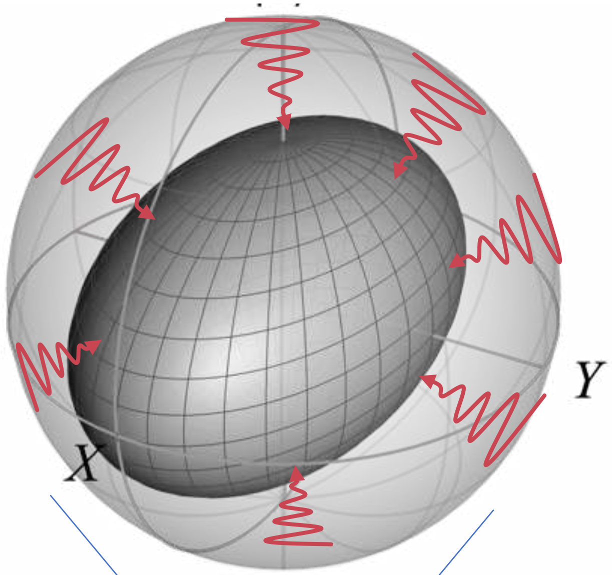
Near-Term Quantum Algorithms Summer School

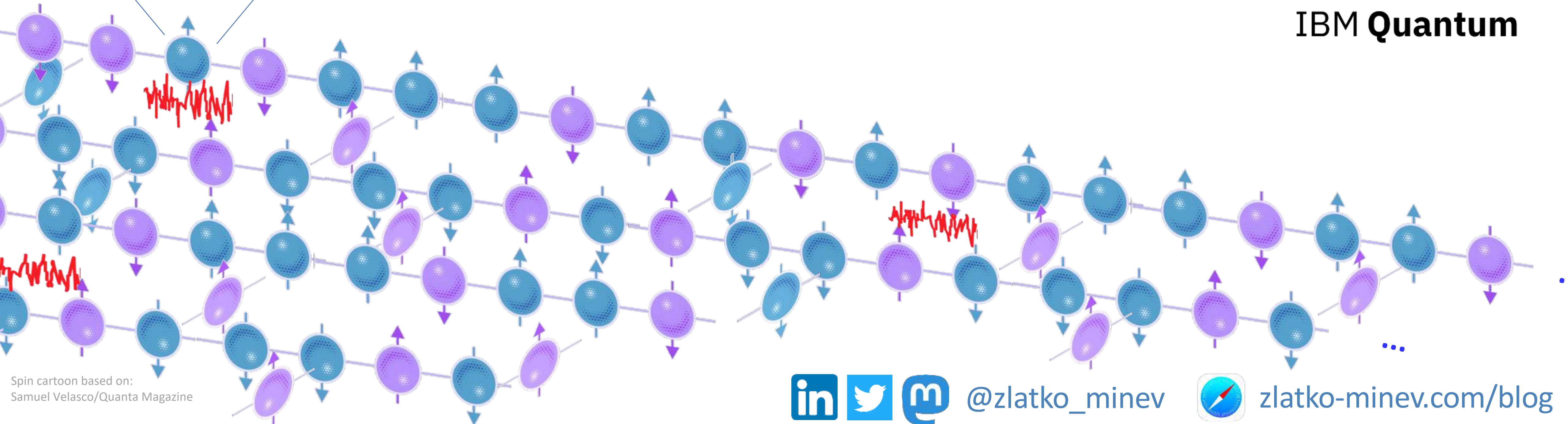
Introduction to Noise and Quantum Error Mitigation for Quantum Computers

Lecture 2

Zlatko K. Minev

IBM **Quantum**





Spin cartoon based on:
Samuel Velasco/Quanta Magazine

@zlatko_minev zlatko-minev.com/blog

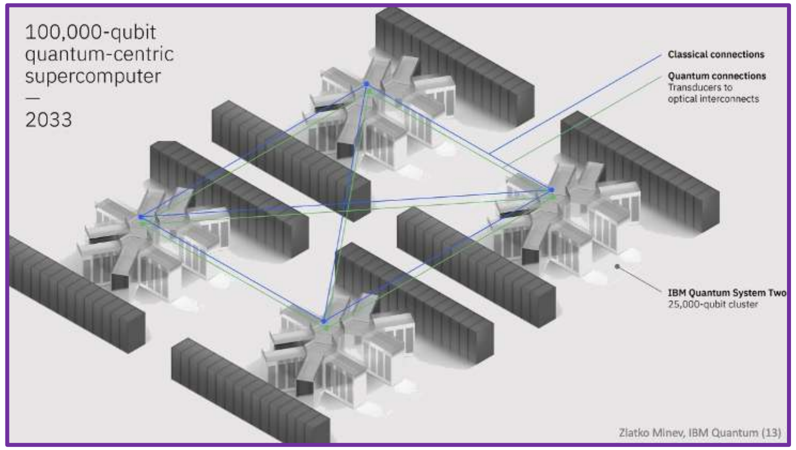
What is one thing you learned in Lecture 1?

Image: Freepik

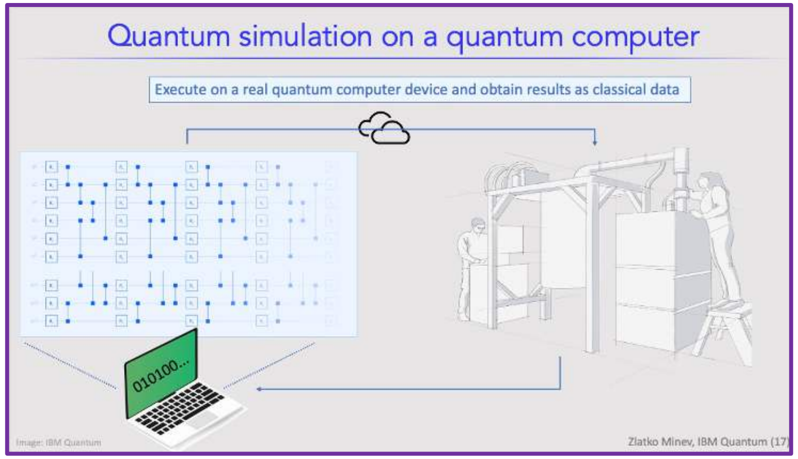
Review of Lecture 1

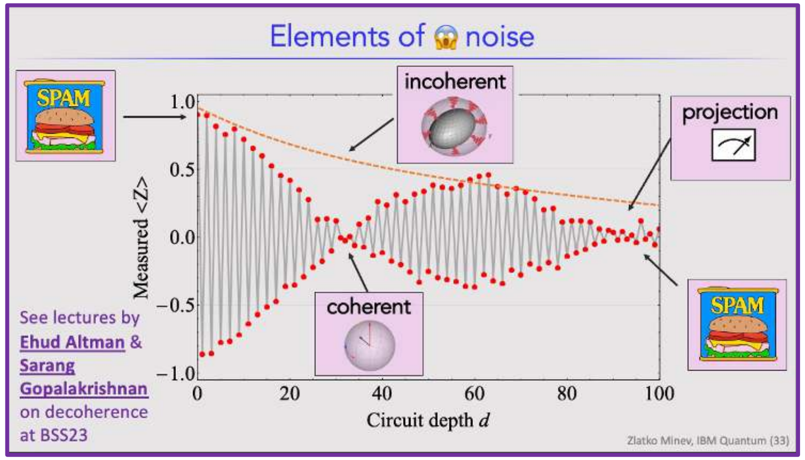
Image: Freepik

Quantum computers

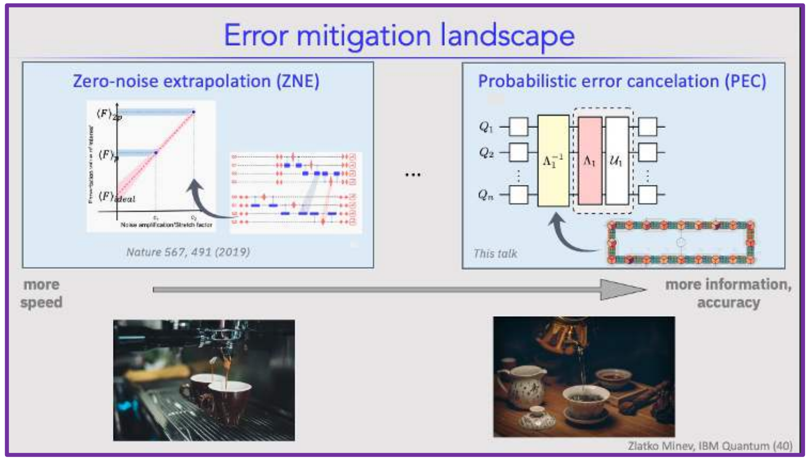


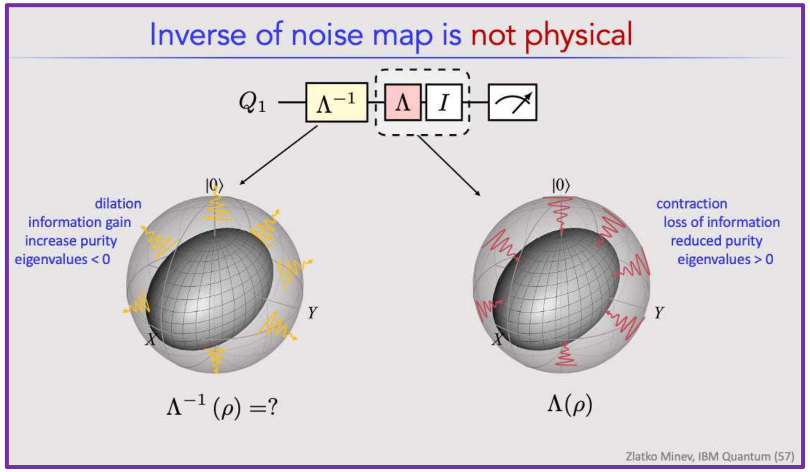
Biggest Problem: Noise





Mitigation





Quantum Error Mitigation

Lecture 2

Mitigation fundamentals

Probabilistic error cancelation (PEC)
- Summarize one qubit example
  - Analogy to random walks
- Error bars & confidence

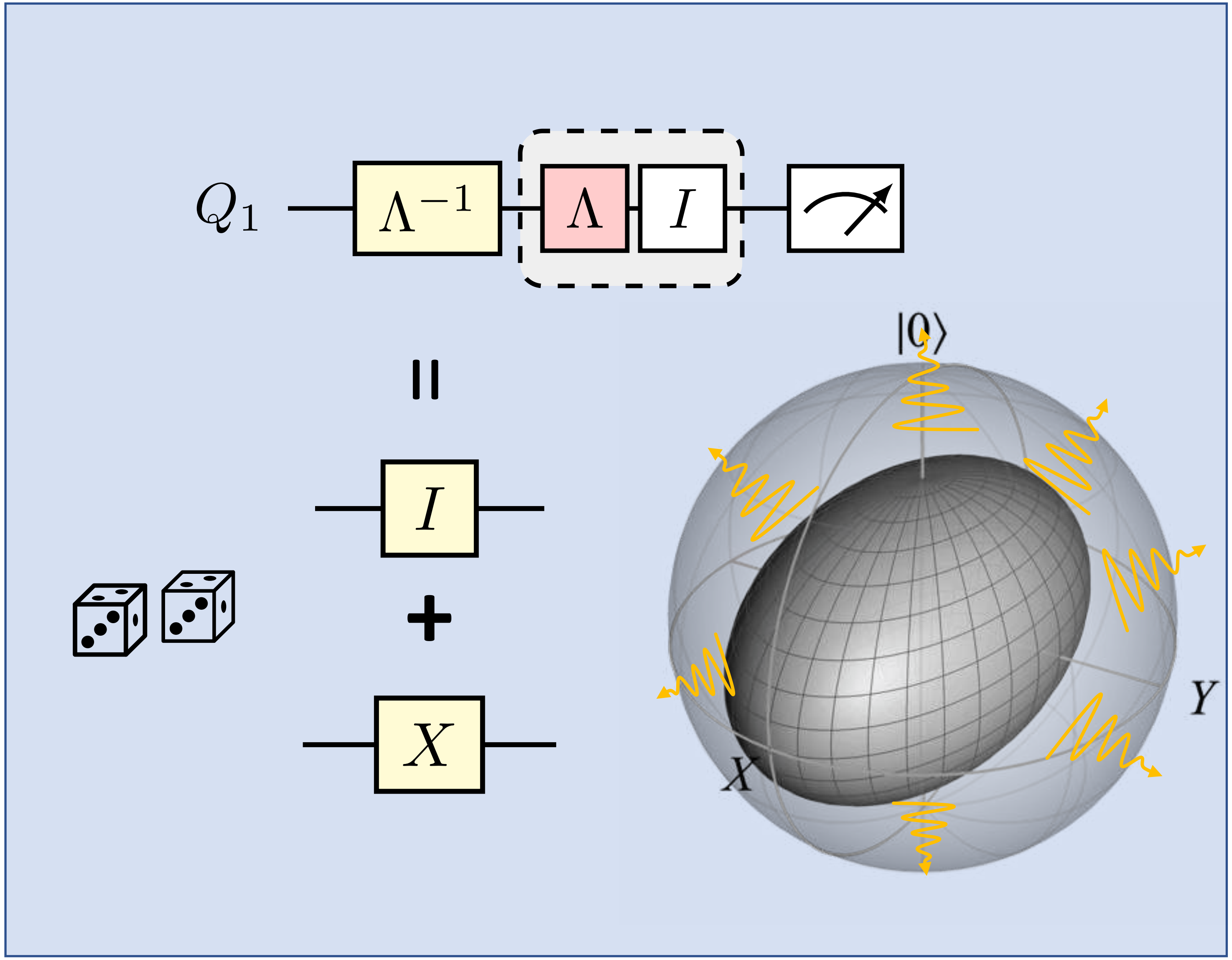
Appendix:
- Generalized
- Show unbiased estimator

Learning quantum noise
- Challenge
- Overcoming
- Sparse model

Putting it together
- Experiments – Ising model
- Consequences for the big picture

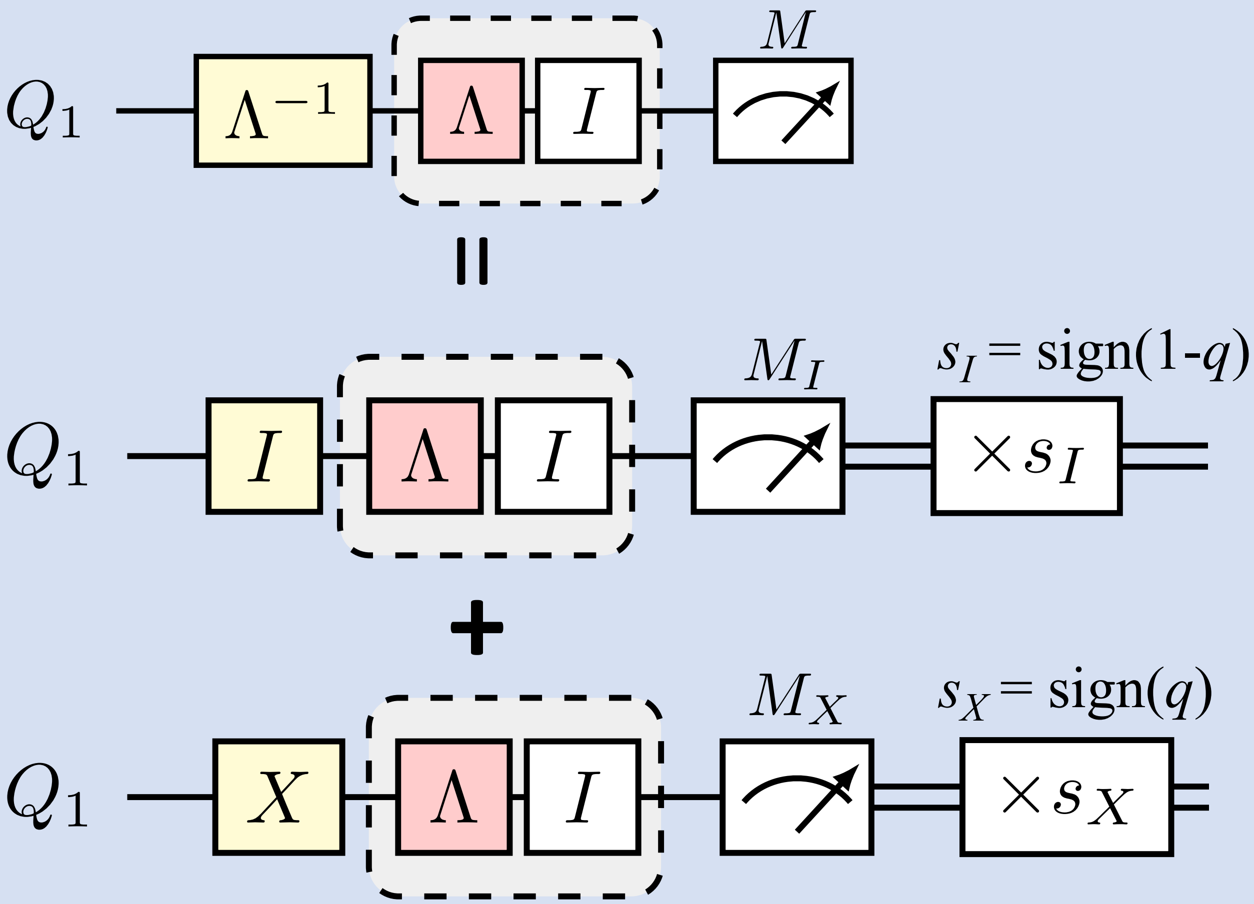
Hardware progress

Canceling noise with noise



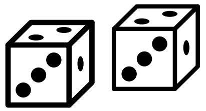
How to implement?

random circuits
(quantum trajectories)



probability

$P_I = |1-q|/\gamma$

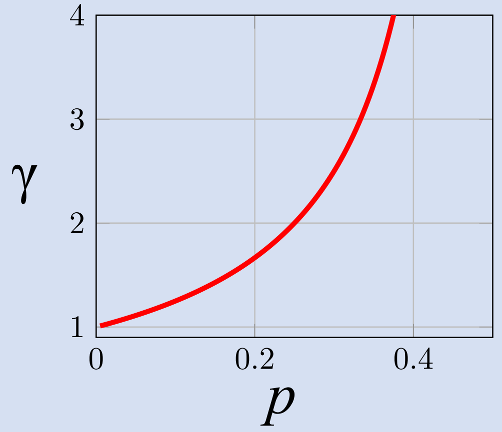


$P_X = |q|/\gamma$

$P_I + P_X = 1$

sampling overhead

$\gamma = |1-q| + |q|$



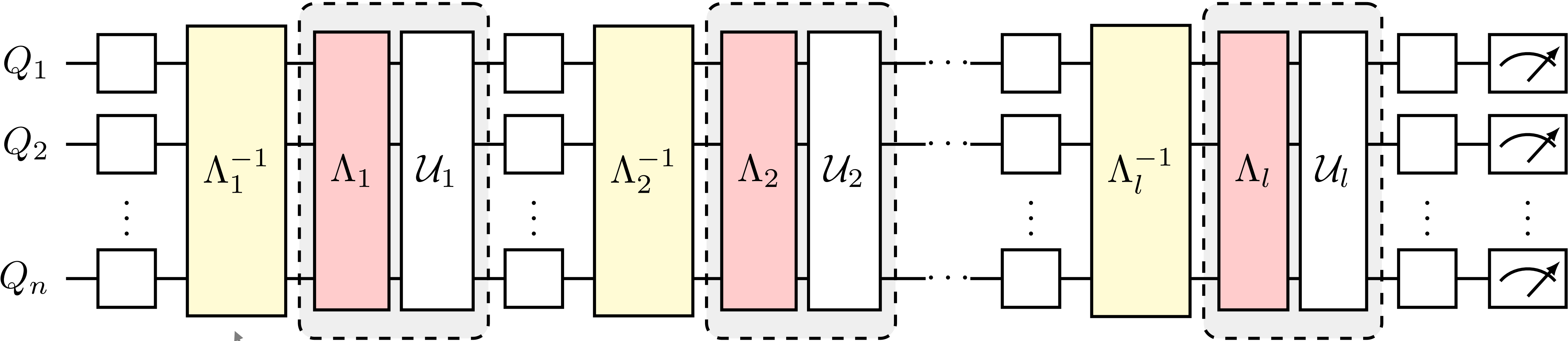
mitigated expectation

$$\langle M \rangle = \gamma\left(s_I P_I M_I + s_X P_X M_X\right)$$

Gain: Bias-free estimate!

Cost: Variance

More qubits, gates, and channels



inverse operation of noise channel

Quasi probability

Generalizing: Raveling trajectories with quasiprobabilities

Channel we *want* to implement

CPTP operation we *can* implement

$$\mathcal{C}(\cdot) = \sum_i a_i \mathcal{F}_i(\cdot)$$

Real coefficients, turn into quasi-probability

Putting the following techniques all on the same footing

Technique	**Channel $\mathbb{C}$**
Prob. error cancelation (PEC)	noise inverse
Cricut cutting (knitting) of gates	non-local gate
Cricut cutting of wires	large unitary
Classical sim. algorithms (QP)	unitary

Discussions with David Sutter

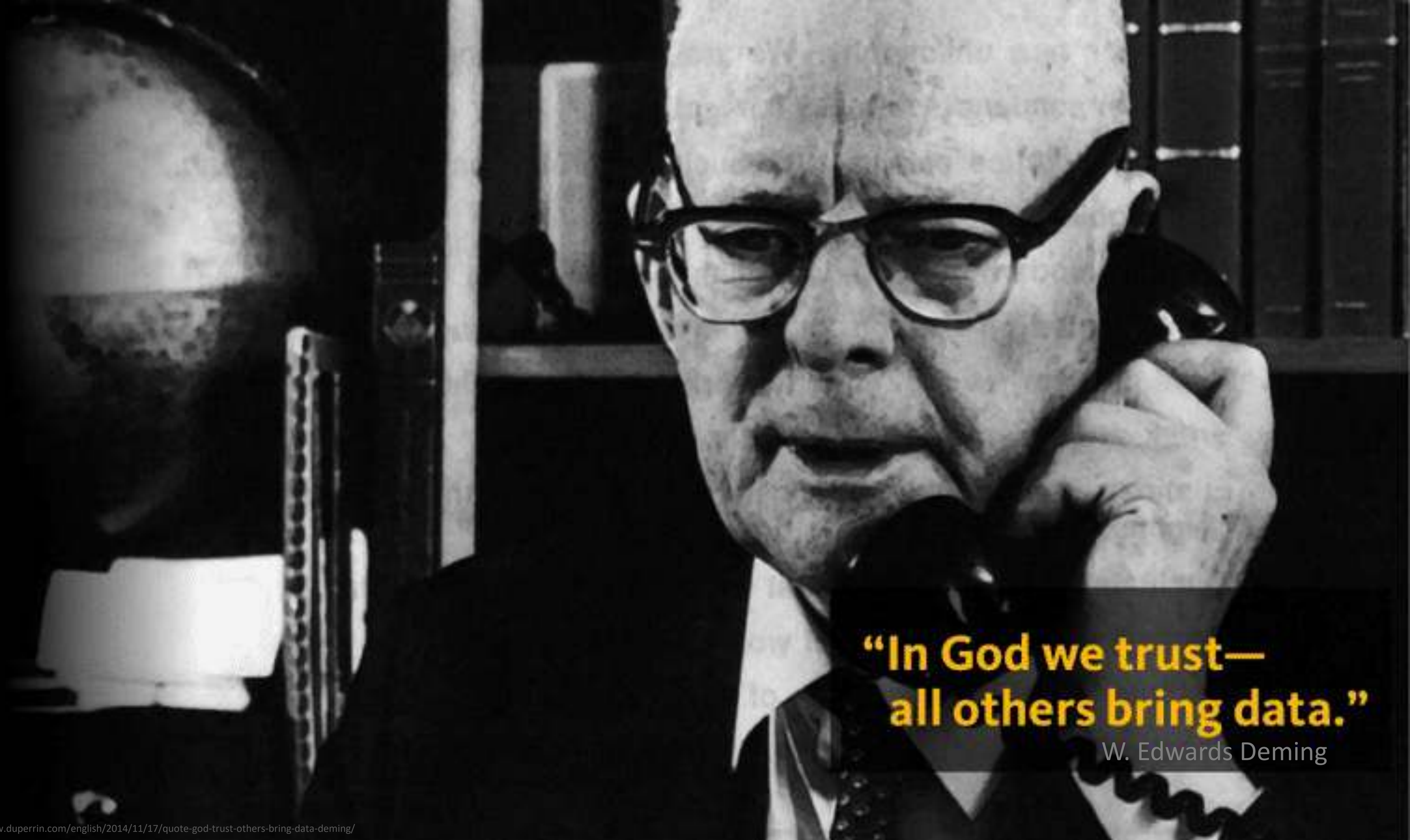
Deep dive:
Probabilistic error cancellation (PEC)
To learn and cancel quantum noise

Got Slides?

SCAN ME

Paper: Nature Physics (2023)

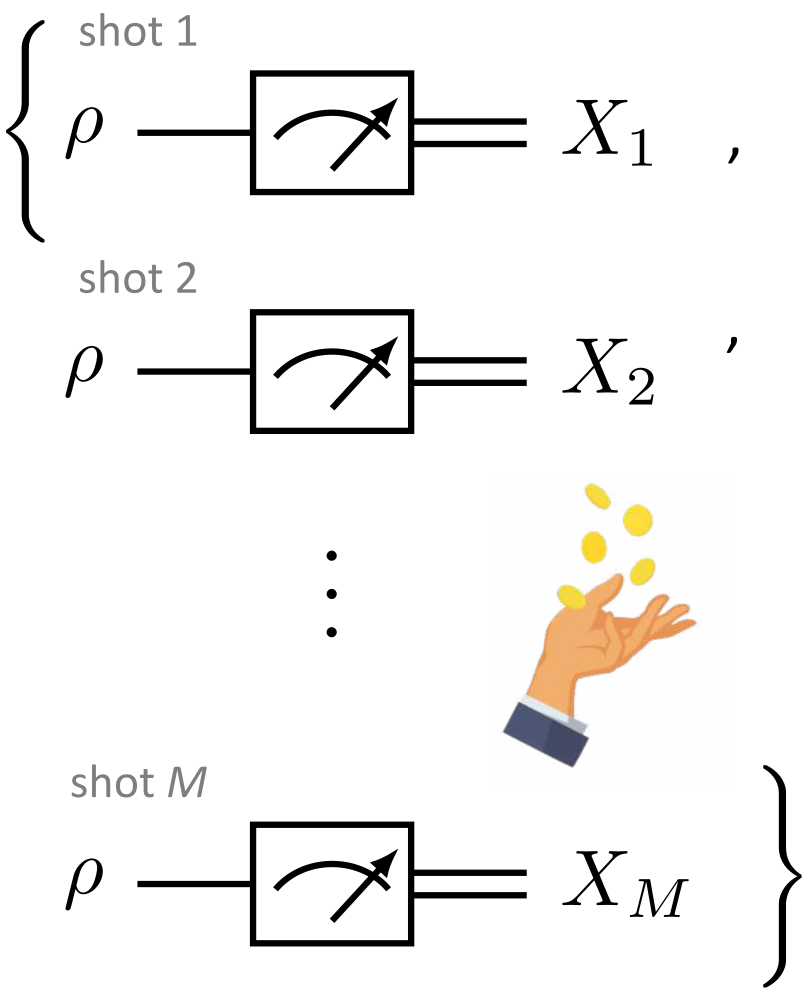
Ewout van den Berg, Zlatko K. Minev, Abhinav Kandala, Kristan Temme



See https://www.duperrin.com/english/2014/11/17/quote-god-trust-others-bring-data-deming/

Hand Image by iconicbestiary on Freepik

Ideal single qubit measurement with *M* shots



M shots with IID distribution

M outcomes: independent and identically distributed (iid) random variables

$$X_1, X_2, \ldots, X_M \in \Sigma$$

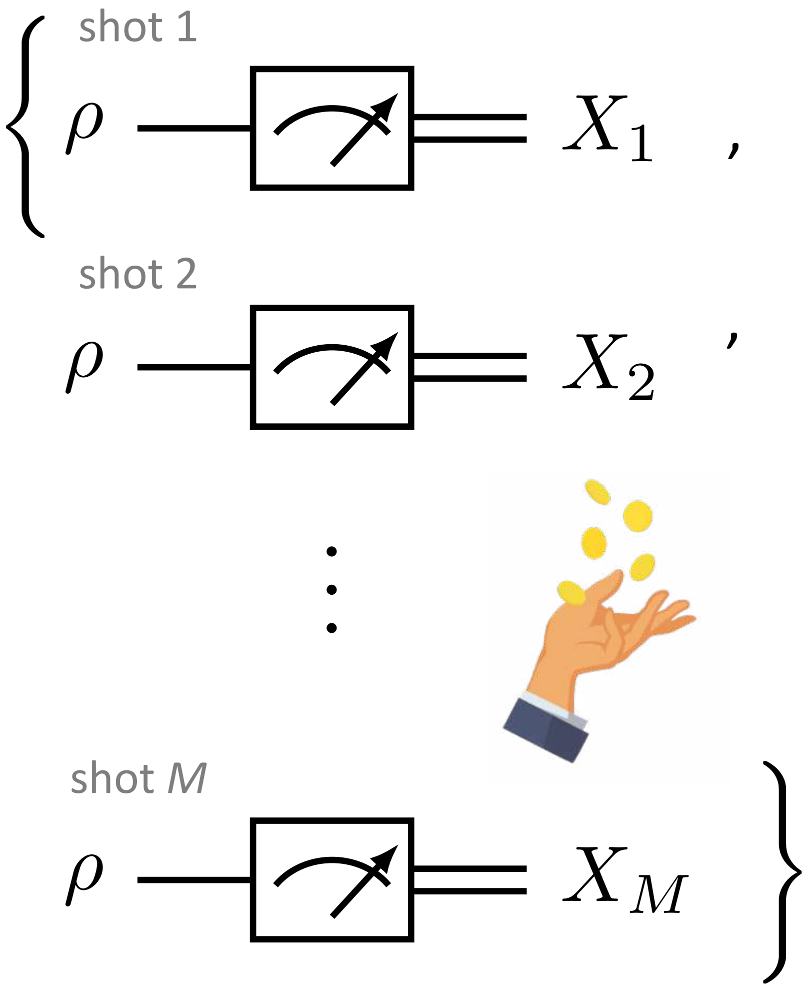
$$X_1, X_2, \ldots, X_M \sim \Pr\left[X = x\right] = \langle \hat{\mu}\left(x\right) \rangle$$

Empirical mean random variable (sample mean statistic)

$$S = \frac{1}{M} \sum_{m=1}^{M} X_m$$

Find the expectation value and variance of the empirical mean

Empirical mean: an unbiased estimator



$$S = \frac{1}{M} \sum_{m=1}^{M} X_m$$

Find the expectation value and variance of the empirical mean

$$\mathrm{E}[S] = \mathrm{E}\left[\frac{1}{M} \sum_{m=1}^{M} X_m \right]$$

$$\mathrm{E}[X_m] = \mathrm{E}[X] = p \quad \forall m \in \{1, \ldots, M\}$$

$$\mathrm{E}[aX_m + bX_n] = a\mathrm{E}[X_m] + b\mathrm{E}[X_n]$$

$$\forall m, n, \quad a, b \in \mathbb{C}$$

$$= \frac{1}{M} \sum_{m=1}^{M} \mathrm{E}[X]$$

linear functional

$$= \mathrm{E}[X]$$

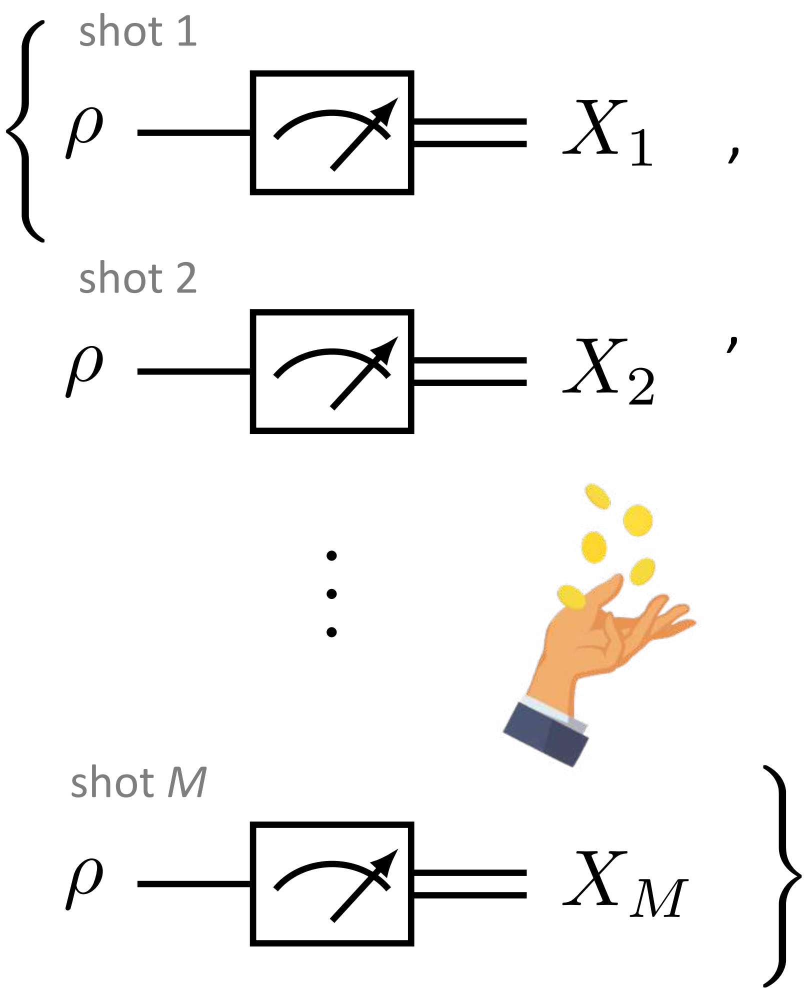
expectation value of a single shot

$$\boxed{\begin{aligned} &= \left\langle \hat{M} \right\rangle \\ &= p \end{aligned}}$$

relation to quantum operator we derived earlier (unbiased estimator)

relation to probability to measure 1

How noisy is our estimate of the empirical mean?



$$S = \frac{1}{M}\sum_{m=1}^{M} X_m$$

$$\mathbb{V}[S] = \mathbb{V}\left[\frac{1}{M}\sum_{m=1}^{M} X_m\right]$$

$$= \frac{1}{M^2}\sum_{m=1}^{M} \mathbb{V}[X]$$

$$\boxed{= \frac{1}{M}\mathbb{V}[X] = \frac{p(1-p)}{M}}$$

$$\boxed{\sigma_S = \sqrt{\mathbb{V}[S]} = \sqrt{\frac{\mathbb{V}[X]}{M}} = \sqrt{\frac{p(1-p)}{M}}}$$

Find the expectation value and variance of the empirical mean

Use key identity for variance

$$\mathbb{V}[aX_m + bX_n] = a^2\mathbb{V}[X_m] + b^2\mathbb{V}[X_n]$$

(you can derive this from the definition)

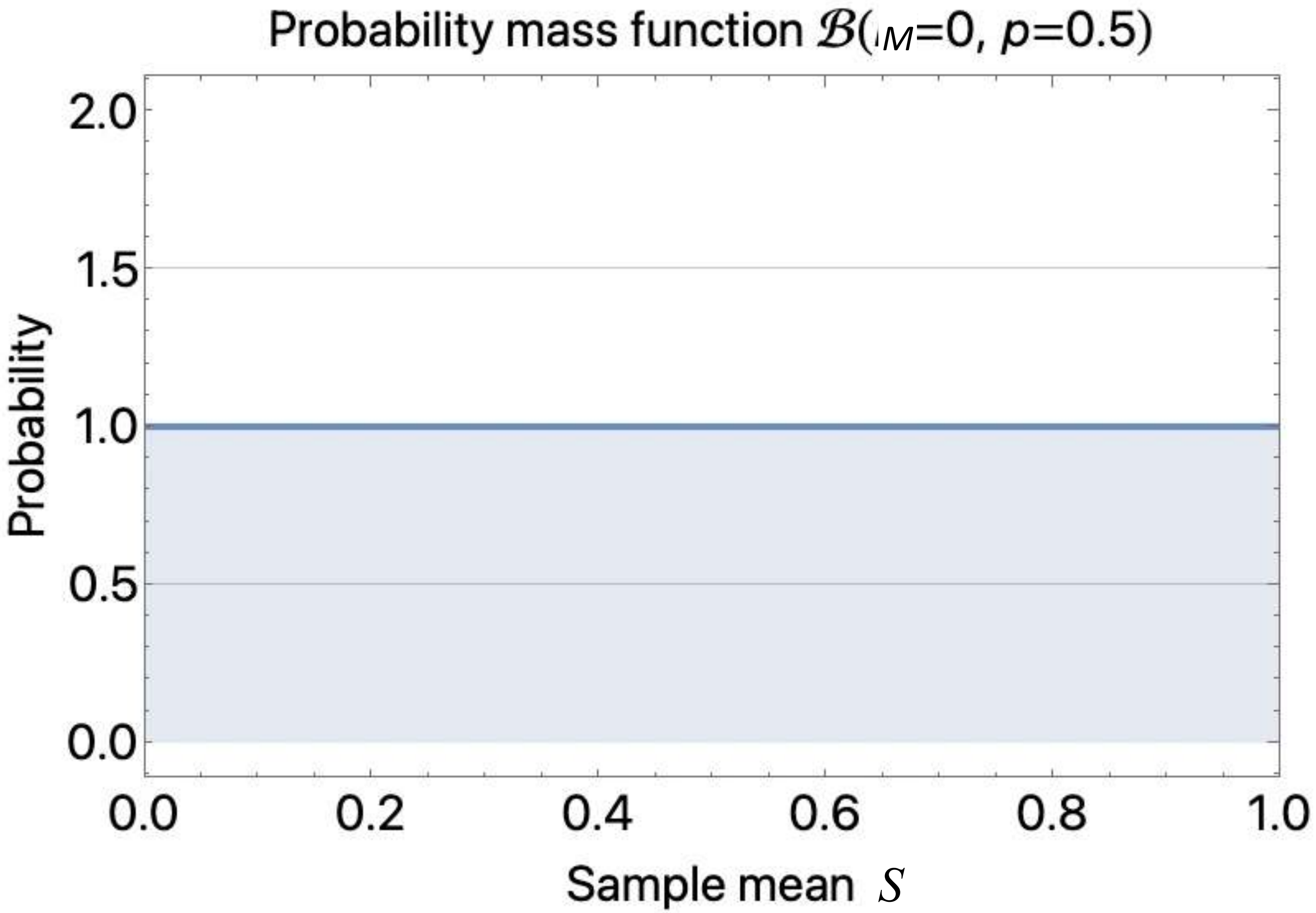
$$\mathbb{V}[X_m] = \mathbb{V}[X] = p(1-p) \quad \forall\, m \in \{1,\ldots,M\}$$

The variance is reduced by the number of sampler we take!

Thus we can suppress the quantum projection noise with enough shots.

The standard deviation drops as one over square root of the number of shots.

