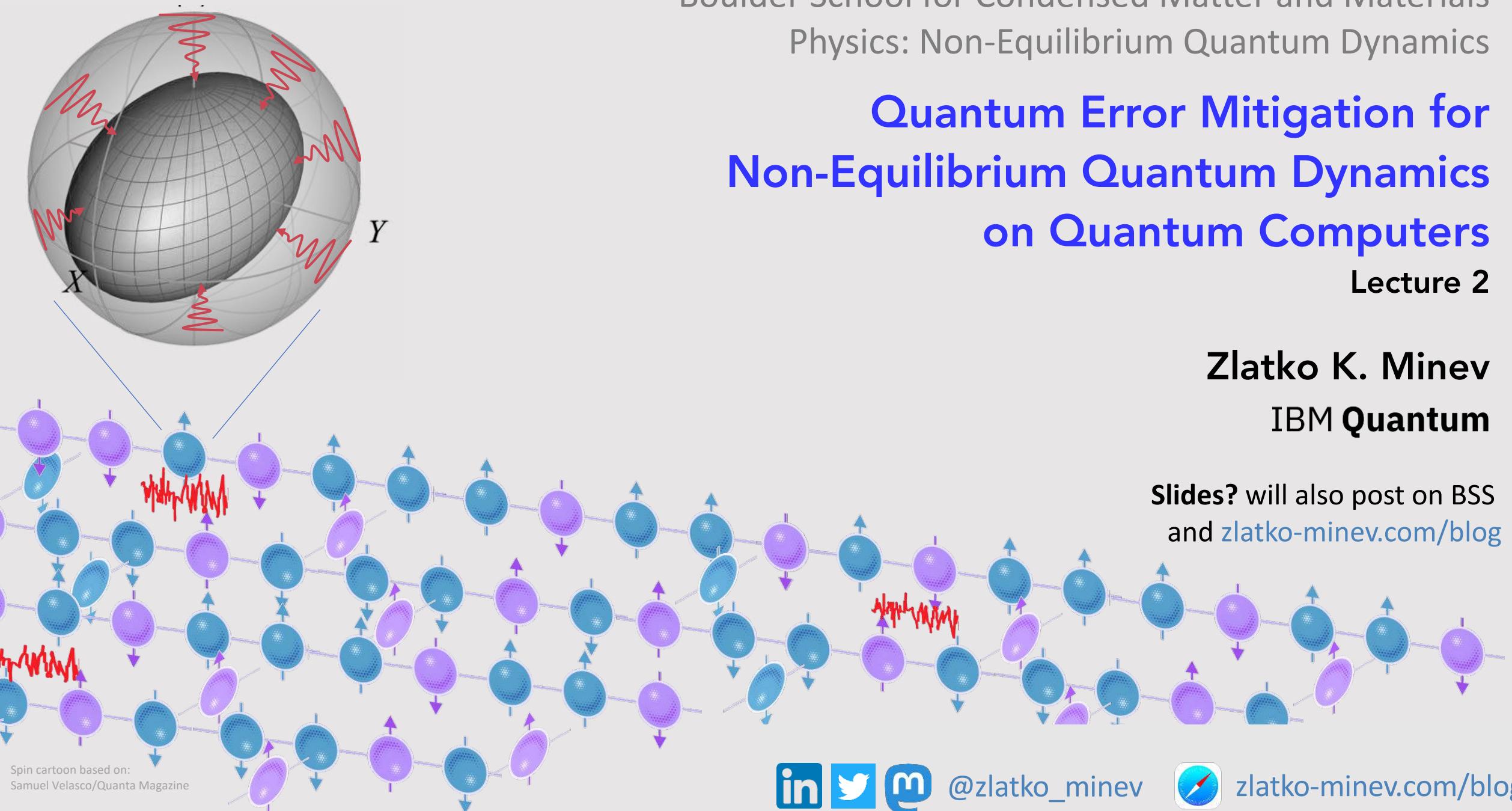


Quantum Error Mitigation for Non-Equilibrium Quantum Dynamics on Quantum Computers

Lecture 2

Zlatko K. Minev
IBM Quantum

Slides? will also post on BSS
and zlatko-minev.com/blog



Spin cartoon based on:
Samuel Velasco/Quanta Magazine



@zlatko_minev



zlatko-minev.com/blog

Where can you find things?

Lecture Slides

Boulder School for Condensed Matter and Materials Physics
boulderschool.yale.edu/2023/boulder-school-2023-lecture-notes



Will also post on zlatko-minev.com/education

Tutorials and additional lecture notes

Twirling, Measurements and Walsh-Hadamard

Cheat sheets, Videos, ...

zlatko-minev.com/blog

See also lectures on qiskit.org/learn

Tutorials and additional lecture notes

Latest seminar qiskit.org/events/seminar-series

A collage of several documents and diagrams related to quantum computing:

- A diagram titled "7. Digital quantum circuits (pictorial)" showing basic elements like Quantum wire, Classical wire, and various types of gates (Control gate U, Control gate U-bar, Control-X (NOT)).
- A diagram titled "Primer on Pauli Twirling" showing a sequence of operations $P_a \rightarrow A \rightarrow P_a^\dagger$ and a 2D grid visualization.
- A diagram titled "learn and cancel quantum noise cancellation with sparse Pauli-Lindblad models on quantum processors" showing a complex circuit with many qubits and controls.
- A "Cheat sheet: Digital quantum circuits - pictorial 101" document with the subtitle "Making a qubit cheatsheet, On this festive Christmas day."
- A "Tutorial on tailoring quantum noise - Twirling 101 (Parts I-IV)" document.
- A "Nutshell introduction to tailoring quantum noise by twirling into stochastic Pauli or Pauli-Lindblad Models on Noisy Quantum Processors" document.

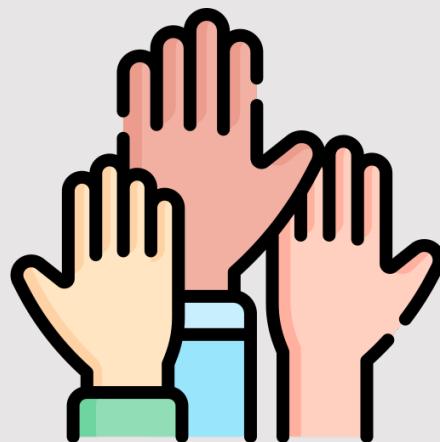


Zlatko Minev, IBM Quantum (2)

What is one thing you
learned in Lecture 1?



Review of Lecture 1



Quantum Error Mitigation for Non-Equilibrium Quantum Dynamics

Lecture 1

Big picture

Why quantum computers?
Status and outlook

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

Mitigation fundamentals

Probabilistic error cancelation (PEC)
Introduction
One qubit example



Quantum computers

My experience circa 2010

Maybe 1 or 2 qubits working some small fraction of the time in select labs

Photo with dilution fridge called Sunshine from Michel Devoret's lab at Yale during my Ph.D.



Hopes for a working qubit in here

Minev, IBM Quantum (9)

This year 2023

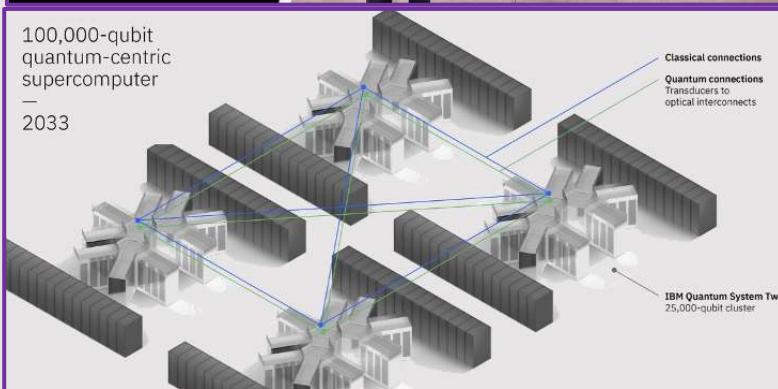
A 127-qubit quantum computer installed in the lobby cafeteria of a research building dutifully executing jobs almost all the time.



at Cleveland Clinic

100,000-qubit quantum-centric supercomputer

— 2033



Biggest Problem: Noise

Quantum simulation on a quantum computer

Execute on a real quantum computer device and obtain results as classical data

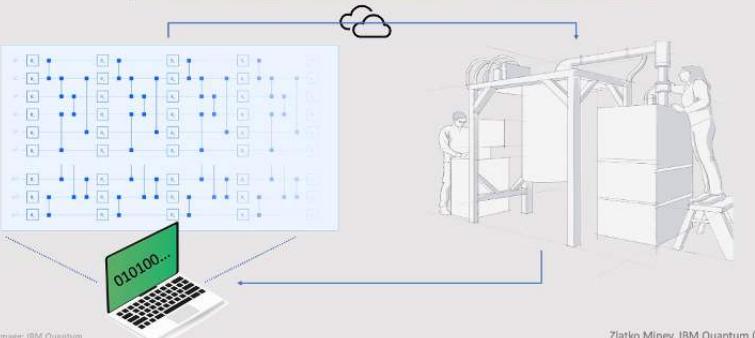
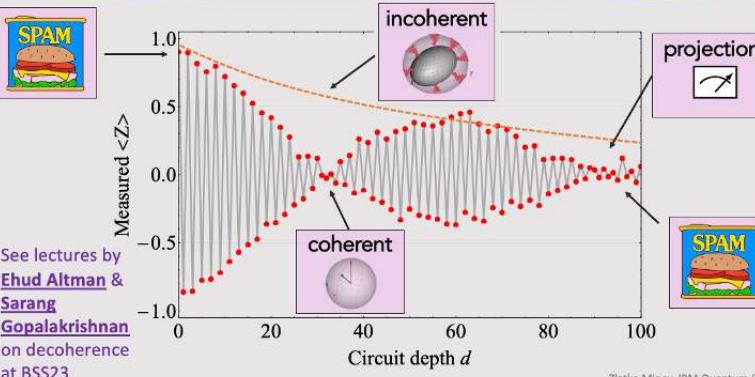


