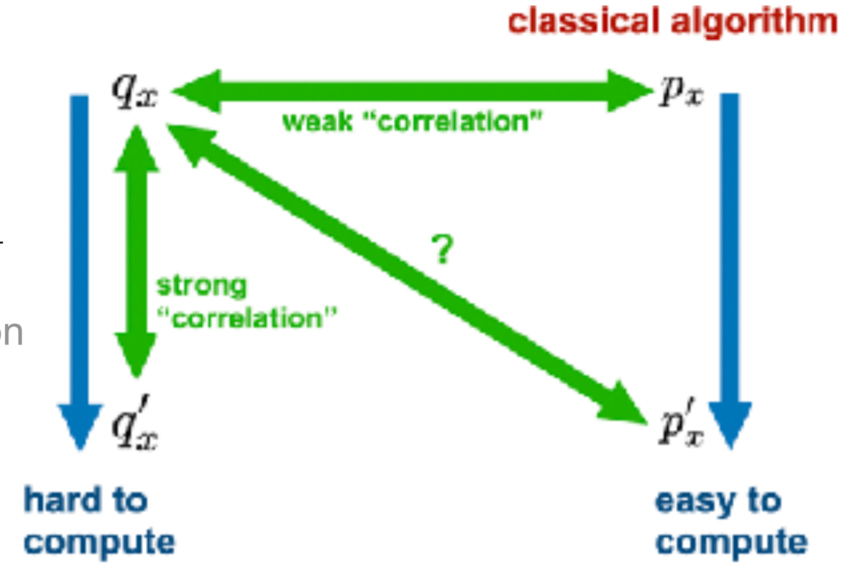
Power-k post-processing

To amplify the average XEB further by heuristic post-processing

rescaling

$$q_x \longrightarrow q'_x = \frac{\sigma(q_x)}{\sum_y \sigma(q_x)}$$

normalization



σ select x with large q_x such that:

$$\text{XEB} = 2^N \sum_x q'_x q_x - 1 \gg 1$$

a natural choice: $\sigma(p_x p_y) = \sigma(p_x)\sigma(p_y)$

solution: $\sigma(p) = p^k$

(when k>1, large p will be amplified)

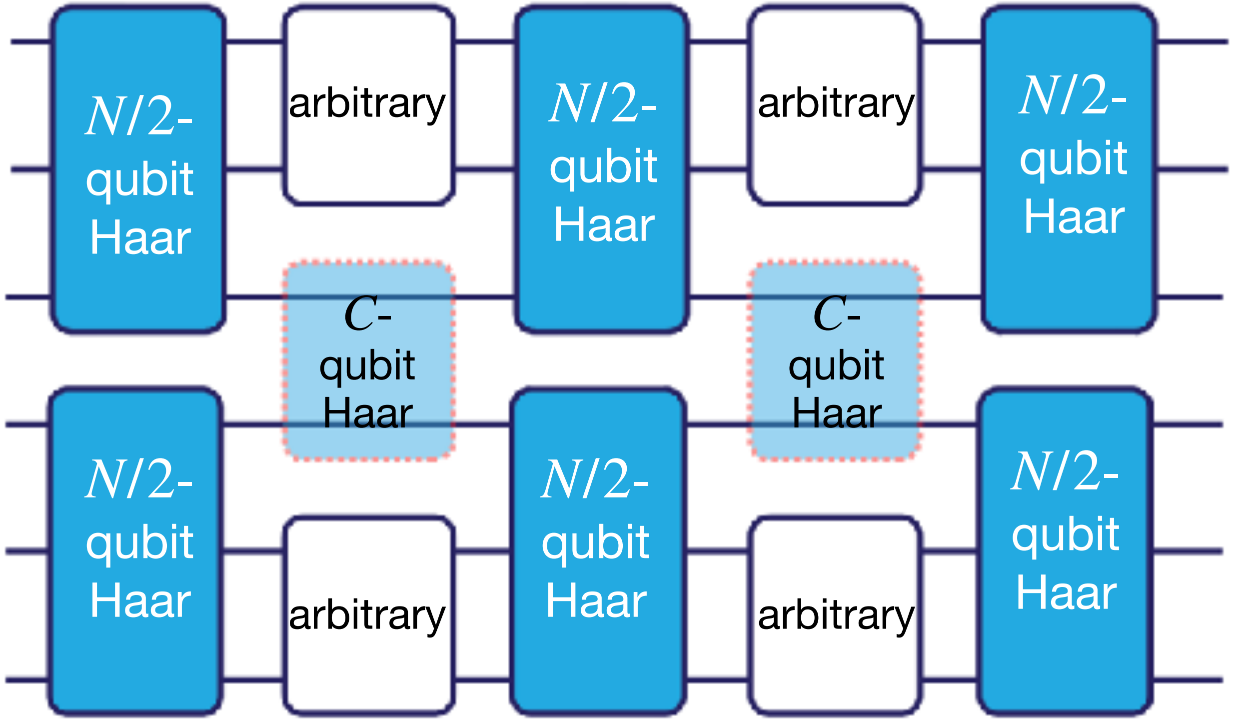
$$p_x^{(k)} = \frac{p_x^k}{\sum_y p_y^k} = \frac{p_{x_L}^k}{\sum_{y_L} p_{y_L}^k} \cdot \frac{p_{x_R}^k}{\sum_{y_R} p_{y_R}^k}$$