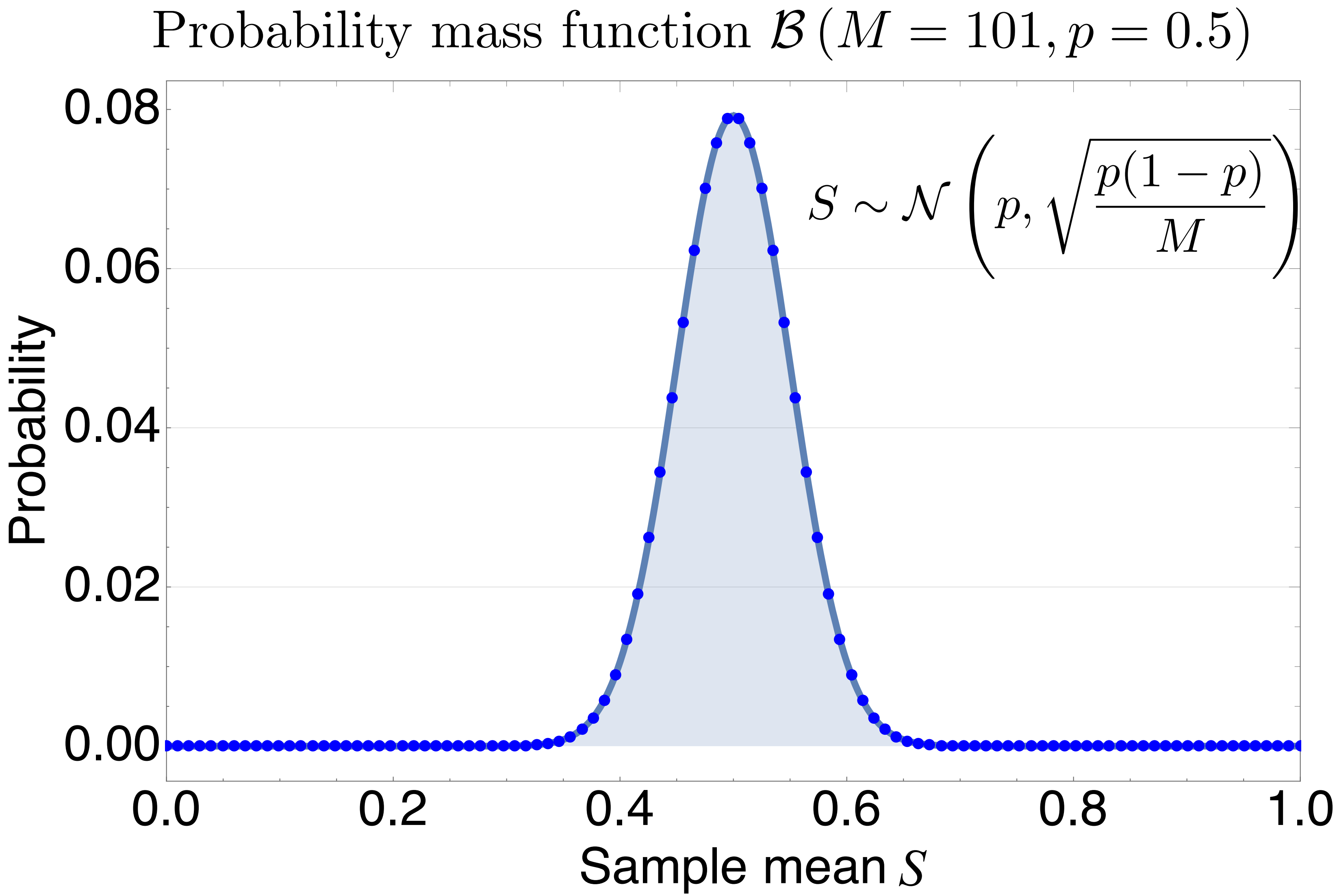
Animation of convergence of shots expectation value and mean



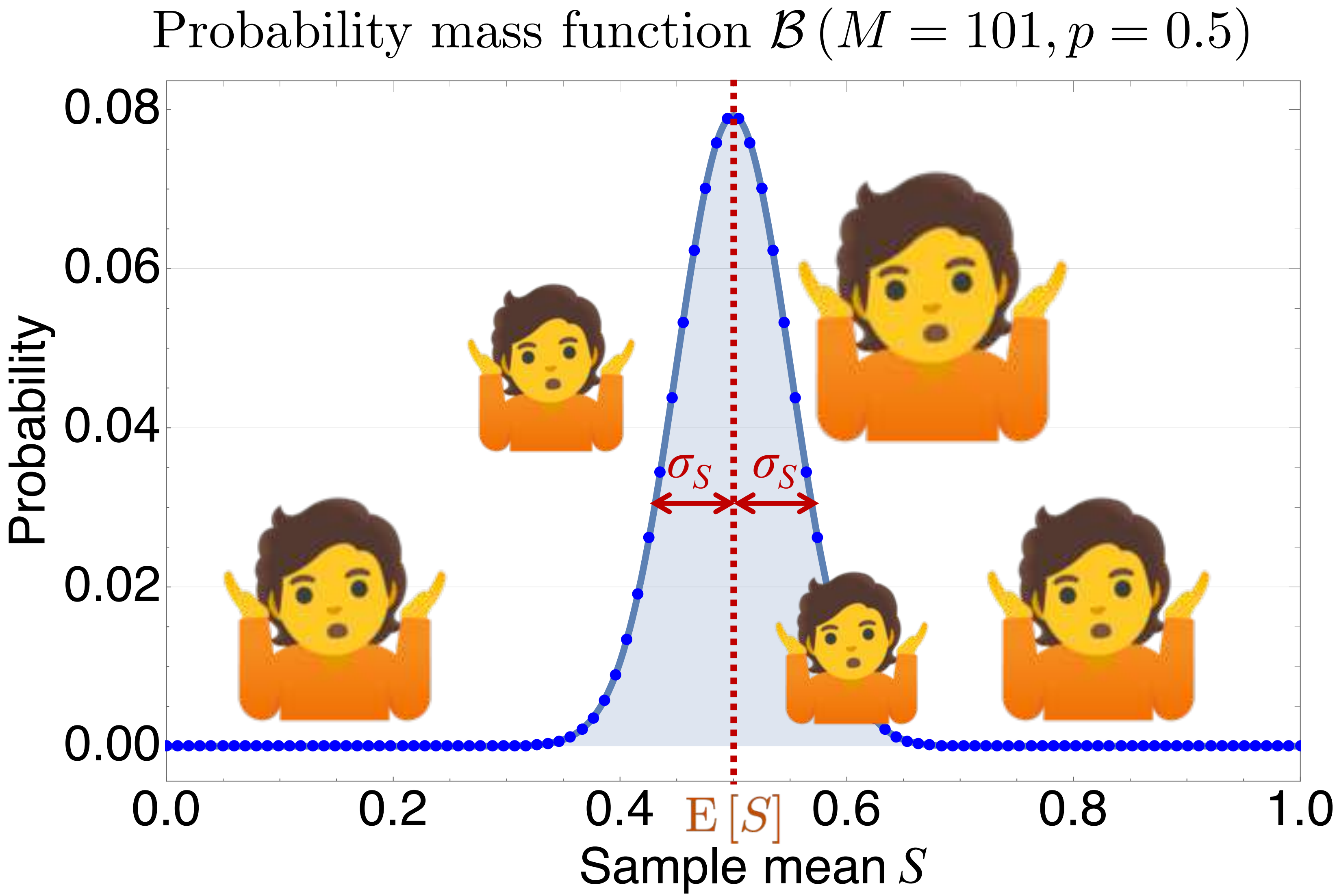
Binomial distribution

$$\binom{M}{k} p^k (1-p)^{M-k}$$

Sampled output distribution



Properties of the output distribution

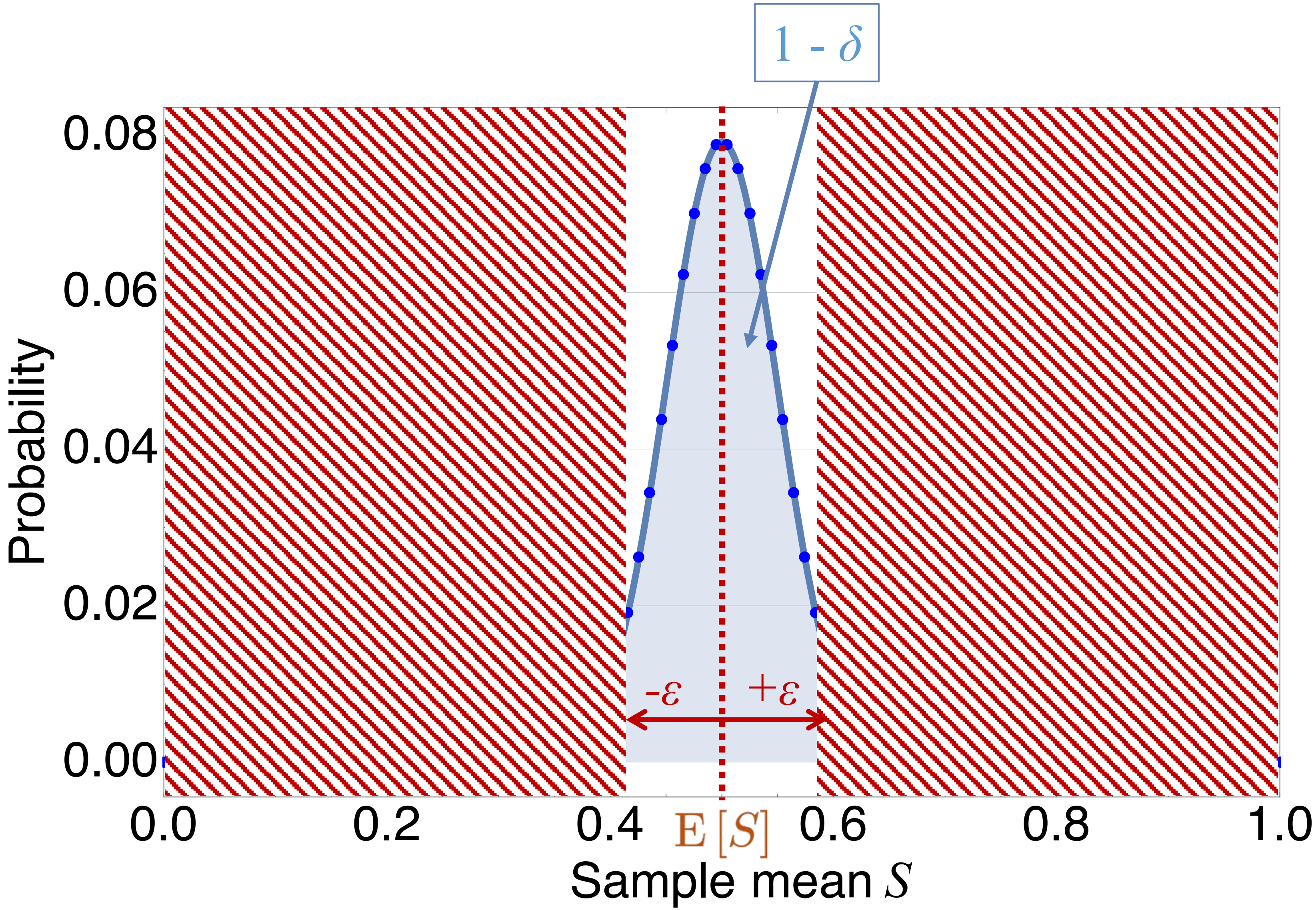


$$\mathrm{E}\,[S] = p$$

$$\sigma_S = \sqrt{\mathrm{Var}\,[S]} = \sqrt{\frac{p(1-p)}{M}}$$

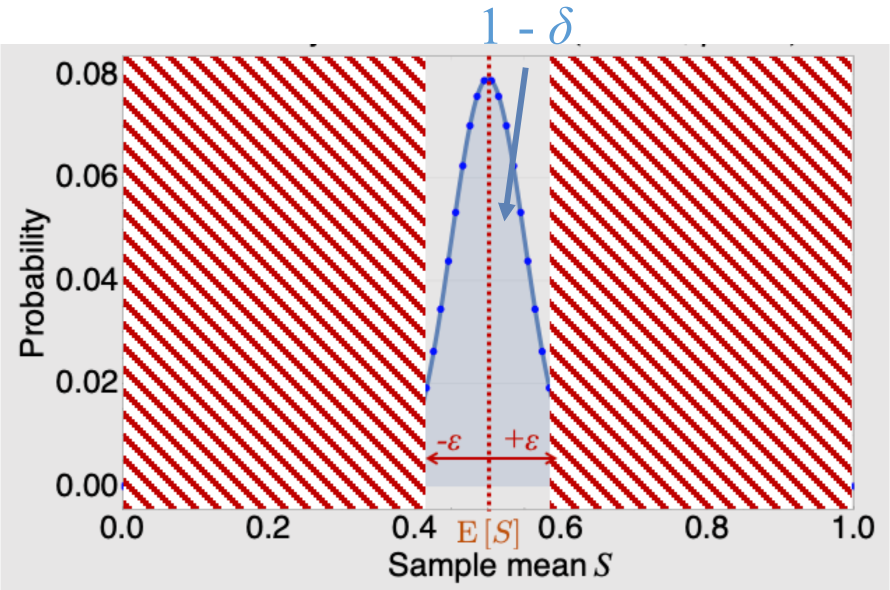
person: facebook

Concentration Measure for Sampling Expectation Values



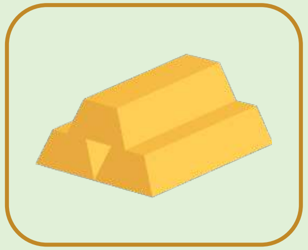
Icon: Flaticon

Error Bound on Quantum Expectation Values
Concentration Measure for Sampling Expectation Values



Chernoff-Hoeffding two-sided tail bound, given $|O(x)| \leq 1$ for

ε specified (additive) precision (worst case additive error) for S with success probability at least $1 - \delta$.



$$\Pr\left[\left|S - \left\langle \hat{O} \right\rangle\right| > \epsilon\right] \leq \delta := 2\exp\left(-\frac{1}{2}M\epsilon^2\right)$$

δ specific failure probability for meeting precision ε empirically.

Empirical mean & sample properties

$$S = \frac{1}{M}\sum_{m=1}^{M} O(X_m)$$

Tf. required number of shots is at least [or with high probability (greater than 2/3)]

$$M \geq 2\epsilon^{-2}\log\left(2\delta^{-1}\right) \qquad \left[M \gtrapprox 4\epsilon^{-2}\right] \ .$$

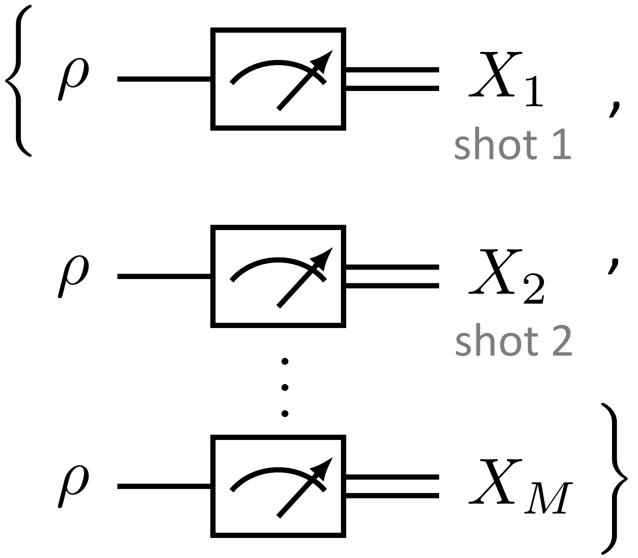
* Can find even tighter bound here owing to smaller [0,1] range

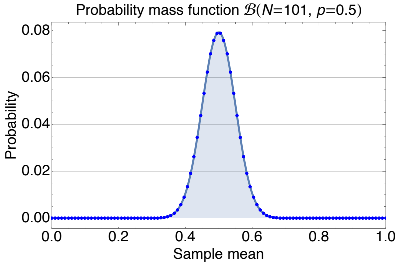
Note that this scales same way (mod δ) as the variance bound with $\varepsilon = \sigma$:

$$M \geq \frac{1}{4}\epsilon^{-2}$$

Very general scaling for sampling problems!

Ideal single qubit measurement with *M* shots



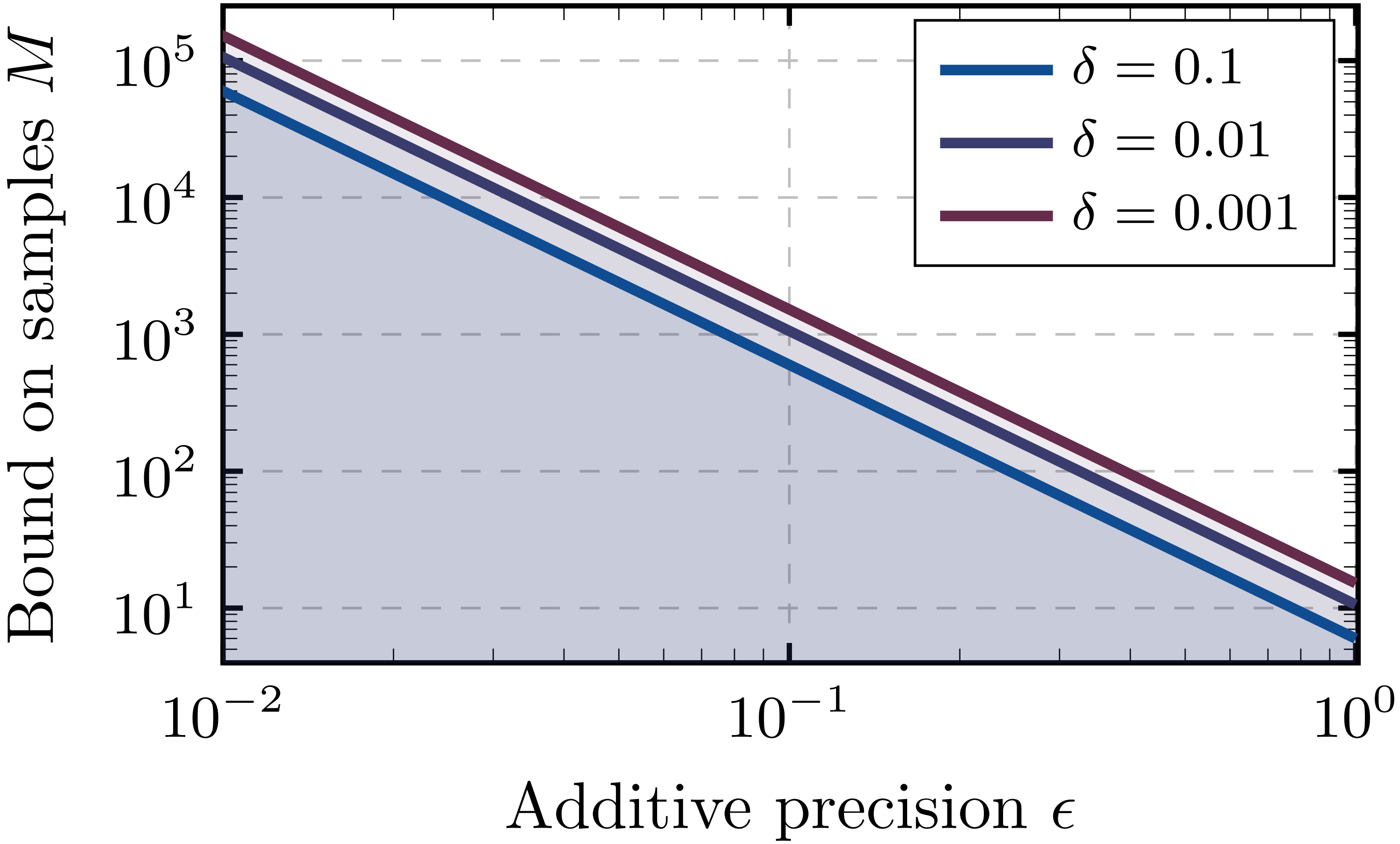


Observe that the **probability** δ is much cheaper than the **precision** ε.
Observe, *n* is not part of the equation.

"Knowing you are *not* wrong is cheaper than knowing you are right."

Tail bound on sample complexity

$$M \geq 2\epsilon^{-2} \log\left(2\delta^{-1}\right) \qquad \left[M \gtrsim 4\epsilon^{-2}\right]$$



Concentration inequalities and tail bounds

Making a list,
checking it twice,
going to see
which inequality
is nice!

Markov? Hoeffding? Jensen? Chebyshev? Chernoff?

https://www.zlatko-minev.com/blog/inequalities

1. Probability (Technical note 11.9 v0.6)

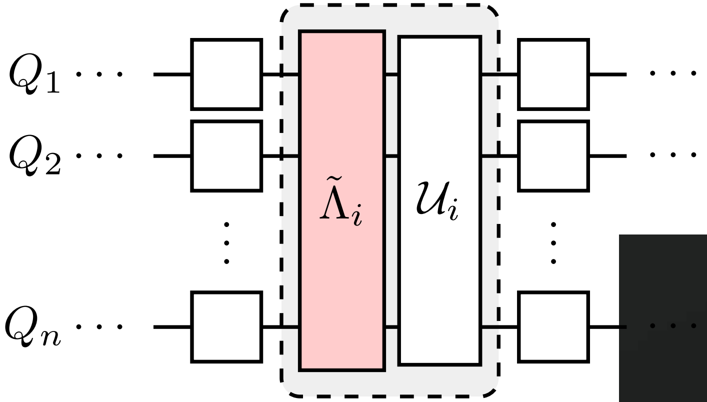
1A. Concentration inequalities and tail bounds

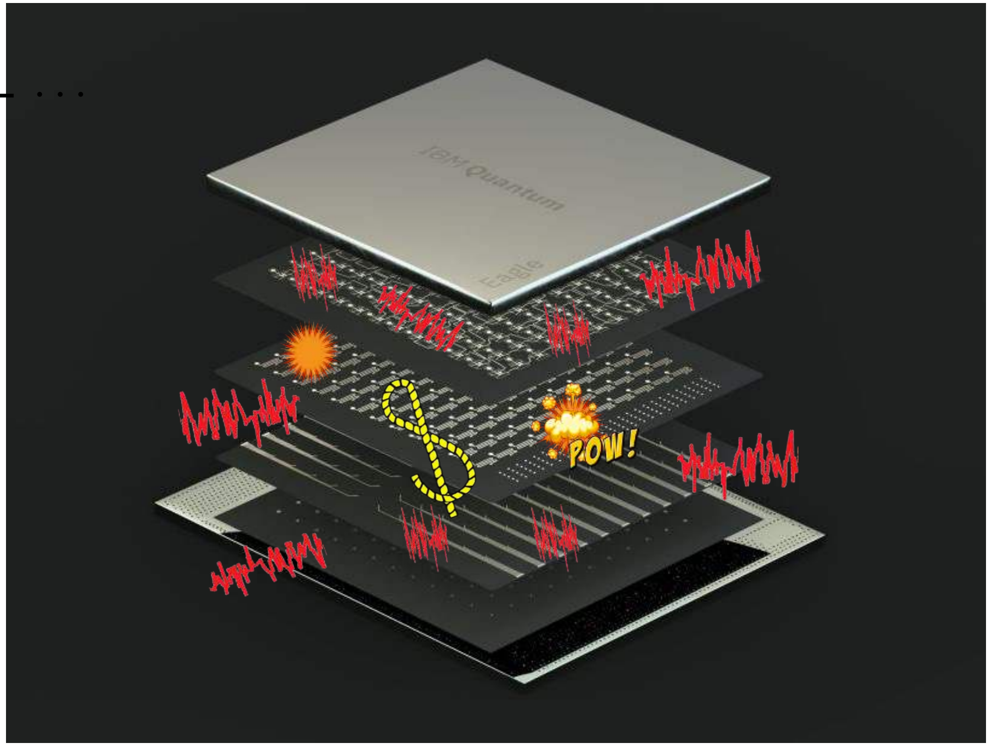
Unless otherwise specified, all variables are real $\mathbb{R}$. Inequalities come as one-sided $\Pr(\cdots \leq \cdots)$ and two-sided $\Pr(\lvert\cdots\rvert \leq \cdots)$. Notation: X is a random variable, $\mu := \mathbb{E}[X]$, $\sigma^2 := \mathrm{Var}[X]$, $S_n := X_1 + \cdots + X_n$.

Inequality	Conditions		Common form	Notes / Alternate form	
			Single random variable		
Markov[1]	Non-negative	$X \geq 0$	$\Pr[X \geq a] \leq \frac{\mathbb{E}[X]}{a}$ $\quad \forall a > 0$	$\Pr[X \geq kE[X]] \leq \frac{1}{k} \quad k > 1$	[3, Sec. 5.1][6, Thm 1.13]
extension	+ non-negative, strictly increasing func Φ	$X \geq 0$ $\Phi(X) \geq 0$ increasing	$\Pr[X \geq a] = \Pr[\Phi(X) \geq \Phi(a)] \leq \frac{\mathrm{E}(\Phi(X))}{\Phi(a)}$	$\forall a > 0$	Wiki
Reverse Markov	upper-bounded by U (can be positive)	$\max X = U$	$\Pr[X \leq a] \leq \frac{U - \mathbb{E}[X]}{U - a}$	$\forall a > 0$	[1, Sec. 3.1]
Chebyshev[2]	Finite mean and variance	$\mathbb{E}[X], \mathrm{Var}[X]$ finite	$\Pr[\lvert X - \mathbb{E}[X]\rvert \geq a] \leq \frac{\sigma^2}{a^2}$	$\Pr[\lvert X - \mathbb{E}[X]\rvert \geq a \cdot \sigma] \leq \frac{1}{a^2}$ $\forall a > 0,\ \sigma^2 = \mathrm{Var}[X]$	[1, Sec. 3.2] [3, Sec. 5.1][2, Thm 18.11]
Cantelli	Improved Chebyshev	(same; but one-sided)	$\Pr[X - \mathbb{E}[X] \geq a]) \leq \dfrac{\sigma^2}{\sigma^2 + a^2}$	$\forall a > 0,\ \sigma^2 = \mathrm{Var}[X]$	Wiki
Chernoff[3]	Generic		$\Pr[X \geq a] = \Pr\left[e^{tX} \geq e^{ta}\right]$	$\forall t > 0, \quad a \in \mathbb{R}$	[1, Sec. 3.3]
Jensen		$f: \mathbb{R} \to \mathbb{R}$; f convex	$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$		[3, Prob. 5.3][6, Thm 1.14]
Hoeffding's lemma		$\mathbb{E}[X] = \mu$ $a \leq X \leq b$	$\mathbb{E}\left[e^{\lambda X}\right] \leq e^{\lambda\mu} e^{\frac{\lambda^2 (b-a)^2}{8}}$	$\lambda \in \mathbb{R}$	[1, Sec. 3.4]
			Sum of random variables		
Chernoff-Hoeffding (one-sided)	n independent random vars	$X_1, \ldots, X_n$ indep $S_n = X_1 + \cdots + X_n$ $X_i \in [a_i, b_i] \quad \forall i$	$\Pr[S_n - \mathbb{E}[S_n] \geq t] \leq \exp\left(\frac{-2t^2 n^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$		[1, Sec. 3.5]
(two-sided)[4]	(same as above)		$\Pr[\lvert S_n - \mathbb{E}[S_n]\rvert > t] \leq 2\exp\left(\frac{-2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$	$\forall t \in \left(0, \frac{1}{2}\right)$	[5, Thm.1.1]
(two-sided iid)	same plus iid, range, mean μ for each	$X_1, \ldots, X_n \in [0, 1]$ $\mathbb{E}[X_i] = \mu$ iid	$\Pr\left[\left\lvert \frac{S_n}{n} - \mu \right\rvert \geq \epsilon\right] \leq 2\exp\left(-2n\epsilon^2\right)$	$\forall \epsilon > 0$	[6, Thm 1.16]
Thm 1.3	n independent random vars	$X_1, \ldots, X_n$ indep $S_n = X_1 + \cdots + X_n$	$\Pr[S_n - \mathbb{E}[S_n] > \epsilon] \leq 2\exp\left(\frac{-\epsilon^2}{4\sum_{i=1}^n \mathrm{Var}[X_i]}\right)$	$\epsilon \in (0, 2\,\mathrm{Var}[S_n] / (\max_i \lvert X_i - \mathbb{E}[X_i]\rvert))$	[5, Thm. 1.3]
Azuma					
Weak law of large numbers	n independent iid random vars	$X_1, \ldots, X_n$ indep $\mathbb{E}[X_i] = \mu$ iid	$\lim_{n\to\infty} \Pr\left[\lvert \frac{1}{n} S_n - \mu \rvert \geq \epsilon\right] = 0$	$\forall \epsilon > 0$	[3, Sec. 5.2][6, Thm 1.15]
Strong law of large numbers	(same)	(same)	$\Pr\left[\lim_{n\to\infty} \frac{1}{n} S_n = \mu\right] = 1$		[3, Sec. 5.5]
	Advanced				
Bennett	n independent zero-mean	$X_1, \ldots, X_n$ indep $\mathbb{E}[X_i] = 0$ iid	$\Pr[S_n > \epsilon] \leq \exp\left(-n\sigma^2 h\left(\frac{\epsilon}{n\sigma^2}\right)\right)$	$\sigma^2 := \frac{1}{n}\sum_{i=1}^n \mathrm{Var}[X_i]$, $\forall \epsilon > 0$, $h(a) := (1+a)\log(1+a) - a$ for $a \geq 0$	[1, 4.1]
Bernstein	(same)	(same)	$\Pr[S_n > \epsilon] \leq \exp\left(\frac{-n\epsilon^2}{2(\sigma^2 + \epsilon/3)}\right)$	(same)	[1, 4.2]
Efron-Stein	scalar func of vars $f: \chi^n \to \mathbb{R}$	$X_1, \ldots, X_n$ indep w/ values in set χ	$\mathrm{Var}[Z] \leq \sum_{i=1}^n \mathbb{E}\left[(Z - \mathbb{E}_i[Z])^2\right]$	$Z := g(X_1, \ldots, X_n)$ $\mathbb{E}_i[Z] := \mathbb{E}[Z \mid X_1, \ldots, X_{i-1}, X_{i+1}, \ldots X_n]$	[1, 4.3]
McDiarmid's	scalar func of vars $f: \chi^n \to \mathbb{R}$	$X_1, \ldots, X_n$ indep w/ values in set χ	$\Pr[f(X_1, \ldots, X_n) - \mathbb{E}[f(X_1, \ldots, X_n)] \geq \epsilon] \leq \exp\left(\frac{-2\epsilon^2}{\sum_{i=1}^n c_i^2}\right)$	condition: c-bounded difference property $\forall \epsilon > 0$ $\lvert f(X_1, \ldots, X_i, \ldots, X_n) - f(X_1, \ldots, X_i', \ldots, X_n)\rvert \leq c_i$	[1, 4.4]

[1] Markov's inequality bounds the first moment of random variable. Use it when a constant probability bound is sufficient [1, Sec. 3.3].
[2] Chebyshev is derived from Markov. It bounds the second moment. It is the appropriate one when the variance σ is known. If σ is unknown, we can use the bounds of $X \in [a, b]$.
[3] Chernoff bound is used to bound the tails of the distribution for a sum of independent random variables. By far the most useful tool in randomized algorithms [1, Sec. 3.3].
[4] This probability can be interpreted as the level of significance ϵ (probability of making an error) for a confidence interval around the mean of size 2ϵ. Therefore, we require at least $\log(2\alpha)/2t^2$ samples to acquire $1 - \alpha$ confidence interval $\mathbb{E}[X] \pm t$.

Is it possible to learn the noise with accuracy, efficiency, and scalability?



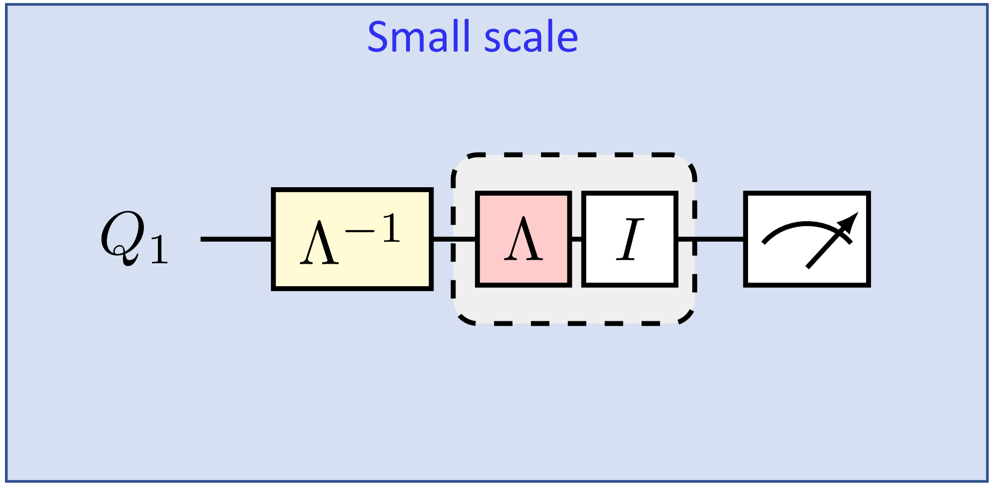


Energy relaxation T_1
Dephasing T_2
Coherent errors ZZ
Classical crosstalk
Quantum crosstalk
State preparation error
Measurement correlated errors
…

Control errors
Photon shot noise
1/f charge noise
1/f flux noise
Nonequilibrium quasiparticles
Leakage
Cosmic rays
…

PEC: Nice, but why hasn't worked so far for experiments?

Practical challenges

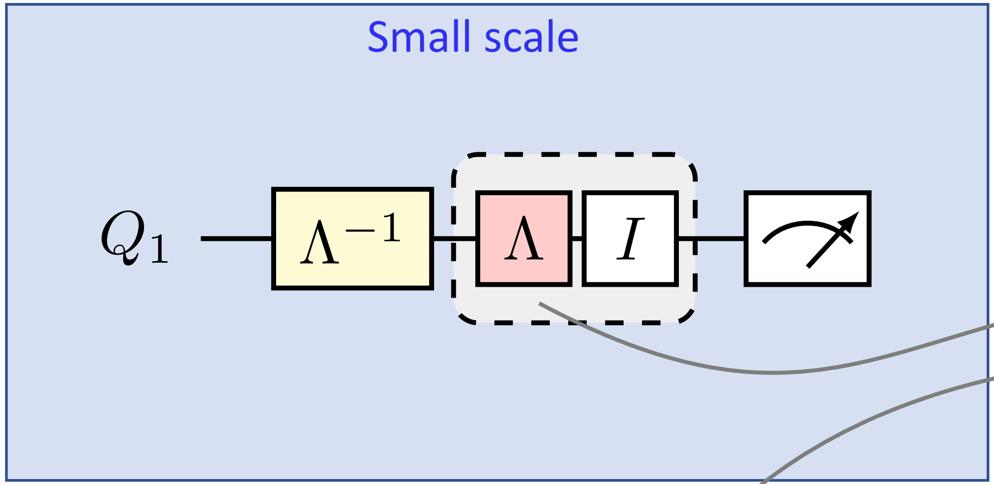


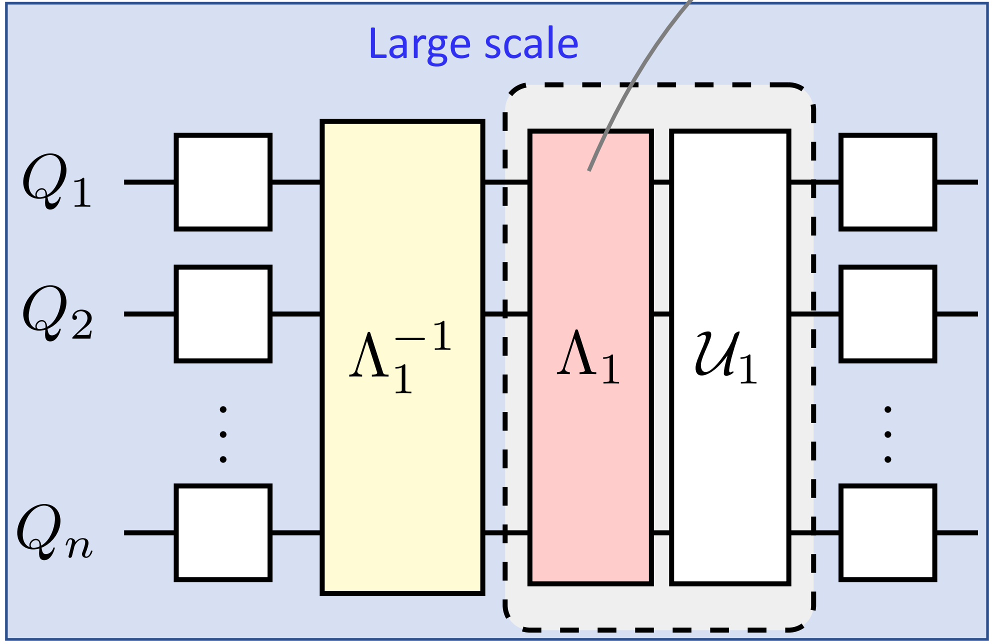
Critically hinges on knowing the full noise near perfectly

Despite the method's theoretical appeal (1-10), practical challenges have limited its demonstration to the one and two-qubit level (2, 3)

1. S. Endo, S. C. Benjamin, Y. Li, Physical Review X 8, 031027 (2018).
2. C. Song, et al., Science Advances 5, arXiv:2109.04457(2019).
3. S. Zhang, et al., Nature Communications 11, 587 (2020).
4. C. Piveteau, D. Sutter, S. Woerner, arXiv:2101.09290 (2021).
5. S. Endo, et al., J. Physi Soc. of Japan 90, 032001 (2021).
6. C. Piveteau, et al., arXiv:2103.04915 (2021).
7. R. Takagi, Phys. Rev. Research 3, 033178 (2021).
8. R. Takagi, S. Endo, S. Minagawa, M. Gu, arXiv:2109.04457 (2021).
9. Y. Guo, S. Yang, arXiv preprint arXiv:2201.00752 (2022).
10. . . .