Image: IBM Quantum

Zlatko Minev, IBM Quantum (17)

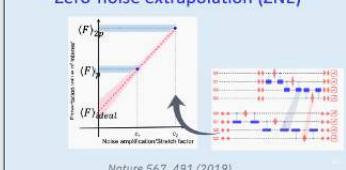
Elements of noise



Zlatko Minev, IBM Quantum (33)

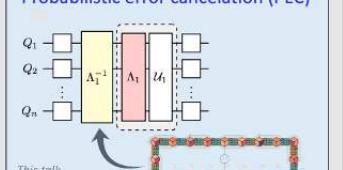
Error mitigation landscape

Zero-noise extrapolation (ZNE)



more speed

Probabilistic error cancellation (PEC)



more information, accuracy

Mitigation

Overview of some key experimental progress in error mitigation:

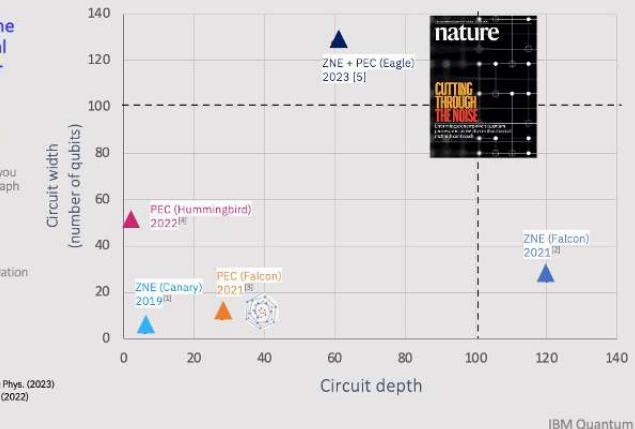
Error mitigation

No matter what you do you have to chop it to this graph

PEC: Probabilistic error cancellation

ZNE: Zero-noise extrapolation

- [1] Kandala, Nature (2019)
- [2] Kim, Nature Phys. (2023)
- [3] van den Berg, Minev, Nature Phys. (2023)
- [4] Temme, IBM Research Blog (2022)
- [5] Kim, Nature (2023)



IBM Quantum

High-level message

Learn
accurate, efficient, scalable



Cancel
noise with noise, practical

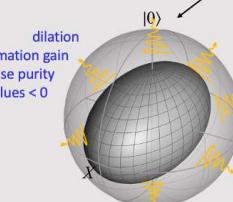
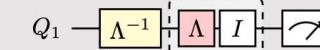


Cost
more noise more cost

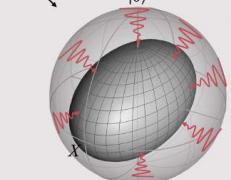


Zlatko Minev, IBM Quantum (48)

Inverse of noise map is not physical



contraction loss of information reduced purity eigenvalues > 0



Quantum Error Mitigation for Non-Equilibrium Quantum Dynamics

Lecture 2

Mitigation fundamentals

Probabilistic error cancelation (PEC)

Summarize one qubit example

Analogy to random walks

Error bars & confidence

Generalize (optional)

Show unbiased estimator



Learning quantum noise

Challenge

Overcoming: sparse model

Putting it together

Experiments – Ising model

Consequences for the big picture

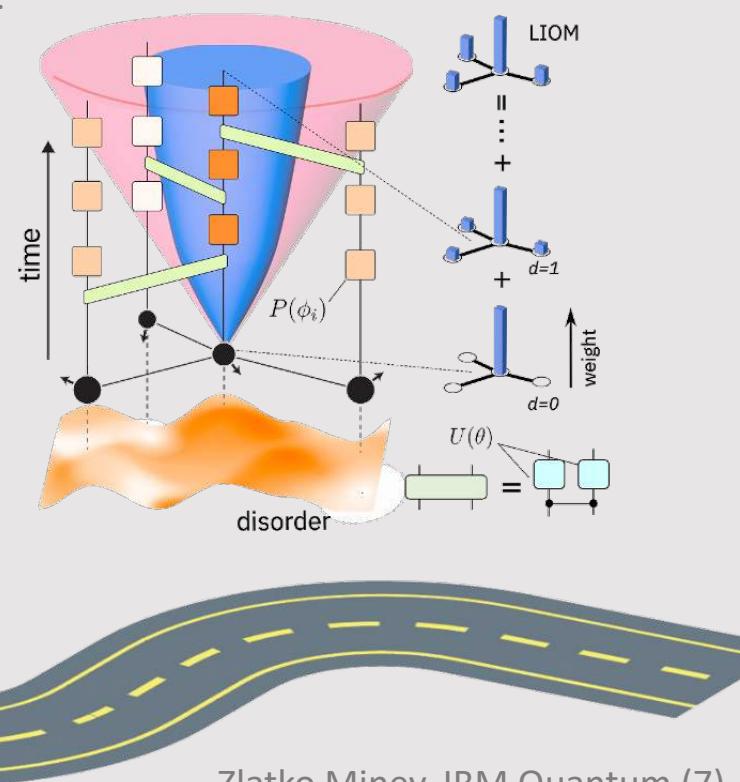
Hardware progress

Lecture 3

Wrap up of Lecture 2, Twirling, ...

State-of-art experiments at the 120Q+, depth 50+: uncovering local integrals of motion

...



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



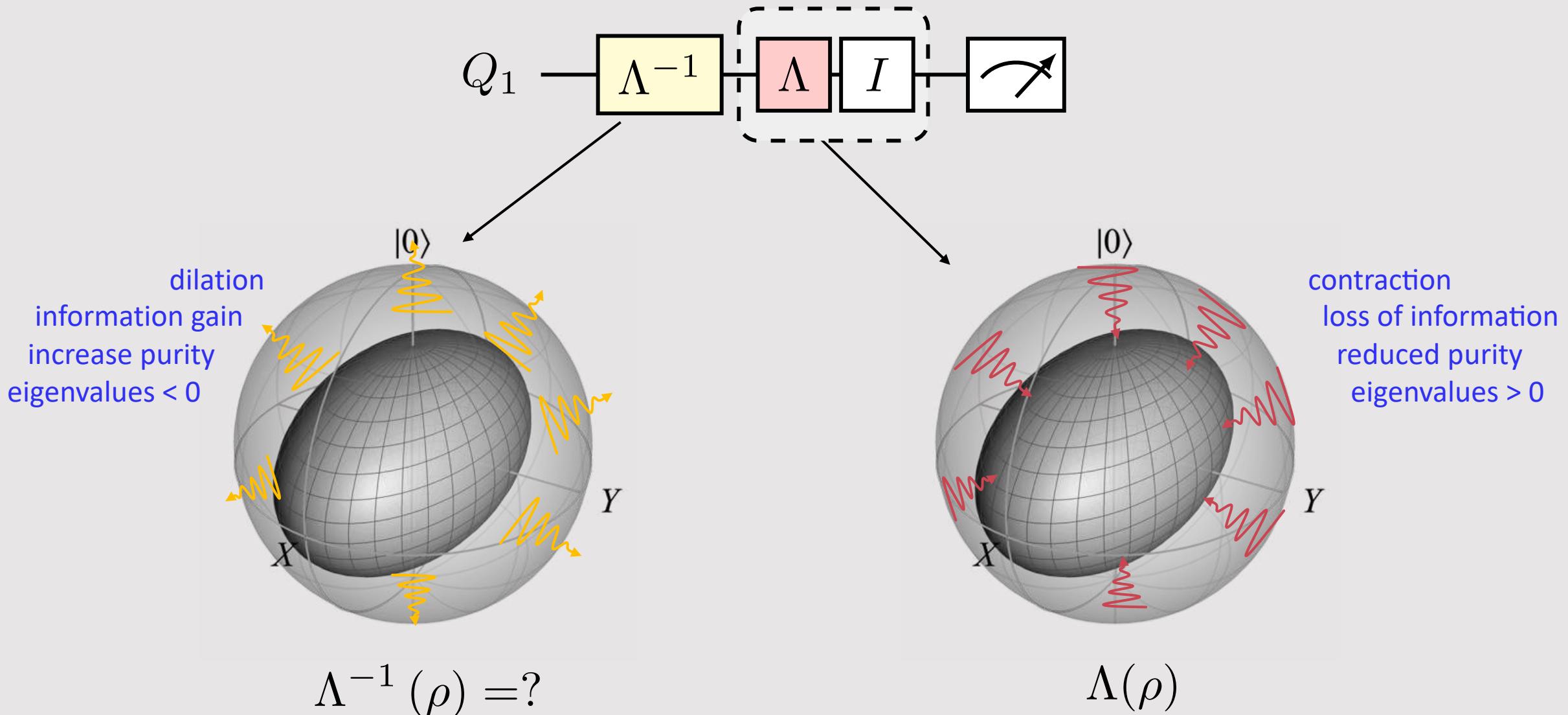
Got Slides?