still easy to simulate:

1. subsystems decouple
2. power-k & summation are easy to parallel in GPU

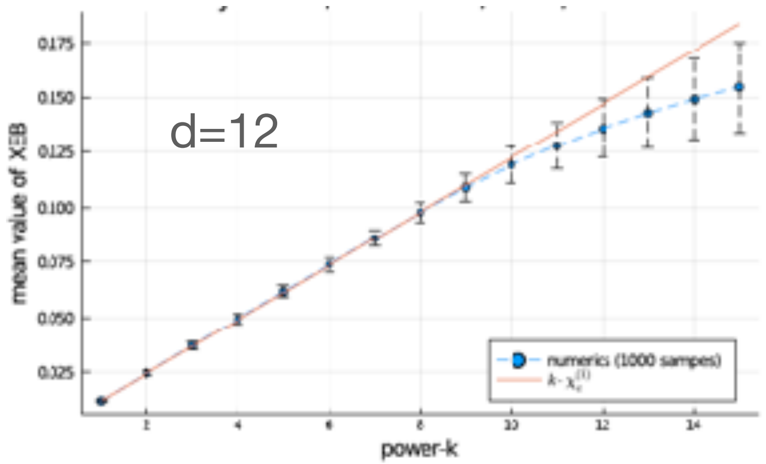
Power-k post-processing

analytics (toy model)

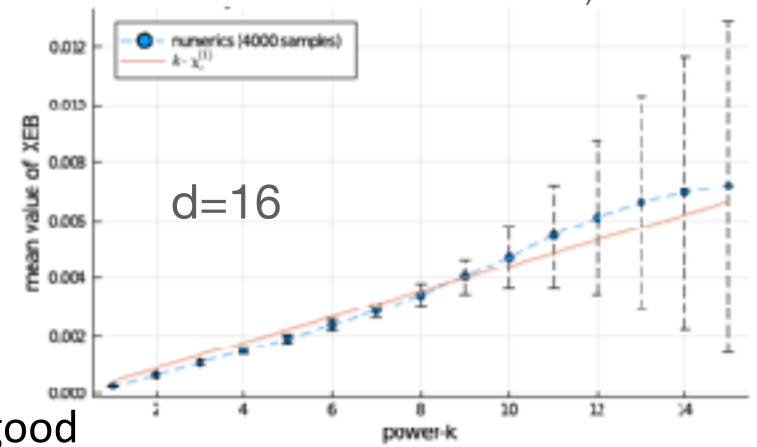


$E_U[\text{XEB}_k] \approx k \cdot E_U[\text{XEB}_1]$ when $k < N$

numerics (Sycamore, N=53)