PEC: Nice, but why hasn't worked so far? Challenges





2 qubits	255 parameters
10 qubits	10^{12} parameters
50 qubits	10^{60} parameters

noise param values 10^{-2} - 10^{-5}

additive error sampling cost (>10^2 – 10^{10})

Challenges

learning complexity

- efficient
- scalable
- accurate
- compact, tractable representation

noise in full device

- cross-talk
- correlated errors
- parallel gates

Outline

Idea

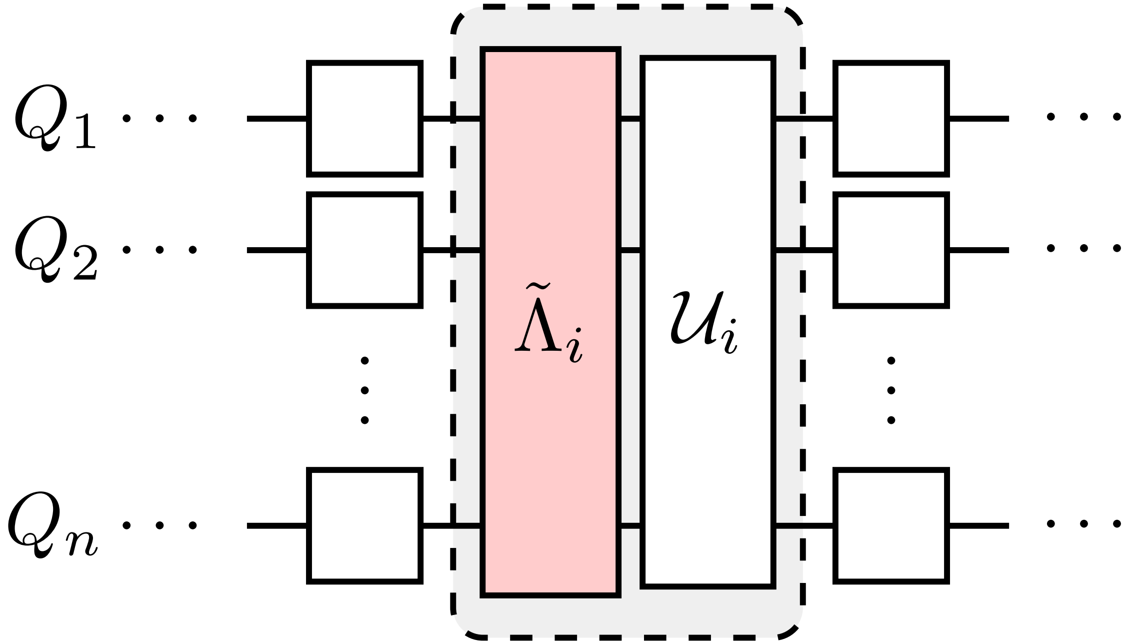
Learn

Cancel

(realization)

Icon: Flaticon

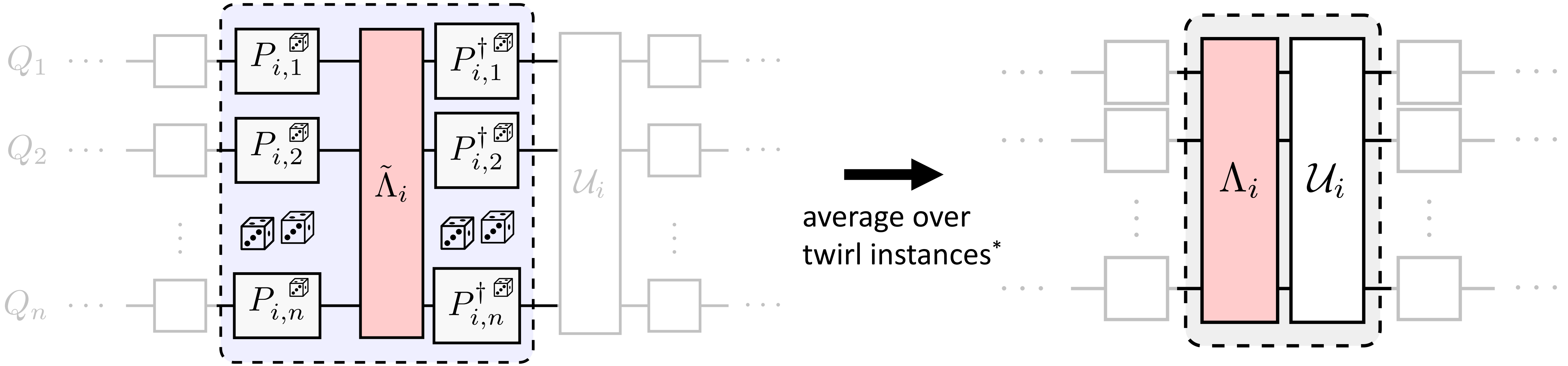
Step 1: Simplify the noise



noise that includes cross-talk errors, etc.
characterized by some 4^n x 4^n matrix

Step 1: Simplify the noise: twirl

twirl reduces to noise 4^n x 4^n matrix to diagonal one with 4^n entries in Pauli basis



Twirling references

1. C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, W. K. Wootters, et al., Phys. Rev. Lett. 76, 722 (1996).
2. E. Knill, arXiv:0404104 (2004).
3. O. Kern, G. Alber, D. L. Shepelyansky, EPJ D 32, 153 (2005).
4. M. R. Geller, Z. Zhou, Physical Review A 88, 012314 (2013).
5. J. J. Wallman, J. Emerson, Physical Review A 94, 052325 (2016)
6. Hashim *et al*., Phys. Rev. X 11, 041039 (2021)
7. Tutorial: zlatko-minev.com/blog/twirling (2022)
8. ...

Tutorial

SCAN ME

Stochastic Pauli channel

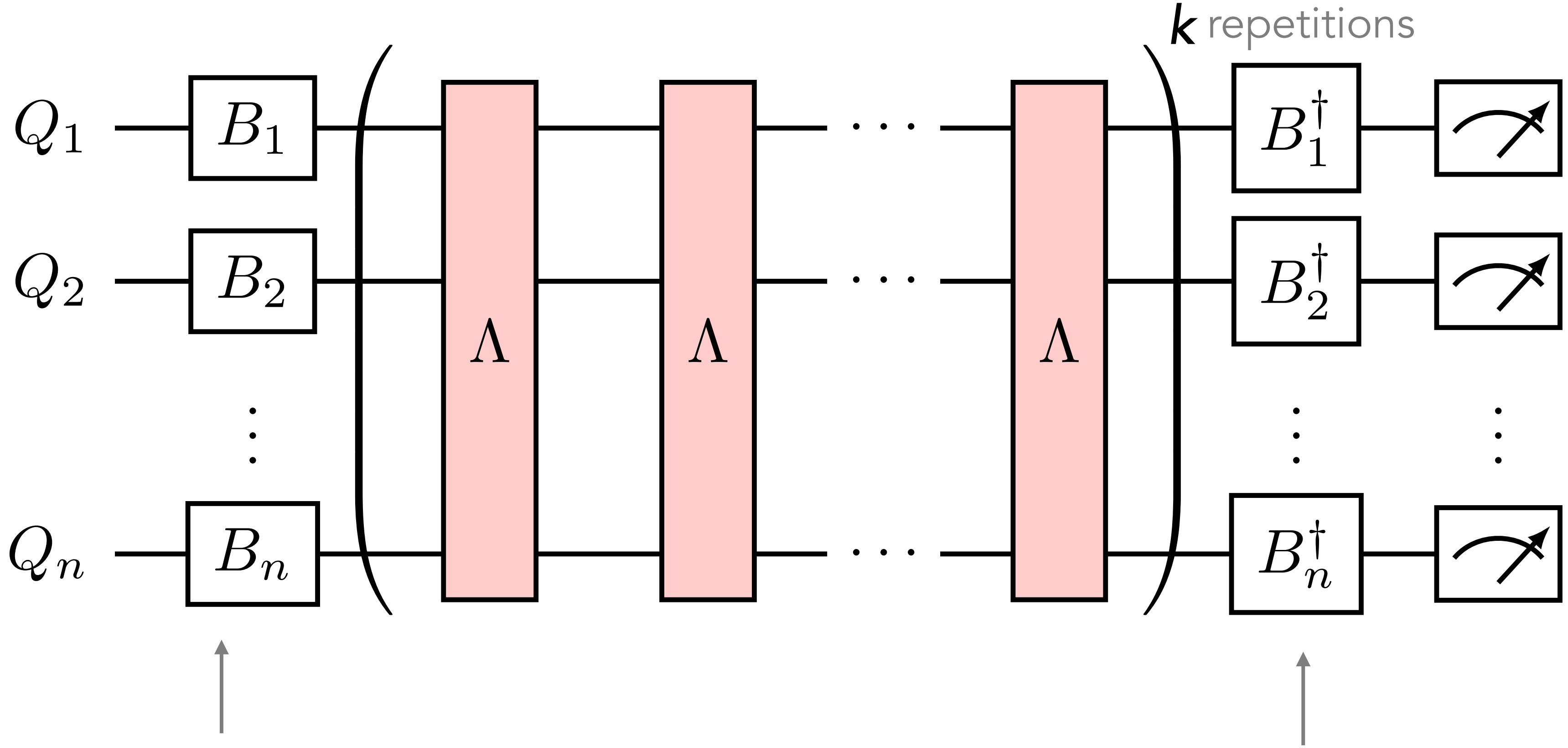
$$\Lambda_i\left(\rho\right) = \sum_{a=0}^{4^n-1} c_{ia} P_a \rho P_a^\dagger$$

$$\Lambda(P_a) = f_a P_a$$ eigenvecs are Paulis

* some sub-Clifford twirl group (use Paulis)

Step 2 wish: amplify noise

for the *i*-th layer



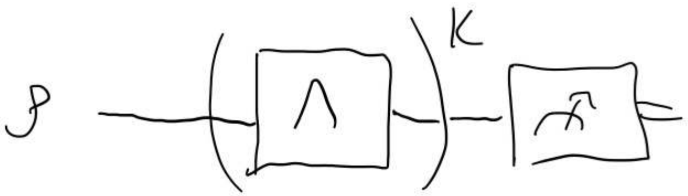
prepare circuit in pre-determined Pauli basis

Since diagonal channel
will amplify eigenvalues
learn with multiplicative precision

measure circuit in same pre-determined Pauli basis

Akin to:
RB, Cycle RB, K-body noise reconstruction, ...
S.T. Flammia and J.J. Wallman ACM Trans QC 1, 3 (2020), ...
For something of a review of protocols, see Helsen, *et al.*, *A general framework for randomized benchmarking* (arXiv:2010.07974)

Notes



$$\Lambda_i\left(\rho\right) = \sum_{a=0}^{4^n-1} c_{ia} P_a \rho P_a^\dagger$$

$$\Lambda(P_a) = f_a P_a$$

For a qubit

$$\rho = \frac{1}{2}\left(\hat{I} + \rho_X \hat{X} + \rho_Y \hat{Y} + \rho_Z \hat{Z}\right)$$

In general:

$$\rho = \frac{1}{2^n}\sum_b \rho_b \hat{P}_b$$

Pauli decomposition of a density matrix is a powerful tool - offering a versatile representation. It expresses a density matrix as a linear combination of Paulis, often referred to as the Pauli basis.

$$\Lambda(\rho) = \rho'$$

Posterior state: Action of the channel on the input state

$$\rho' = \frac{1}{2^n}\sum_b \rho'_b \hat{P}_b$$

Pauli decomposition of posterior state

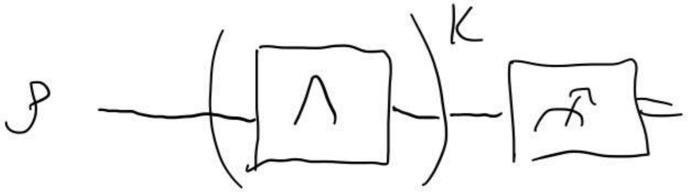
Since the channel is linear

$$\Lambda(a\hat{A} + b\hat{B}) = a\Lambda(\hat{A}) + b\Lambda(\hat{B})$$

$$\Lambda(\rho) = \Lambda\left(\frac{1}{2^n}\sum_b \rho_b \hat{P}_b\right)$$

$$= \frac{1}{2^n}\sum_b \rho_b \underbrace{\Lambda(\hat{P}_b)}$$

just need to know action on P_b :)



$$\Lambda_i\left(\rho\right) = \sum_{a=0}^{4^n-1} c_{ia} P_a \rho P_a^\dagger$$

$$\Lambda(P_a) = f_a P_a$$

$$\Lambda(\rho) = \Lambda\Big(\frac{1}{2^n}\sum_b \rho_b \hat{P}_b\Big)$$

$$= \frac{1}{2^n}\sum_b \rho_b \underbrace{\Lambda(\hat{P}_b)}_{\text{just need to know action on } P_b}$$

$$= \rho'$$

$$= \frac{1}{2^n}\sum_c \rho'_c \hat{P}_c$$

since $\mathrm{Tr}(\hat{P}_a^\dagger P_b) = \delta_{ab} 2^n$

orthogonal

Solve for ρ'_c by equate & orthogonal use:

$$\mathrm{Tr}\Big(P_c^\dagger \cancel{\tfrac{1}{2^n}} \sum_b \rho_b \Lambda(\hat{P}_b)\Big) = \mathrm{Tr}\Big(P_c^\dagger \cancel{\tfrac{1}{2^n}} \sum_{c'} \rho'_{c'} \hat{P}_{c'}\Big)$$

$$\sum_b \rho_b \mathrm{Tr}\big(P_c^\dagger \Lambda(\hat{P}_b)\big) = \sum_{c'} \rho'_{c'} \mathrm{Tr}\big(\hat{P}_c^\dagger \hat{P}_{c'}\big)$$

$$= \sum_{c'} \rho'_{c'} \delta_{cc'} 2^n$$

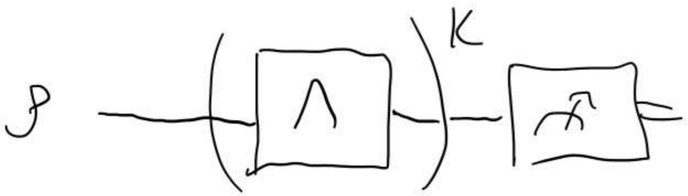
$$\frac{1}{2^n}\sum_b \rho_b \mathrm{Tr}\big[P_c^\dagger \Lambda(\hat{P}_b)\big] = \rho'_c \quad \checkmark$$

or written another way:

$$\rho_c' = \sum_b \rho_b \frac{1}{2^n}\mathrm{Tr}\big(\hat{P}_c^\dagger \Lambda(\hat{P}_b)\big)$$

$$= \sum_b \rho_b \frac{\langle\langle P_c | \Lambda | P_b \rangle\rangle}{\langle\langle 2 P_c | P_c \rangle\rangle}$$

Notes



$$\Lambda_i(\rho) = \sum_{a=0}^{4^n-1} c_{ia} P_a \rho P_a^\dagger$$

$$\Lambda(P_a) = f_a P_a$$

The measurement outcome then is (say measured in the $\hat{P}_c$ basis):

$$\langle \hat{P}_c \rangle = \mathrm{Tr}(\hat{P}_c \, \rho')$$

For a single step

$$= \mathrm{Tr}\Big(\hat{P}_c \sum_{c'} \frac{1}{2^n} \rho_{c'} \hat{P}_{c'}\Big)$$

$$= \sum_{c'} \rho_{c'} \frac{1}{2^n} \overbrace{\mathrm{Tr}(\hat{P}_c \hat{P}_{c'})} \rightarrow 2^n \delta_{cc'}$$

$$= \rho_c$$

now sub in

$$= \sum_b \rho_b \frac{\langle\langle P_c | \Lambda | P_b \rangle\rangle}{2^n}$$

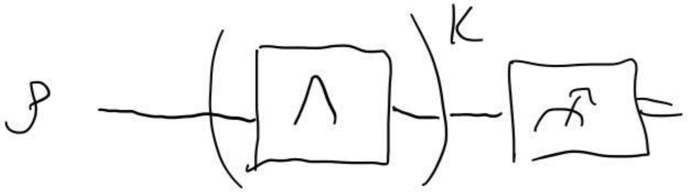
$\rightarrow$ for a pauli channel $\langle\langle P_c | \Lambda | P_b \rangle\rangle = f_c \, \delta_{bc} \, 2^n$

$$= f_c \, \rho_c$$

← initial state coefficient

← fidelity c of Λ

Notes



$$\Lambda_i(\rho) = \sum_{a=0}^{4^n-1} c_{ia} P_a \rho P_a^\dagger$$

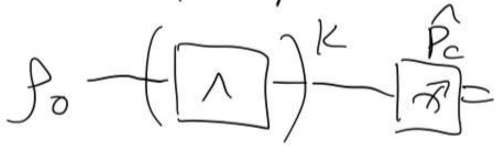
$$\Lambda(P_a) = f_a P_a$$

What is ρ_c for an eigenstate of $\hat{P}_c \rightarrow 1$.

e.g. for $c=Z$: $\hat{P}_c = \hat{Z}$ and $\rho = |0\rangle = \frac{1}{2}(I + Z)$

$$\Rightarrow \quad \rho_{c=Z} = \mathrm{Tr}(Z\rho) = +1$$

For repeating the channel K times



$$\rho' = \Lambda^K \rho$$

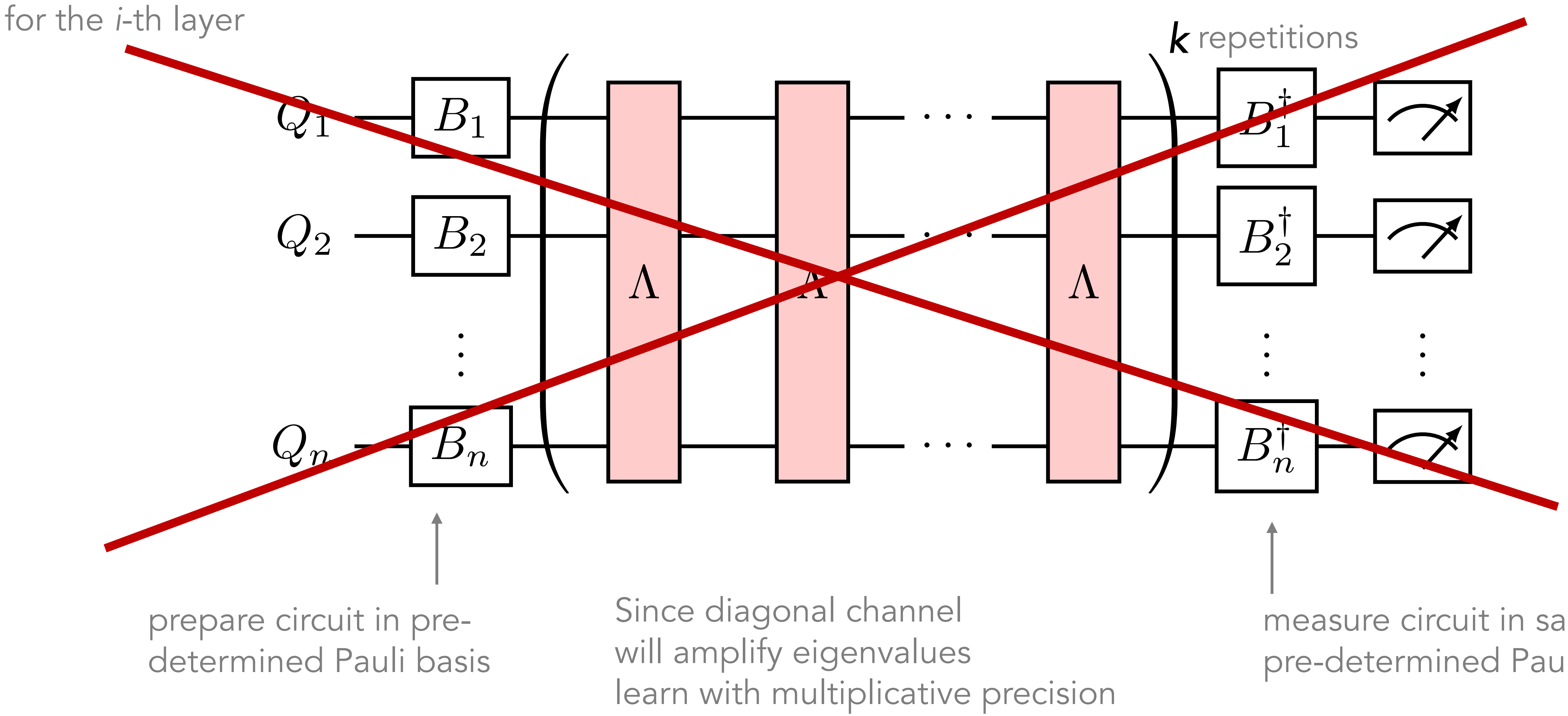
$$\Rightarrow \quad \rho'_c = \sum_b \rho_b \frac{\langle\langle P_c|\Lambda^K|P_b\rangle\rangle}{2^n}$$

$$= \sum_b f_b^K \rho_b$$

$$\Rightarrow \quad \langle \hat{P}_c \rangle = f_c^K \rho_c$$

Can include all diagonal SPAM

Step 2 wish: amplify noise



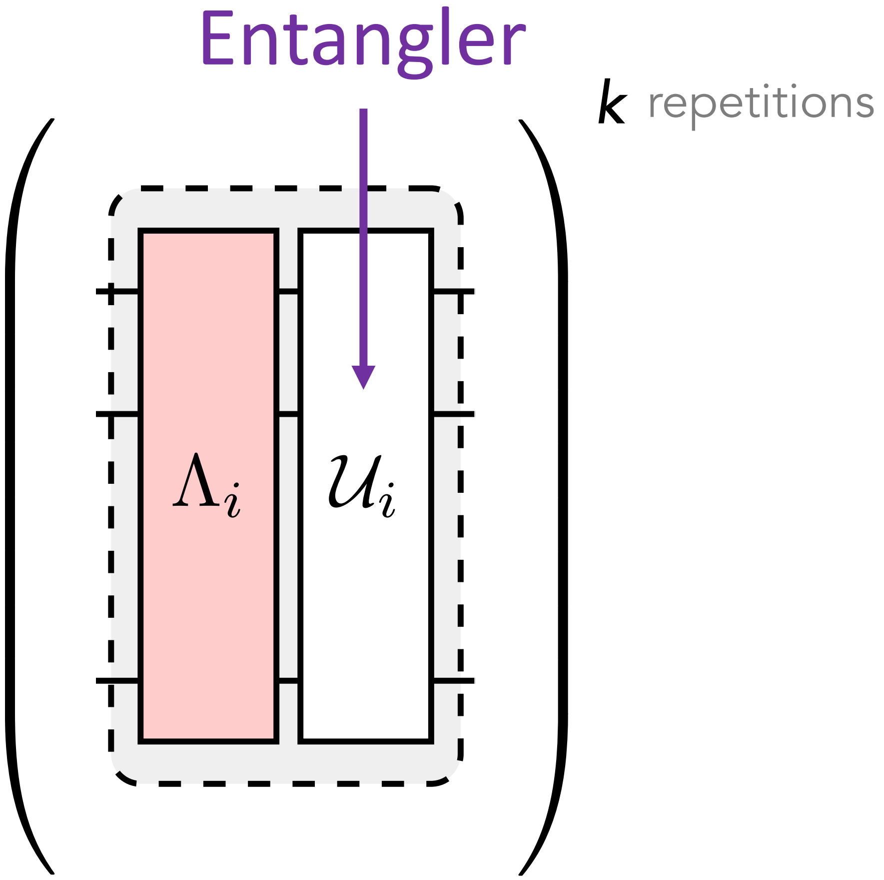
Akin to:

RB, Cycle RB, K-body noise reconstruction, ...

S.T. Flammia and J.J. Wallman ACM Trans QC 1, 3 (2020), ...

For something of a review of protocols, see Helsen, *et al.*, *A general framework for randomized benchmarking* (arXiv:2010.07974)

Step 2: Ideally, amplify the noise and learn



Ideally wish

$$\cancel{\Lambda_i^k(P_a) = f_{ia}^k P_a}$$

Akin to:

RB, Cycle RB, K-body noise reconstruction, ...

S.T. Flammia and J.J. Wallman ACM Trans QC 1, 3 (2020), ...

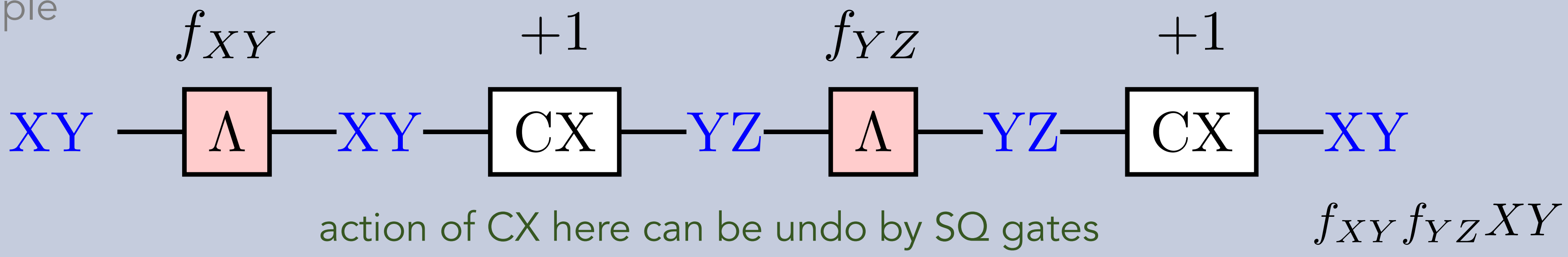
Erhard *et al.*, arXiv:1902.08543; Ferracin *et al.*, arXiv:2201.10672, ...

For something of a review of protocols, see Helsen, *et al.*, arXiv:2010.07974

Let's see how the amplification works with gates: no-go theorem

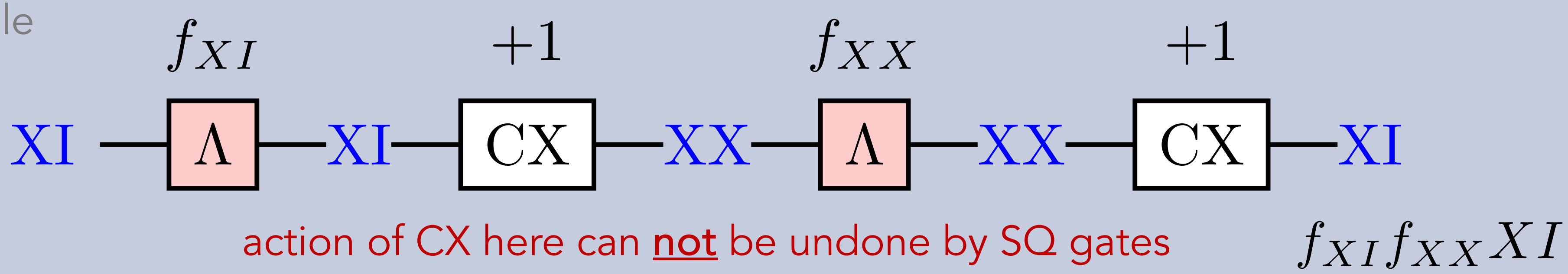
$$\Lambda(P_a) = f_a P_a$$

2Q example



action of CX here can be undo by SQ gates

2Q example

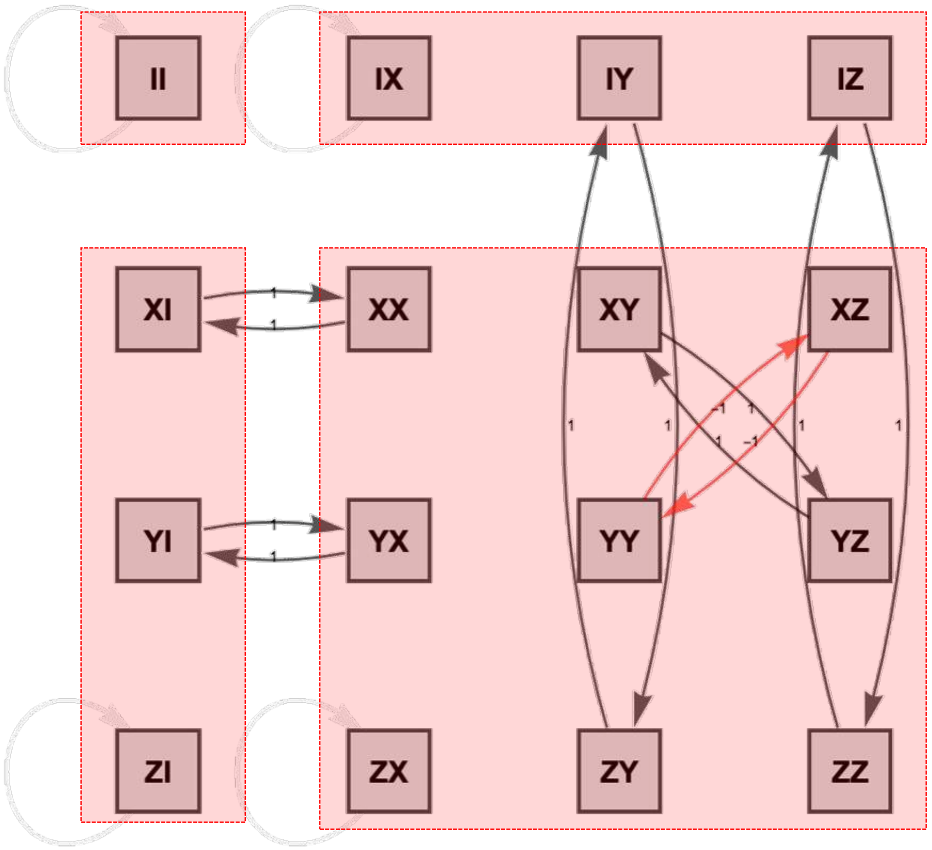


action of CX here can __**not**__ be undone by SQ gates

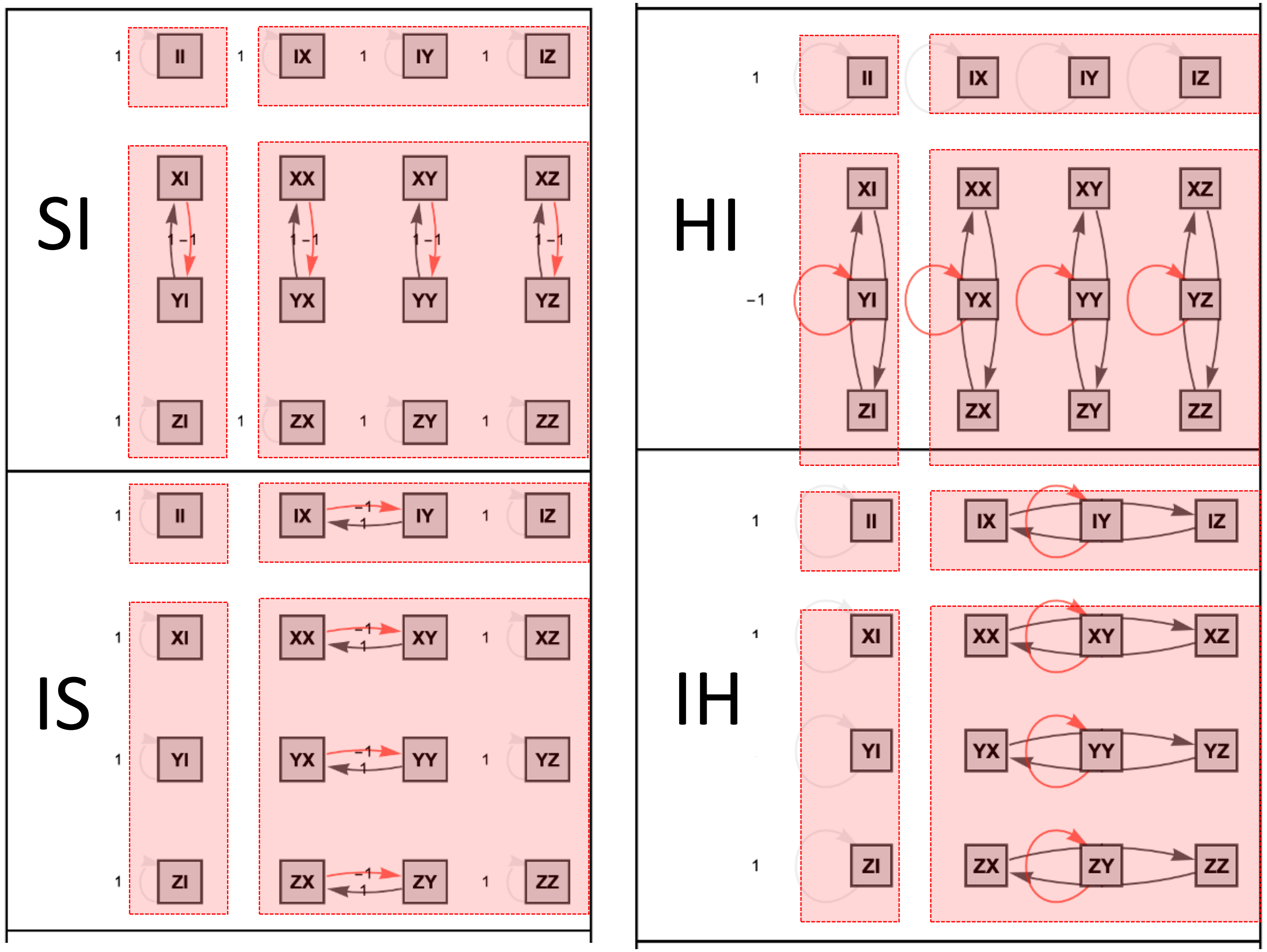
fundamental degeneracy – can not undo some non-local – need entangling operation

How to gates move state Paulis around?

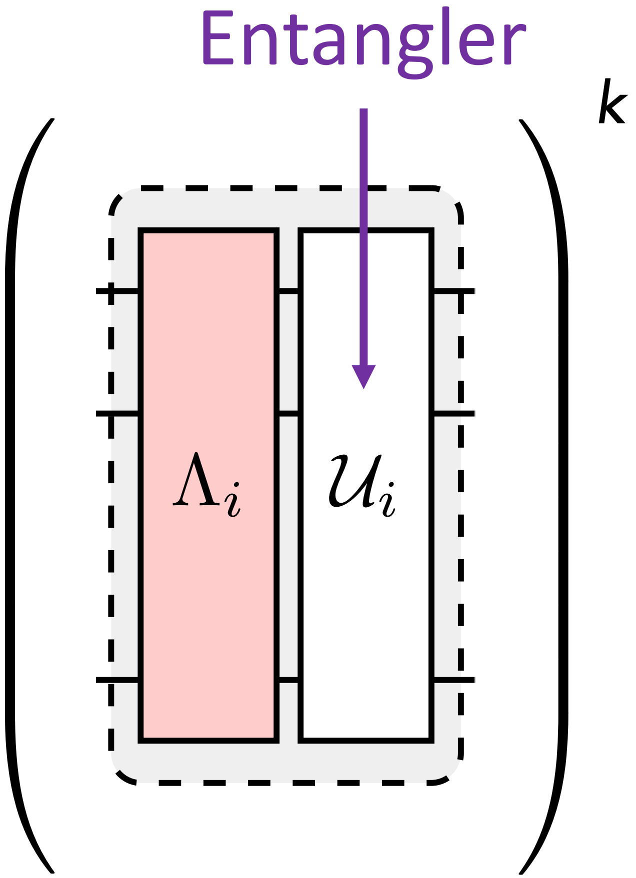
cNOT



Example single qubit gates



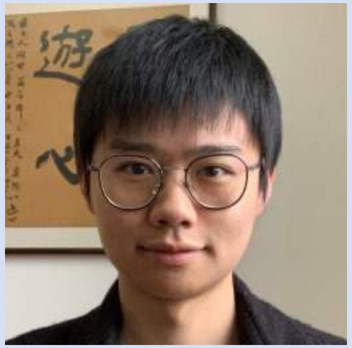
Not so simple to learn noise of entangling gates



Ideally wish

$$\cancel{\Lambda_i^k(P_a) = f_{ia}^k P_a}$$

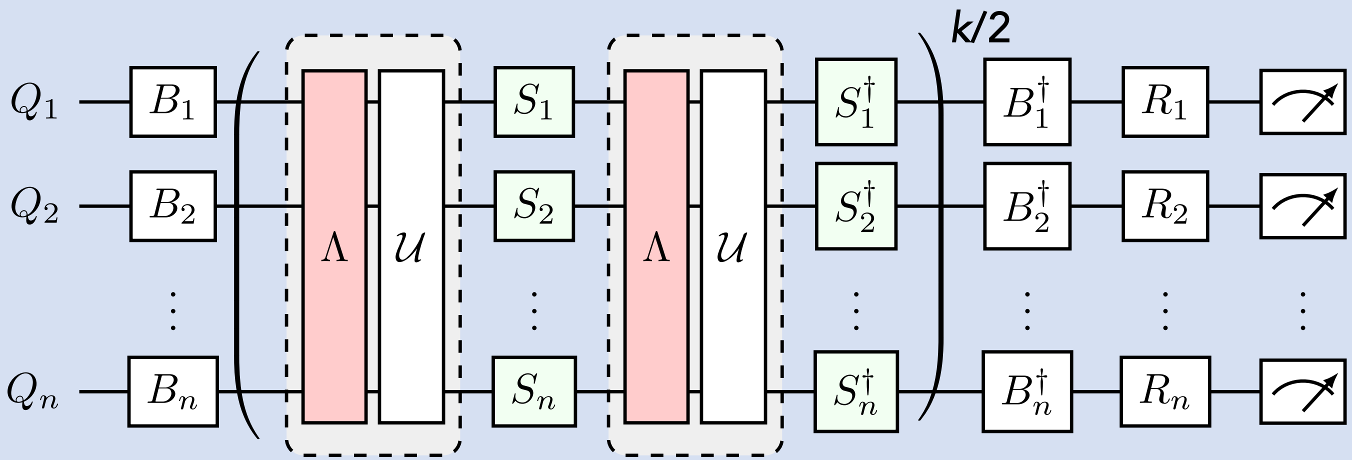
Fundamental no-go theorem on learning



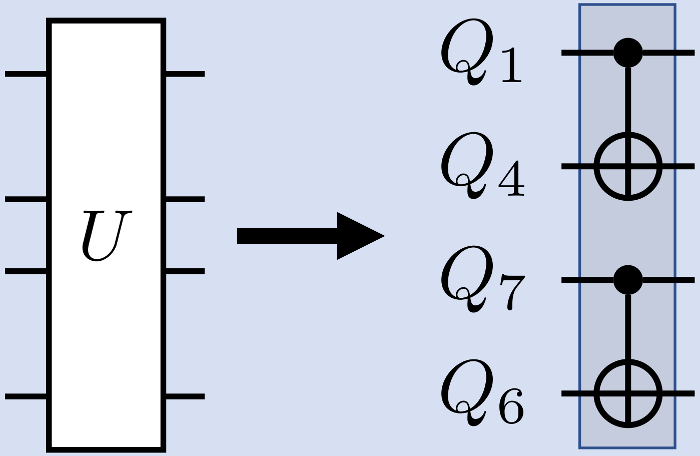
For general and in-depth
Senrui Chen, Y. Liu, M. Otten, A. Seif, B. Fefferman, L. Jiang
arXiv:2206.06362 (2022)
or supplement of our paper for qubit version and work by S. Flammia, S Benjamin, and teams.