Paper: Nature Physics (2023)

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

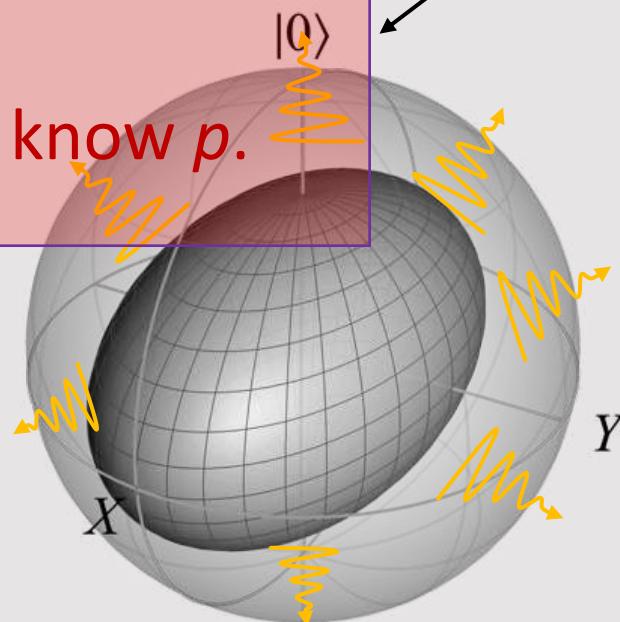
Inverse of noise map is not physical



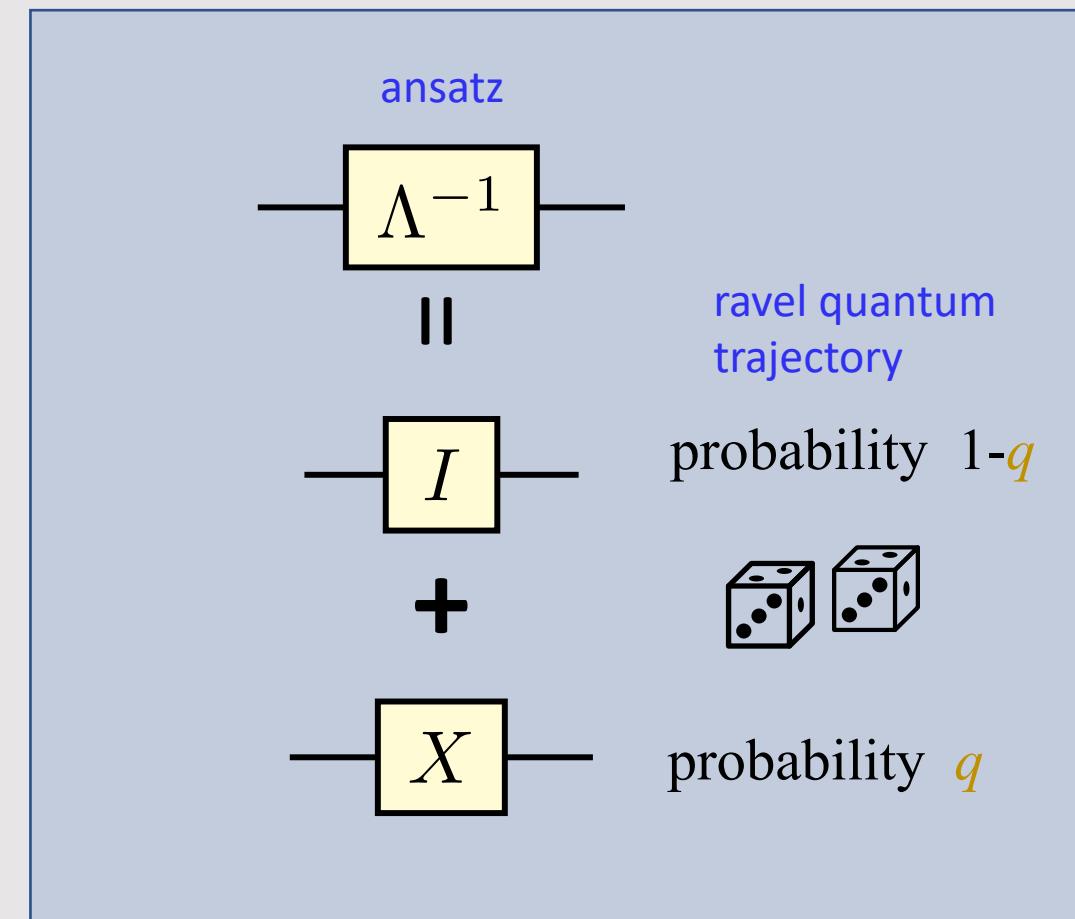
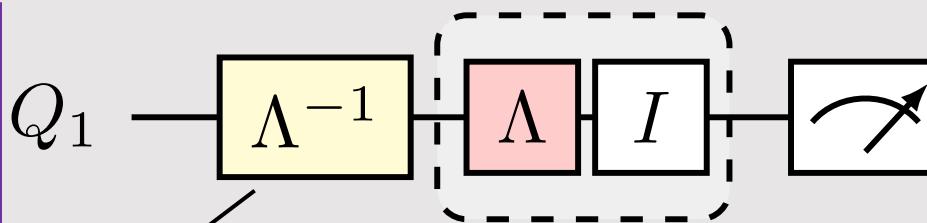
Inverse of noise map is not physical

Critically hinges on
knowing the noise
exactly!