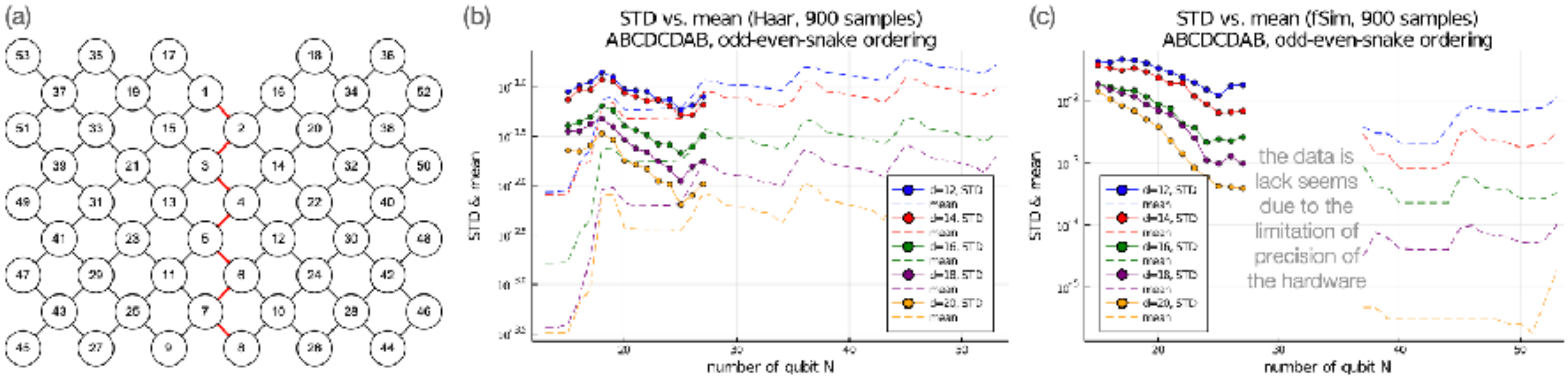


(PS: error bar is not due to fluctuation of different U)