Solution: Custom protocol + weak assumption

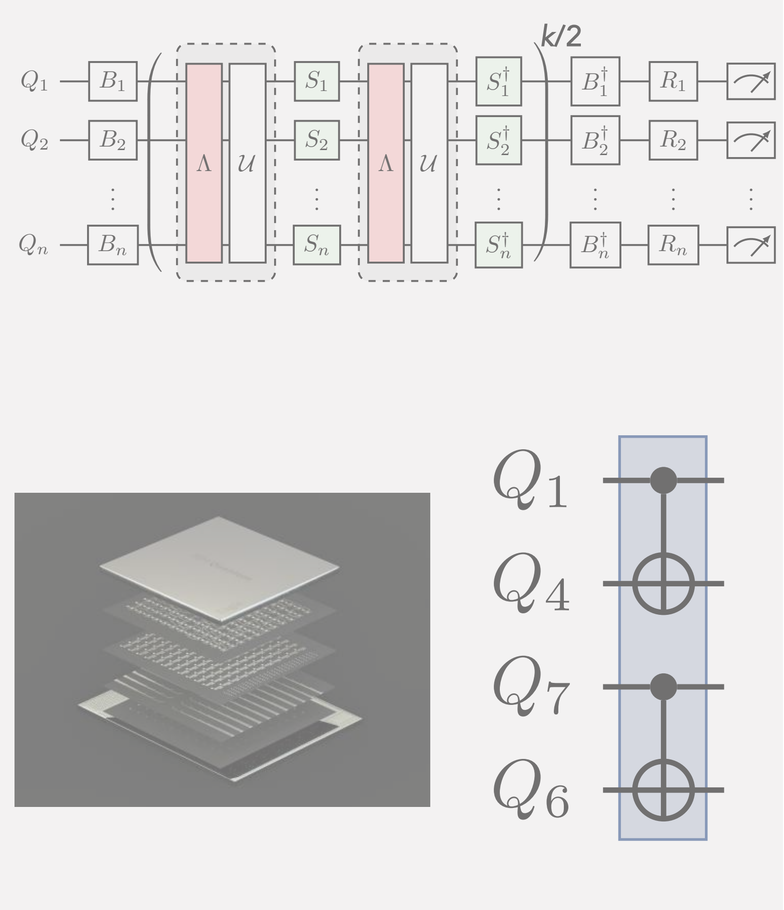
Learning circuits



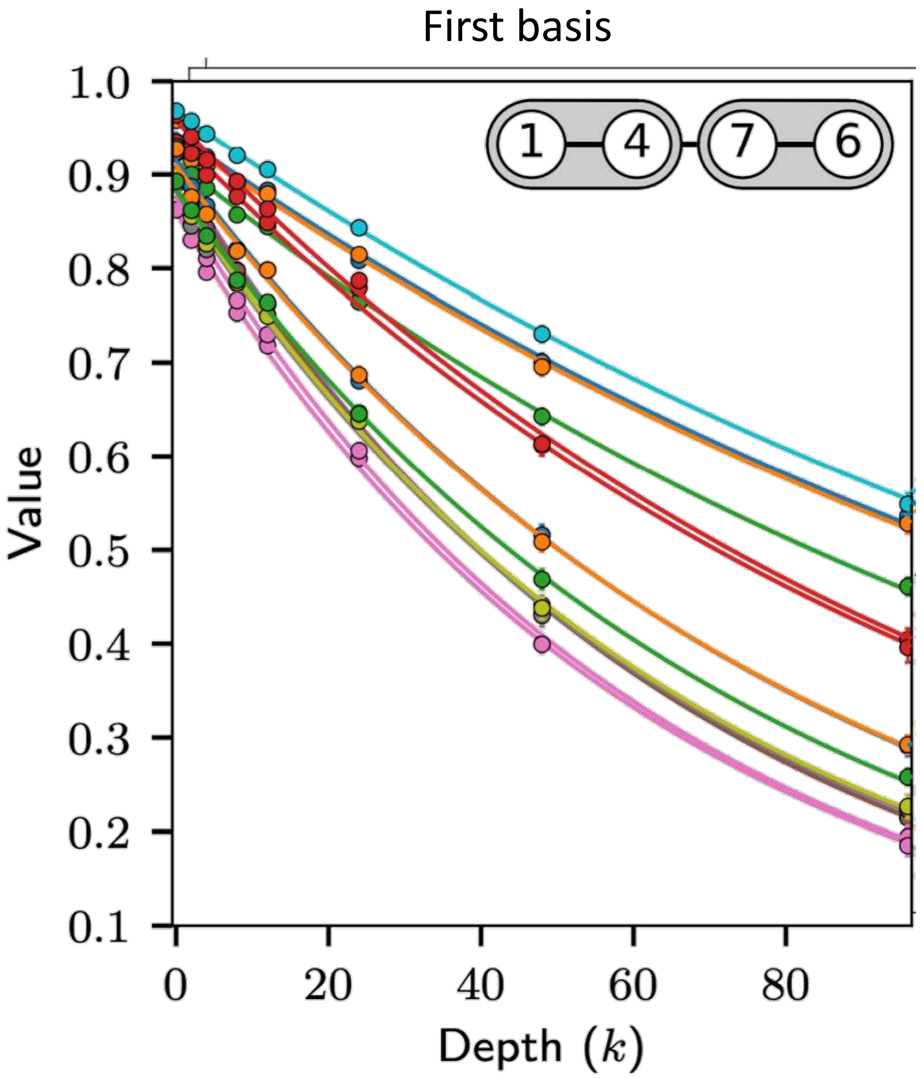
Example



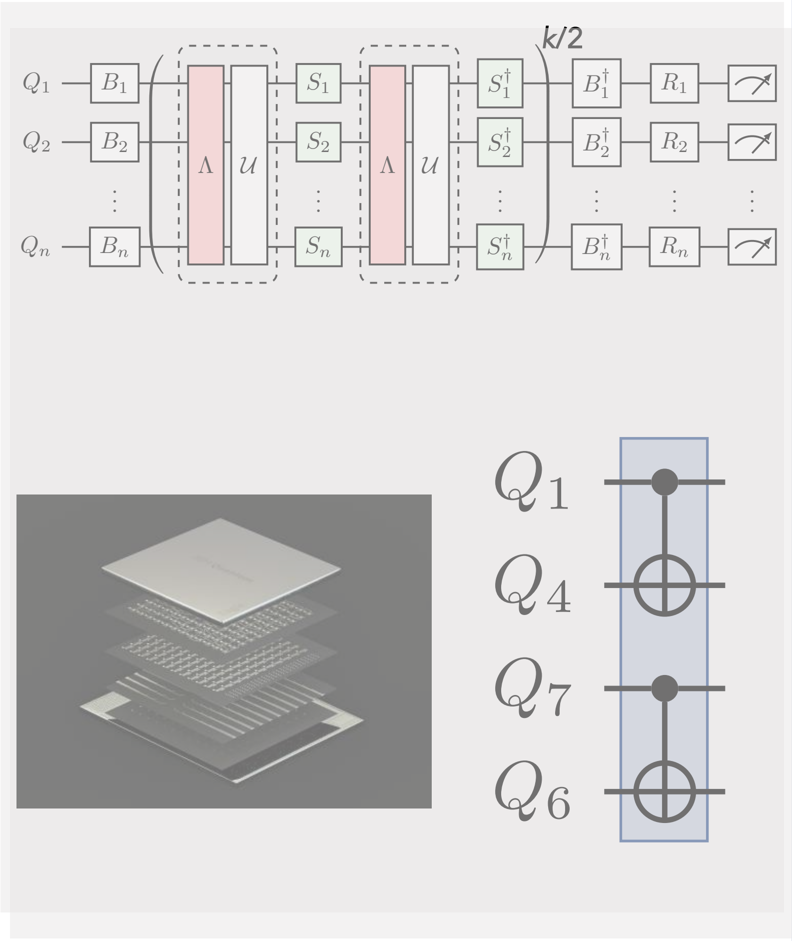
Learning the noise: raw data



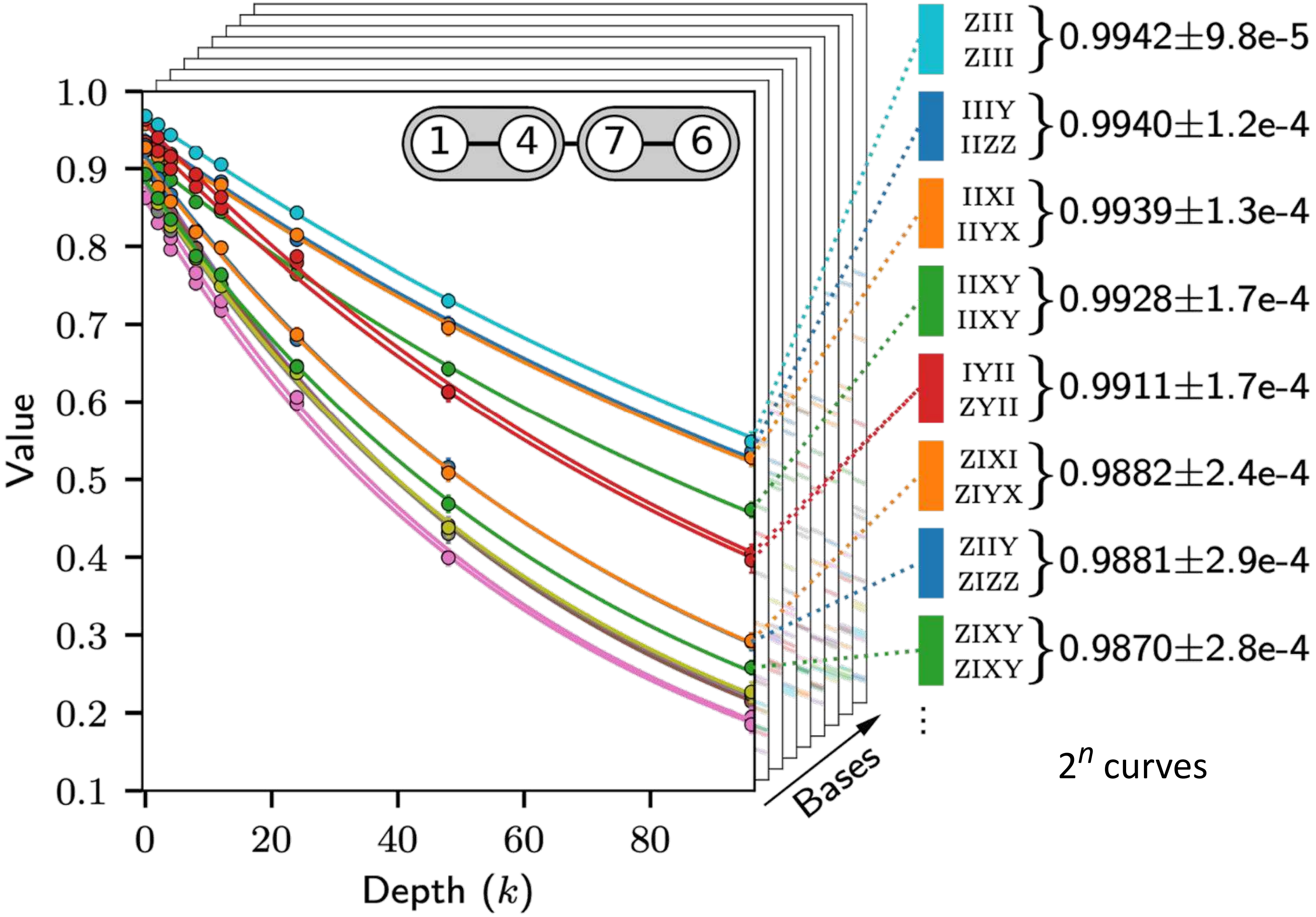
9 bases, 100 randomizations, 216 shots



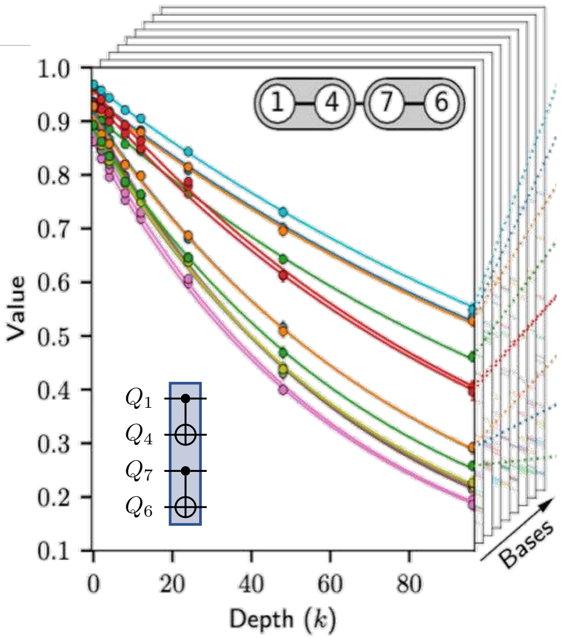
Learning the noise: raw data



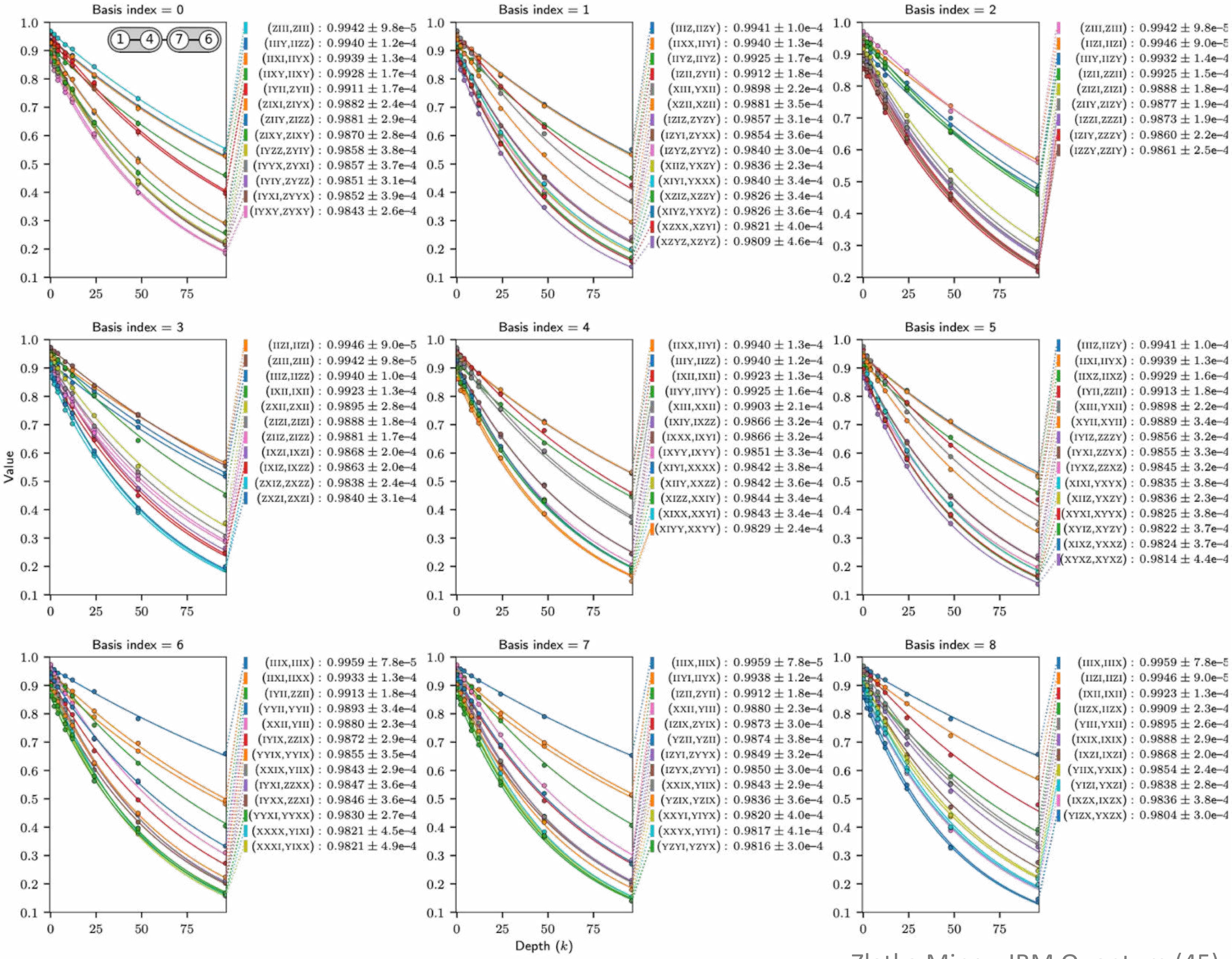
9 bases, 100 randomizations, 216 shots



Raw data

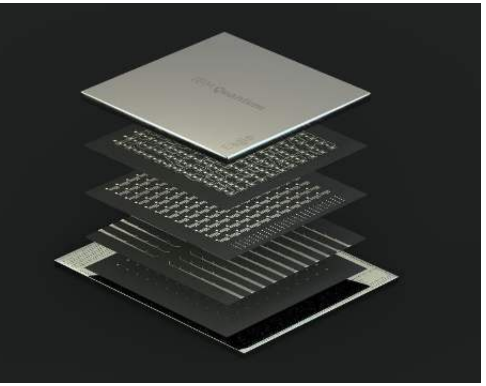


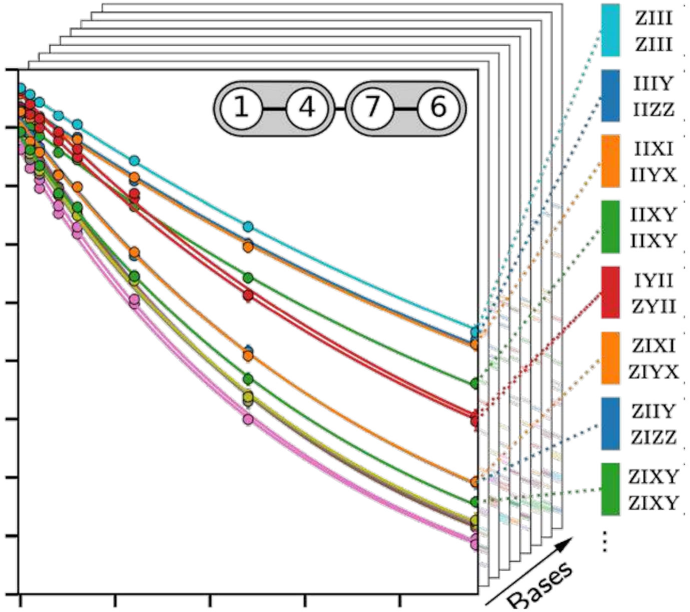
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Value
0
20
40
60
80
Depth (k)
Bases
1
4
7
6
Q_1
Q_4
Q_7
Q_6



Basis index = 0
(ZIII,ZIII) : 0.9942 ± 9.8e–5
(IIIY,IIZZ) : 0.9940 ± 1.2e–4
(IIXI,IIYX) : 0.9939 ± 1.3e–4
(IIXY,IIXY) : 0.9928 ± 1.7e–4
(IYII,ZYII) : 0.9911 ± 1.7e–4
(ZIXI,ZIYX) : 0.9882 ± 2.4e–4
(ZIIY,ZIZZ) : 0.9881 ± 2.9e–4
(ZIXY,ZIXY) : 0.9870 ± 2.8e–4
(IYZZ,ZYIY) : 0.9858 ± 3.8e–4
(IYYX,ZYXI) : 0.9857 ± 3.7e–4
(IYIY,ZYZZ) : 0.9851 ± 3.1e–4
(IYXI,ZYYX) : 0.9852 ± 3.9e–4
(IYXY,ZYXY) : 0.9843 ± 2.6e–4
Basis index = 1
(IIIZ,IIZY) : 0.9941 ± 1.0e–4
(IIXX,IIYI) : 0.9940 ± 1.3e–4
(IIYZ,IIYZ) : 0.9925 ± 1.7e–4
(IZII,ZYII) : 0.9912 ± 1.8e–4
(XIII,YXII) : 0.9898 ± 2.2e–4
(XZII,XZII) : 0.9881 ± 3.5e–4
(IZIZ,ZYZY) : 0.9857 ± 3.1e–4
(IZYI,ZYXX) : 0.9854 ± 3.6e–4
(IZYZ,ZYYZ) : 0.9840 ± 3.0e–4
(XIIZ,YXZY) : 0.9836 ± 2.3e–4
(XIYI,YXXX) : 0.9840 ± 3.4e–4
(XZIZ,XZZY) : 0.9826 ± 3.4e–4
(XIYZ,YXYZ) : 0.9826 ± 3.6e–4
(XZXX,XZYI) : 0.9821 ± 4.0e–4
(XZYZ,XZYZ) : 0.9809 ± 4.6e–4
Basis index = 2
(ZIII,ZIII) : 0.9942 ± 9.8e–5
(IIZI,IIZI) : 0.9946 ± 9.0e–5
(IIIY,IIZY) : 0.9932 ± 1.4e–4
(IZII,ZZII) : 0.9925 ± 1.5e–4
(ZIZI,ZIZI) : 0.9888 ± 1.8e–4
(ZIIY,ZIZY) : 0.9877 ± 1.9e–4
(IZZI,ZZZI) : 0.9873 ± 1.9e–4
(IZIY,ZZZY) : 0.9860 ± 2.2e–4
(IZZY,ZZIY) : 0.9861 ± 2.5e–4
Basis index = 3
(IIZI,IIZI) : 0.9946 ± 9.0e–5
(ZIII,ZIII) : 0.9942 ± 9.8e–5
(IIIZ,IIZZ) : 0.9940 ± 1.0e–4
(IXII,IXII) : 0.9923 ± 1.3e–4
(ZXII,ZXII) : 0.9895 ± 2.8e–4
(ZIZI,ZIZI) : 0.9888 ± 1.8e–4
(ZIIZ,ZIZZ) : 0.9881 ± 1.7e–4
(IXZI,IXZI) : 0.9868 ± 2.0e–4
(IXIZ,IXZZ) : 0.9863 ± 2.0e–4
(ZXIZ,ZXZZ) : 0.9838 ± 2.4e–4
(ZXZI,ZXZI) : 0.9840 ± 3.1e–4
Basis index = 4
(IIXX,IIYI) : 0.9940 ± 1.3e–4
(IIIY,IIZZ) : 0.9940 ± 1.2e–4
(IXII,IXII) : 0.9923 ± 1.3e–4
(IIYY,IIYY) : 0.9925 ± 1.6e–4
(XIII,XXII) : 0.9903 ± 2.1e–4
(IXIY,IXZZ) : 0.9866 ± 3.2e–4
(IXXX,IXYI) : 0.9866 ± 3.2e–4
(IXYY,IXYY) : 0.9851 ± 3.3e–4
(XIYI,XXXX) : 0.9842 ± 3.8e–4
(XIIY,XXZZ) : 0.9842 ± 3.6e–4
(XIZZ,XXIY) : 0.9844 ± 3.4e–4
(XIXX,XXYI) : 0.9843 ± 3.4e–4
(XIYY,XXYY) : 0.9829 ± 2.4e–4
Basis index = 5
(IIIZ,IIZY) : 0.9941 ± 1.0e–4
(IIXI,IIYX) : 0.9939 ± 1.3e–4
(IIXZ,IIXZ) : 0.9929 ± 1.6e–4
(IYII,ZZII) : 0.9913 ± 1.8e–4
(XIII,YXII) : 0.9898 ± 2.2e–4
(XYII,XYII) : 0.9889 ± 3.4e–4
(IYIZ,ZZZY) : 0.9856 ± 3.2e–4
(IYXI,ZZYX) : 0.9855 ± 3.3e–4
(IYXZ,ZZXZ) : 0.9845 ± 3.2e–4
(XIXI,YXYX) : 0.9835 ± 3.8e–4
(XIIZ,YXZY) : 0.9836 ± 2.3e–4
(XYXI,XYYX) : 0.9825 ± 3.8e–4
(XYIZ,XYZY) : 0.9822 ± 3.7e–4
(XIXZ,YXXZ) : 0.9824 ± 3.7e–4
(XYXZ,XYXZ) : 0.9814 ± 4.4e–4
Basis index = 6
(IIIX,IIIX) : 0.9959 ± 7.8e–5
(IIXI,IIXX) : 0.9933 ± 1.3e–4
(IYII,ZZII) : 0.9913 ± 1.8e–4
(YYII,YYII) : 0.9893 ± 3.4e–4
(XXII,YIII) : 0.9880 ± 2.3e–4
(IYIX,ZZIX) : 0.9872 ± 2.9e–4
(YYIX,YYIX) : 0.9855 ± 3.5e–4
(XXIX,YIIX) : 0.9843 ± 2.9e–4
(IYXI,ZZXX) : 0.9847 ± 3.6e–4
(IYXX,ZZXI) : 0.9846 ± 3.6e–4
(YYXI,YYXX) : 0.9830 ± 2.7e–4
(XXXX,YIXI) : 0.9821 ± 4.5e–4
(XXXI,YIXX) : 0.9821 ± 4.9e–4
Basis index = 7
(IIIX,IIIX) : 0.9959 ± 7.8e–5
(IIYI,IIYX) : 0.9938 ± 1.2e–4
(IZII,ZYII) : 0.9912 ± 1.8e–4
(XXII,YIII) : 0.9880 ± 2.3e–4
(IZIX,ZYIX) : 0.9873 ± 3.0e–4
(YZII,YZII) : 0.9874 ± 3.8e–4
(IZYI,ZYYX) : 0.9849 ± 3.2e–4
(IZYX,ZYYI) : 0.9850 ± 3.0e–4
(XXIX,YIIX) : 0.9843 ± 2.9e–4
(YZIX,YZIX) : 0.9836 ± 3.6e–4
(XXYI,YIYX) : 0.9820 ± 4.0e–4
(XXYX,YIYI) : 0.9817 ± 4.1e–4
(YZYI,YZYX) : 0.9816 ± 3.0e–4
Depth (k)
Basis index = 8
(IIIX,IIIX) : 0.9959 ± 7.8e–5
(IIZI,IIZI) : 0.9946 ± 9.0e–5
(IXII,IXII) : 0.9923 ± 1.3e–4
(IIZX,IIZX) : 0.9909 ± 2.3e–4
(YIII,YXII) : 0.9895 ± 2.6e–4
(IXIX,IXIX) : 0.9888 ± 2.9e–4
(IXZI,IXZI) : 0.9868 ± 2.0e–4
(YIIX,YXIX) : 0.9854 ± 2.4e–4
(YIZI,YXZI) : 0.9838 ± 2.8e–4
(IXZX,IXZX) : 0.9836 ± 3.8e–4
(YIZX,YXZX) : 0.9804 ± 3.0e–4

Reconstructing quantum channel from measurement data

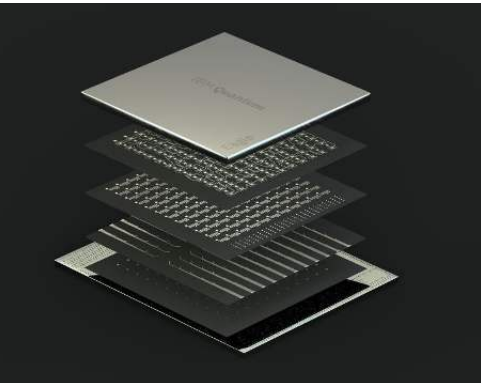


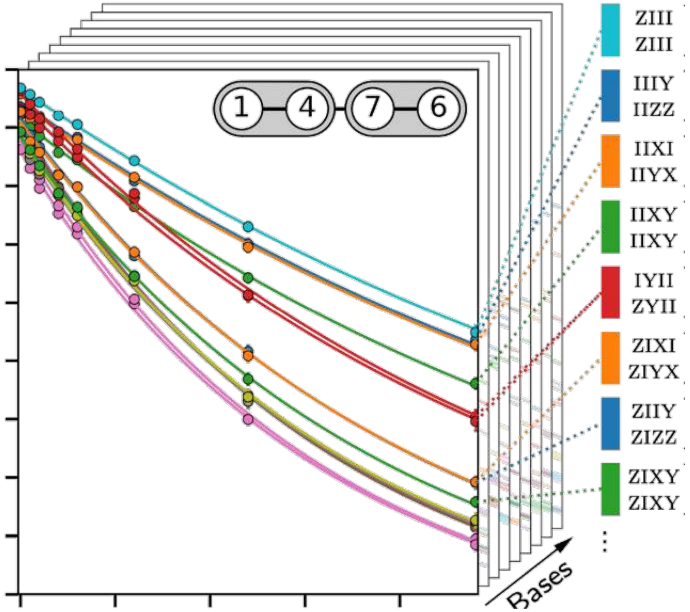


$$\Lambda(\rho) = \sum_{a=0}^{4^n-1} \boxed{c_a} P_a \rho P_a^\dagger$$

Still 4^n

Sparse Pauli-Lindblad model

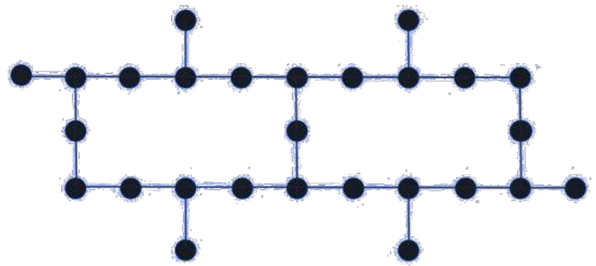




$$\Lambda(\rho) = \sum_{a=0}^{4^n-1} \boxed{c_a} P_a \rho P_a^\dagger$$

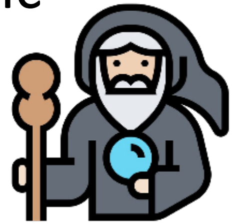
$$\Lambda(\rho) = \exp[\mathcal{L}](\rho)$$

$$\mathcal{L}(\rho) = \sum_{k \in \mathcal{K}} \lambda_k (P_k \rho P_k - \rho)$$



Highlight: Ewout van den Berg

Magic



icon: Eucalyp

Notes

Dissipator for a given Pauli

$$\mathcal{D}[P](\rho) = P\rho P^\dagger - \frac{1}{2}\left(PP^\dagger\rho + \rho PP^\dagger\right)$$

$$= P\rho P^\dagger - \rho$$

$$= \left(P \cdot P - I\right)\rho$$

$$\hat{\hat{\mathcal{L}}} = \sum_a \gamma_a \left(\hat{\hat{\mathcal{P}}}_a - \hat{\hat{I}}\right)$$

$$= \sum_a \gamma_a \mathcal{D}[\hat{P}_a]$$

$$e^{\hat{\hat{\mathcal{L}}}} = \prod_a \exp\left(\gamma_a \mathcal{D}[\hat{P}_a]\right)$$

$$= \prod_a \Lambda[P_a]$$

but $\left[\mathcal{D}[P_a], \mathcal{D}[P_b]\right] = 0$

Each sub channel

$$\Lambda[P_a] = \exp(\gamma \mathcal{D}[\rho])$$

$$= \exp(\gamma[\hat{\vec{P}}_a + \hat{\vec{I}}])$$

$$= \exp(\gamma \hat{\vec{P}}_a)\exp(\gamma \hat{\vec{I}})\exp(\tfrac{-1}{2}\gamma^2 \underbrace{\cancel{[\hat{\vec{P}}_a, \hat{\vec{I}}]}}_{0}) \cdots$$

$$= \exp(\gamma \hat{\vec{P}}_a)\exp(\gamma \hat{\vec{I}})$$

$$\hookrightarrow \sum_n \frac{\gamma^n \hat{\vec{I}}^n}{n!} = \sum_n \frac{\gamma^n}{n!} \hat{\vec{I}}$$

$$= \exp(\gamma)\hat{\vec{I}} = \exp(\gamma)\hat{I} \cdot \hat{I}$$

Each sub channel

$$= \exp(\gamma \hat{P}_a)$$

$$\hookrightarrow \sum_{n=0}^{\infty} \frac{\gamma^n}{n!} (\hat{P}_a)^n$$

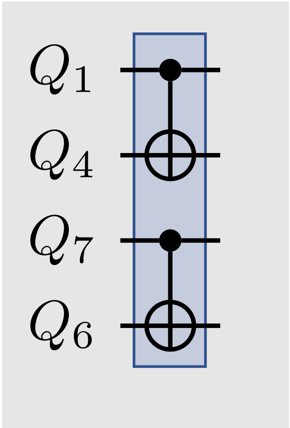
note $\hat{P}_a^{\,n-\text{even}} = \hat{P}_a$

$\hat{P}_a^{\,n-\text{odd}} = \hat{I}$

$$= \left(\sum_{\substack{n=0 \\ \text{even}}}^{\infty} \frac{\gamma^n}{n!} \right) \hat{I} + \left(\sum_{\substack{n=1 \\ \text{odd}}}^{\infty} \frac{\gamma^n}{n!} \right) \hat{P}_a$$

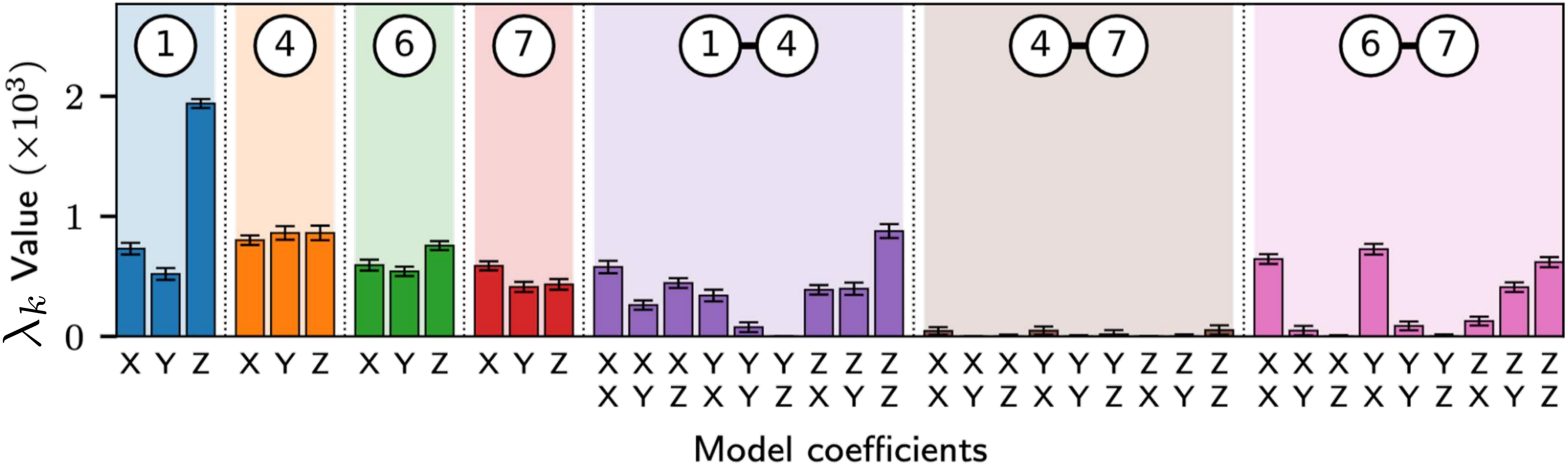
$$= \cosh(\gamma)\, \hat{I} + \sinh(\gamma)\, \hat{P}_a$$

$$\Lambda[P_a] = \cosh(\gamma_a)\, \hat{I} + \sinh(\gamma_a)\, \hat{P}_a$$

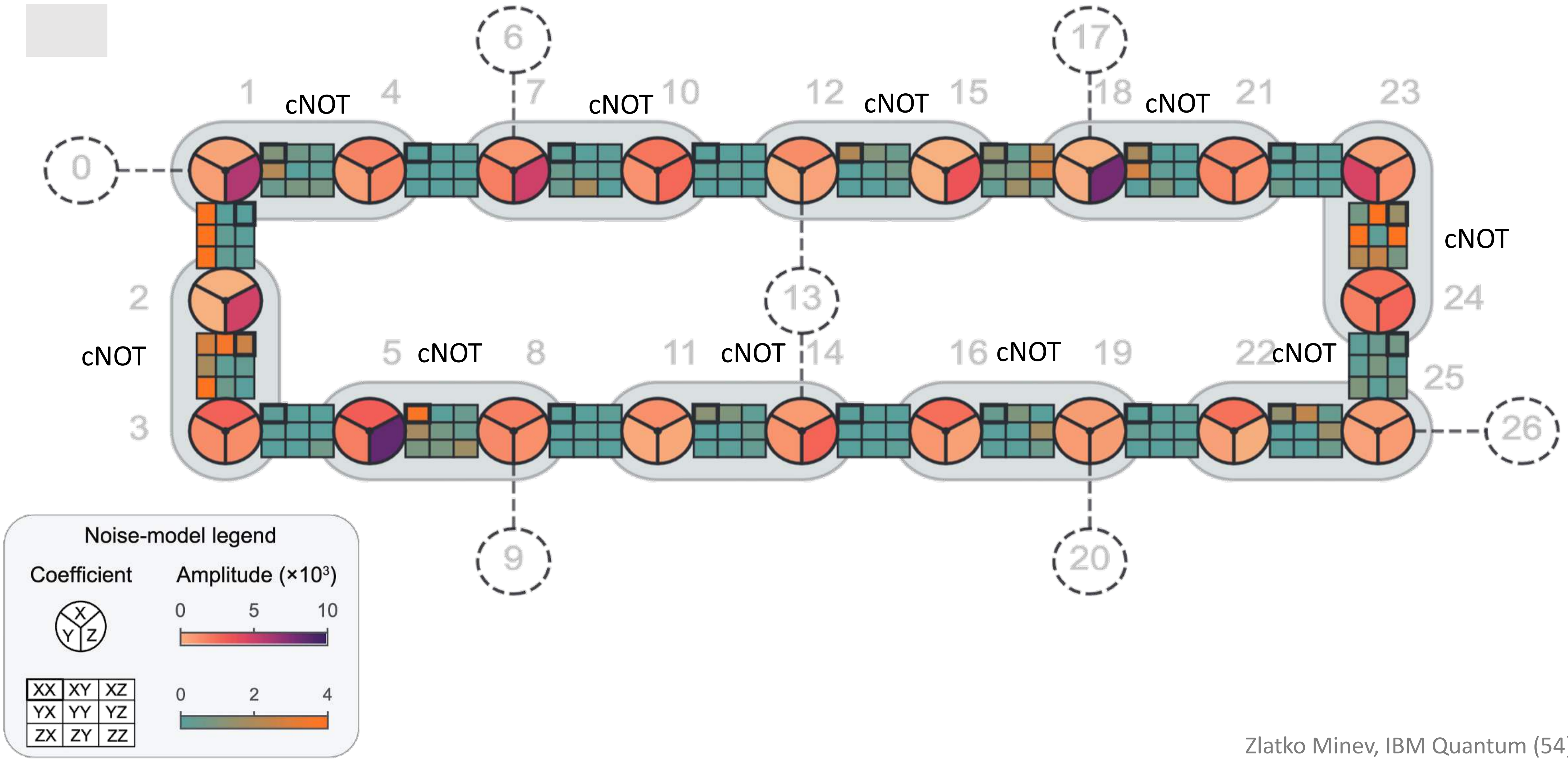


Sparse Lindblad tomogram

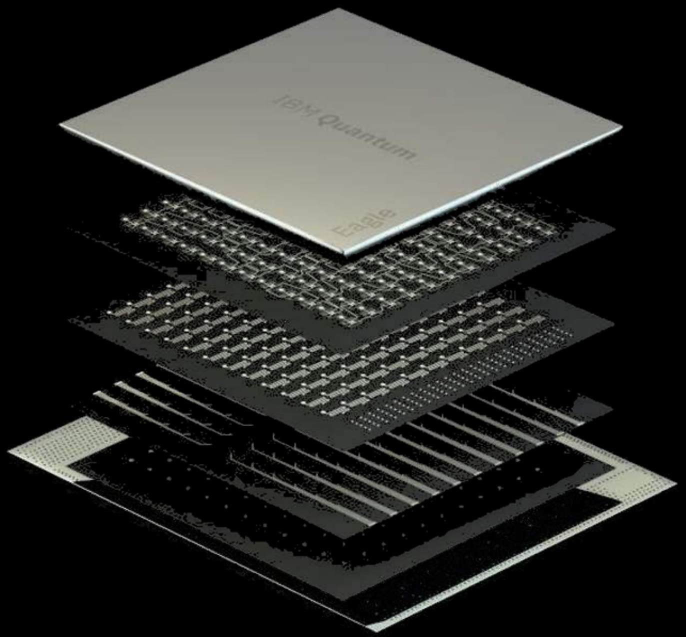
$$\mathcal{L}(\rho) = \sum_{k \in \mathcal{K}} \boxed{\lambda_k} \left(P_k \rho P_k^\dagger - \rho \right)$$

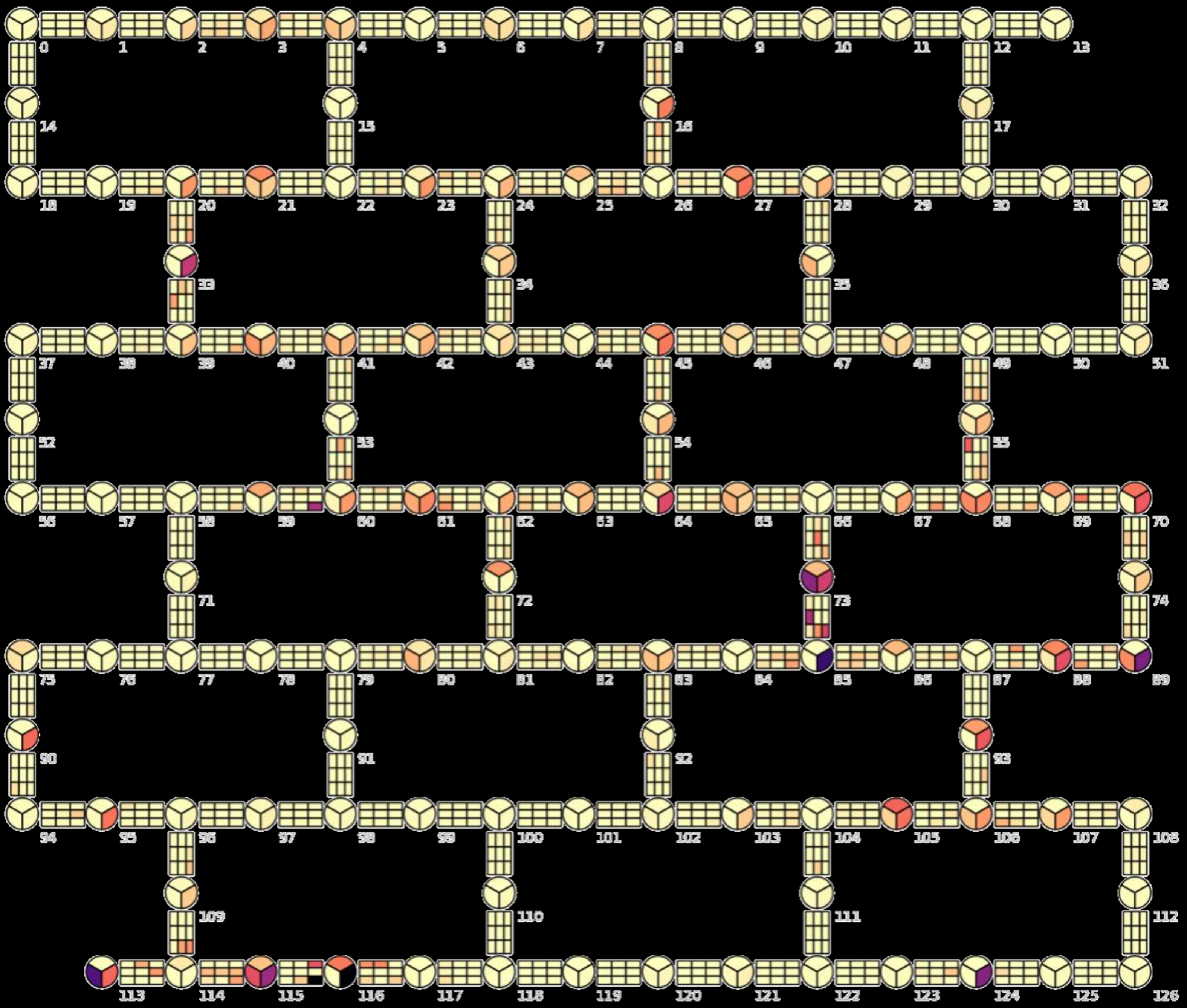


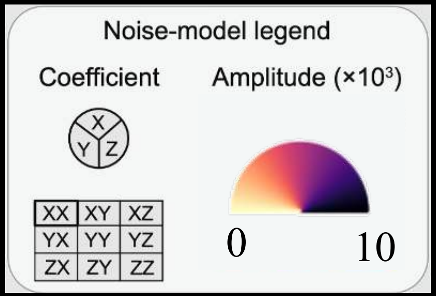
Noise tomogram for 20Q Ising-ring Trotter layer



Noise tomogram for 127Q Trotter layer







Same number of learning circuits as for 4Q

Zlatko Minev, IBM Quantum

Simulating transverse-field Ising model time evolution with PEC

Spins

$$H = -J\sum_j Z_j Z_{j+1} + h\sum_j X_j$$

J: exchange coupling between neighboring spins
h: transverse magnetic field

Average magnetization density?