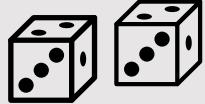
Need to know $p.$

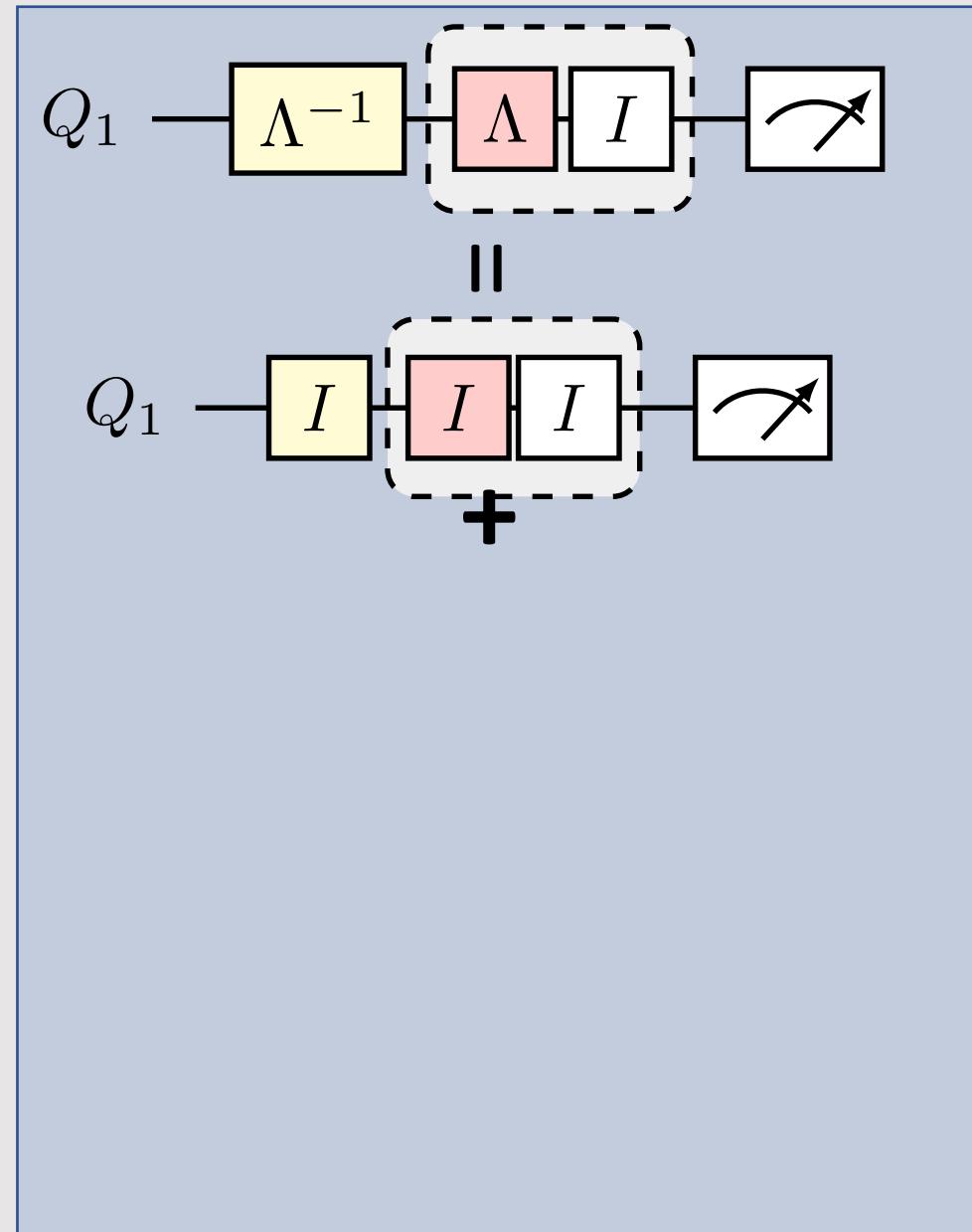


$$\Lambda^{-1}(\rho) = ?$$



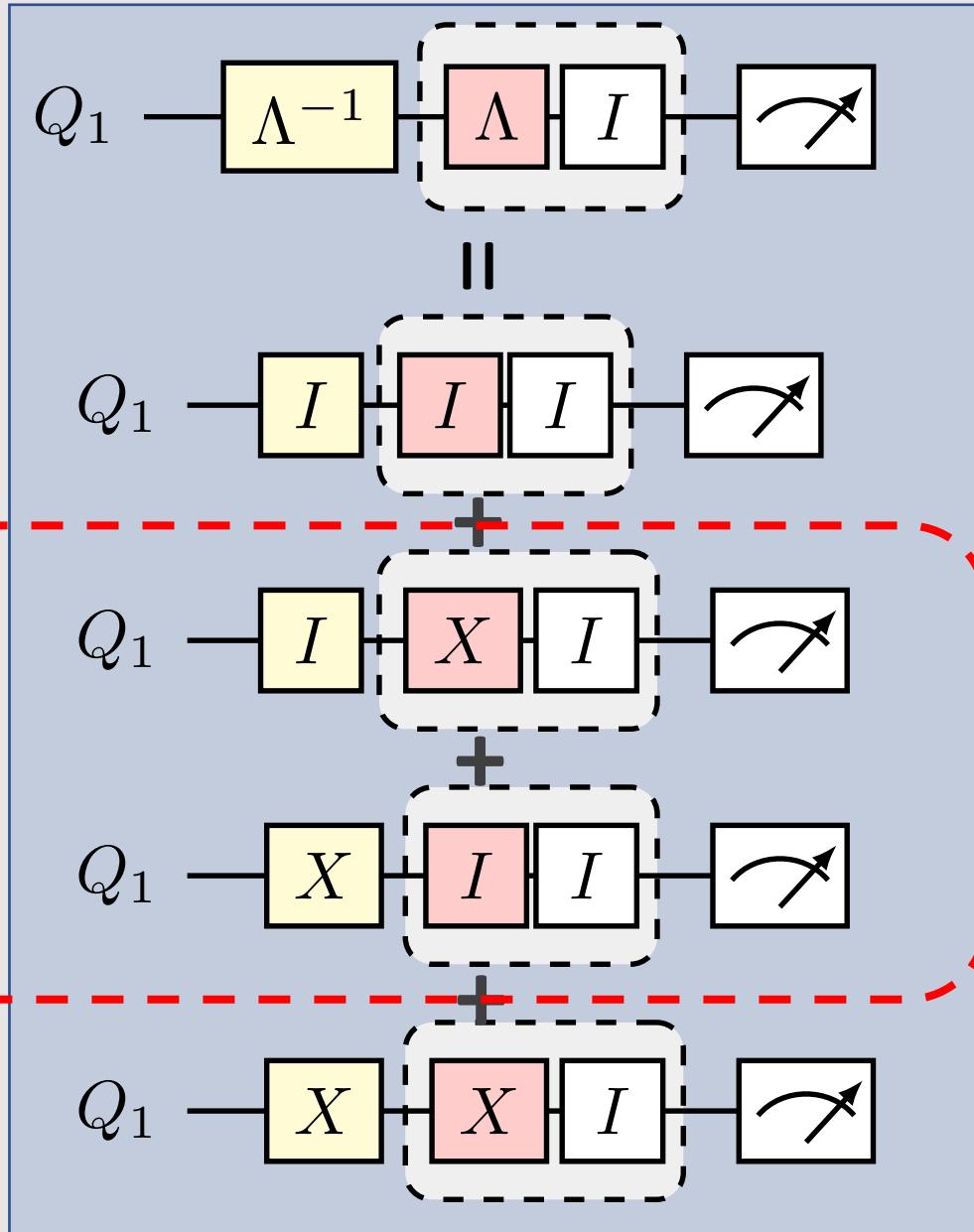
Raveling quantum trajectories to undo noise

No error probability
 $(1-q)(1-p)$




Raveling quantum trajectories to undo noise

No error	probability $(1-q)(1-p)$
ERROR!	$(1-q)p$
ERROR!	$q(1-p)$
Error CANCELED!	qp

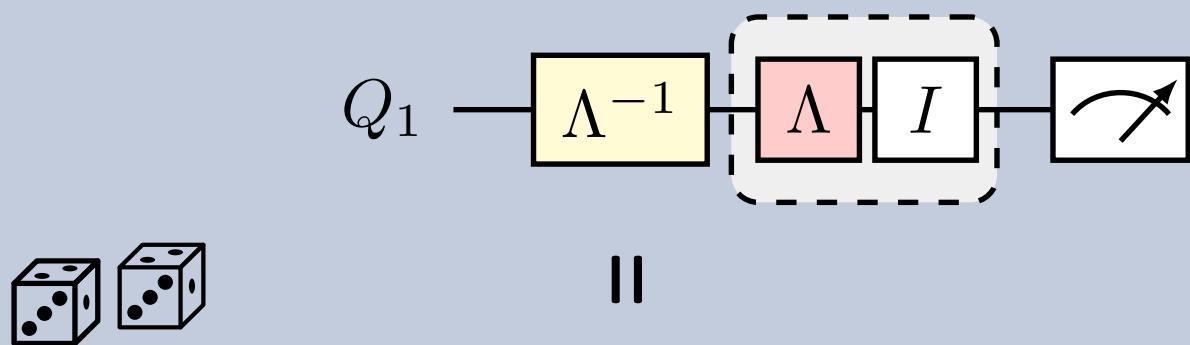


Solution to noise free!

$$q = \frac{-p}{1 - 2p}$$

Interfere
Destructively

Raveling quantum trajectories to undo noise



probability

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

sign

$$\begin{array}{c} \text{---} \quad \text{---} \\ \boxed{I} \\ \text{---} \quad \text{---} \end{array} \quad s_I = \text{sign}(1-q)$$

+

$$P_X = |q|/\gamma$$

$$\begin{array}{c} \text{---} \quad \text{---} \\ \boxed{X} \\ \text{---} \quad \text{---} \end{array} \quad s_X = \text{sign}(q)$$

$$P_I + P_X = 1$$

Solution to noise free!