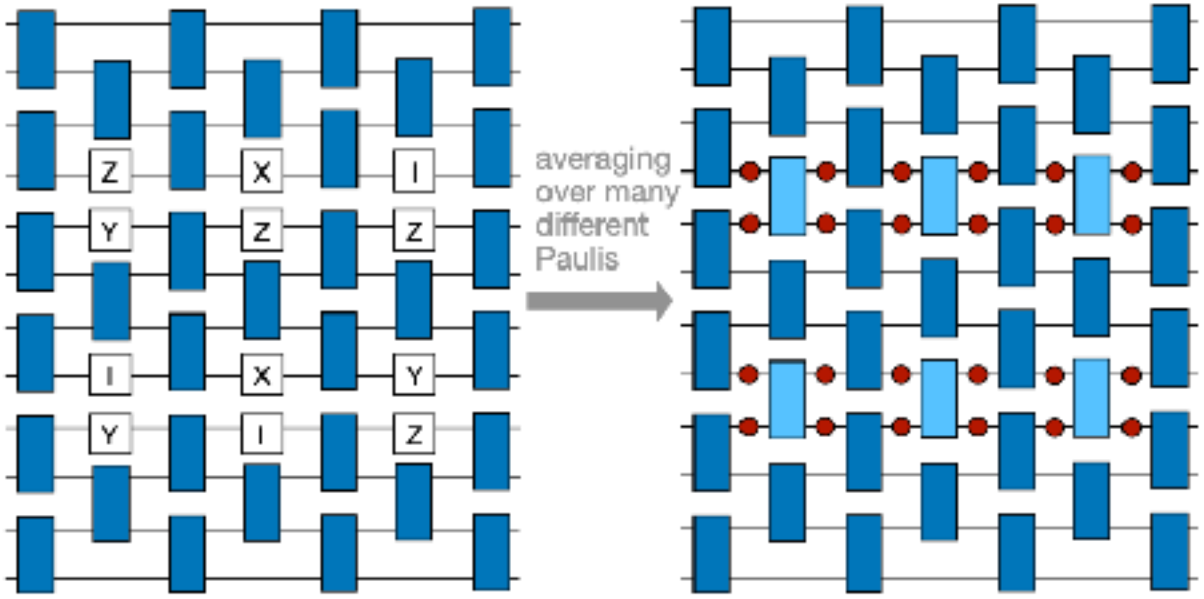


$k \approx 10$ is good

STD of our algorithm

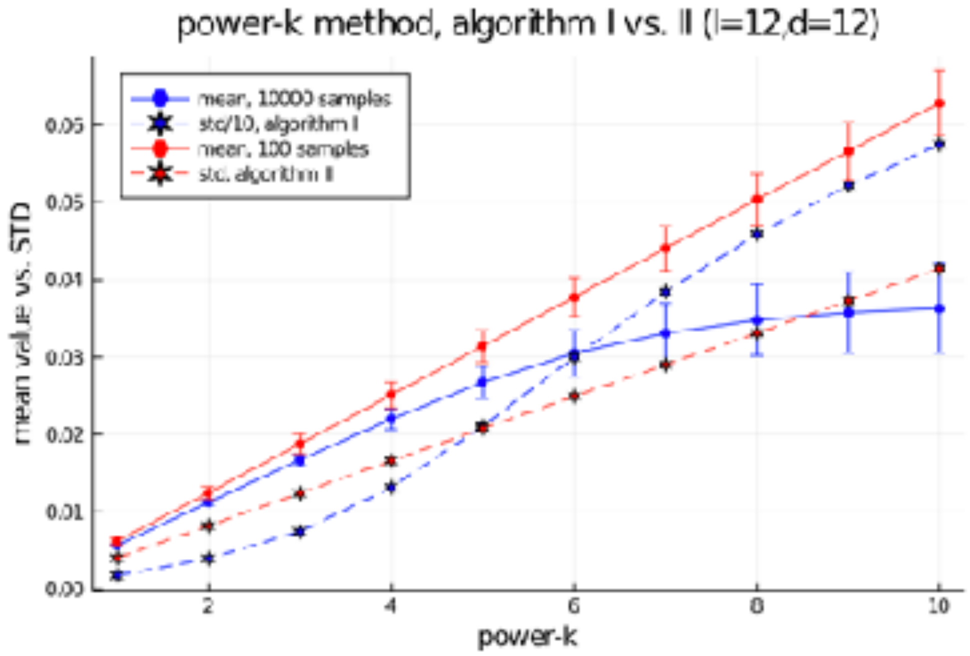