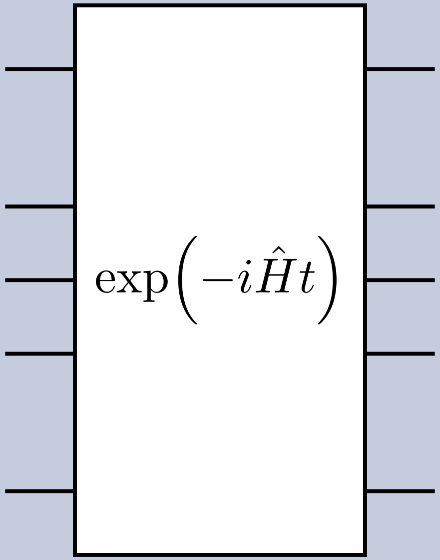
$$\vec{M} := \sum_n \left(\langle X\rangle_n, \langle Y\rangle_n, \langle Z\rangle_n\right)/N$$

Top image: Samuel Velasco/Quanta Magazine

Step 1: Map to quantum circuit

$$H = -J\sum_j Z_j Z_{j+1} + h\sum_j X_j$$

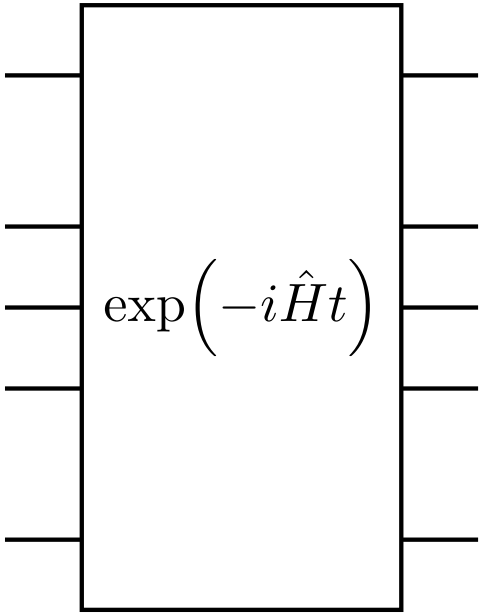
Time evolution



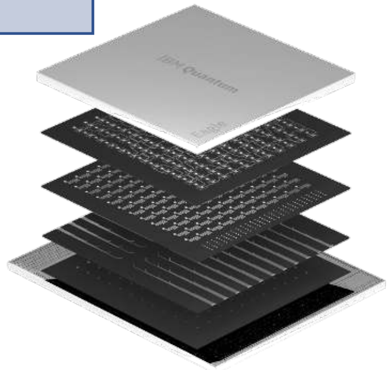
See lecture by
Frank Pollmann

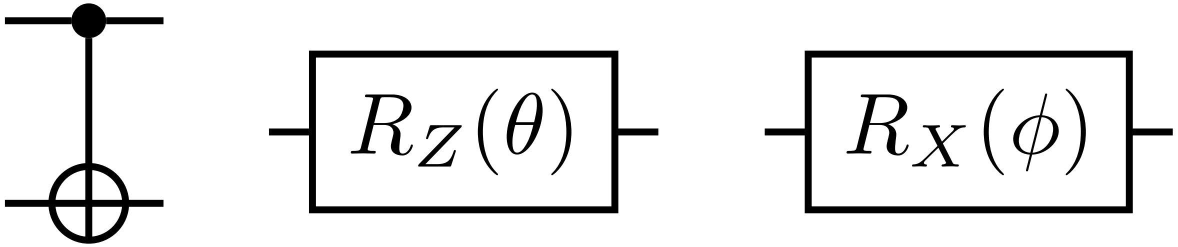
Step 1: Map to quantum circuit

$$H = -J\sum_j Z_j Z_{j+1} + h\sum_j X_j$$



Decompose into native gates that
we can actually implement on the QPU

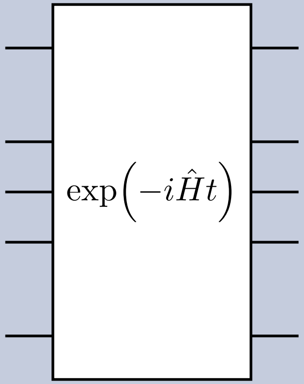




Step 1: Trotter circuit (30 sec)

The product formula describing the time evolution of a quantum system over time *t* is

See lecture by Frank



$$\exp\left(-i\hat{H}t\right) \approx \prod_{d=1}^{D} \hat{U}_k\left(\Delta t\right) \ , \qquad \hat{U}_1(\Delta t) = \prod_{j=1}^{N} e^{-i\hat{H}_j \Delta t} \ ,$$

U_k: Unitary time evolution over a finite Trotter time-step delta t for k-th order Trotter-Suzuki product formula

First order expansion

$$\hat{H} = -J\sum_j Z_j Z_{j+1} + h\sum_j X_j = -J\hat{H}_{ZZ} + h\hat{H}_X$$

$$\hat{H}_{ZZ} = \sum_j Z_j Z_{j+1}$$

$$\hat{H}_X = \sum_j \hat{X}_j$$

$$\exp\left(-i\hat{H}t\right) \approx \left[\exp\left(iJ\hat{H}_{ZZ}t/d\right)\exp\left(-ih\hat{H}_X t/d\right)\right]^d$$

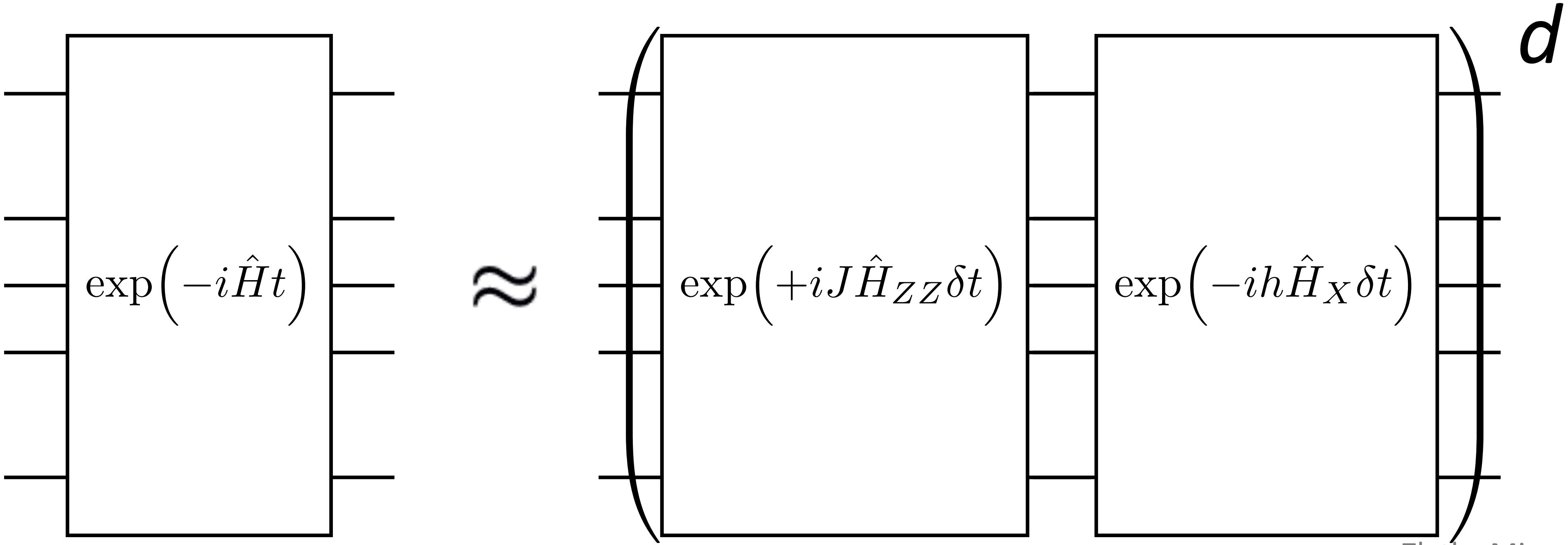
$$\delta t := t/d$$

First order Trotter

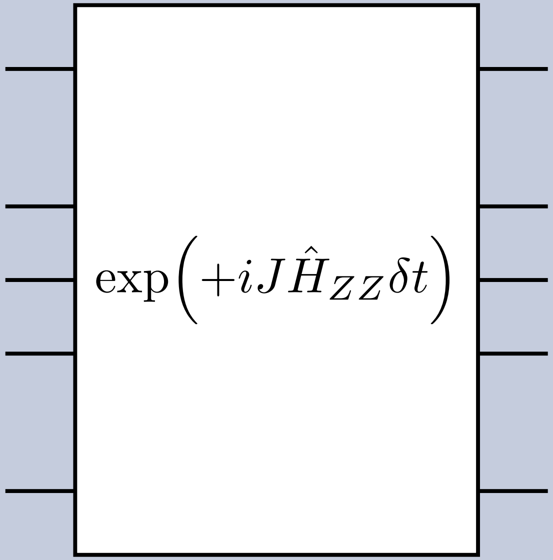
For error bounds, see arxiv:2302.14592

Step 1: Map to quantum circuit

$$H = -J\sum_j Z_j Z_{j+1} + h\sum_j X_j$$



Step 1: Trotter circuit (30 s)

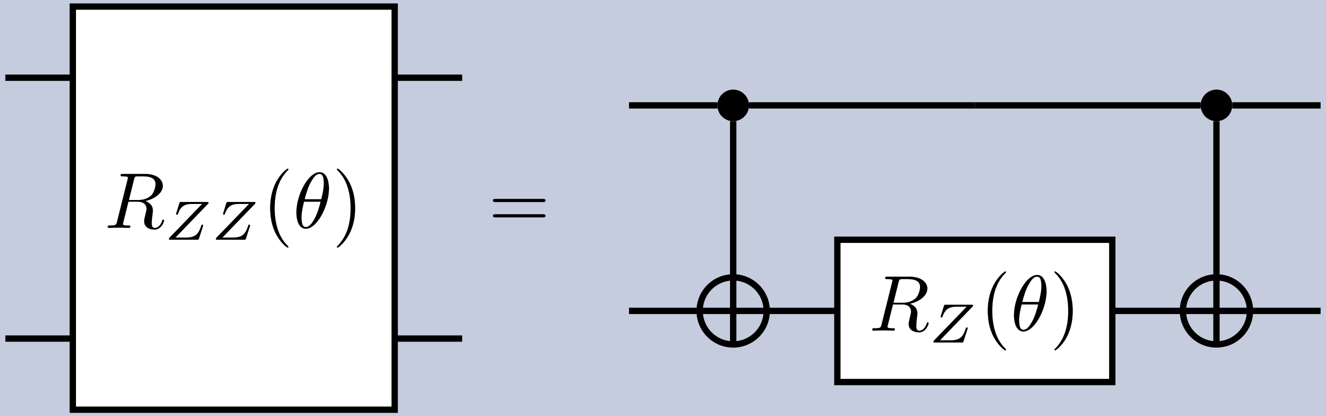


$$\exp\left(-i\frac{\theta}{2}\sum_{j=0}^{n-1} Z_j Z_{j+1}\right) = \quad (\text{use } [Z_j Z_{j+1}, Z_{j'} Z_{j'+1}] = 0)$$

$$\theta := -2J\delta t$$

$$= \prod_{j=0}^{n-1} \exp\left(-i\frac{\theta}{2} Z_j Z_{j+1}\right)$$

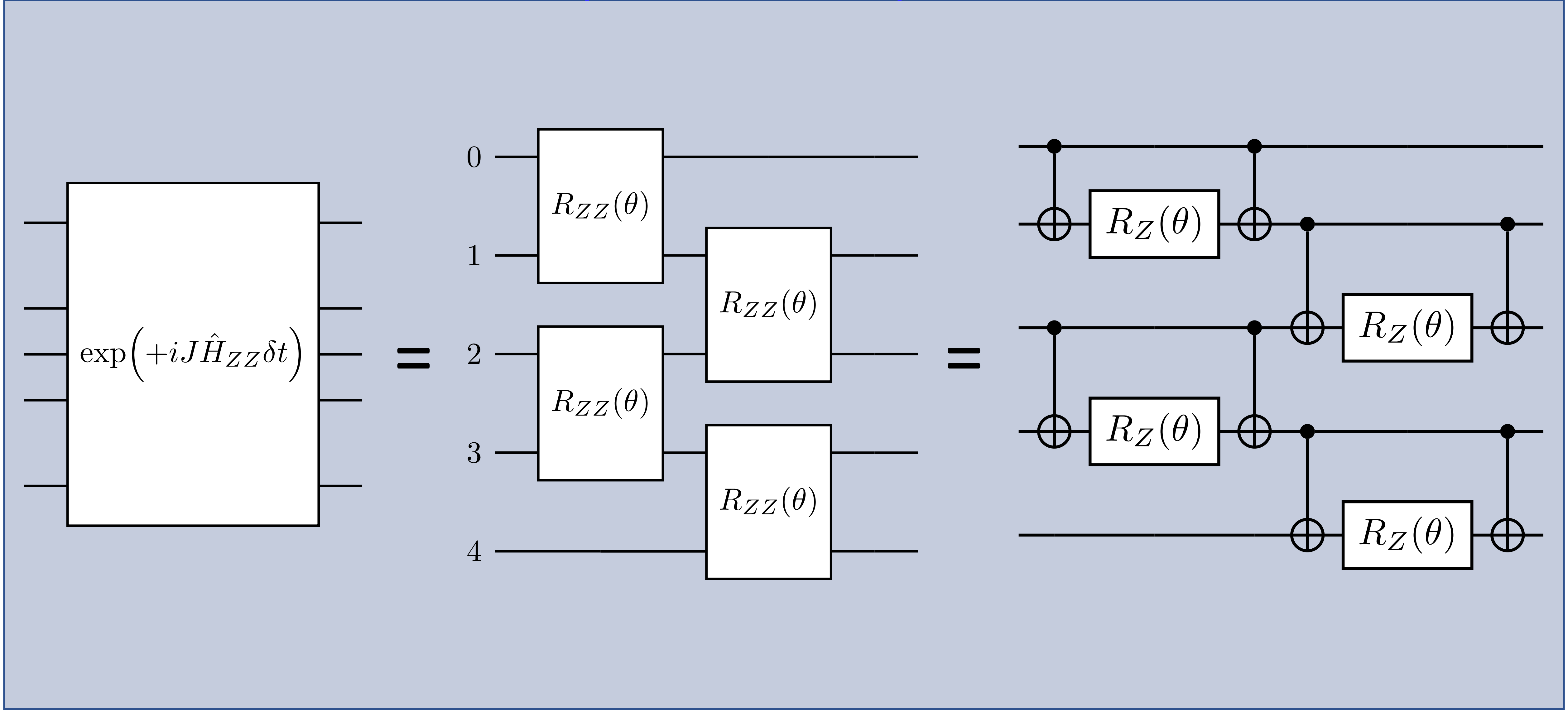
$$= \prod_{j=0}^{n-1} R_{ZZ}(\theta)$$



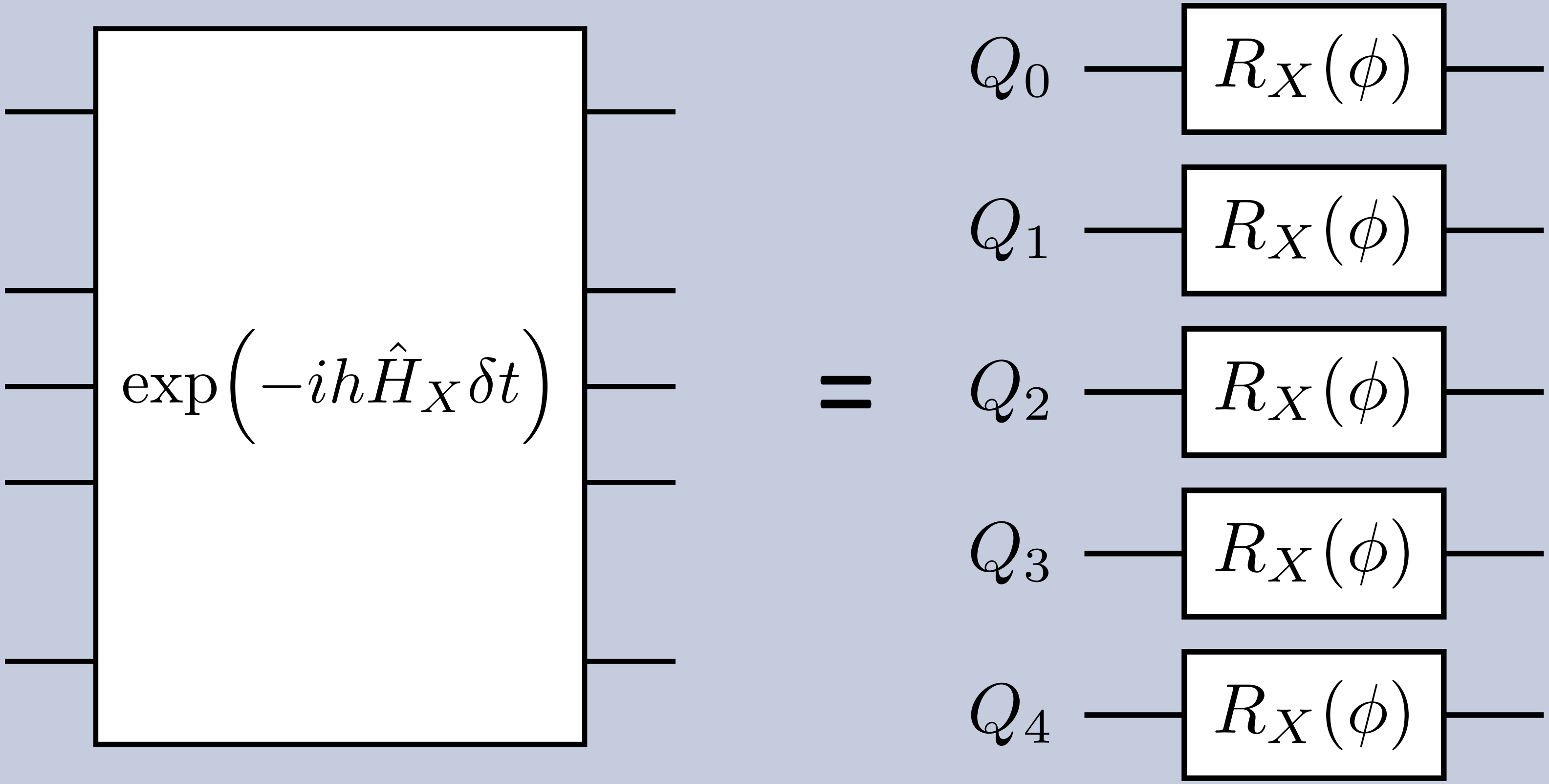
$$R_{ZZ}(\theta) = \exp\left(-i\frac{\theta}{2} ZZ\right)$$

$$= \mathrm{cXRz}(\theta)\,\mathrm{cX}$$

Step 1: Decompose into native gates



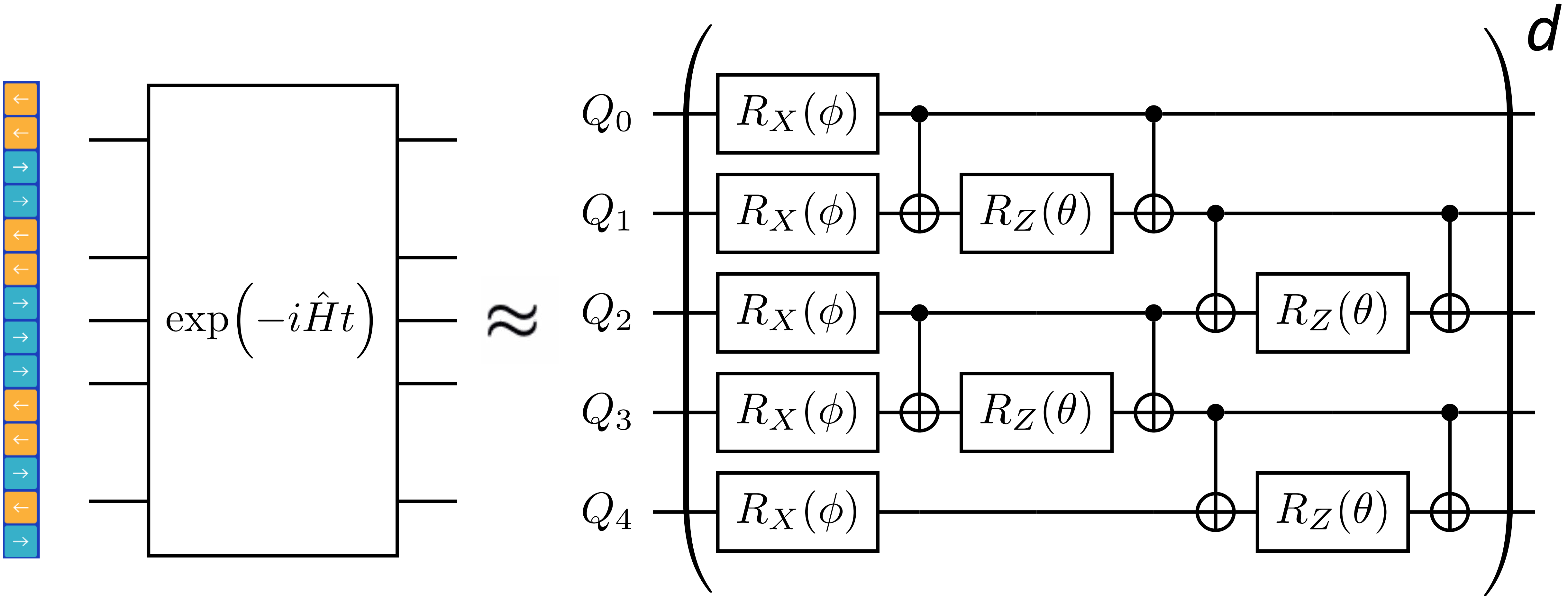
Step 1: Decompose into native gates



$$R_X(\phi) := \exp\left(-\frac{1}{2}\phi X\right)$$

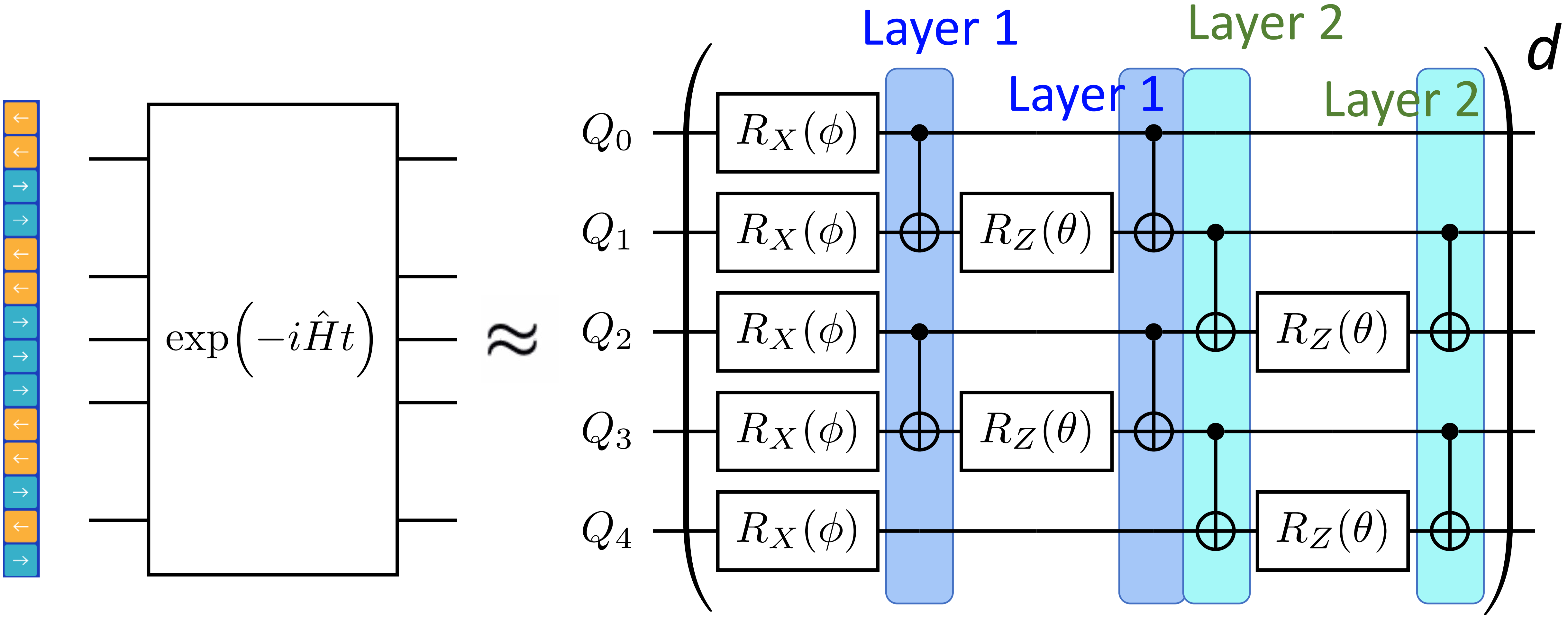
$$\phi := +2h\delta t$$

Step 1: Quantum Hamiltonian time evolution



Decomposed into quantum computer native gates

Step 2: Decompose into layers

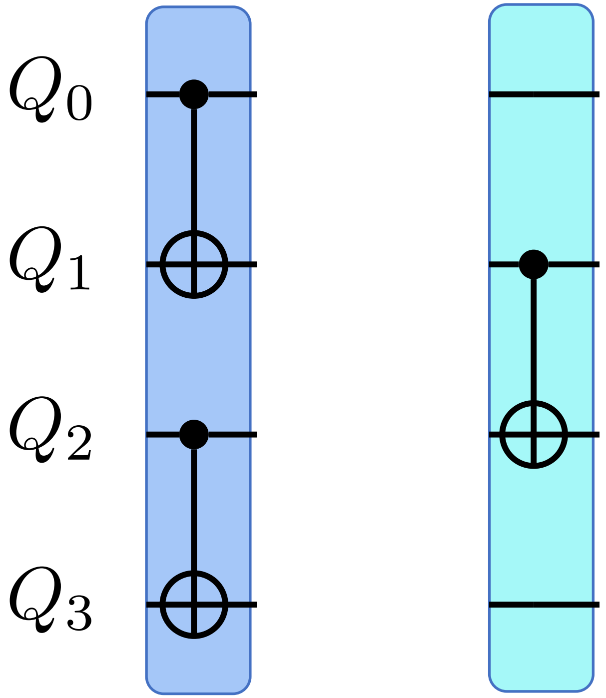


Decomposed into 2Q gate layers

Step 2: Learn the noise on each layer

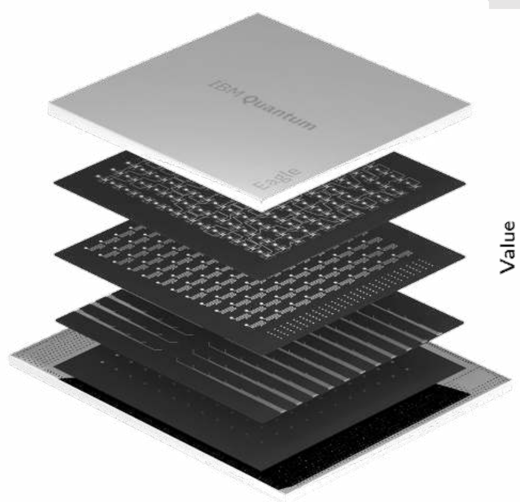
Let's make even simpler, and make it 4 qubits (instead of 5 for our first example)