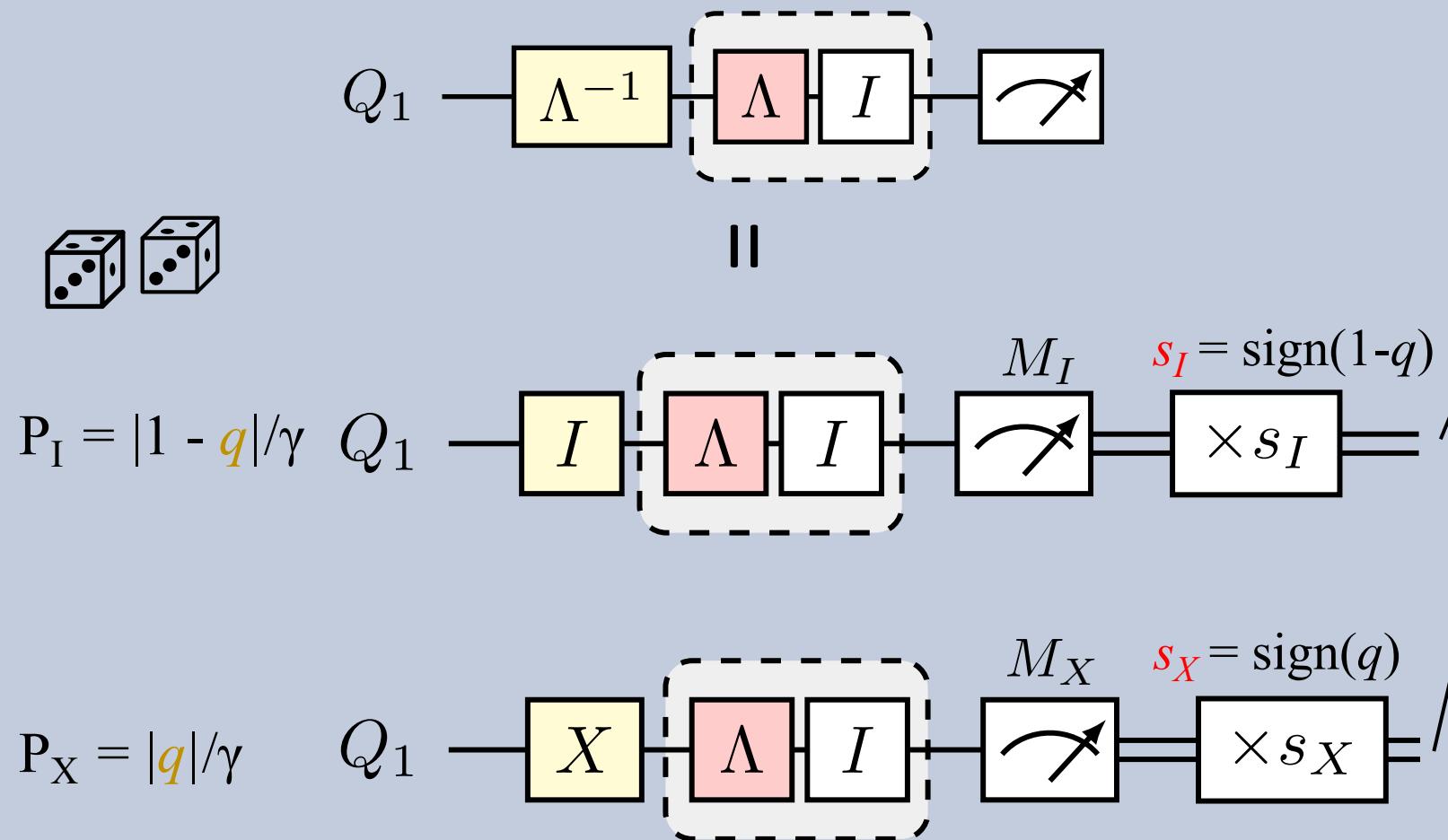
$$q = \frac{-p}{1 - 2p}$$

norm

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

Quasi-probability

Raveling quantum trajectories to undo noise



mitigated expectation

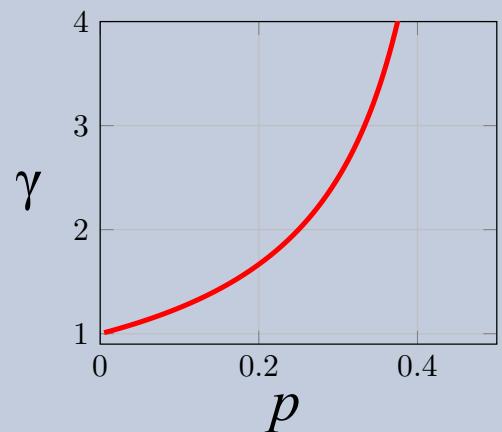
$$\langle M \rangle = \gamma(s_I P_I M_I + s_X P_X M_X)$$

Gain: Bias-free estimate!