STD of our algorithm

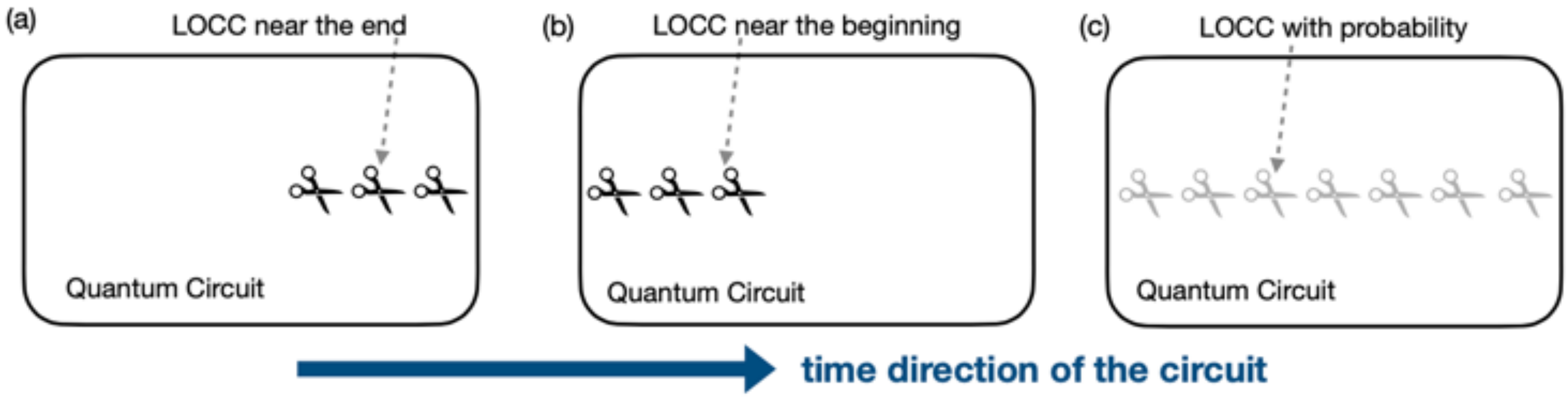


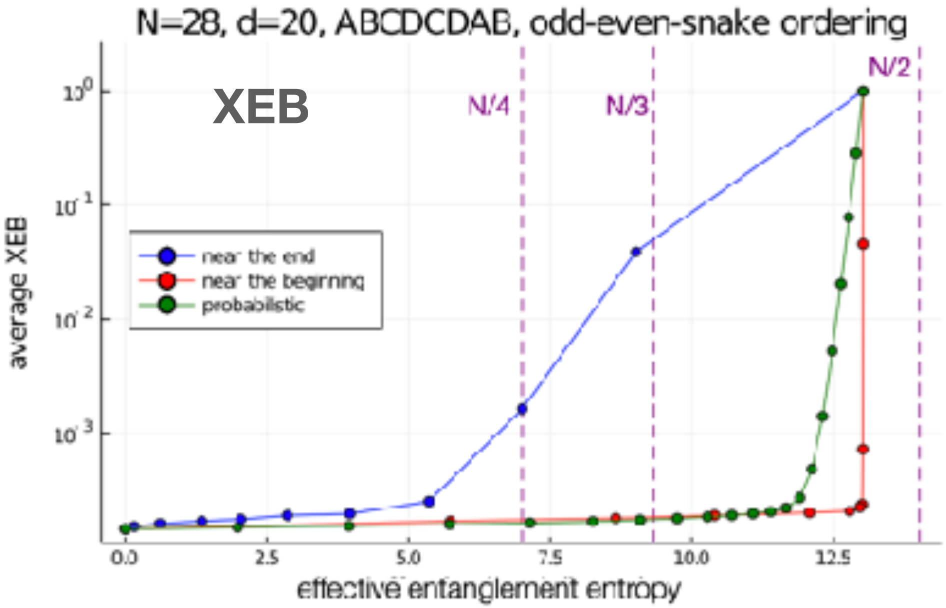
algorithm I
(the original)