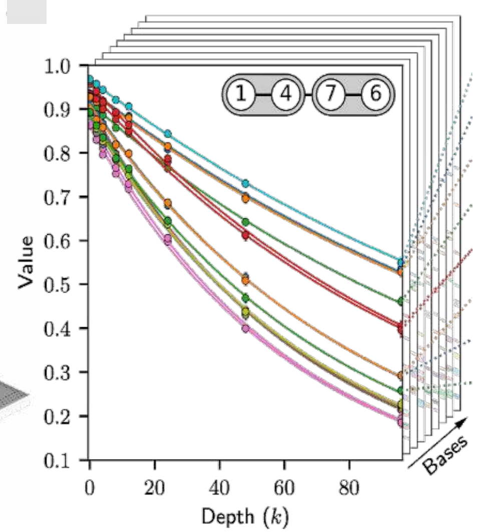
Layer 1 Layer 2

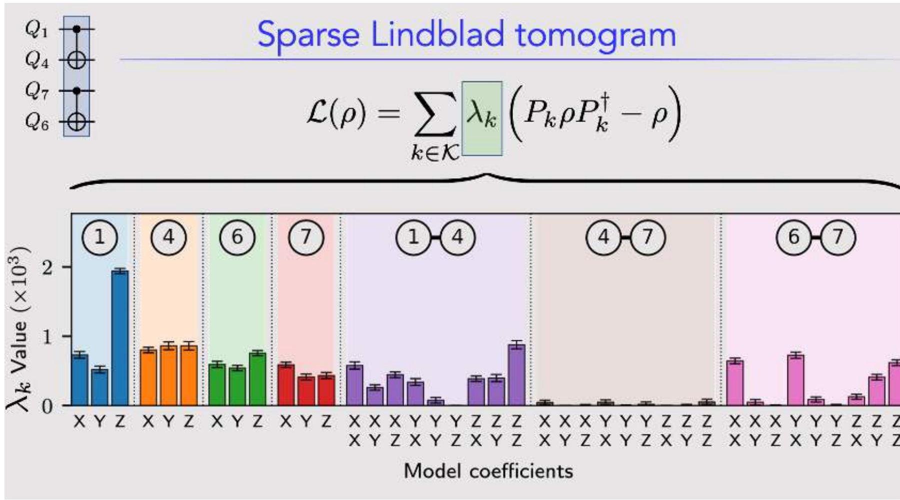


Λ_1 Λ_2

See lecture 2 for learning the noise



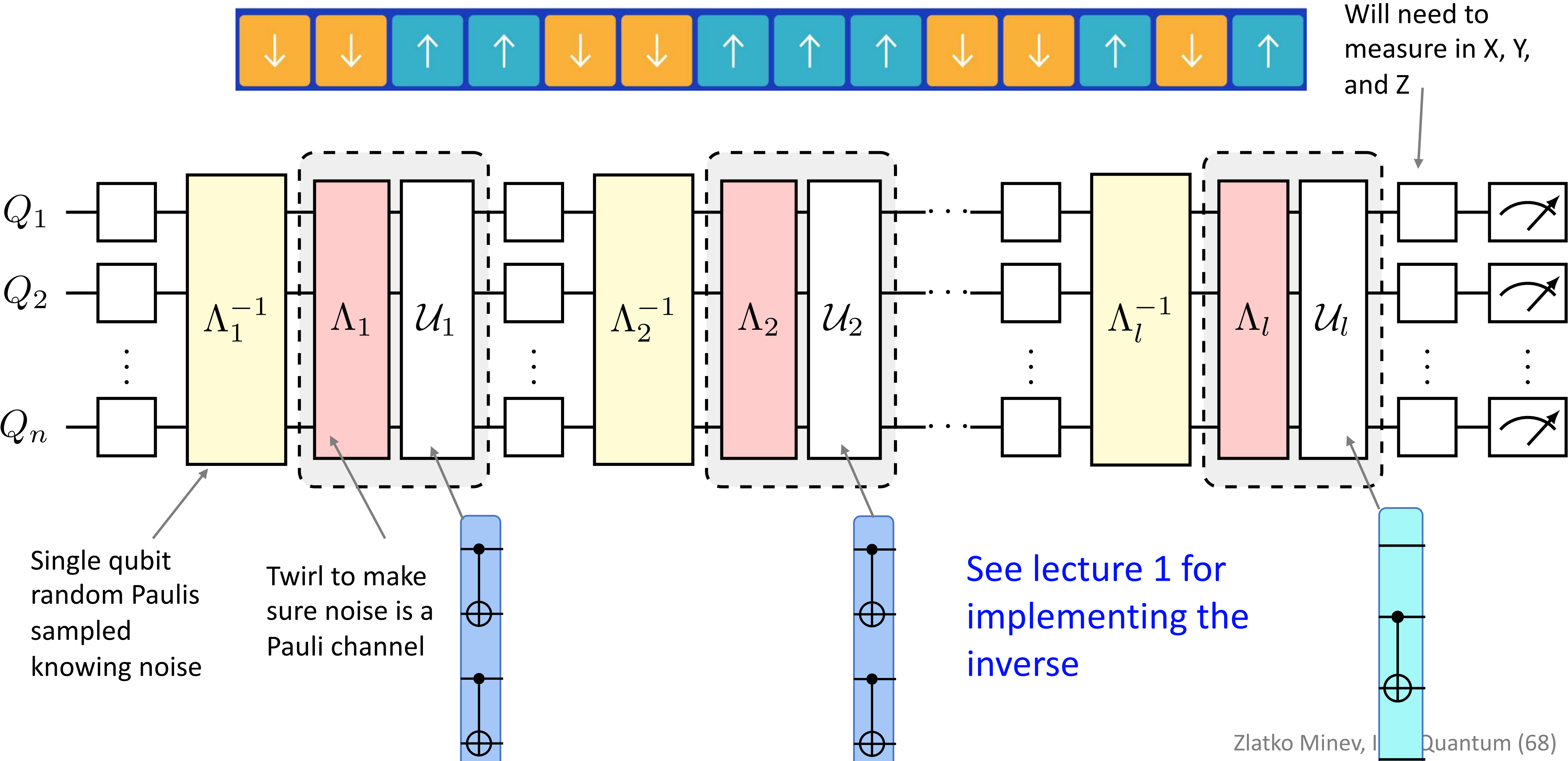




$$\gamma_1 = 1.0309 \pm 8.40 \cdot 10^{-5}$$

$$\gamma_2 = 1.0384 \pm 2.20 \cdot 10^{-4}$$

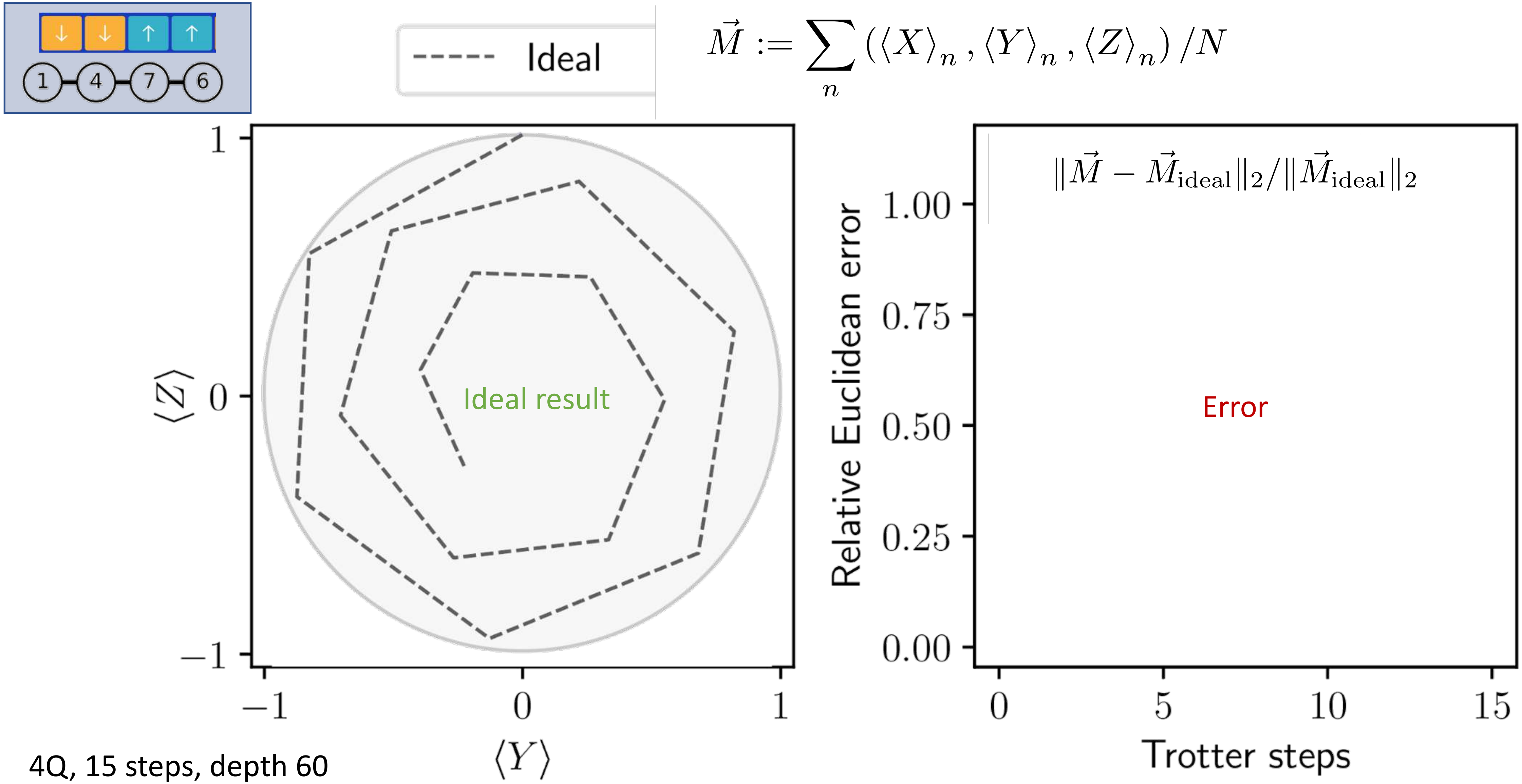
Step 3: Cancel the noise



See lecture 1 for implementing the inverse

Ideal Ising model evolution

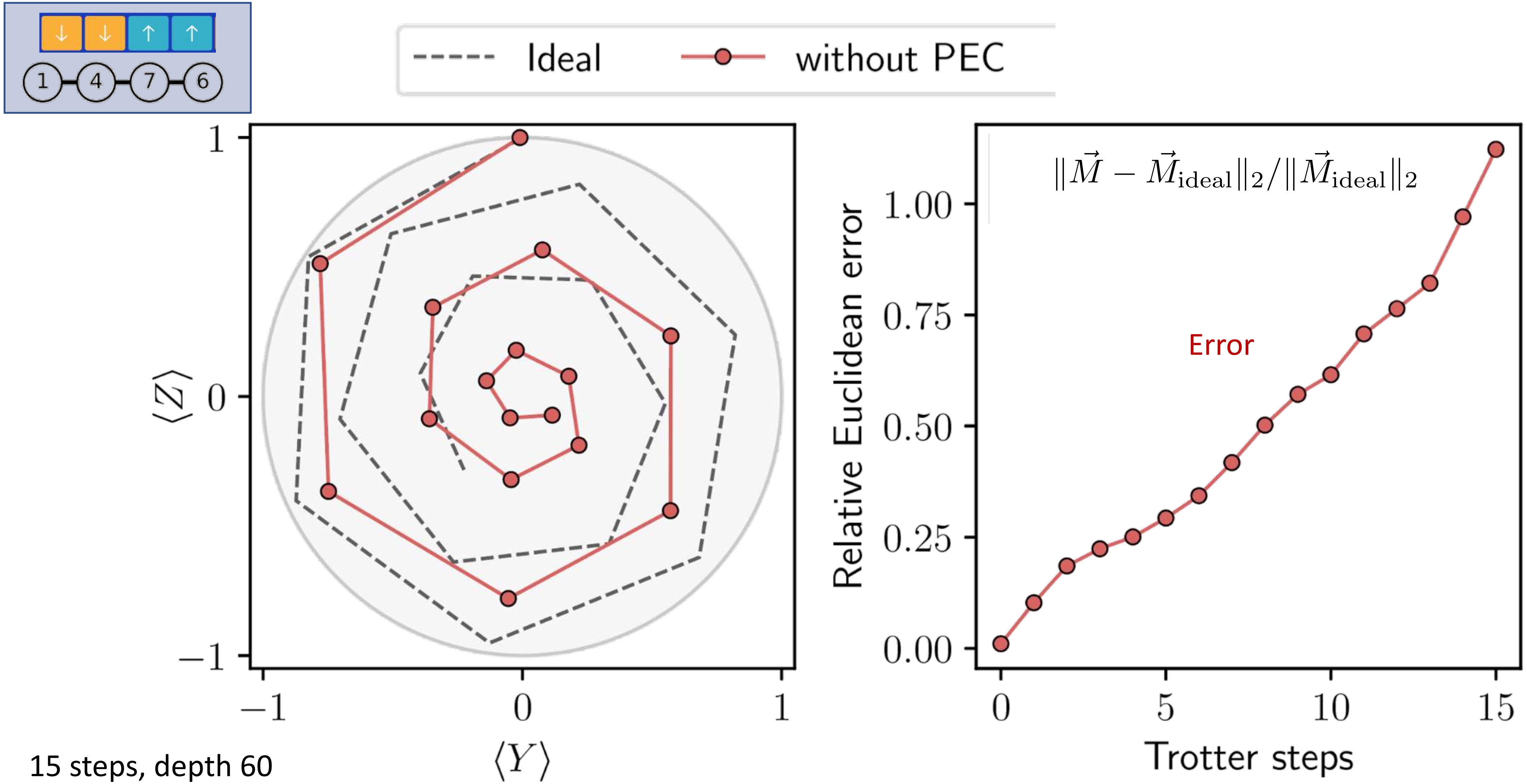
$$\vec{M} := \sum_n \left(\langle X \rangle_n , \langle Y \rangle_n , \langle Z \rangle_n \right) / N$$



4Q, 15 steps, depth 60

$h = 1, J = -0.15, \delta t = 1/4$

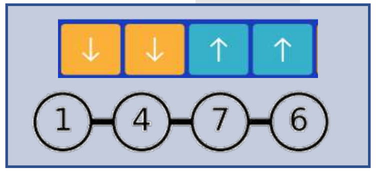
Without PEC: but with DD & twirl readout mitigation

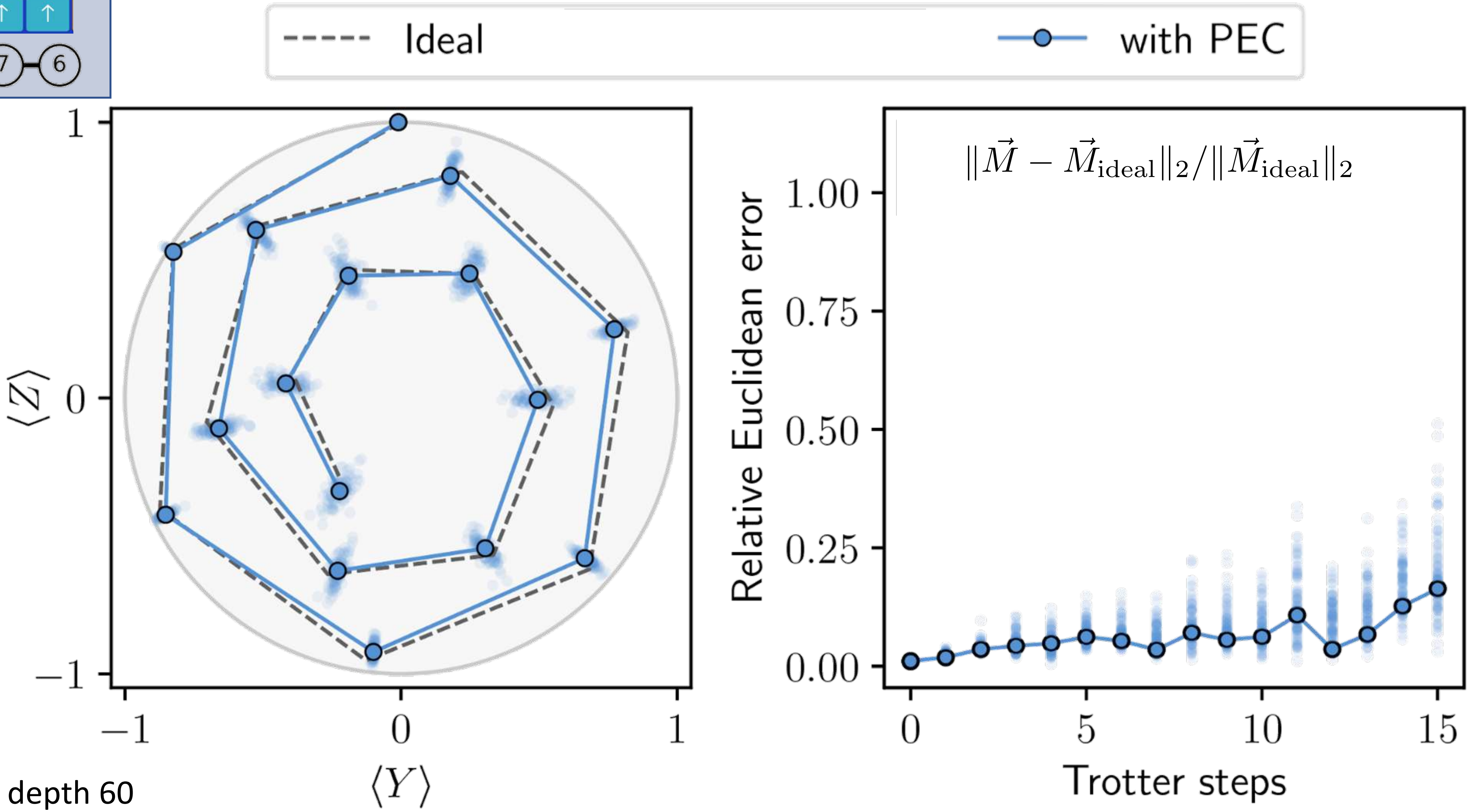


15 steps, depth 60

$h = 1, J = -0.15, \delta t = 1/4$

With PEC

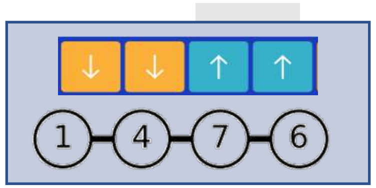


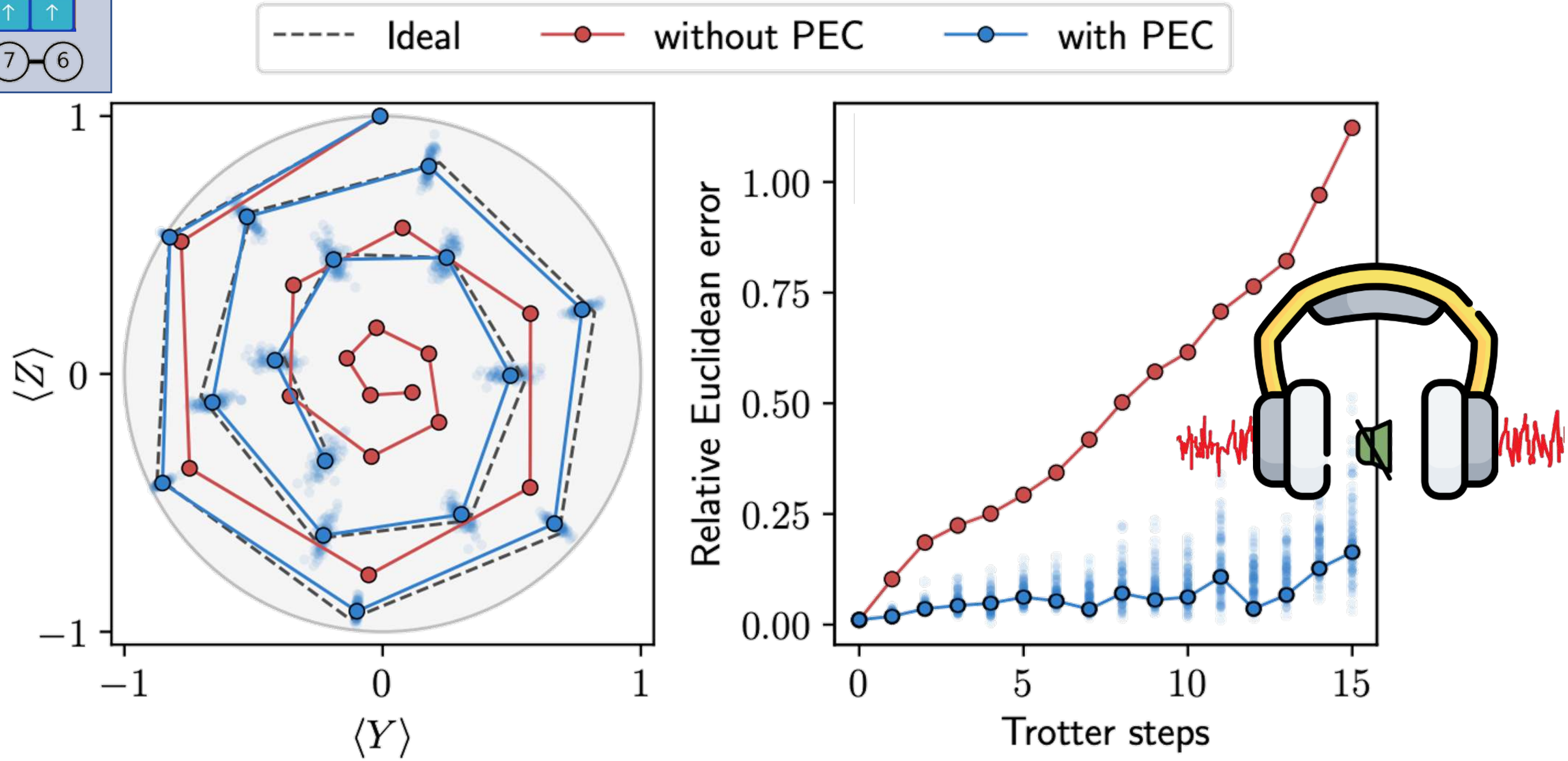


15 steps, depth 60

$h = 1, J = -0.15, \delta t = 1/4$

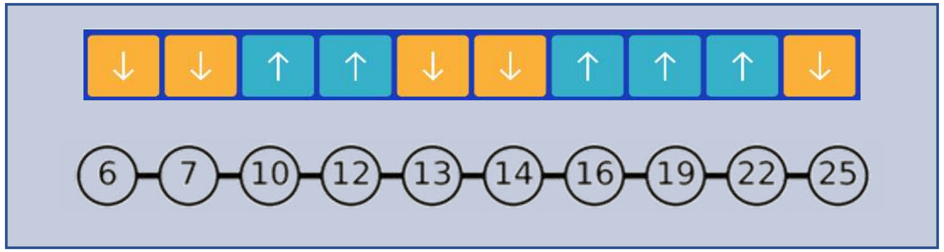
With vs. without PEC

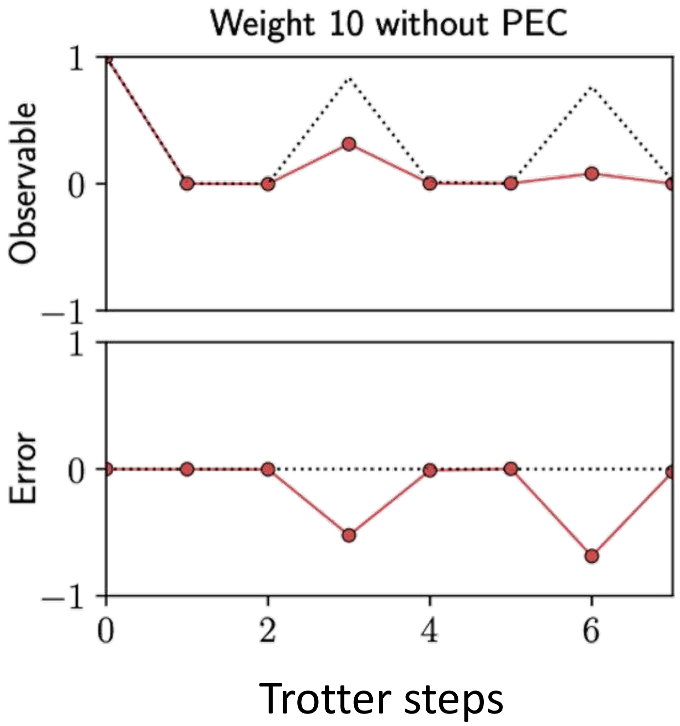




$h = 1, J = -0.15, \delta t = 1/4$

With vs without PEC: 10 qubit high-weight observables

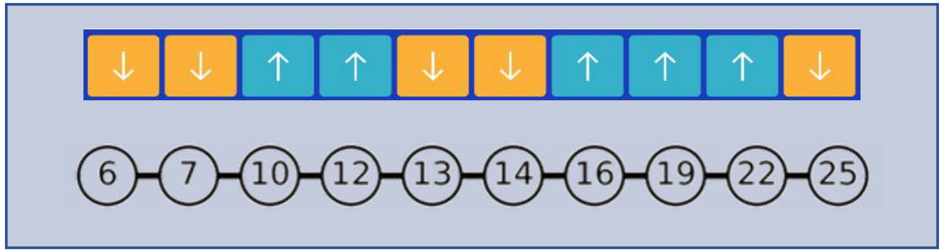


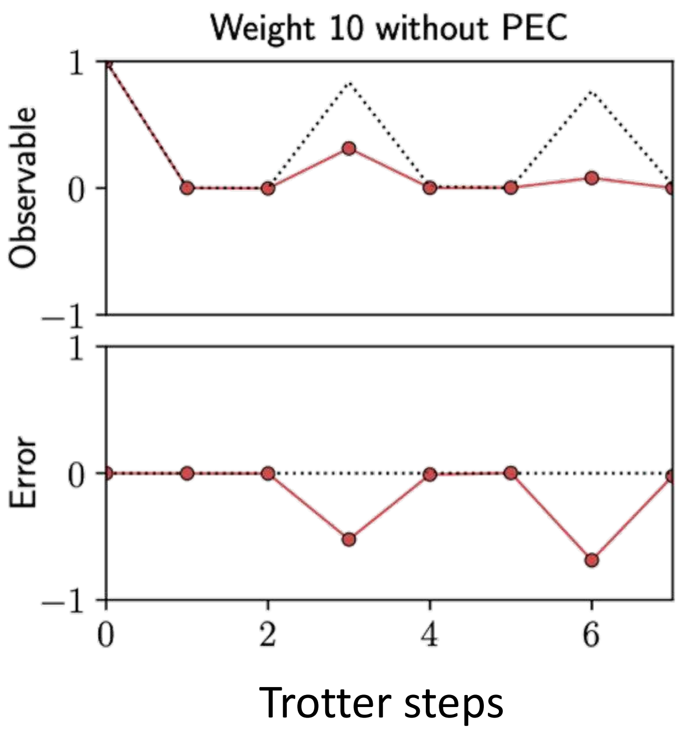


ZZZZZZZZZZ

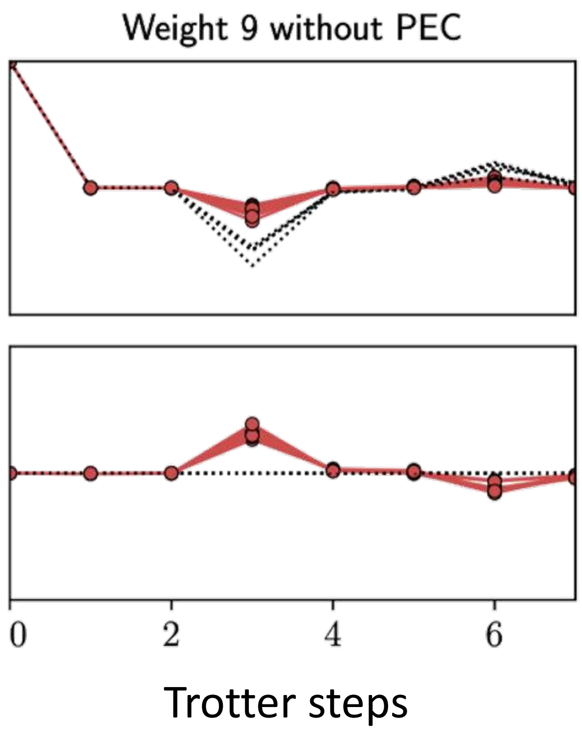
$h = 1, J = -0.5236, \delta t = 1/4$

With vs without PEC: 10 qubit high-weight observables





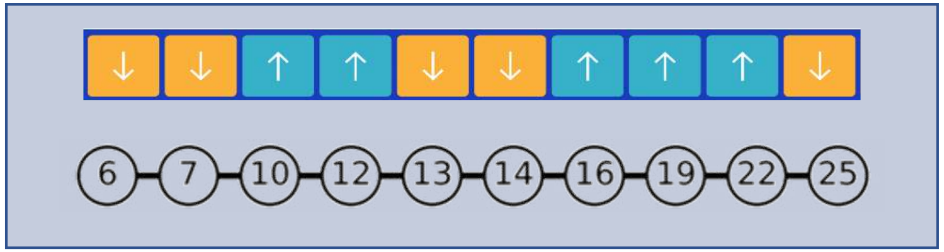
ZZZZZZZZZZ

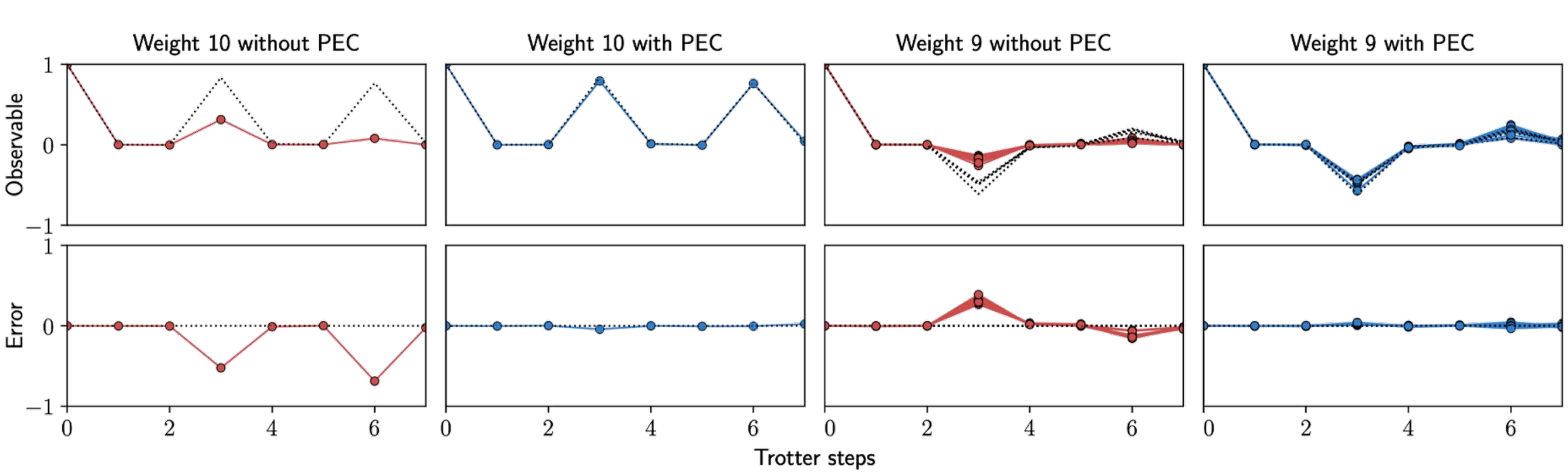


IZZZZZZZZZ
ZIZZZZZZZZ
ZZIZZZZZZZ
…
ZZZZZZZZZI

$h = 1, J = -0.5236, \delta t = 1/4$

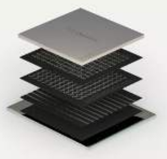
With vs without PEC: 10 qubit high-weight observables



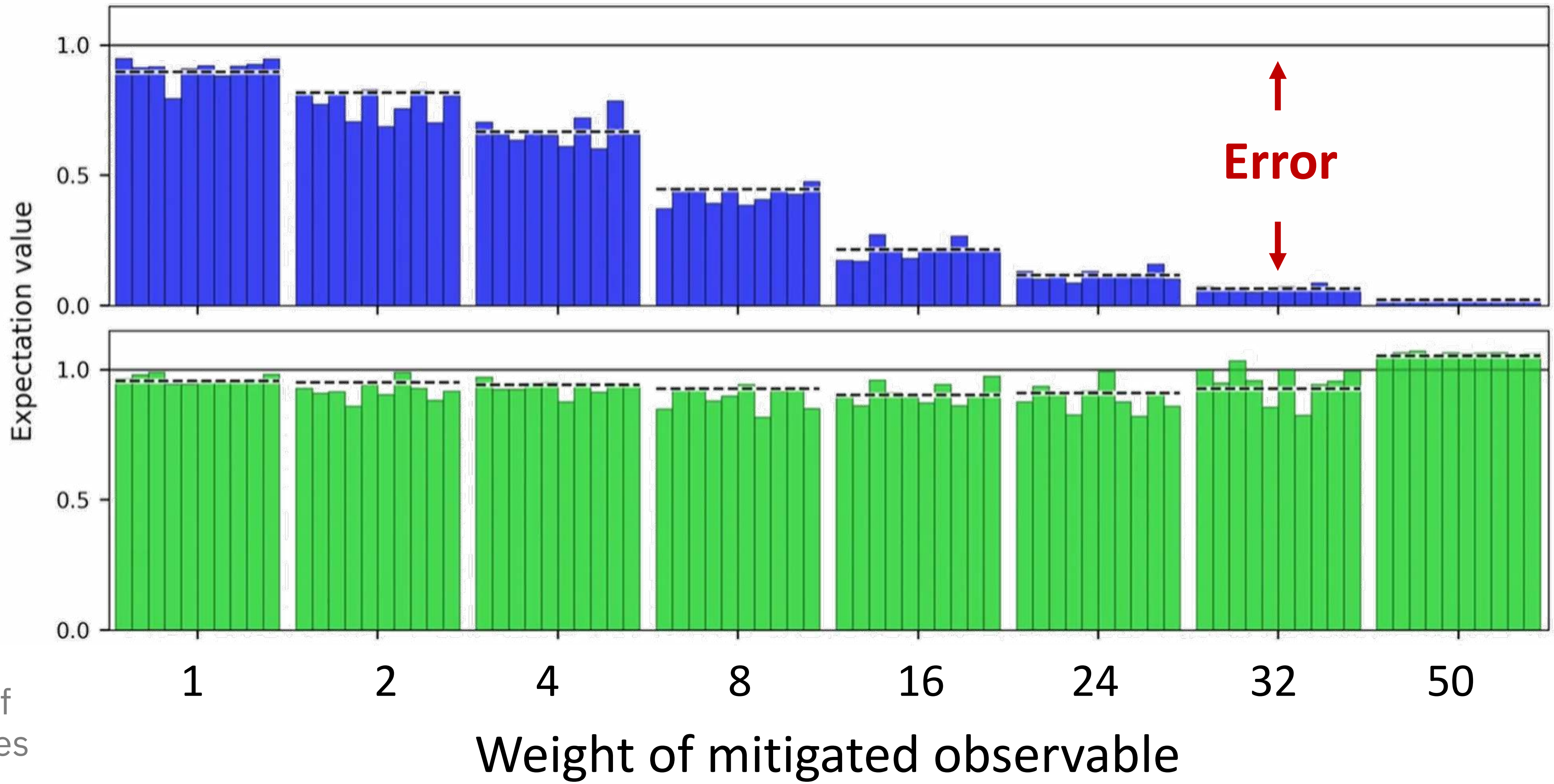


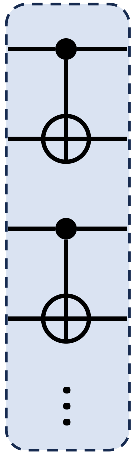
$h = 1, J = -0.5236, \delta t = 1/4$

Scaling and error budget



PEC on 50 qubit observables



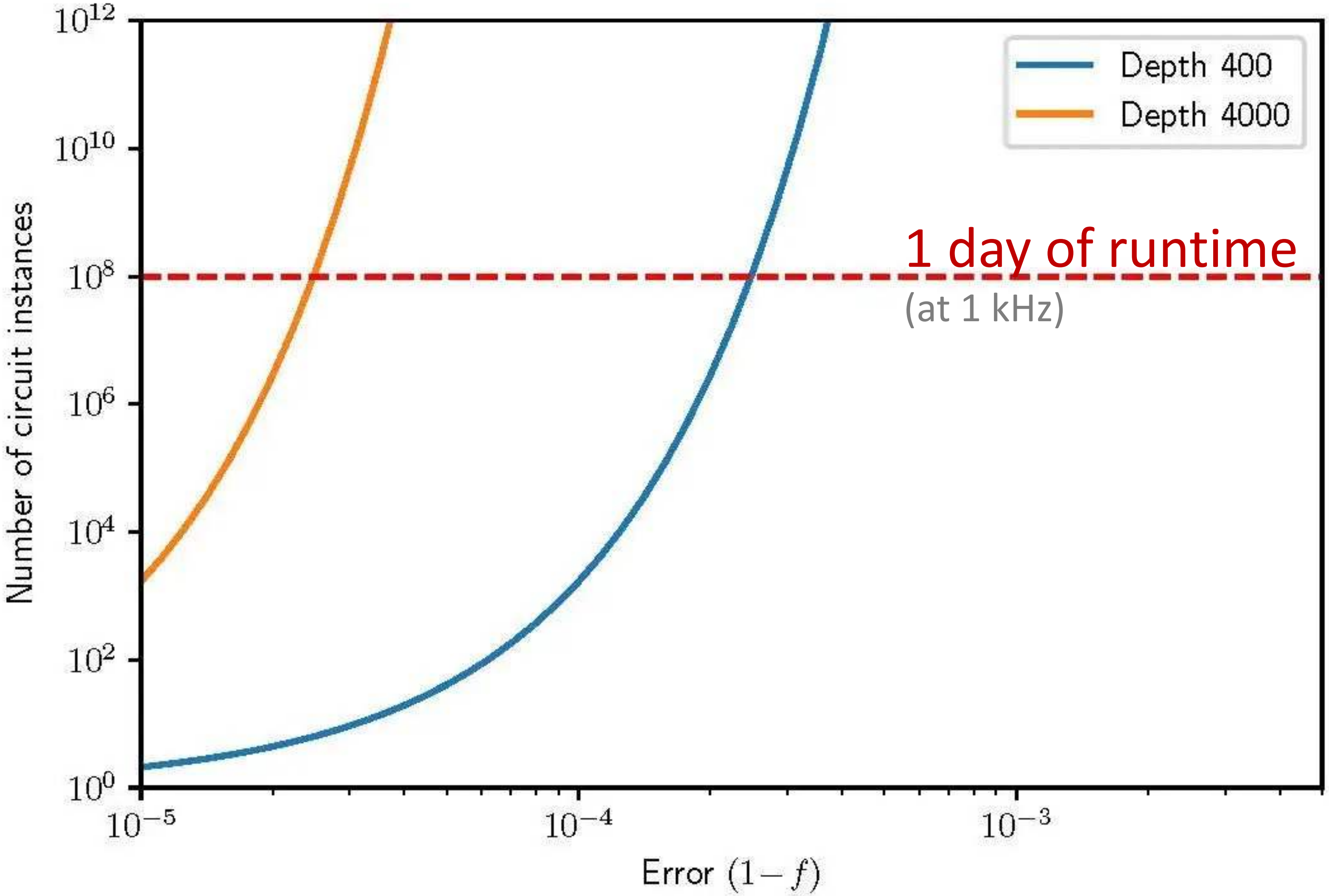


50Q
2 layers of
cNOT gates

See research.ibm.com/blog/gammabar-for-quantum-advantage

Path to 100+ qubits?

Estimating PEC overhead for Trotter circuits comprising 100 qubits



See also on speed: A. Wack, et al., Quality, speed, and scale: three key attributes to measure the performance of near-term quantum computers (2021).

Path to quantum computing

<u>Noise-free</u> estimators can be obtained from noisy quantum computers TODAY, at a runtime cost that is exponential in number of qubits *n* and circuit depth *d*

$$\text{Runtime} = \beta\, d\ (\bar{\gamma})^{n d}\ \text{seconds}$$

You can further reduce runtime using light cones and other strategies.

d is the depth of the quantum circuit

β is a measure of the time per circuit layer operation (CLOPS) (increase by pushing speed)

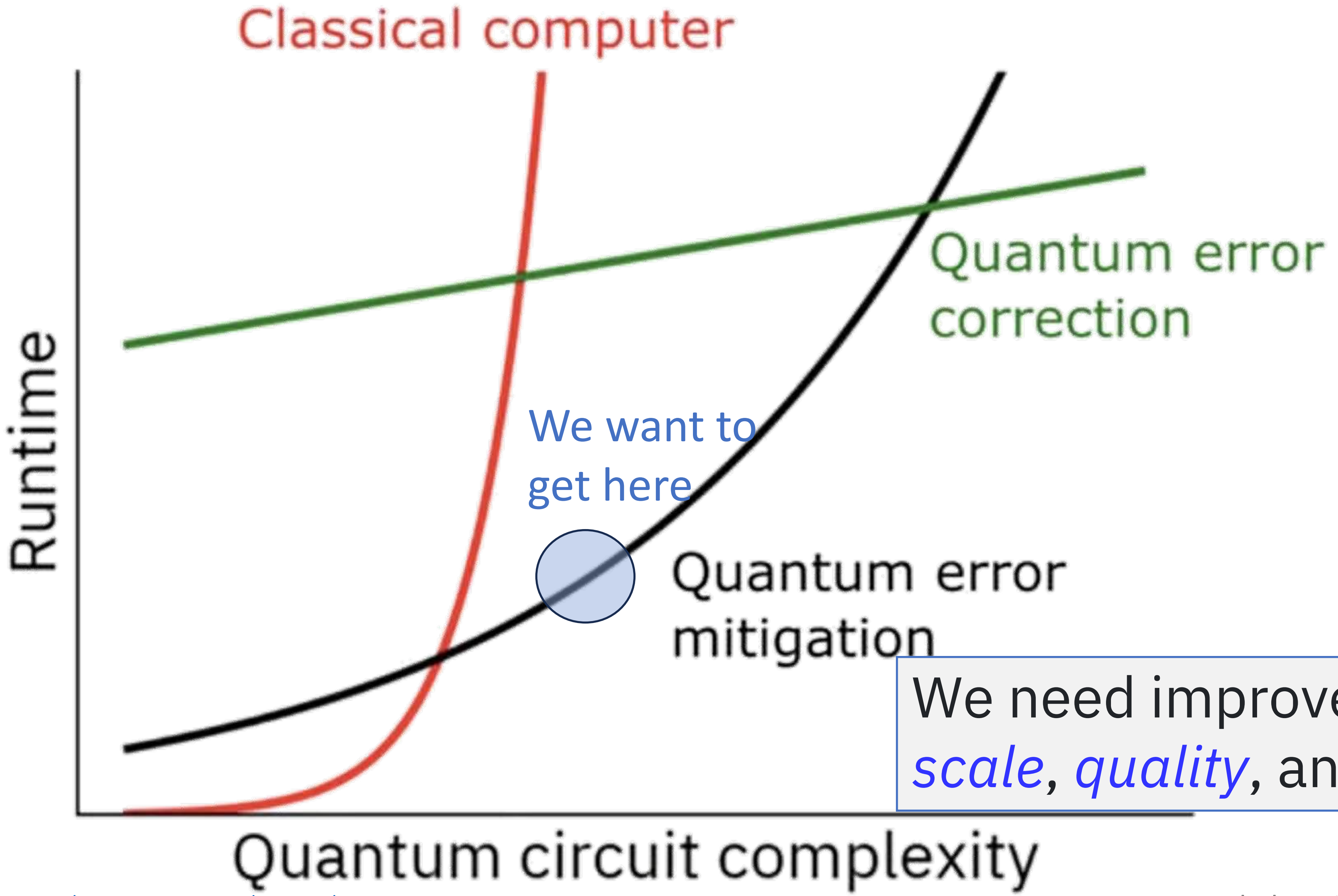
$\bar{\gamma}$ is a measure of the collective quantum noise (increasing quality brings it closer to 1)

n is the number of operational qubits (increase by pushing scale)

	Improvements	$\bar{\gamma}$
Hummingbird r2 (Brooklyn, 65Q)		1.038
Hummingbird r3 (Ithaca, 65Q)	2-3x coherence improvements over r2	1.024
Falcon r10 (Prague, 32Q)	State-of-the-art two-qubit gates, reduced crosstalk	1.012

https://research.ibm.com/blog/gammabar-for-quantum-advantage

Path to utility of quantum computers before error correction



We need improvements in *scale*, *quality*, and *speed*.