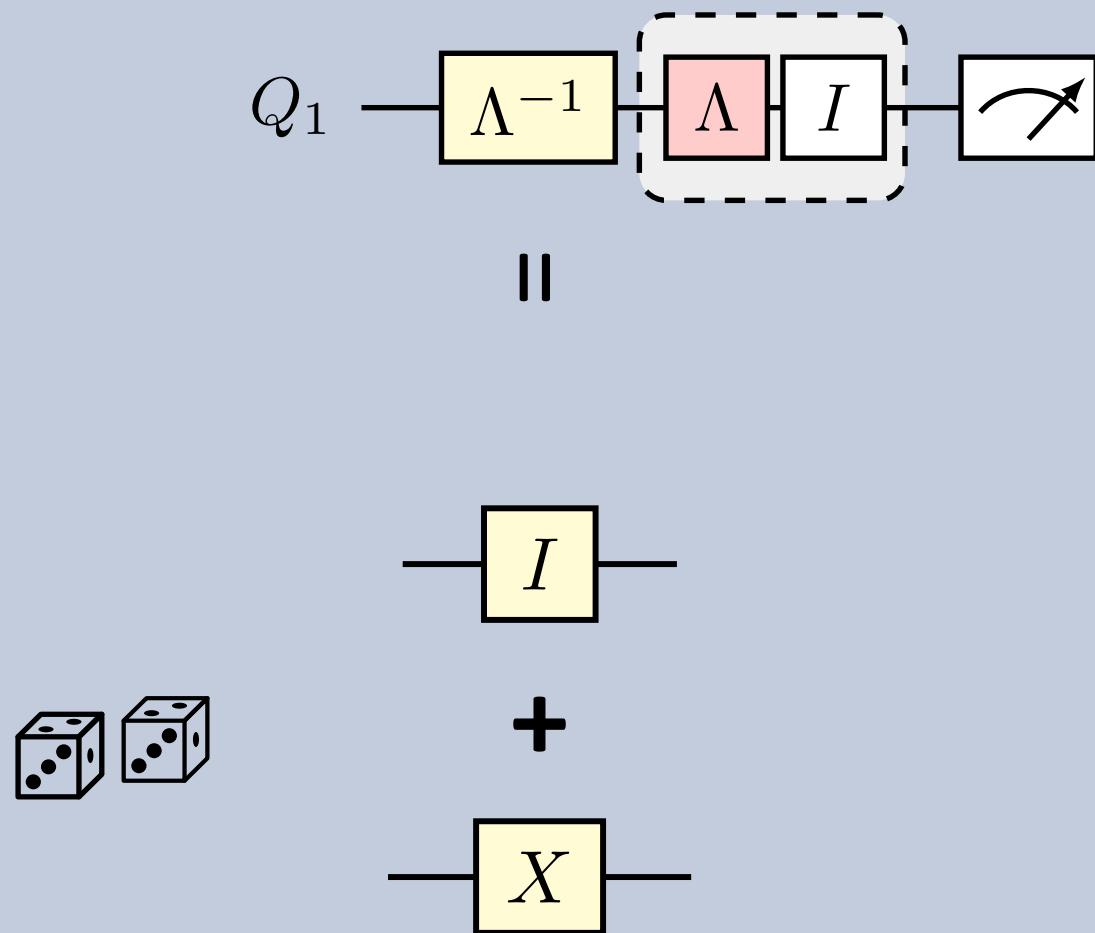
Cost: Variance

Sampling overhead

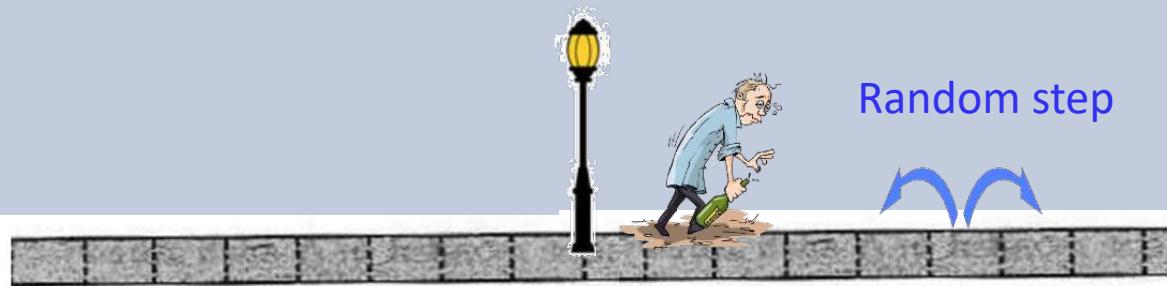
$$\gamma = |1 - q| + |q|$$



Cancelling noise with noise



Cancelling noise with noise: Drunkard's classical random walk analogy



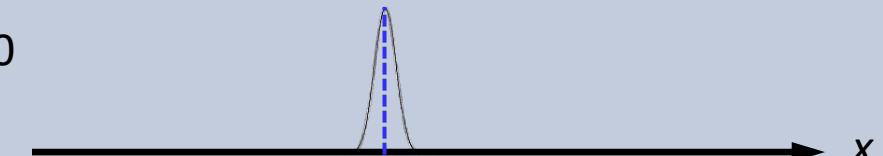
Random step

$$P(1 \text{ step left}) = \frac{1}{2}-p$$

$$P(1 \text{ step right}) = \frac{1}{2}+p$$

Distribution of random walk

$t = 0$



$t > 0$



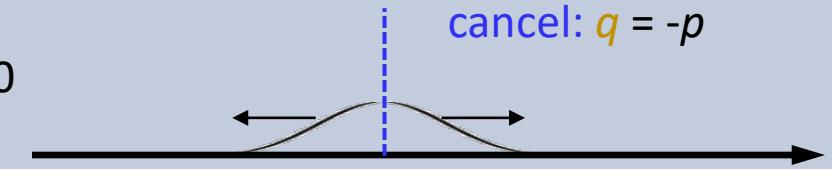
add 2nd random process
wind blows

$$P(1 \text{ step left}) = \frac{1}{2}-q$$

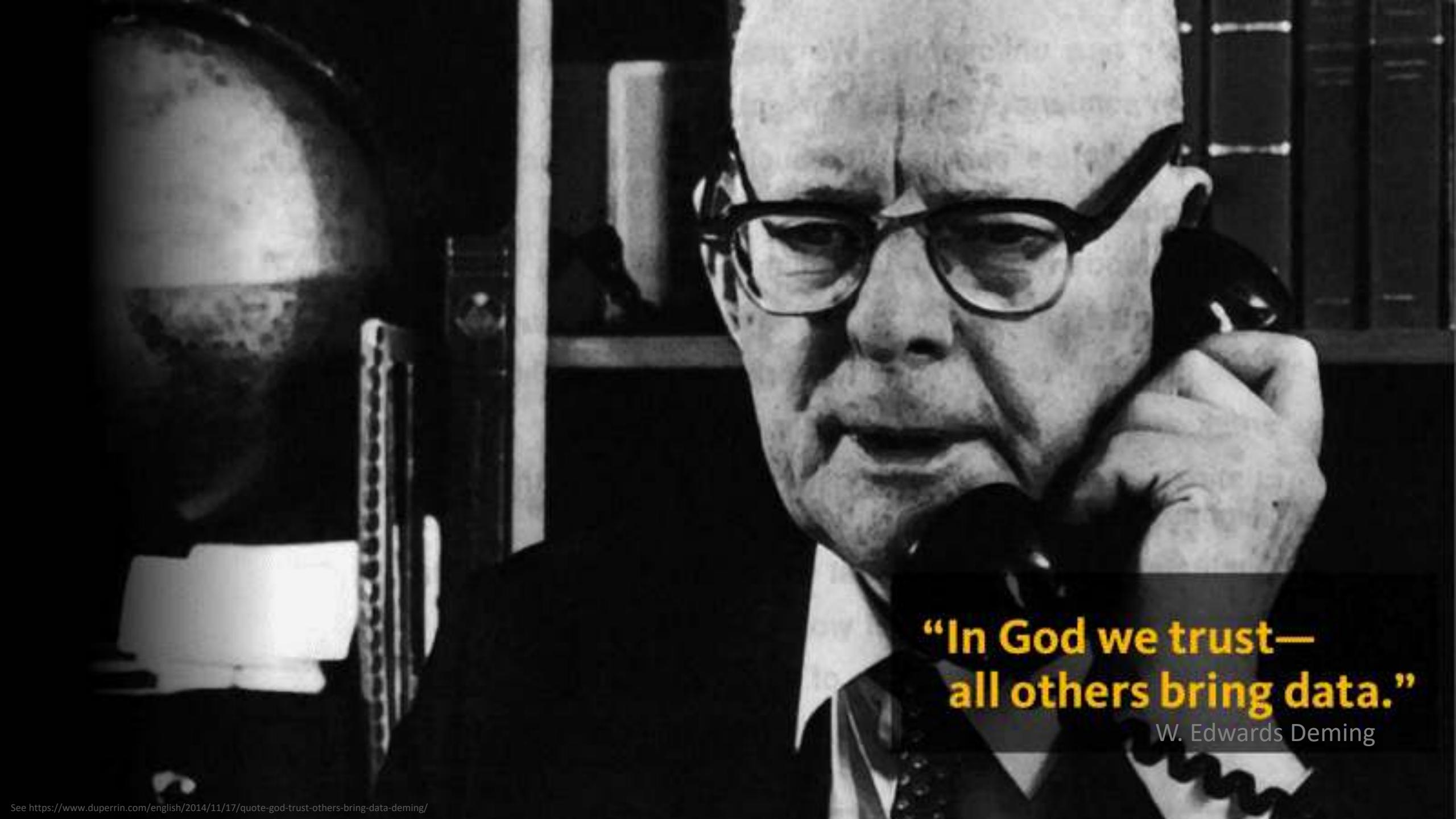
$$P(1 \text{ step right}) = \frac{1}{2}+q$$

Distribution of random walk with wind

$t > 0$



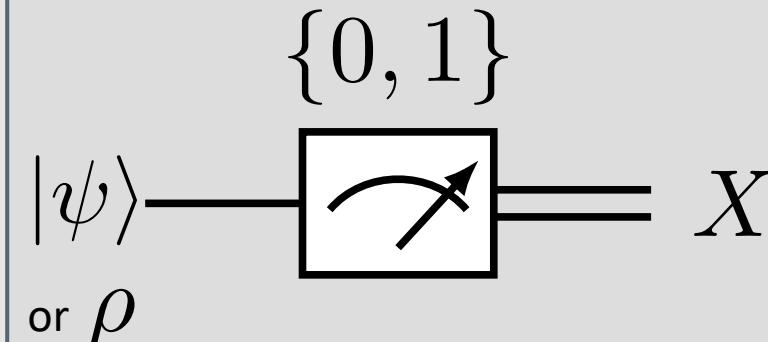
Gain: Bias-free estimate!
Cost: Variance



**“In God we trust—
all others bring data.”**

W. Edwards Deming

Summary: Qubit measured in the computational basis



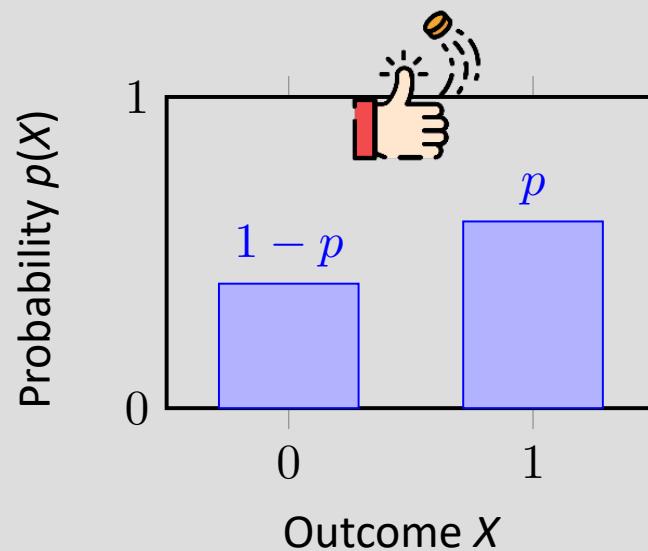
Set of possible outcomes

$$\Sigma = \{0, 1\}$$

$$X \in \Sigma$$

Measurement operators

$$\begin{array}{ll} x & \hat{\mu}(x) \\ 0 : & \hat{\mu}(0) = |0\rangle\langle 0| \\ 1 : & \hat{\mu}(1) = |1\rangle\langle 1| \end{array}$$



Probability to measure outcome

$$\begin{cases} X = 0 : & p(X = 0) = \text{Tr}(\hat{\mu}(0)^\dagger \rho) = \text{Tr}(|0\rangle\langle 0| \rho) = \frac{1}{2}(1 + \langle Z \rangle) \\ X = 1 : & p(X = 1) = \text{Tr}(\hat{\mu}(1)^\dagger \rho) = \text{Tr}(|1\rangle\langle 1| \rho) = \frac{1}{2}(1 - \langle Z \rangle) \end{cases}$$

Bernoulli distribution. Single shot outcome follows a Bernoulli distribution:

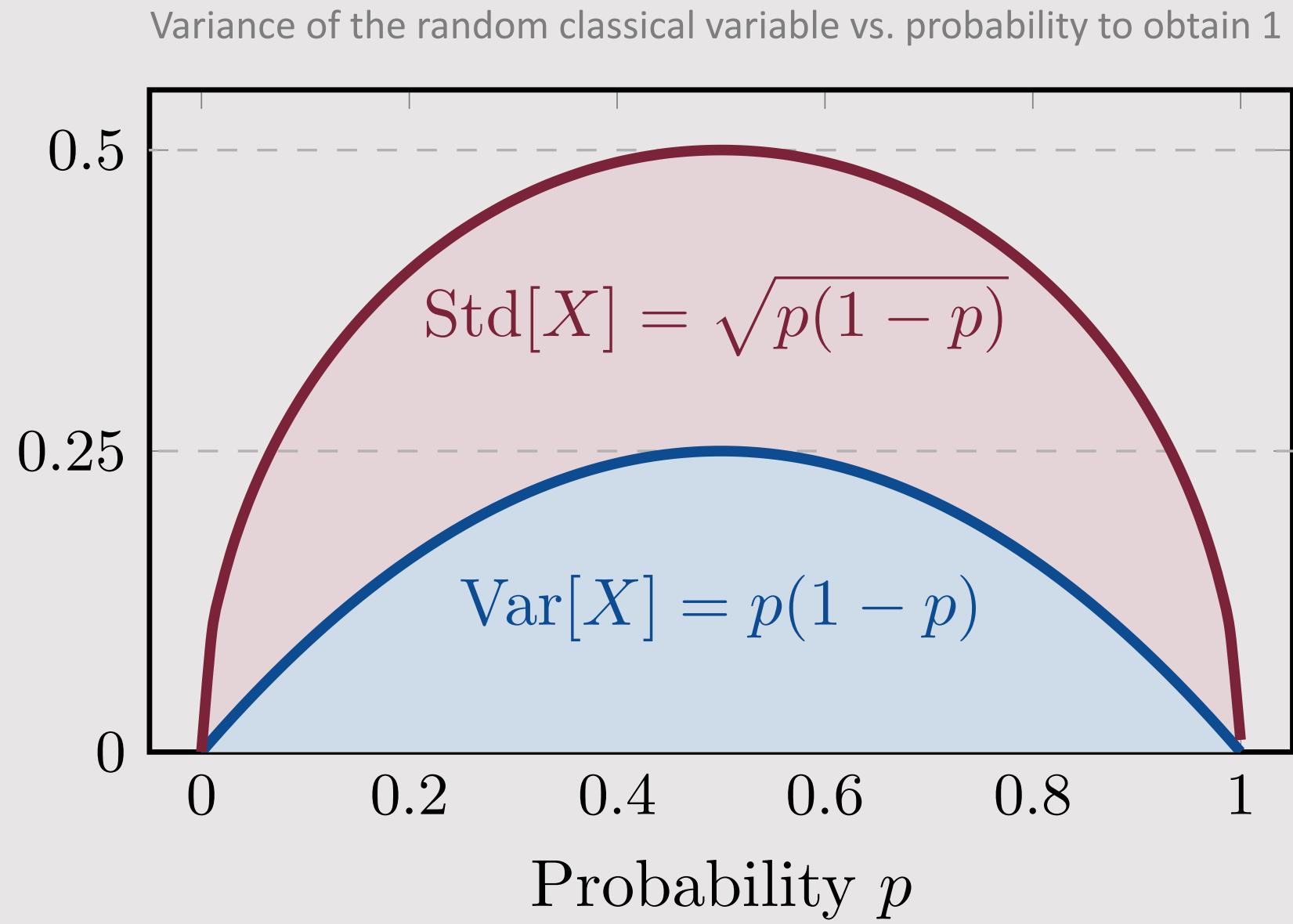
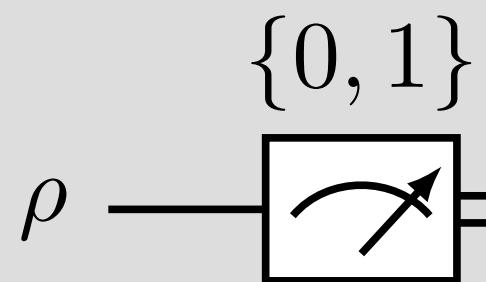
$$X \sim \text{Bernoulli}(p)$$