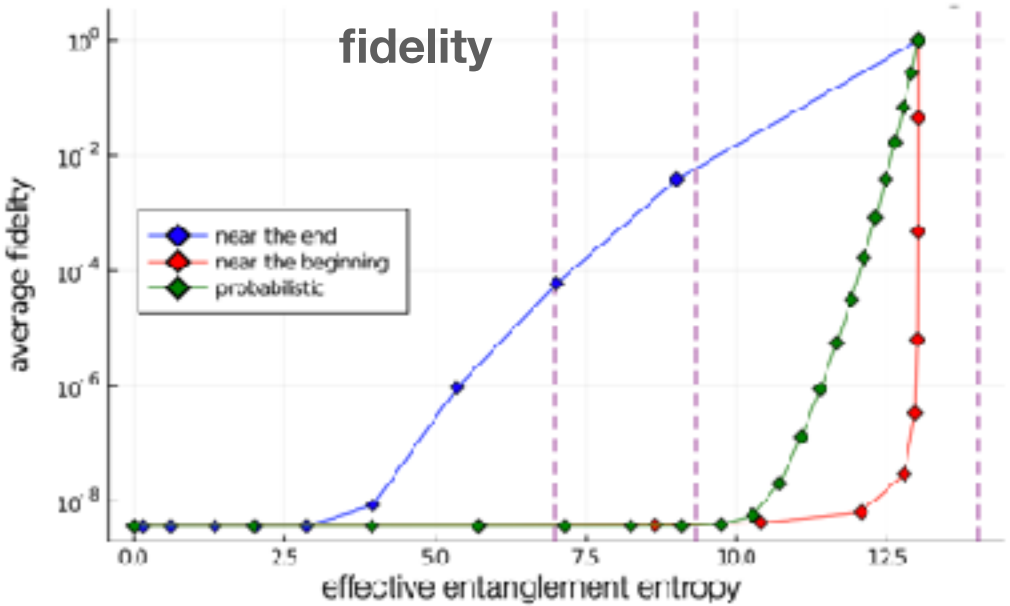
algorithm II



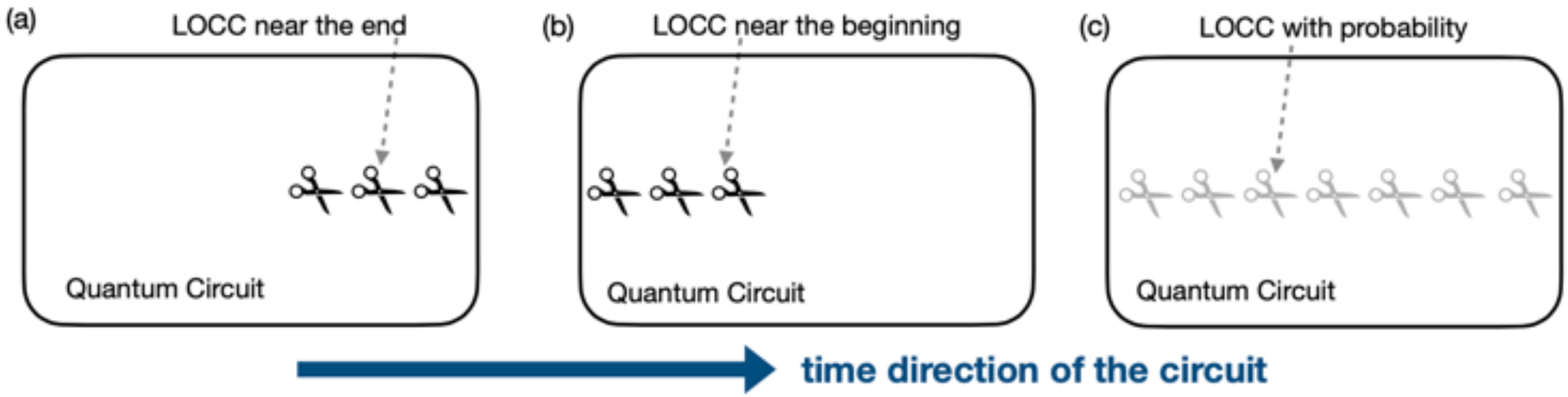
Adding entanglement



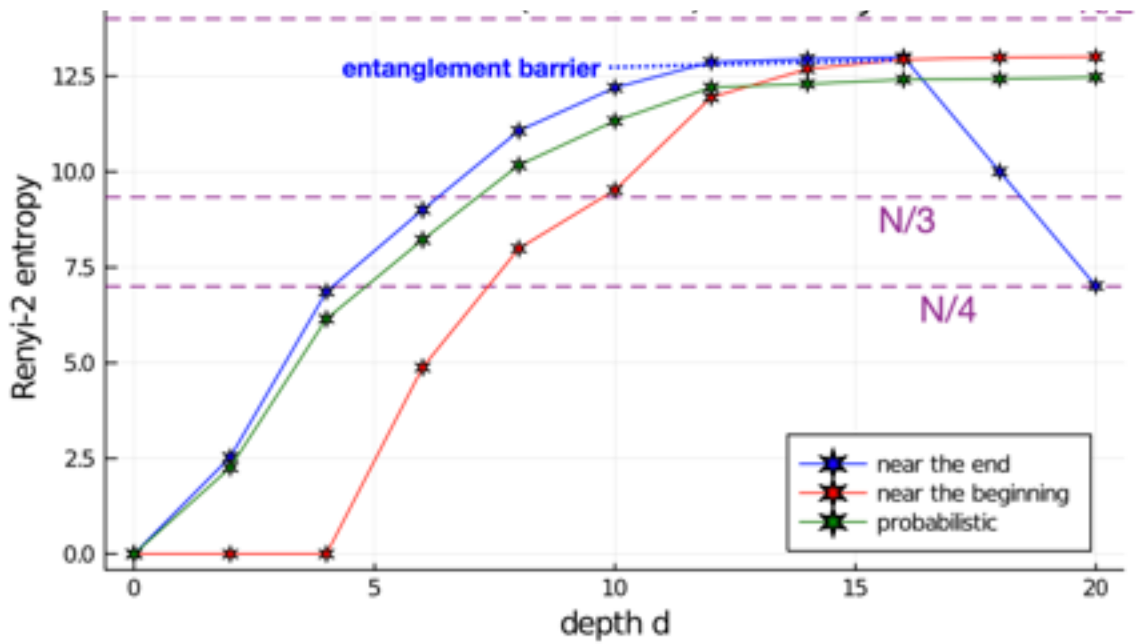




Adding entanglement



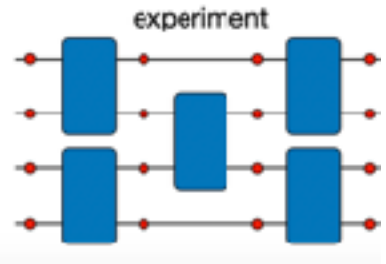
entanglement growth history



Generalizing XEB