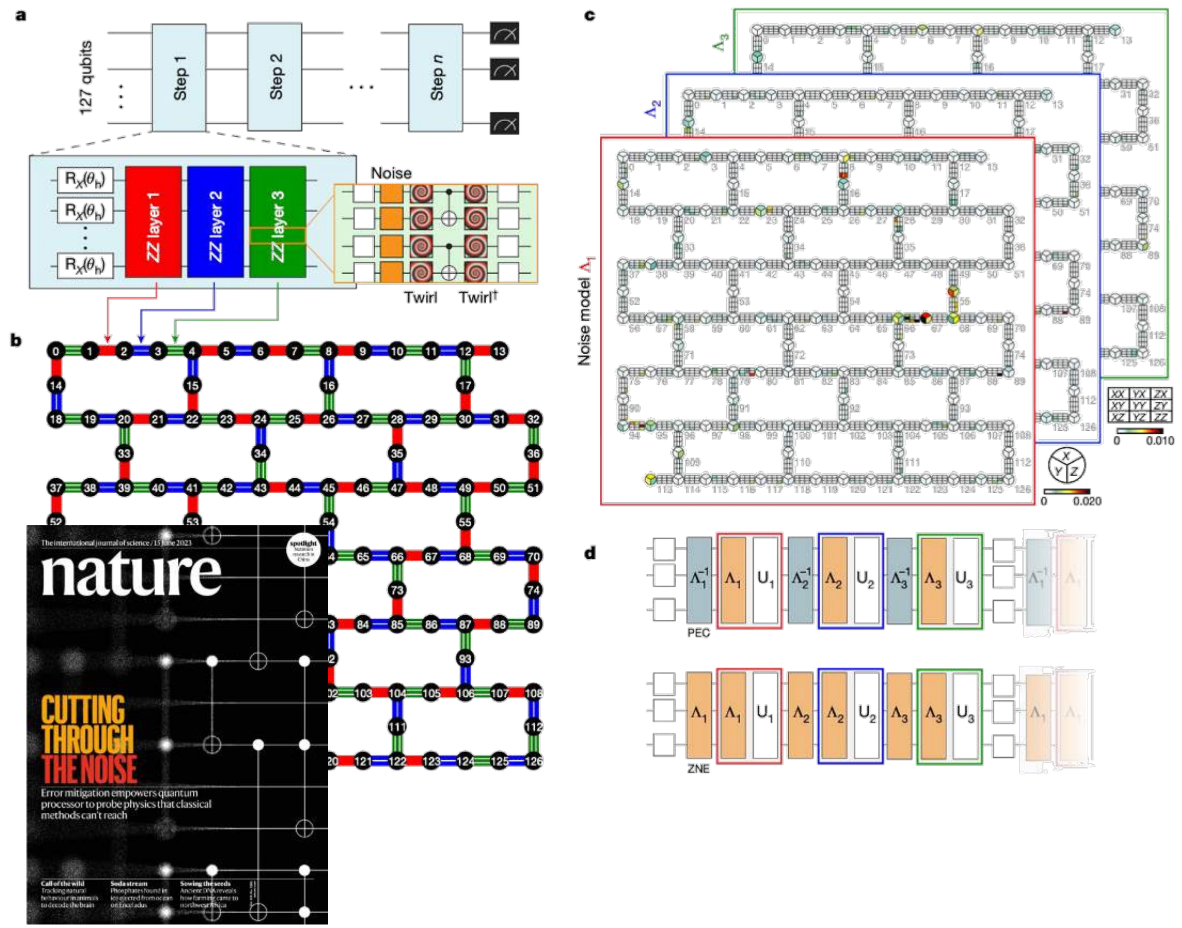
https://qiskit.org/documentation/partners/qiskit_ibm_runtime/tutorials/Error-Suppression-and-Error-Mitigation.html

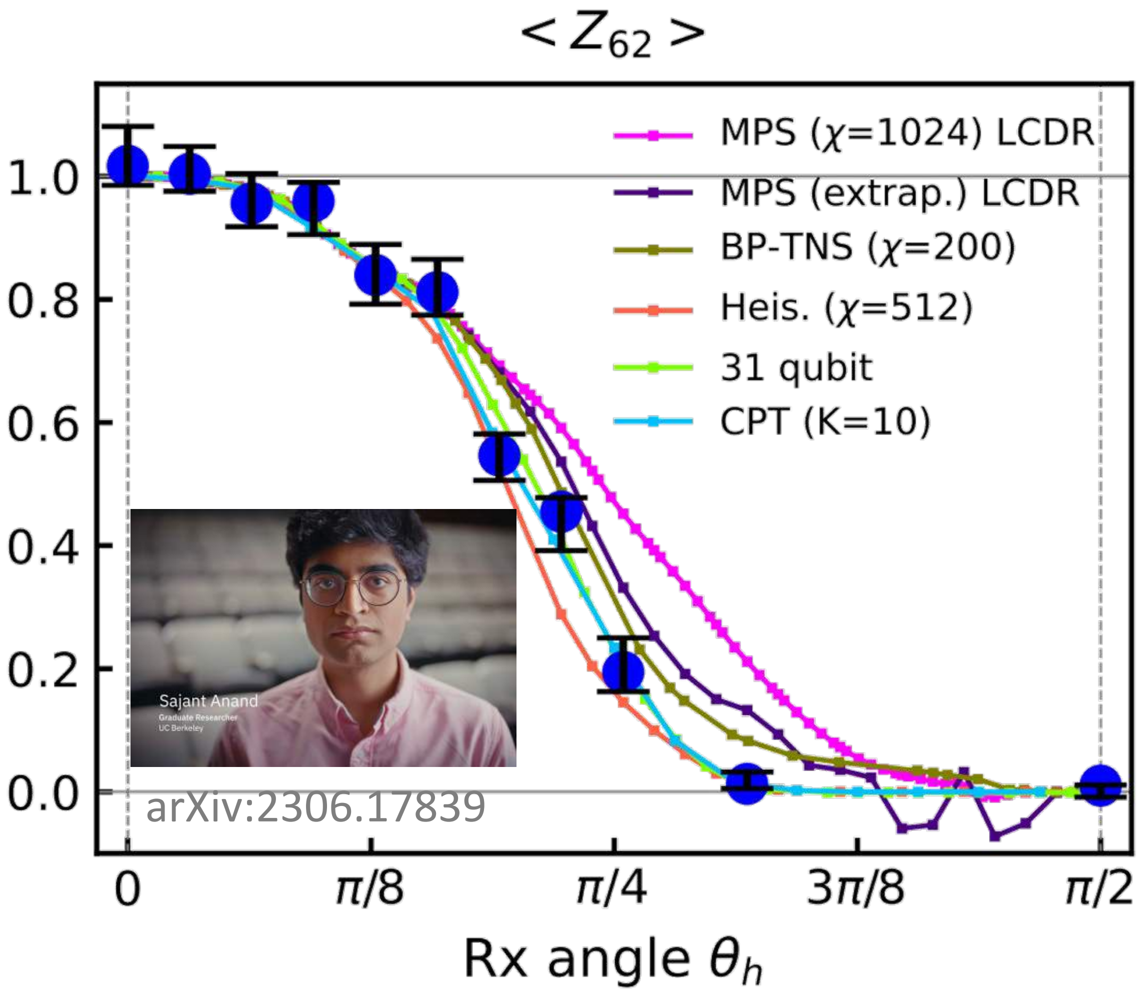
PEC + ZNE for 127 qubits

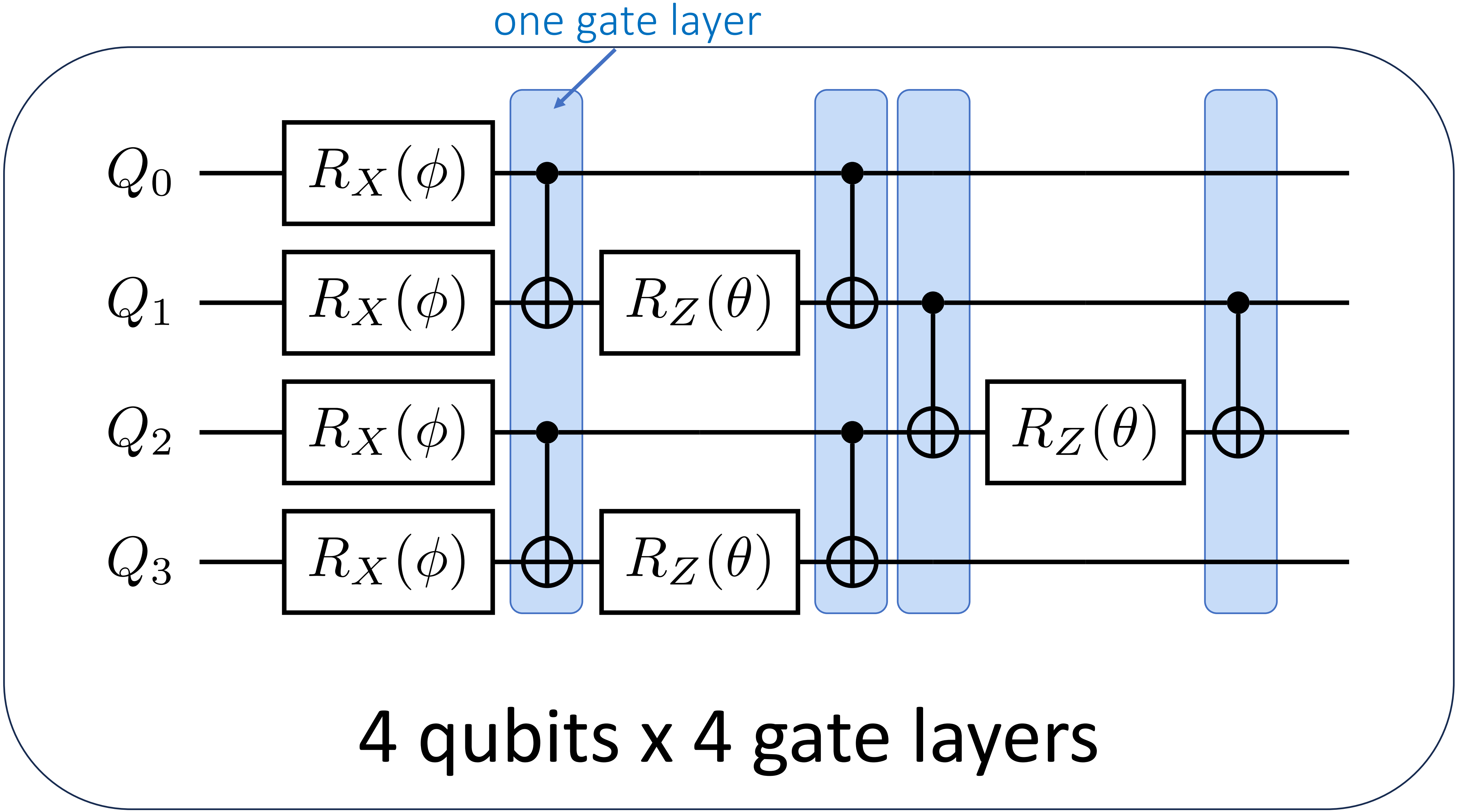
Article

Evidence for the utility of quantum computing before fault tolerance

Youngseok Kim[1,6]✉, Andrew Eddins[2,6]✉, Sajant Anand[3], Ken Xuan Wei[1], Ewout van den Berg[1], Sami Rosenblatt[1], Hasan Nayfeh[1], Yantao Wu[3,4], Michael Zaletel[3,5], Kristan Temme[1] & Abhinav Kandala[1]✉

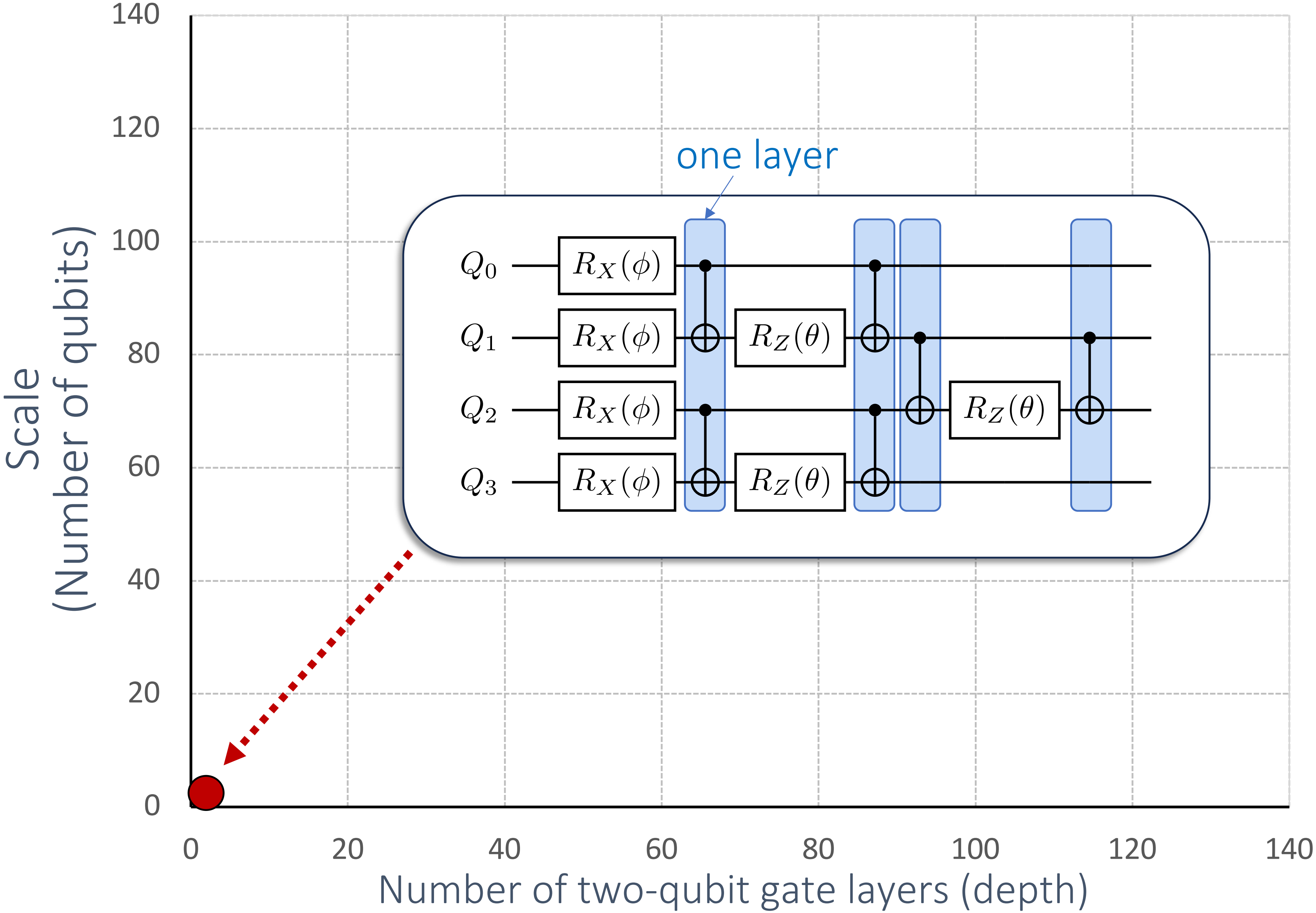




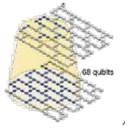


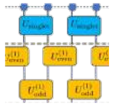
4 qubits x 4 gate layers

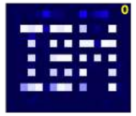
Circuit width × depth

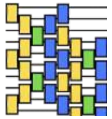


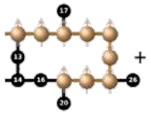
Some early utility-scale experiments


[0] Kim, Eddins, ..., Temme, Kandala
Nature 618, 500–505 (2023)

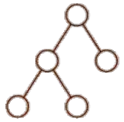

[1] Yu, Zhao, Wei.
arXiv: 2207.09994 (2022)


[2] Shtanko, Wang, Zhang, Harle, Seif, Movassagh, Minev.
arXiv: 2307.07552 (2023)

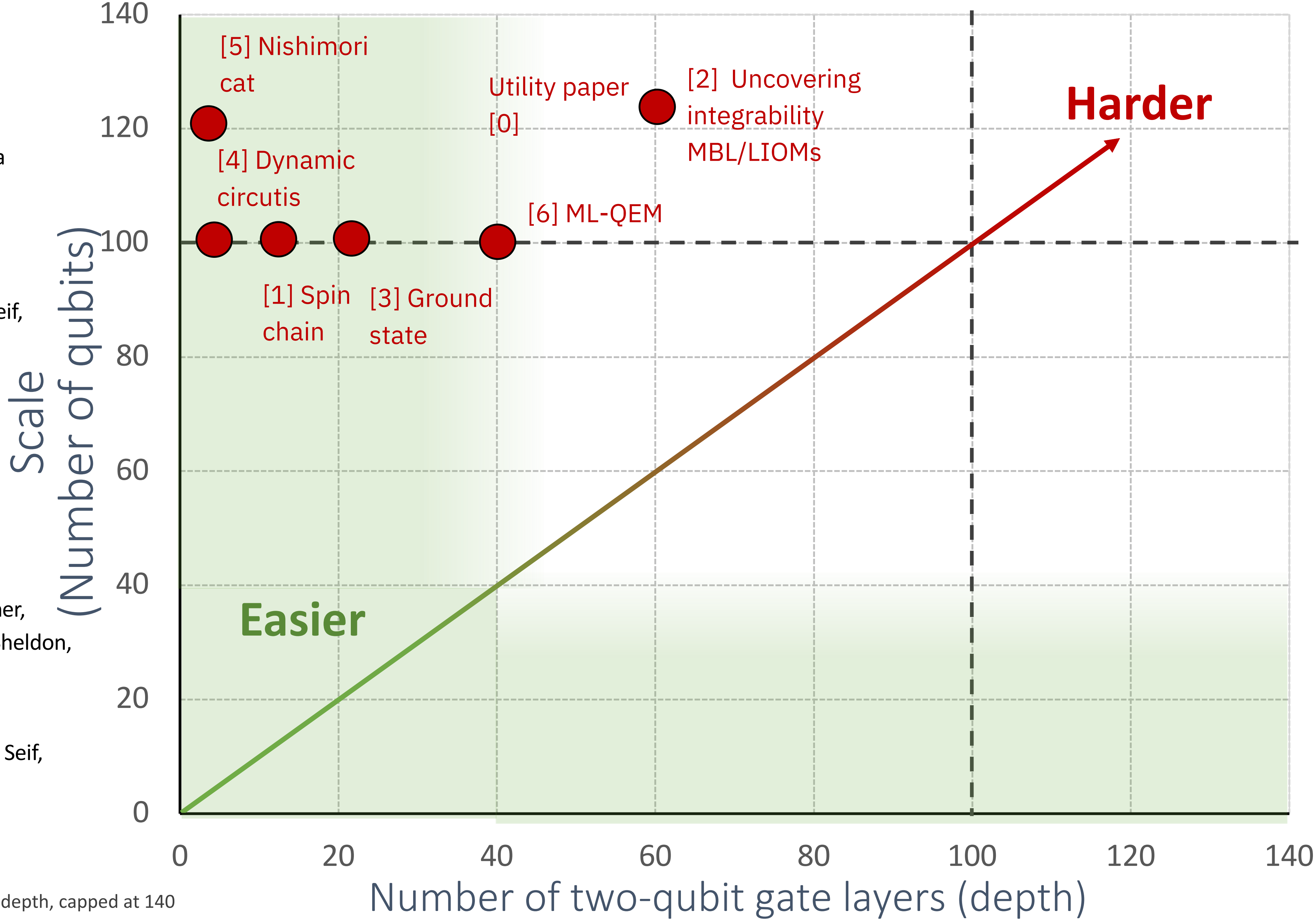

[3] Farrell, Illa, Ciavarella, Savage.
arXiv: 2308.04481 (2023)


[4] Bäumer, Tripathi, Wang, Rall, Chen, Majumder, Seif, Minev.
arXiv: 2308.13065 (2023)


[5] Chen, Zhu, Verresen, Seif, Baümer, Layden, Tantivasadakarn, Zhu, Sheldon, Vishwanath, Trebst, Kandala.
arXiv: 2309.02863 (2023)


[6] Liao, Wang, Sitdikov, Salcedo, Seif, Minev.
arXiv: 2308.13065 (2023)

...



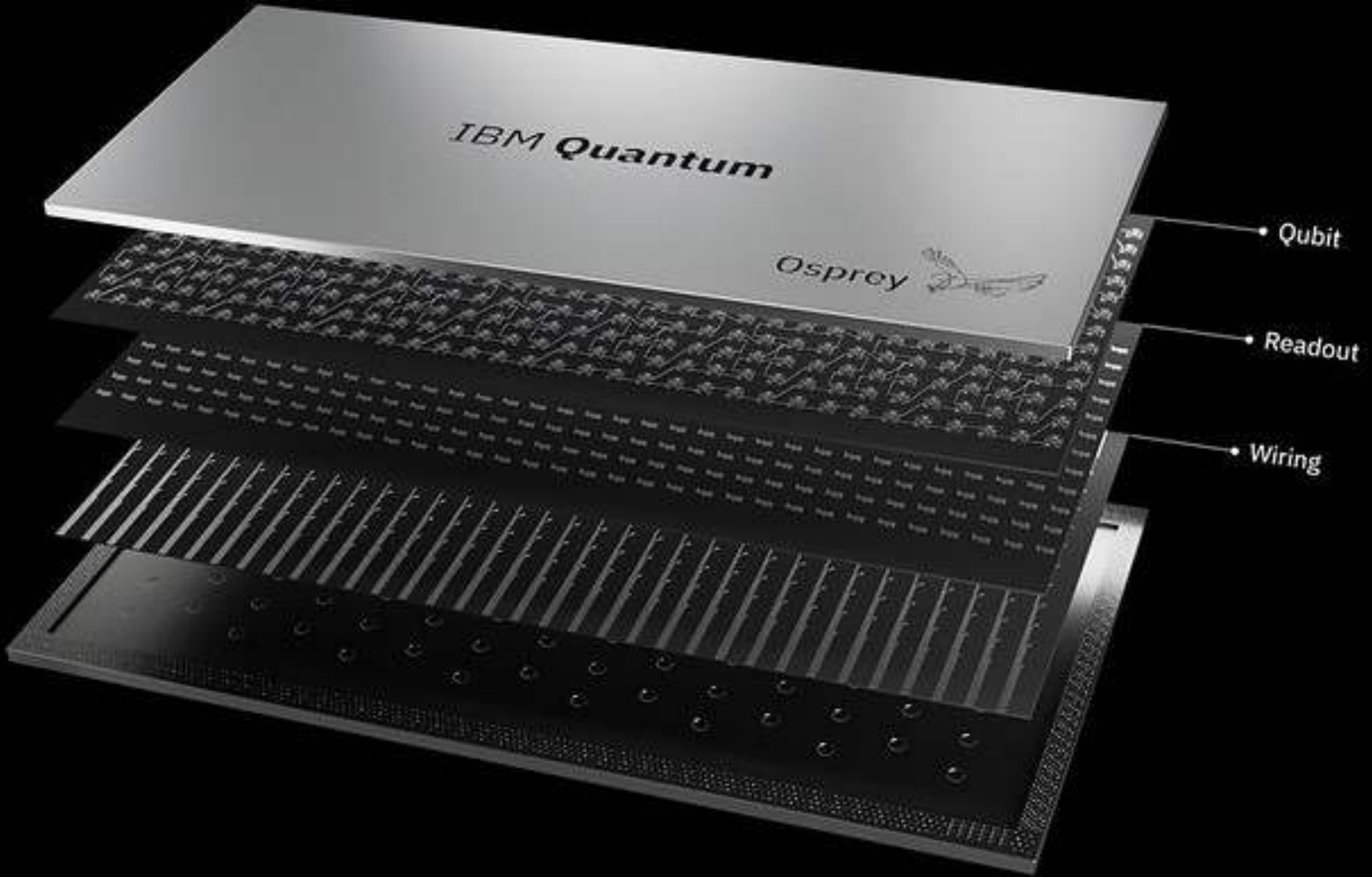
* signal stops or decays more than 50% beyond this depth, capped at 140
* note: quantum advantage with shallow circuits, Bravyi, Gosset, Konig, and related

2019
Falcon
27 Qubits

2020
Hummingbird
65 Qubits

2021
Eagle
127 Qubits

2022
Osprey
433 Qubits


IBM Quantum
Osprey
Qubit
Readout
Wiring

Quantum Noise and Error Mitigation

Lectures 1 and 2

Big picture

Quantum computers status

Why error mitigation?
- Noise in quantum computers
- Overview of error mitigation

Mitigation fundamentals

Probabilistic error cancelation (PEC)
- Introduction
- One qubit example

Lecture 2

Learning noise
State-of-art mitigation experiments
Hardware
Outlook

road based on: freepik; Monster image by jcomp on Freepik

The important thing is not to stop questioning. Curiosity has its own reason for existence.

One cannot help but be in awe when they contemplate the mysteries of eternity, of life, of the marvelous structure of reality.

It is enough if one tries merely to comprehend a little of this mystery each day.

Albert Einstein

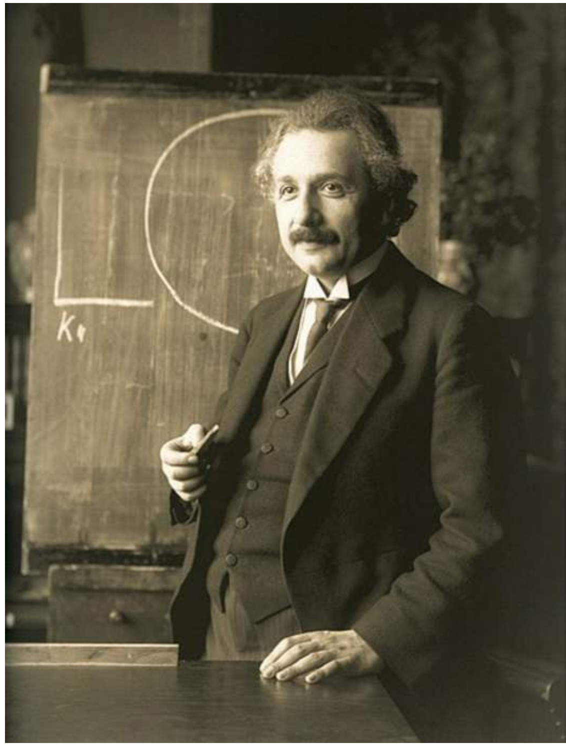


Photo: F. Schmutzer

Bonus content

Comprehension questions

QUESTION 1. Which of the following is a primary goal of quantum error mitigation techniques?
A) To completely eliminate all errors in quantum computations, thus giving zero bias for any depth circuit within reasonable shot budget.
B) To reduce the effective error rate in quantum computations to acceptable levels without requiring additional qubits, but at the cost of more sampling.
C) To replace faulty qubits with error-free qubits.
D) To enhance the speed of quantum gates.

QUESITON 2. Which of the following statements best describes the concept of probabilistic error cancellation in quantum computing?
A) It involves using classical error correction codes to detect and correct errors in quantum circuits.
B) It refers to the use of redundant qubits to store quantum information, thus preventing errors from occurring.
C) It is a technique to estimate and mitigate the impact of noise in near-term quantum computations by applying the noise channel on average.
D) It uses quantum error correction codes, such as the surface code, to identify and correct errors directly on quantum bits.

Scaling PEC to *n* qubits and larger circuits (ADVANCED – OPTINAL)

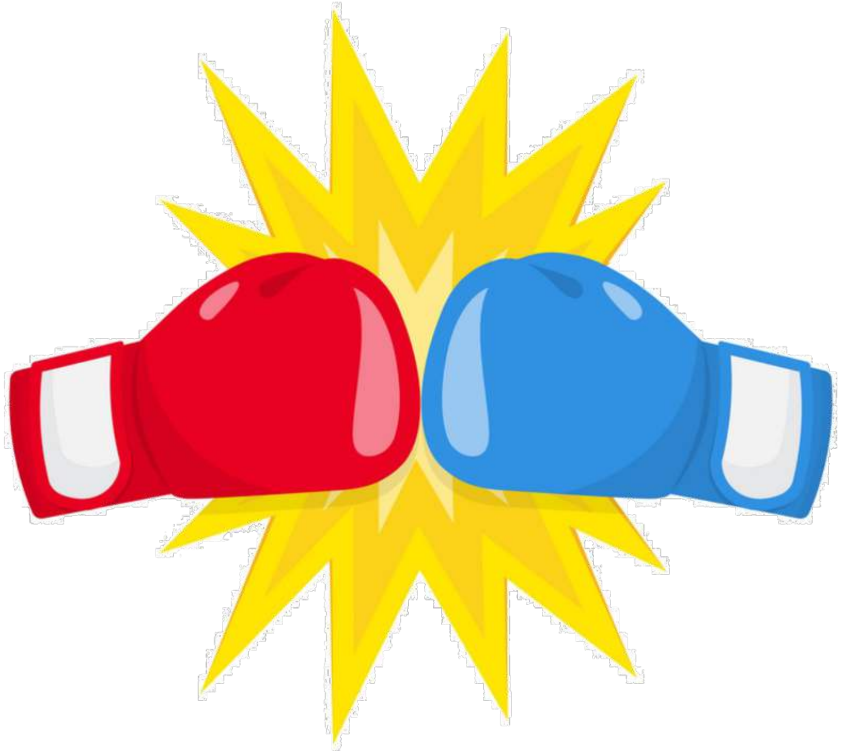


image: makc76

Language of errors and error mitigation

$$\frac{\cancel{d}}{\cancel{d}x}\frac{1}{x} = -\frac{1}{x^2}$$

Notation: Super bra ket

$$\rho \in L(\mathcal{H})$$

$$\hat{A}, \hat{B} \in L(\mathcal{H})$$

$$\mathcal{S} \in L(L(\mathcal{H}))$$

$$|\hat{A}\rangle\!\rangle \leftrightarrow \hat{A}$$

$$\mathcal{S}|\hat{A}\rangle\!\rangle \leftrightarrow \mathcal{S}\left(\hat{A}\right)$$

Vectorization isomorphism

$$\langle\!\langle \hat{A}|\hat{B}\rangle\!\rangle \leftrightarrow \left\langle \hat{A}, \hat{B} \right\rangle = \mathrm{Tr}\left(\hat{B}\hat{A}^\dagger\right)$$

$$\begin{aligned}\langle\!\langle \hat{A}|\mathcal{S}|\hat{B}\rangle\!\rangle &= \langle\!\langle \hat{A}|\mathcal{S}(\hat{B})\rangle\!\rangle \\ &= \left\langle \hat{A}, \mathcal{S}\left(\hat{B}\right) \right\rangle \\ &= \mathrm{Tr}\left(\hat{A}^\dagger \mathcal{S}(\hat{B})\right)\end{aligned}$$

$$\langle\!\langle \hat{A}| \cdot \leftrightarrow \left\langle \hat{A}, \cdot \right\rangle = \mathrm{Tr}\left(\hat{A}^\dagger \cdot\right)$$

$$|\hat{A}\rangle\!\rangle\langle\!\langle \hat{B}| \leftrightarrow \hat{A}\left\langle \hat{B}, \cdot \right\rangle \leftrightarrow \hat{A}\,\mathrm{Tr}\left(\hat{B}^\dagger \cdot\right)$$

Familiar ideas revisited with super notation

$$\left\langle \hat{P}_a \middle| \hat{P}_b \right\rangle = \mathrm{Tr}\left(\hat{P}_a^\dagger \hat{P}_b\right) = d\delta_{ab}$$

$$\Gamma = \{I, X, Y, Z\}^{\otimes n}$$

$$|\Gamma| = 4^n = d^2,\ d = 2^n$$

$$a, b \in \Gamma$$

$$\mathcal{I} = \sum_{a\in\Gamma} \frac{\left|\hat{P}_a\right\rangle \left\langle \hat{P}_a\right|}{\left\langle \hat{P}_a \middle| \hat{P}_a \right\rangle}$$

$$\Lambda = \mathcal{I}\Lambda\mathcal{I} = \sum_{a,b\in\Gamma} \frac{\left\langle \hat{P}_a \middle| \Lambda \middle| \hat{P}_b \right\rangle}{\left\langle \hat{P}_a \middle| \hat{P}_a \right\rangle \left\langle \hat{P}_b \middle| \hat{P}_b \right\rangle} \left|\hat{P}_a\right\rangle \left\langle \hat{P}_b\right| = \frac{1}{d} \sum_{a,b\in\Gamma} \Lambda_{ab} \left|\hat{P}_a\right\rangle \left\langle \hat{P}_b\right| \ ,$$

Stochastic Pauli channel

$$\Lambda : L\left(\mathcal{H}\right) \to L\left(\mathcal{H}\right)$$

$$0 \leq p_a \leq 1 \,,$$

$$\Lambda\left(\rho\right) = \sum_{a\in\Gamma} p_a P_a \rho P_a \;,$$

$$\sum_{a\in\Gamma} p_a = 1 \,,$$

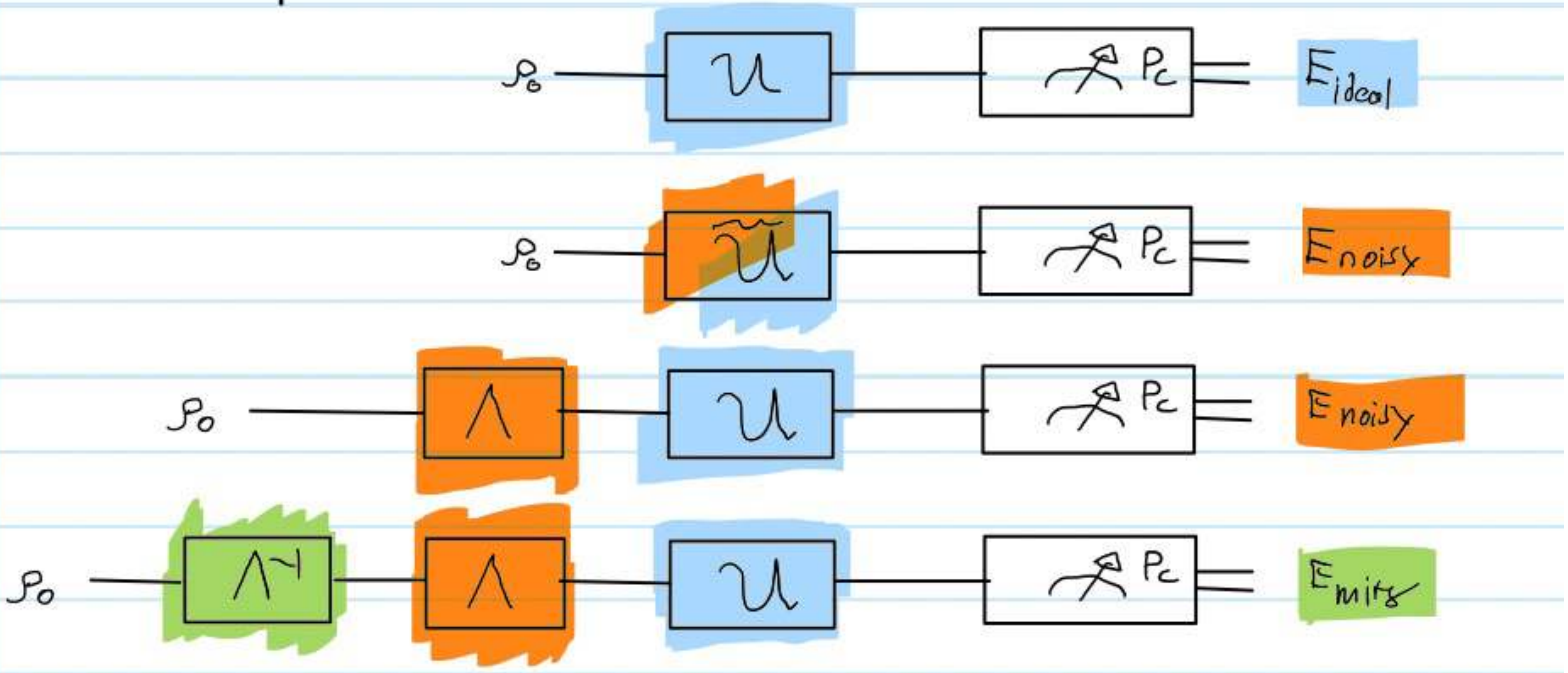
$$\Lambda\left(\rho\right) = \frac{1}{\sqrt{|\Gamma|}} \sum_{b\in\Gamma} f_b \operatorname{Tr}\left(P_b\rho\right) P_b = \sum_{b\in\Gamma} f_b \frac{|P_b\rangle\!\rangle\langle\!\langle P_b|}{\langle\!\langle P_b|P_b\rangle\!\rangle} \;,$$

$$\Lambda\left(P_b\right) = f_b P_b \;, \quad \forall a \in \Gamma \;,$$

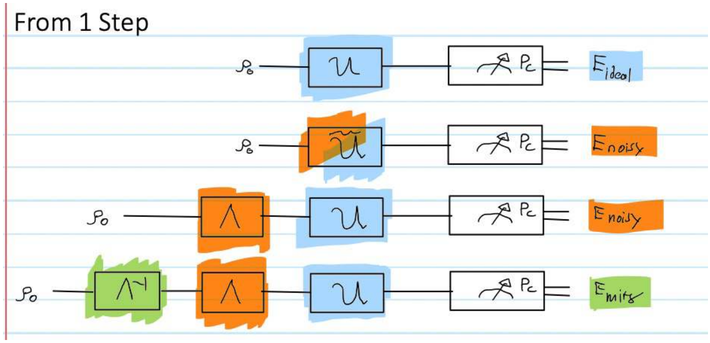
$$-1 \leq f_b \leq 1 \quad f_I = 1 \,,$$

Probabilistic error cancelation: Derivation

From 1 Step



Probabilistic error cancelation: Derivation



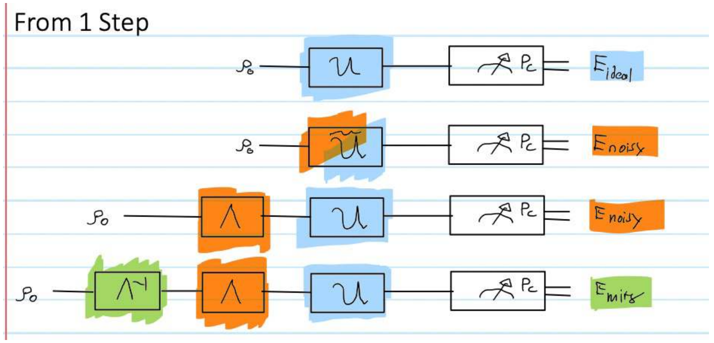
Channel definitions

$$\mathcal{U} = U \cdot U^{\dagger}$$

$$\Lambda = \sum_a f_a |P_a\rangle\rangle\langle\langle P_a| \cdot$$

$$= \sum_b c_b P_a \cdot$$

Probabilistic error cancelation: Derivation



Channel definitions

$$\mathcal{U} = U \cdot U^\dagger$$

$$\langle \hat{P}_c \rangle(\cdot) = \langle\langle P_c | \circ$$

Need

$$\Lambda = \sum_a f_a |P_a\rangle\rangle\langle\langle P_a| \cdot$$

$$= \sum_b c_b \mathcal{P}_a \cdot$$

$$-1 \le f_a \le 1$$

$$c_b \ge 0 \, , \sum_b c_b = 1$$

$$c_b = \frac{1}{2^n} \sum_a (-1)^{\langle a,b \rangle_{sp}} f_a$$

$$\vec{c_b} = W^{-1} \vec{f_a} \qquad \vec{f_a} = W c_b$$

$$\Lambda^{-1} = \sum_a f_a^{-1} |P_a\rangle\rangle\langle\langle P_a|$$

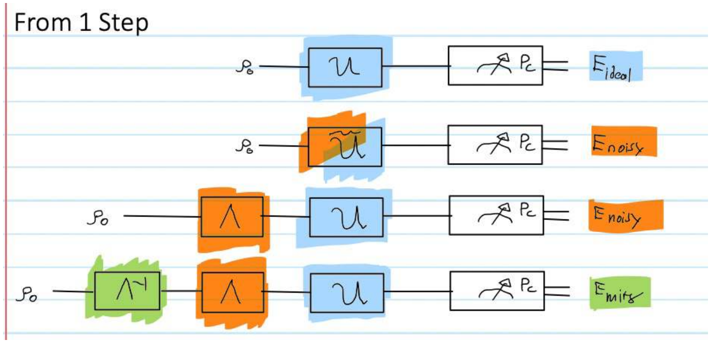
$$= \sum_b c_b^{inv} \mathcal{P}_b \cdot$$

$$\vec{c_b^{inv}} = W \vec{f_a^{-1}}$$

$$c_b^{inv} = \frac{1}{2^n} \sum_a (-1)^{\langle a,b \rangle_{sp}} f_a$$

$$c_b^{inv} \in \mathbb{R}$$

Probabilistic error cancelation: Derivation



Channel definitions

$$\mathcal{U} = U \cdot U^{\dagger}$$

$$\langle \hat{P}_c \rangle(\cdot) = \langle\langle P_c | \cdot$$

Need

$$\Lambda = \sum_a f_a |P_a\rangle\rangle\langle\langle P_a| \cdot$$

$$= \sum_b c_b P_a \cdot$$

$$-1 \le f_a \le 1$$

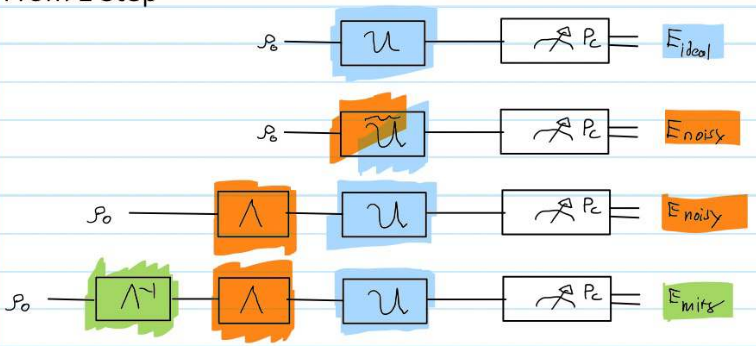
$$c_b \ge 0 \,, \quad \sum_0 c_b = 1$$

$$c_b = \frac{1}{2^n} \sum_a (-1)^{\langle a,b \rangle_{sp}} f_a$$

$$\vec{c}_b = W^{-1} \vec{f}_a \qquad \vec{f}_a = W c_b$$

Circuit expectation values

From 1 Step



Channel definitions

$\mathcal{U} = u \cdot u^{+}$

$\langle \hat{P}_c \rangle(\cdot) = \langle\langle P_c | \cdot$

Need to introduce
- Super-op bra-ket notation ②
- Pauli channel χ & PTM ③
  - super-op notation ④
  - WH ⑤
  - Pauli class representation ①
    - SQ X→Y
    - $\langle a,b \rangle_{sp}$