$$p := \text{Tr}(|1\rangle\langle 1| \rho) \in [0, 1]$$

$$\text{E}[X] = p$$

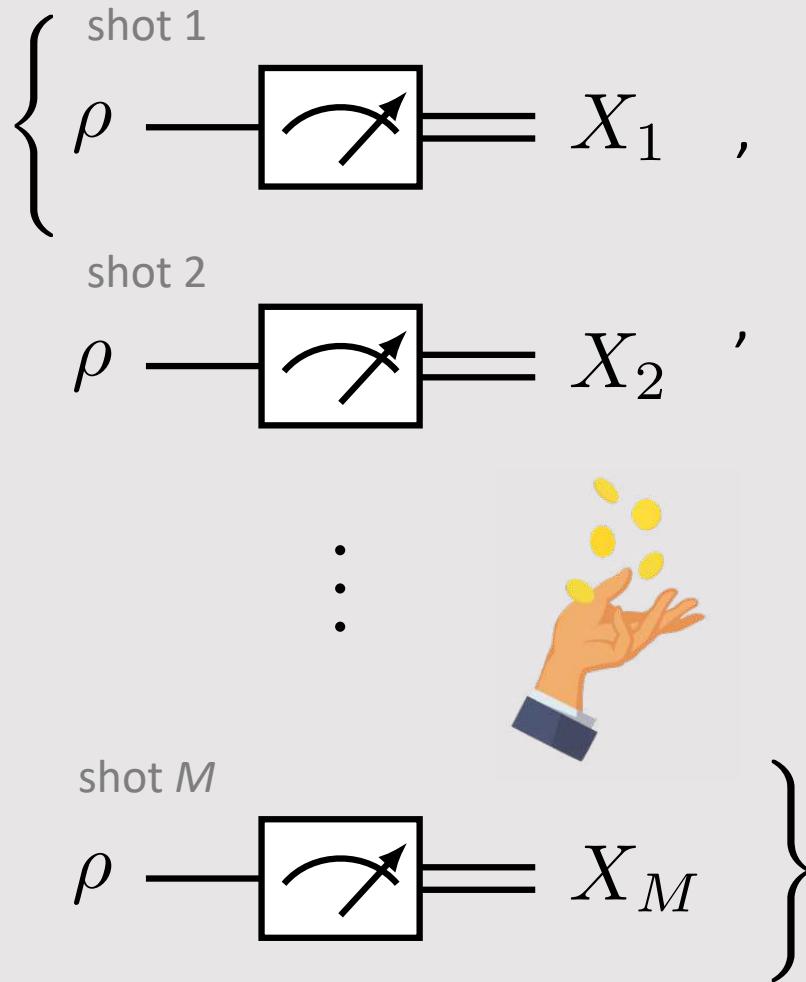
$$\text{Var}[X] = p(1 - p)$$

Quantum projection noise for a single shot





Ideal single qubit measurement with M shots



M shots with IID distribution

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

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

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

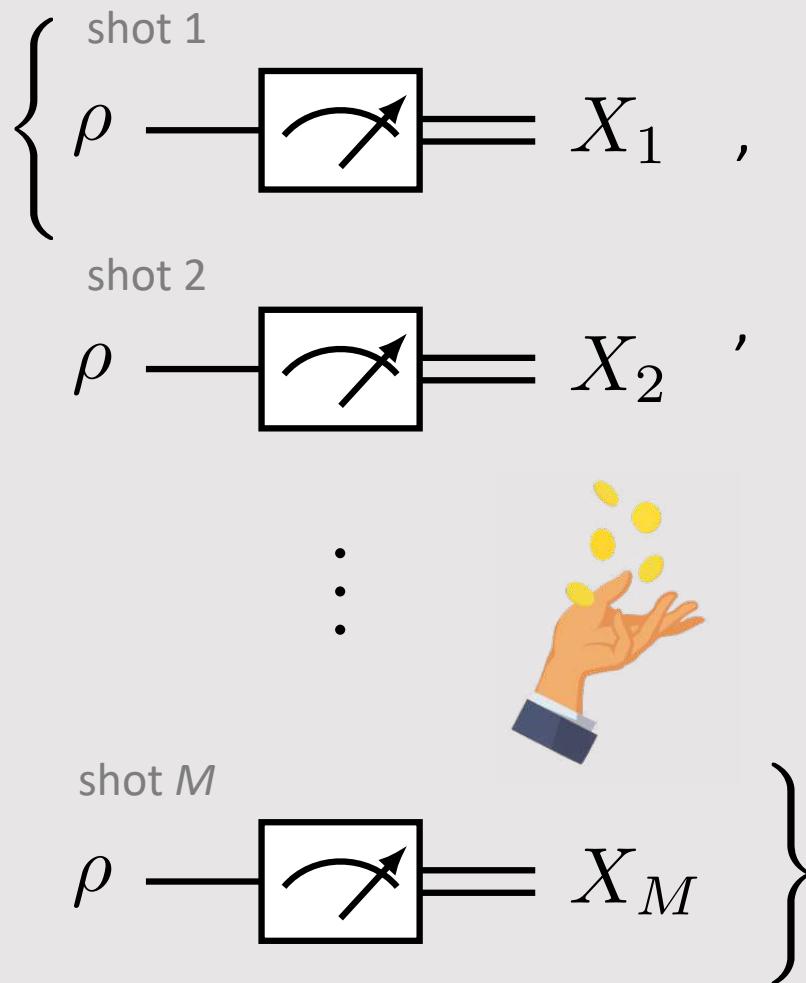
Empirical mean random variable (sample mean statistic)

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

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



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

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

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

$$= \langle \hat{M} \rangle$$

$$= p$$

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

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

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

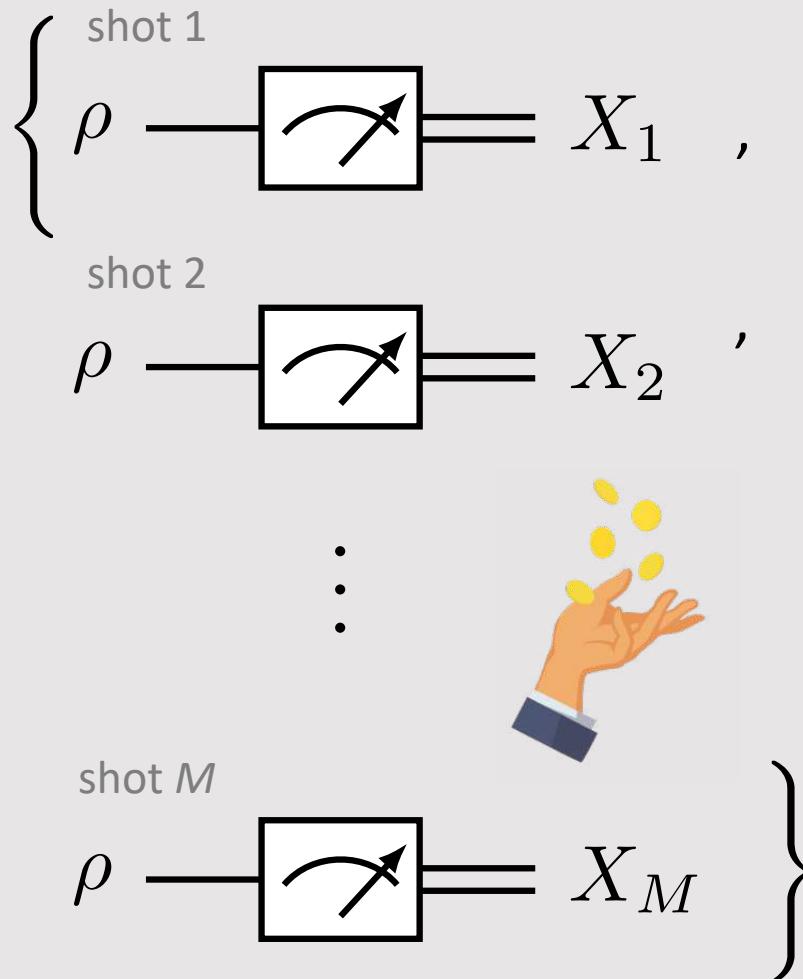
linear functional

expectation value of a single shot

relation to quantum operator we derived earlier (unbiased estimator)

relation to probability to measure 1

How noisy is our estimate of the empirical mean?