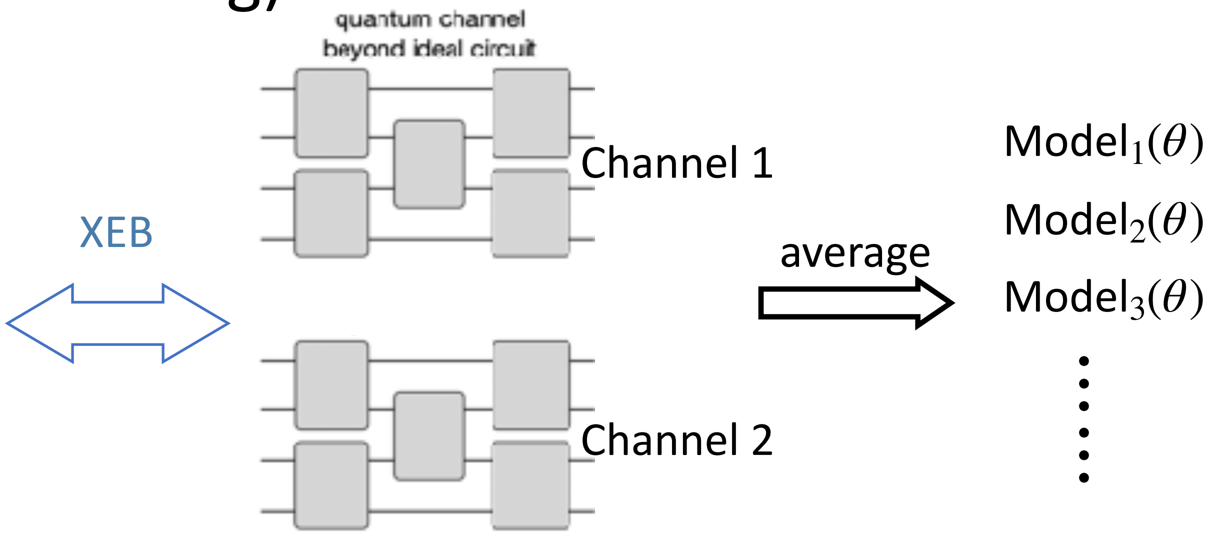
Extension of XEB —>

other statistical classical models (many available tools: Monte Carlo, Bayesian network in machine learning)



θ properties of circuits:

strength and type of noise,
scrambling speed, etc.



Learn θ from
different models