$\Lambda = \sum_a f_a |P_a\rangle\rangle\langle\langle P_a| \cdot$

$= \sum_b c_b \mathcal{P}_a \cdot$

$-1 \le f_a \le 1$

$c_b \ge 0 \,, \quad \sum_0 c_b = 1$

$c_b = \frac{1}{2^n} \sum_a (-1)^{\langle a,b \rangle_{sp}} f_a$

$\vec{c}_b = W^{-1} \vec{f}_a \qquad \vec{f}_a = W c_b$

$\Lambda^{-1} = \sum_a f_a^{-1} |P_a\rangle\rangle\langle\langle P_a|$

$= \sum_b c_b^{inv} \mathcal{P}_b \cdot$

$\overrightarrow{c_b^{inv}} = W \overrightarrow{f_a^{-1}}$

$c_b^{inv} = \frac{1}{2^n} \sum_a (-1)^{\langle a,b \rangle_{sp}} f_a$

$c_b^{inv} \in \mathbb{R}$

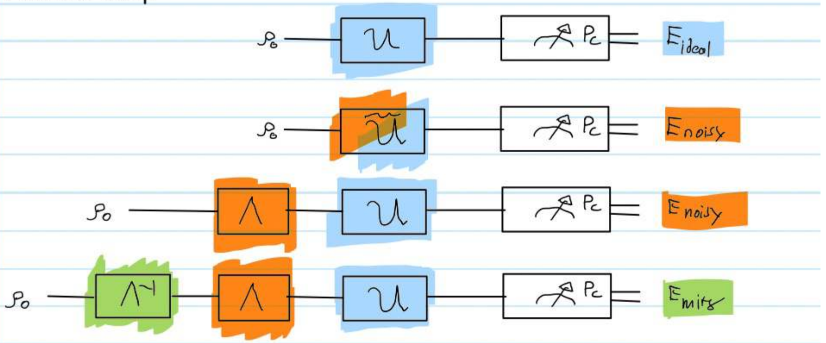
$$E_{ideal} := \langle \hat{P}_c \rangle_{\mathcal{U}}$$

$$= \langle\langle P_c | \mathcal{U} | \rho_0 \rangle\rangle$$

ideal exp value with noiseless unitary

Circuit expectation values

From 1 Step



Channel definitions

$\mathcal{U} = U \cdot U^{\dagger}$

$\langle \hat{P}_c \rangle (\cdot) = \langle\langle P_c | \cdot$

Need to introduce
- Super-op bra-ket notation (2)
- Pauli channel χ + PTM (3)
  - super-op notation (4)
  - WH (5)
  - Pauli class representation (1)
    - SQ X, Y, Z
  - $\langle a, b \rangle_{sp}$

$\Lambda = \sum_a f_a |P_a\rangle\rangle\langle\langle P_a | \cdot$

$= \sum_b c_b P_a \cdot$

$-1 \leq f_a \leq 1$

$c_b \geq 0 \;, \sum_b c_b = 1$

$c_b = \frac{1}{2^n} \sum_a (-1)^{\langle a,b \rangle_{sp}} f_a$

$\vec{c}_b = W^{-1} \vec{f}_a \qquad \vec{f}_a = W c_b$

$\Lambda^{-1} = \sum_a f_a^{-1} |P_a\rangle\rangle\langle\langle P_a|$

$= \sum_b c_b^{inv} P_b \cdot$

$\vec{c}_b^{\,inv} = W \vec{f_a^{-1}}$

$c_b^{inv} = \frac{1}{2^n} \sum_a (-1)^{\langle a,b \rangle_{sp}} f_a$

$c_b^{inv} \in \mathbb{R}$

$$E_{ideal} := \langle \hat{P}_c \rangle_{\mathcal{U}}$$

$$= \langle\langle P_c | \mathcal{U} | \rho_0 \rangle\rangle$$

ideal exp value with noiseless unitary

$$E_{noisy} := \langle \hat{P}_c \rangle_{\tilde{\mathcal{U}}}$$

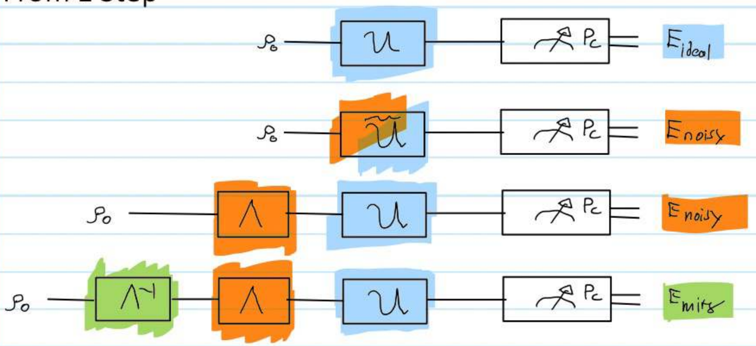
$$= \langle\langle P_c | \mathcal{U} \Lambda | \rho_0 \rangle\rangle$$

$$= \langle\langle P_c | \tilde{\mathcal{U}} | \rho_0 \rangle\rangle$$

noisy-gate expectation value

Circuit expectation values

From 1 Step



Channel definitions

$\mathcal{U} = U \cdot U^{\dagger}$

$\langle \hat{P}_c \rangle(\cdot) = \langle\langle P_c | \cdot$

Need to introduce
- Super-op bra-ket notation ②
- Pauli channel χ & PTM ③
  - super-op notation ④
  - WH ⑤
  - Pauli class representation ①
    - SQ [illegible]
  - $\langle a,b \rangle_{sp}$

$\Lambda = \sum_a f_a |P_a\rangle\rangle\langle\langle P_a| \cdot$

$= \sum_b c_b P_a \cdot$

$-1 \le f_a \le 1$

$c_b \ge 0, \quad \sum_b c_b = 1$

$c_b = \frac{1}{2^n} \sum_a (-1)^{\langle a,b \rangle_{sp}} f_a$

$\vec{c}_b = W^{-1} \vec{f}_a \qquad \vec{f}_a = W c_b$

$\Lambda^{-1} = \sum_a f_a^{-1} |P_a\rangle\rangle\langle\langle P_a|$

$= \sum_b c_b^{inv} P_b \cdot$

$\vec{c}_b^{\,inv} = W \vec{f_a^{-1}}$

$c_b^{inv} = \frac{1}{2^n} \sum_a (-1)^{\langle a,b \rangle_{sp}} f_a$

$c_b^{inv} \in \mathbb{R}$

$$E_{ideal} := \langle \hat{P}_c \rangle_{\mathcal{U}}$$

$$= \langle\langle P_c | \mathcal{U} | \rho_0 \rangle\rangle$$

ideal exp value with noiseless unitary

$$E_{noisy} := \langle \hat{P}_c \rangle_{\tilde{\mathcal{U}}}$$

$$= \langle\langle P_c | \mathcal{U} \Lambda | \rho_0 \rangle\rangle$$

$$= \langle\langle P_c | \tilde{\mathcal{U}} | \rho_0 \rangle\rangle$$

noisy-gate expectation value

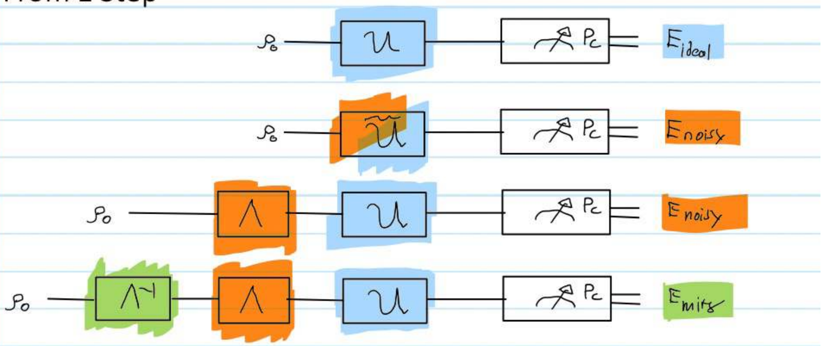
$$E_{mitg} := \langle\langle P_c | \mathcal{U} \Lambda \Lambda^{-1} | \rho_0 \rangle\rangle$$

$$= \langle\langle P_c | \mathcal{U} \Lambda \left(\sum_a f_a^{-1} |P_a\rangle\rangle\langle\langle P_a| \right) | \rho_0 \rangle\rangle$$

$$= \langle\langle P_c | \mathcal{U} \Lambda \left(\sum_b c_b^{inv} P_b \right) | \rho_0 \rangle\rangle$$

Circuit expectation values

From 1 Step



Channel definitions

$\mathcal{U} = U \cdot U^{\dagger}$

$\langle \hat{P}_c \rangle(\cdot) = \langle\langle P_c | \cdot$

Need to introduce
- Super-op bra-ket notation ②
- Pauli channel χ & PTM ③
  - super-op notation ④
  - WH ⑤
  - Pauli class representation ①
    - SQ
  - $\langle a,b \rangle_{sp}$

$\Lambda = \sum_a f_a |P_a\rangle\rangle\langle\langle P_a| \cdot$

$= \sum_b c_b P_a \cdot$

$-1 \le f_a \le 1$

$c_b \ge 0$, $\sum_b c_b = 1$

$c_b = \frac{1}{2^n} \sum_a (-1)^{\langle a,b \rangle_{sp}} f_a$

$\vec{c}_b = W^{-1} \vec{f}_a \qquad \vec{f}_a = W c_b$

$\Lambda^{-1} = \sum_a f_a^{-1} |P_a\rangle\rangle\langle\langle P_a|$

$= \sum_b c_b^{inv} P_b \cdot$

$\vec{c}_b^{inv} = W \vec{f}_a^{-1}$

$c_b^{inv} = \frac{1}{2^n} \sum_a (-1)^{\langle a,b \rangle_{sp}} f_a$

$c_b^{inv} \in \mathbb{R}$

$E_{ideal} := \langle \hat{P}_c \rangle_{\mathcal{U}}$

$= \langle\langle P_c | \mathcal{U} | \rho_0 \rangle\rangle$ ideal exp value with noiseless unitary

$E_{noisy} := \langle \hat{P}_c \rangle_{\tilde{\mathcal{U}}}$

$= \langle\langle P_c | \mathcal{U} \Lambda | \rho_0 \rangle\rangle$

$= \langle\langle P_c | \tilde{\mathcal{U}} | \rho_0 \rangle\rangle$ noisy-gate expectation value

$E_{mitg} := \langle\langle P_c | \mathcal{U} \Lambda \Lambda^{-1} | \rho_0 \rangle\rangle$

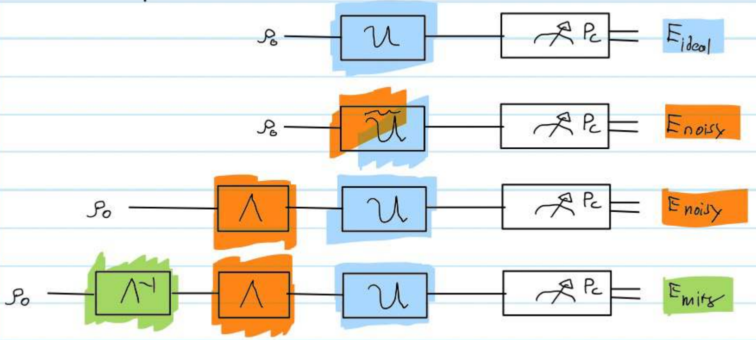
$= \langle\langle P_c | \mathcal{U} \Lambda \left(\sum_a f_a^{-1} |P_a\rangle\rangle\langle\langle P_a| \right) |\rho_0\rangle\rangle$

$= \langle\langle P_c | \mathcal{U} \Lambda \left(\sum_b c_b^{inv} P_b \right) | \rho_0 \rangle\rangle$

$= \sum_b c_b^{inv} \langle\langle P_c | \mathcal{U} \Lambda P_b | \rho_0 \rangle\rangle$ sum of trajectories with weight $c_b^{inv} \in \mathbb{R}$

$= \underbrace{\sum_b c_b^{inv}}_{\text{classical post process}} \underbrace{\langle P_c \rangle(\tilde{\mathcal{U}}, b)}_{\text{quantum}}$

From 1 Step



Channel definitions

$\mathcal{U} = U \cdot U^{\dagger}$

$\Lambda = \sum_a f_a |P_a\rangle\rangle\langle\langle P_a| \cdot$

$\quad = \sum_b c_b \mathcal{P}_a \cdot$

$\Lambda^{-1} = \sum_a f_a^{-1} |P_a\rangle\rangle\langle\langle P_a|$

$\quad = \sum_b c_b^{inv} \mathcal{P}_b \cdot$

$\langle \hat{P}_c \rangle(\cdot) = \langle\langle P_c | \cdot$

$-1 \leq f_a \leq 1$

$c_b \geq 0 \; , \sum_b c_b = 1$

$c_b = \frac{1}{2^n} \sum_a (-1)^{\langle a,b \rangle_{sp}} f_a$

$\vec{c}_b = W^{-1} \vec{f}_a \qquad \vec{f}_a = W c_b$

$\vec{c}_b^{inv} = W \vec{f}_a^{-1}$

$c_b^{inv} = \frac{1}{2^n} \sum_a (-1)^{\langle a,b \rangle_{sp}} f_a$

$c_b^{inv} \in \mathbb{R}$

Need to introduce
- Super-op bra-ket notation ②
- Pauli channel χ + PTM ③
  - super-op notation ④
  - WH ⑤
  - Pauli class representation ①
    - SQ X Z Y
  - $\langle a,b \rangle_{sp}$

$$E_{mitg} := \langle\langle P_c | \mathcal{U} \Lambda \Lambda^{-1} | \rho_0 \rangle\rangle$$

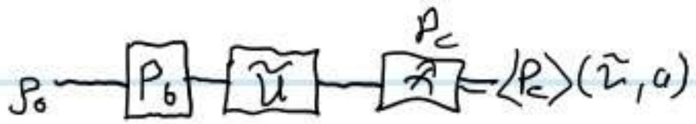
$$= \langle\langle P_c | \mathcal{U} \Lambda \left(\sum_a f_a^{-1} |P_a\rangle\rangle\langle\langle P_a| \right) |\rho_0\rangle\rangle$$

$$= \langle\langle P_c | \mathcal{U} \Lambda \left(\sum_b c_b^{inv} \mathcal{P}_b \right) |\rho_0\rangle\rangle$$

$$= \sum_b c_b^{inv} \langle\langle P_c | \mathcal{U} \Lambda \mathcal{P}_b | \rho_0 \rangle\rangle$$

sum of trajectories with weight $c_b^{inv} \in \mathbb{R}$

$$= \underbrace{\sum_b c_b^{inv}}_{\text{classical post process}} \underbrace{\langle P_c \rangle(\tilde{\mathcal{U}}, b)}_{\text{quantum}}$$



Quantum circuit we can execute on HW and find exp. value from.

$\therefore$ To find noise-free val all we have to do is to compute exp-val of all 4^n b-modified circuits! This would give us ideal exp value.

However, $|\{b\}| = 4^n$ grows exponentially, hence, infeasible.

but what if we could sample from it to approximate full sum. But... can't sample directly from c_b^{inv} which does not form a valid prob-distribution. Let's solve:

Quasi probability distribution

c_b^{inv} can be outside $[0,1]$

$\sum_b c_b^{inv} =: \gamma \geq 1$ generally for Λ not unitary

eg bit flip channel $\Lambda = (1-p) I\cdot I + p X\cdot X$

$\Lambda^{-1} = \left(1 + \frac{p}{1-2p}\right) I\cdot I + \frac{-p}{1-2p} X\cdot X$

$c_I^{inv} = 1 + \frac{p}{1-2p}$ $\quad$ $c_X^{inv} = \frac{-p}{1-2p}$

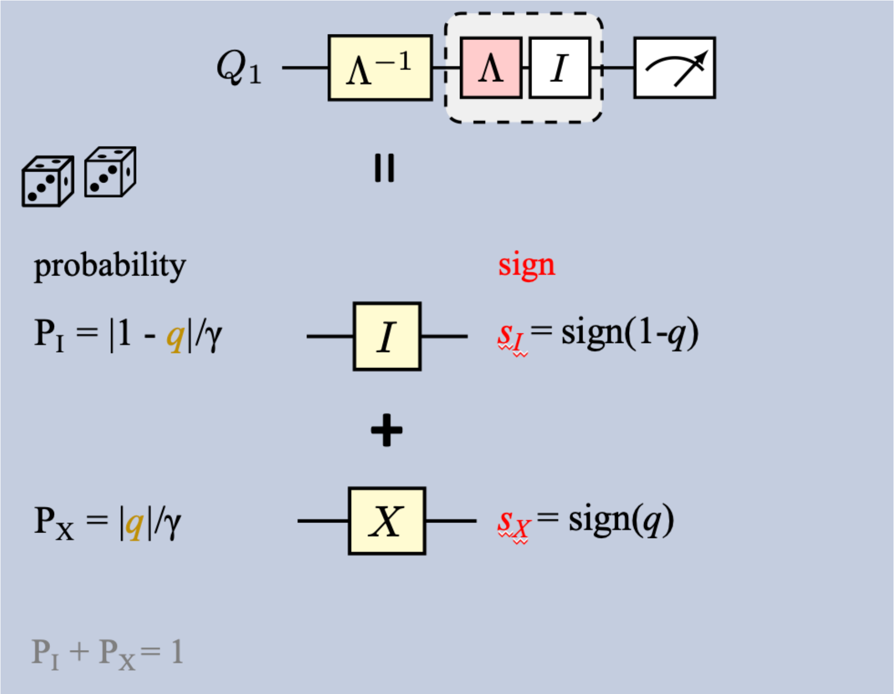
$b=(0,0)$ choose vector $\quad$ $b=(1,0)$

Turn into probabilities

$$c_b^{inv} = \underbrace{\mathrm{sgn}(c_b^{inv})}_{\text{sign} \in \{-1,+1\}} \underbrace{\frac{|c_b^{inv}|}{\gamma}}_{\text{prob} \in [0,1]} \underbrace{\gamma}_{\text{scale}}$$

$$\tilde{c}_b^{inv} := \frac{|c_b^{inv}|}{\|c_b^{inv}\|_1} \quad \text{abs val} \quad \|c_b\|_1 = \sum_b |c_b^{inv}| \iff L_1 \text{ norm}$$

$$= \frac{|c_b^{inv}|}{\gamma}$$



E mitigated

$$E_{mitg} = \sum_b C_b^{inv} \langle \hat{P}_c \rangle (\tilde{\mathcal{U}}, b)$$

$$= \sum_b \operatorname{sgn}(C_b^{inv}) \frac{|C_b^{inv}|}{\gamma} \gamma \langle \hat{P}_c \rangle (\hat{\mathcal{U}}, b)$$

$$= \underbrace{\gamma}_{\text{scale}} \sum_b \underbrace{\operatorname{sgn}(C_b^{inv}) \, \bar{C}_b^{inv}}_{\text{classical post-processing.}} \underbrace{\langle \hat{P}_c \rangle (\hat{\mathcal{U}}, b)}_{\text{valid QC circuit can run \& find value on HW}}$$

C_b^{inv} can be outside $[0,1]$

$\sum_b C_b^{inv} =: \gamma \geq 1$ generally, for Λ not unitary

eg bit flip channel $\Lambda = (1-p) I \cdot I + p X \cdot X$

$\Lambda^{-1} = (1 + p/(1-2p)) I \cdot I + \frac{-p}{1-2p} X \cdot X$

$C_I^{inv} = 1 + \frac{p}{1-2p}$ $C_X^{inv} = \frac{-p}{1-2p}$

$b = (0,0)$ choice vector $b = (1,0)$

Turn into probabilities

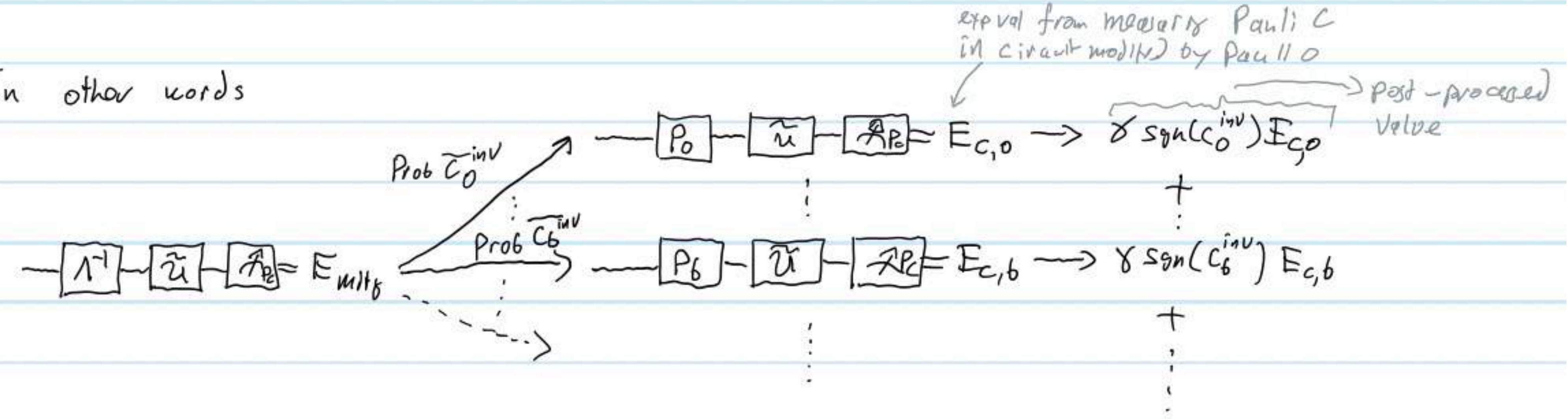
$$C_b^{inv} = \underbrace{\operatorname{sgn}(C_b^{inv})}_{\text{sign} \in \{-1,+1\}} \underbrace{\frac{|C_b^{inv}|}{\gamma}}_{\text{prob} \in [0,1]} \underbrace{\gamma}_{\text{scale}}$$

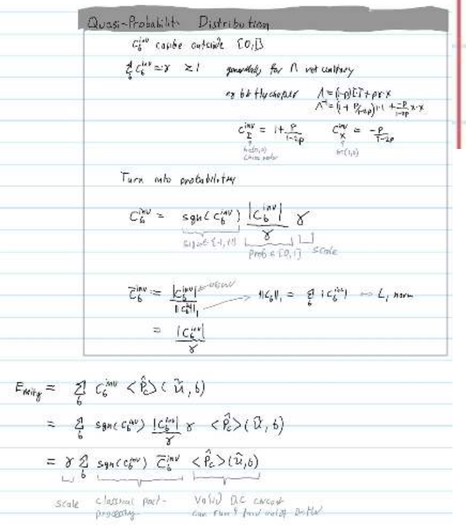
$\bar{C}_b^{inv} := \frac{|C_b^{inv}|}{\|C_b^{inv}\|_1}$ ← abs val $\|C_b\|_1 = \sum_b |C_b^{inv}| \Leftrightarrow L_1$ norm

$= \frac{|C_b^{inv}|}{\gamma}$

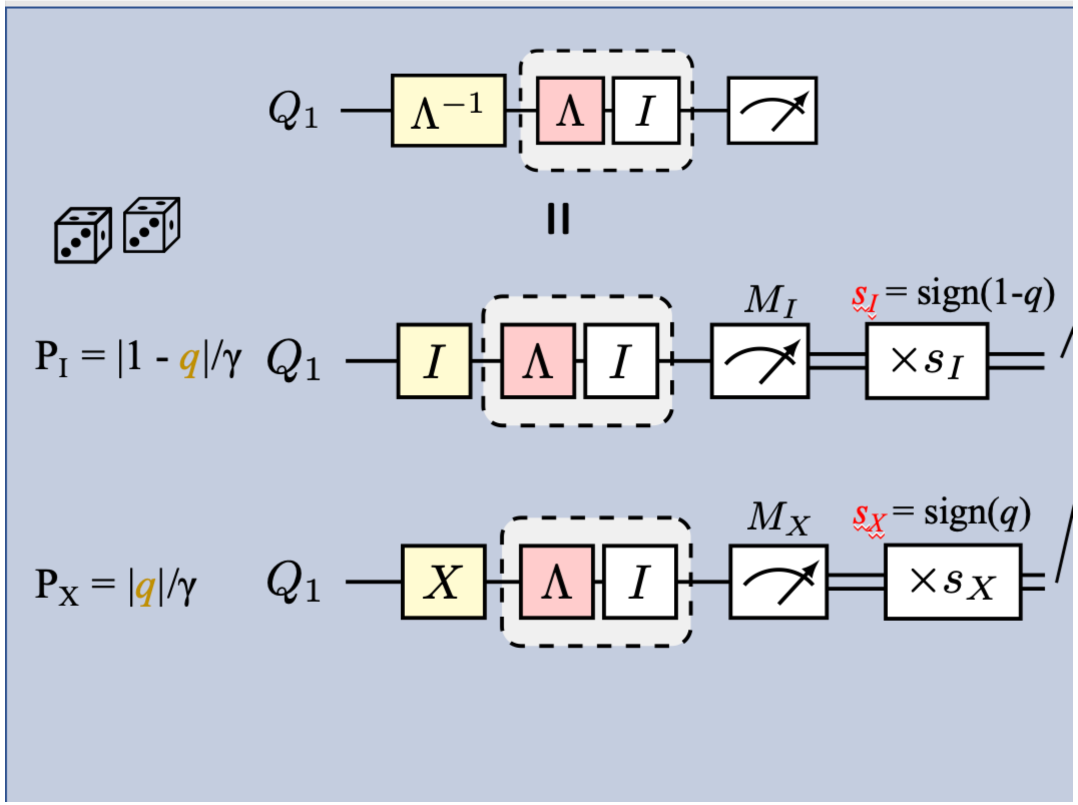
In this form, the decomposition of the error error mitigated expectation value is simply a sum over expectation values of Pauli-gate modified circuits, whose value can be obtained from direct quantum computer execution, weighted by a probability, c bar b inverse, and the sign.
The elements that perform the weighing and rescaling can all be done in classical post-processing.

In other words





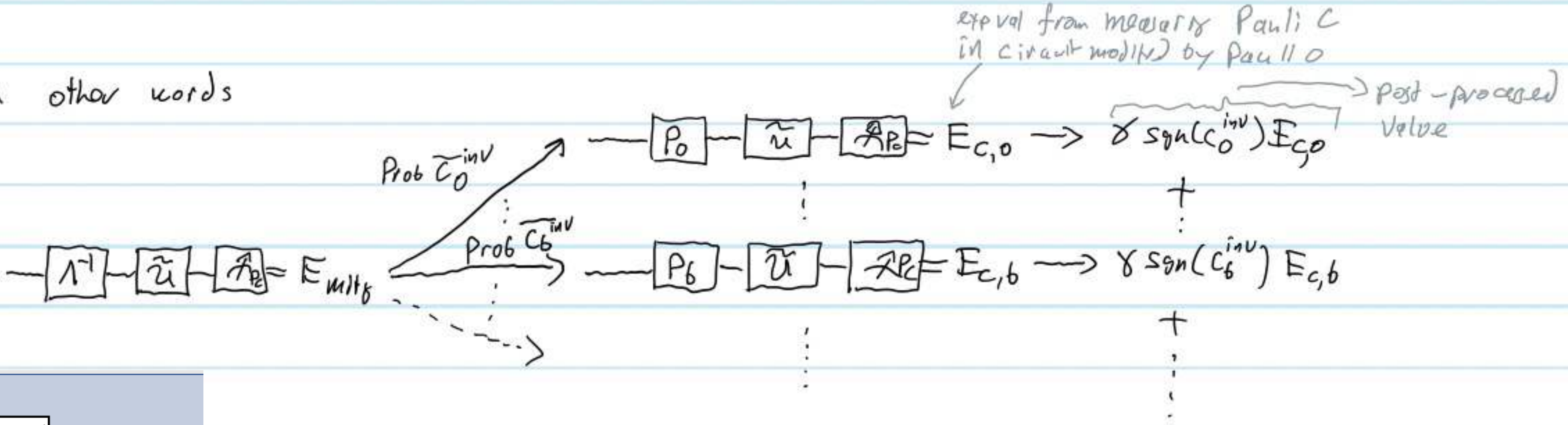
In this form, the decomposition of the error error mitigated expectation value is simply a sum over expectation values of Pauli-gate modified circuits, whose value can be obtained from direct quantum computer execution, weighted by a probability, c bar b inverse, and the sign.
The elements that perform the weighing and rescaling can all be done in classical post-processing.



Estimator

post processing:

In other words

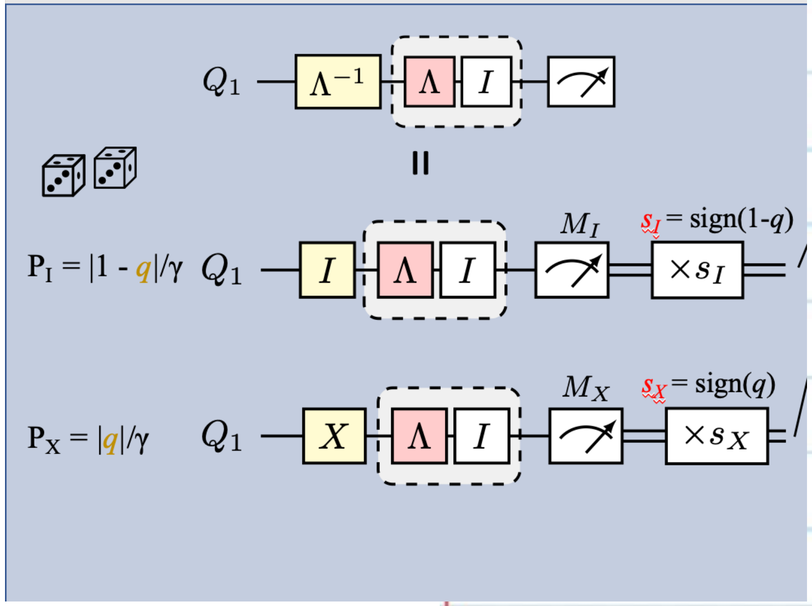


$$\underbrace{E_{c,mitg}}_{} = \sum_b \gamma\, \mathrm{sgn}(c_b^{inv})\, E_{c,b}$$

mitigated value for Pauli C obtained from the quasi-prob distribution

From above, we know this is an unbiased estimator,

but what about the error and sampling



mitigated value for Pauli C obtained from the quasi-prob distribution

From above, we know this is an unbiased estimator,
but what about the error and sampling

Estimator, Sampling, and Error Bounds

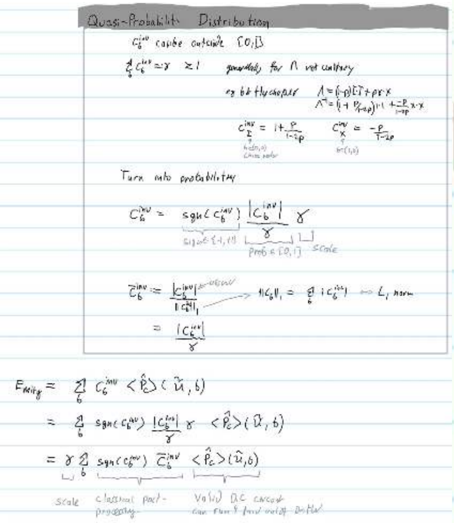
Sample circuits of the form

$$\left\{ \underbrace{\bar{C}_b^{inv}}_{\text{prob}} : -\boxed{P_b}-\boxed{\tilde{\mathcal{U}}}-\boxed{\tilde{\mathcal{X}} P_c} = \underbrace{Y \in \{-1,+1\}}_{\substack{\text{Pauli observable} \\ \text{single-shot outcome} \\ \text{ie e-vals}}} \quad \longrightarrow \underbrace{X = \text{sgn}(C_b^{inv})\, Y}_{\substack{\text{post processed value } X \\ X \in \{+1,-1\}}} \right\}$$

Let's say we sample M instances, randomly sampling a single value for b and obtain one-shot value on the QC, ie one random instance of $Y=1$ or $Y=-1$, which we then post-process.

The results are thus the classical random variables

$$\{X_1, X_2, \ldots, X_M\} \quad \text{or} \quad \{X_m : m=1,\ldots,M\}$$



In this form, the decomposition of the error error mitigated expectation value is simply a sum over expectation values of Pauli-gate modified circuits, whose value can be obtained from direct quantum computer execution, weighted by a probability, c bar b inverse, and the sign.
The elements that perform the weighing and rescaling can all be done in classical post-processing.

Probabilistic error cancelation: Derivation

where each $X_m \in \{1, -1\}$ and is distributed to modified Bernoulli distribution with some probability which can be any valid value and can vary from shot-to-shot m.

Our mitigation estimator is then for M shots:

rand outcome of m-th shot for b-th circ.

noisy circuit

$$E_M := \gamma \frac{1}{M} \sum_{m=1}^{M} X_m = \frac{1}{M} \sum_{m=1}^{M} \gamma \,\mathrm{sgn}(c^{\mathrm{inv}}_{b_m}) \, Y_{m_b}(\tilde{U}, b_m)$$

Pauli chosen for m-th shot

There are now 2 random processes:

b_m : which pauli b we pick for shot m

Y_{b_m} : which outcome ± 1 we get for b_m circuit of shot m

Unbiased Estimator of the Ideal, noise-free circuit expectation

$$\mathbb{E}[E_M] = \frac{1}{M} \sum^{M} \mathbb{E}[\gamma X_m] \qquad \text{iid rand vars}$$

Unbiased estimator

$$\mathbb{E}[E_M] = \frac{1}{M}\sum_{m=1}^{M} \mathbb{E}[\gamma X_m]$$ iid rand vars

$$= \mathbb{E}[\gamma X_m]$$ no X_m is different

$$= \mathbb{E}\left[\gamma \, \mathrm{sgn}(c^{inv}_{b_m}) \, Y_{b_m}(\tilde{u}, b_m)\right]$$ where rand var is b_m now, not just m, so

$$= \mathbb{E}_{b,Y}\left[\gamma \, \mathrm{sgn}(c^{inv}_b) \, \underbrace{Y_b(\tilde{u}, b)}_{\text{drop notation}}\right]$$ $\mathrm{Prob}[b] = \bar{c}^{inv}_b$

$$= \sum_b \mathbb{E}_Y\left[\gamma \, \mathrm{sgn}(c^{inv}_b) \, Y_b\right] \mathrm{Prob}[b]$$

$$= \sum_b \underbrace{\gamma \, \mathrm{sgn}(c^{inv}_b)}_{\text{post-process of outcome}} \underbrace{\bar{c}^{inv}_b}_{\text{sample prob}} \mathbb{E}_Y[\underbrace{Y_b}_{\text{rand outcome}}]$$ $\longrightarrow$

note: $\langle \hat{P}_c \rangle(\tilde{u}, b) = \langle\langle P_c | \mathcal{U} \Lambda \mathcal{P}_b | \rho_0 \rangle\rangle$
$= \mathbb{E}[Y_b]$
Rand variable ± 1 for output of the b-th pauli circuit.

$\therefore$ for some classical func of b $f(b)$ which does not depend on the value Y_b but only on the label b:

$\mathbb{E}[f(b) Y_b] = f(b) \langle \hat{P}_c \rangle(\tilde{u}, b)$

$$= \sum_b \gamma \, \mathrm{sgn}(c^{inv}_b) \frac{c^{inv}_b}{\gamma} \langle\langle P_c | \mathcal{U} \Lambda \mathcal{P}_b | \rho_0 \rangle\rangle$$

Probabilistic error cancelation: Derivation

$$\approx \langle\langle P_c | \mathcal{U} \Lambda \left(\sum_{\sigma} c_{\sigma}^{inv} \mathcal{P}_{\sigma} \right) | \rho_0 \rangle\rangle$$

$$= \langle\langle P_c | \mathcal{U} \Lambda \Lambda^{-1} | \rho_0 \rangle\rangle$$

$$= \langle\langle P_c | \mathcal{U} | \rho_0 \rangle\rangle$$

$$= \langle \hat{P}_c \rangle (\mathcal{U}_{ideal})$$ without noise!

unbiased estimator of the true noise-free, ideal value of the circuit

the value Y_σ but only on the label σ:

$$\mathbb{E}[f(\sigma) Y_\sigma] = f(\sigma) \langle \hat{P}_c \rangle (\tilde{\mathcal{U}}, \sigma)$$

(Optional step) Variance

Variance of E_M

$$\mathbb{V}_{\underbrace{\{\sigma_m, X_m\}}_{\text{with respect to}}}[E_M] = \frac{\gamma^2}{M^2} \sum_{m=1}^{M} \mathbb{V}[X_m]$$

X_m iid
can drop subscript m
and emphasize σ and value X

$$= \frac{\gamma^2}{M} \mathbb{V}[X(\tilde{\mathcal{U}}, \sigma)]$$

note the same variance is just
rescaled by γ^2 due to γ

Generalizing: Raveling trajectories with quasiprobabilities

Channel we *want* to implement

CPTP operation we *can* implement

$$\mathcal{C}(\cdot) = \sum_i a_i \mathcal{F}_i(\cdot)$$

Real coefficients, turn into quasi-probability

Putting the following techniques all on the same footing

Technique	**Channel $\mathbb{C}$**
Prob. error cancelation (PEC)	noise inverse
Cricut cutting (knitting) of gates	non-local gate
Cricut cutting of wires	large unitary
Classical sim. algorithms (QP)	unitary

Discussions with David Sutter