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

$$\begin{aligned} \mathbb{V}[S] &= \mathbb{V}\left[\frac{1}{M} \sum_{m=1}^M X_m\right] \\ &= \frac{1}{M^2} \sum_{m=1}^M \mathbb{V}[X] \\ &= \frac{1}{M} \mathbb{V}[X] \\ &= \frac{p(1-p)}{M} \end{aligned}$$

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

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)

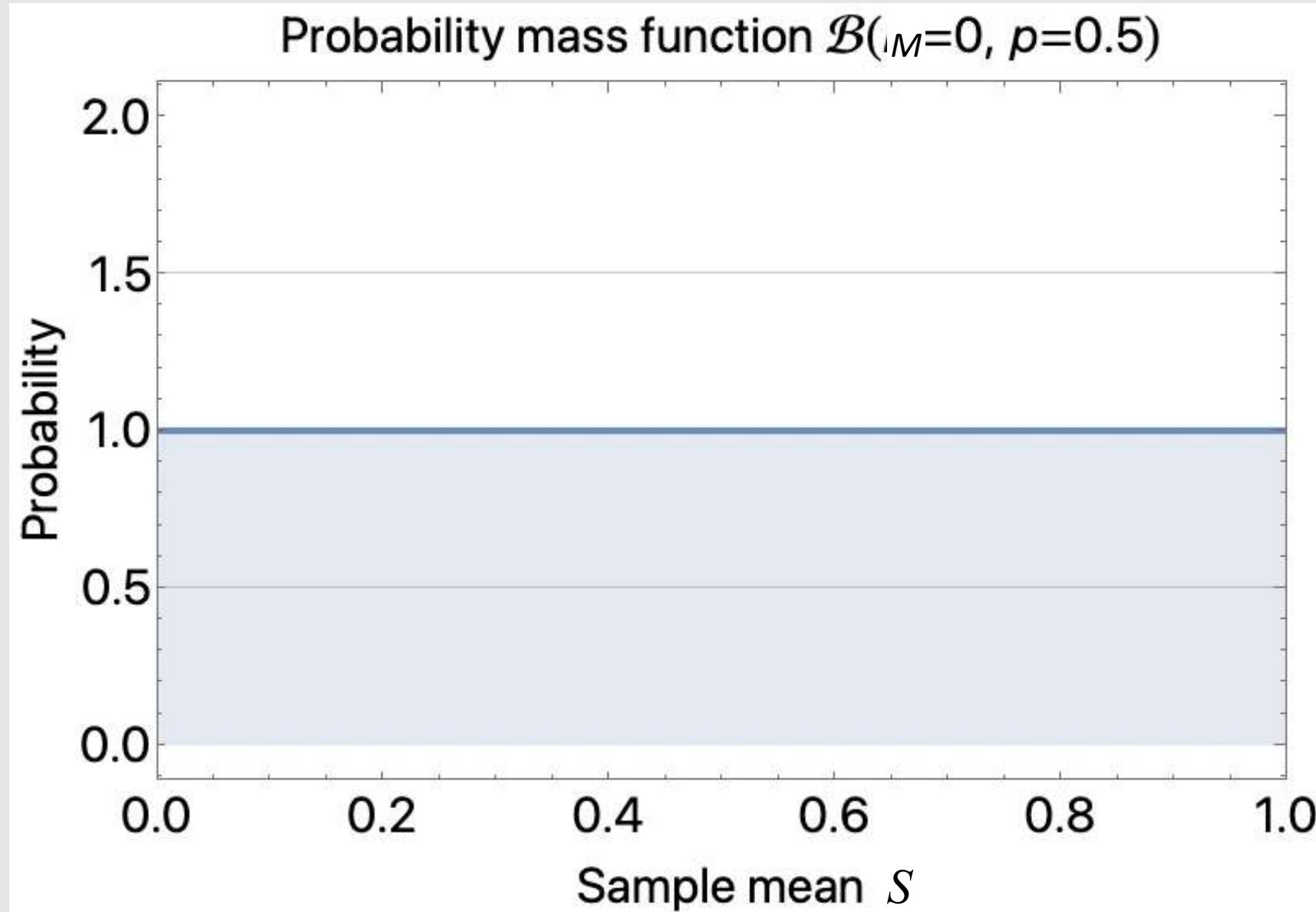
$$\mathbb{V}[X_m] = \mathbb{V}[X] = p(1-p) \quad \forall m \in \{1, \dots, M\}$$

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 drop as one over square root of the number of shots

Animation of convergence of shots expectation value and mean

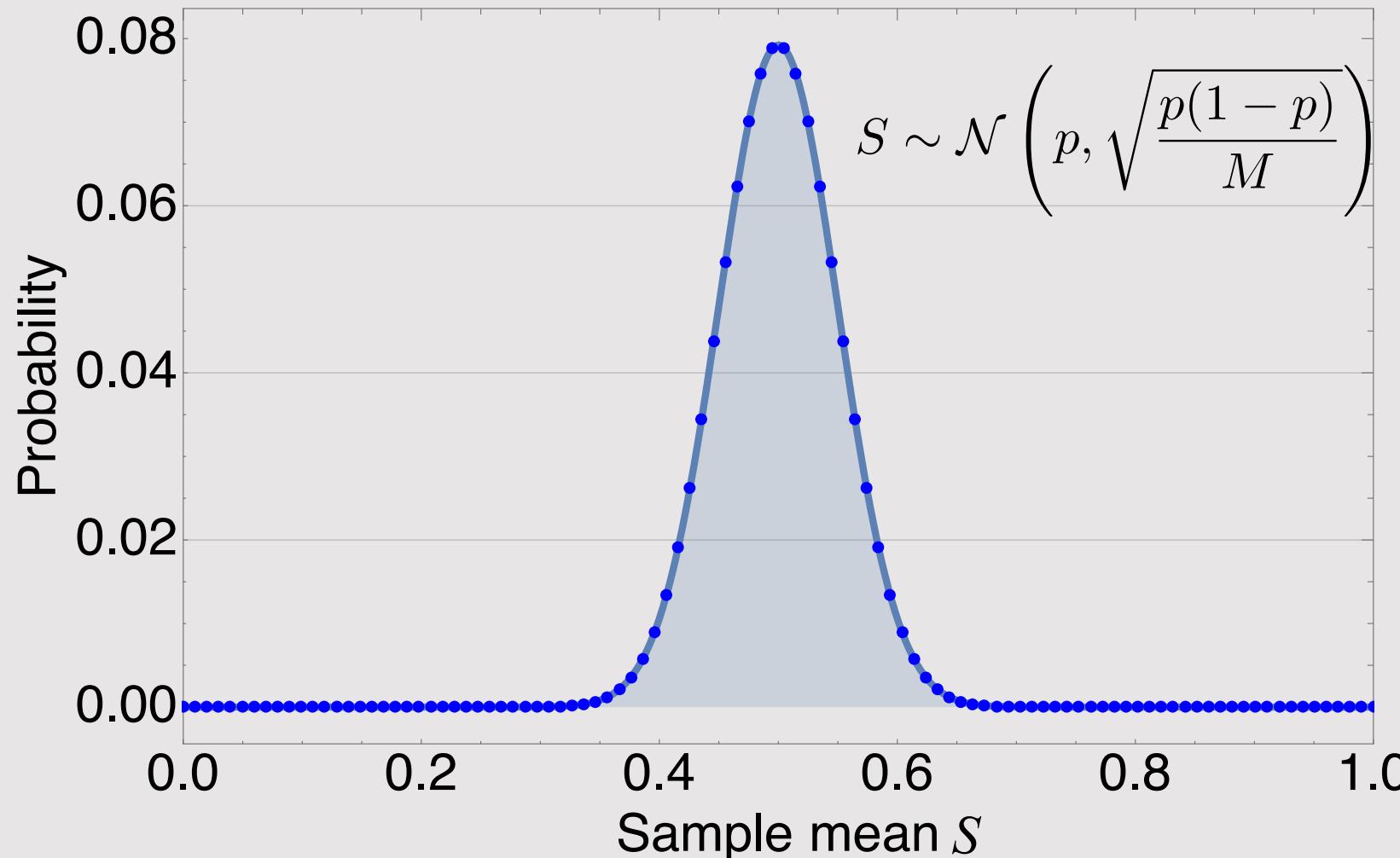


Binomial distribution

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

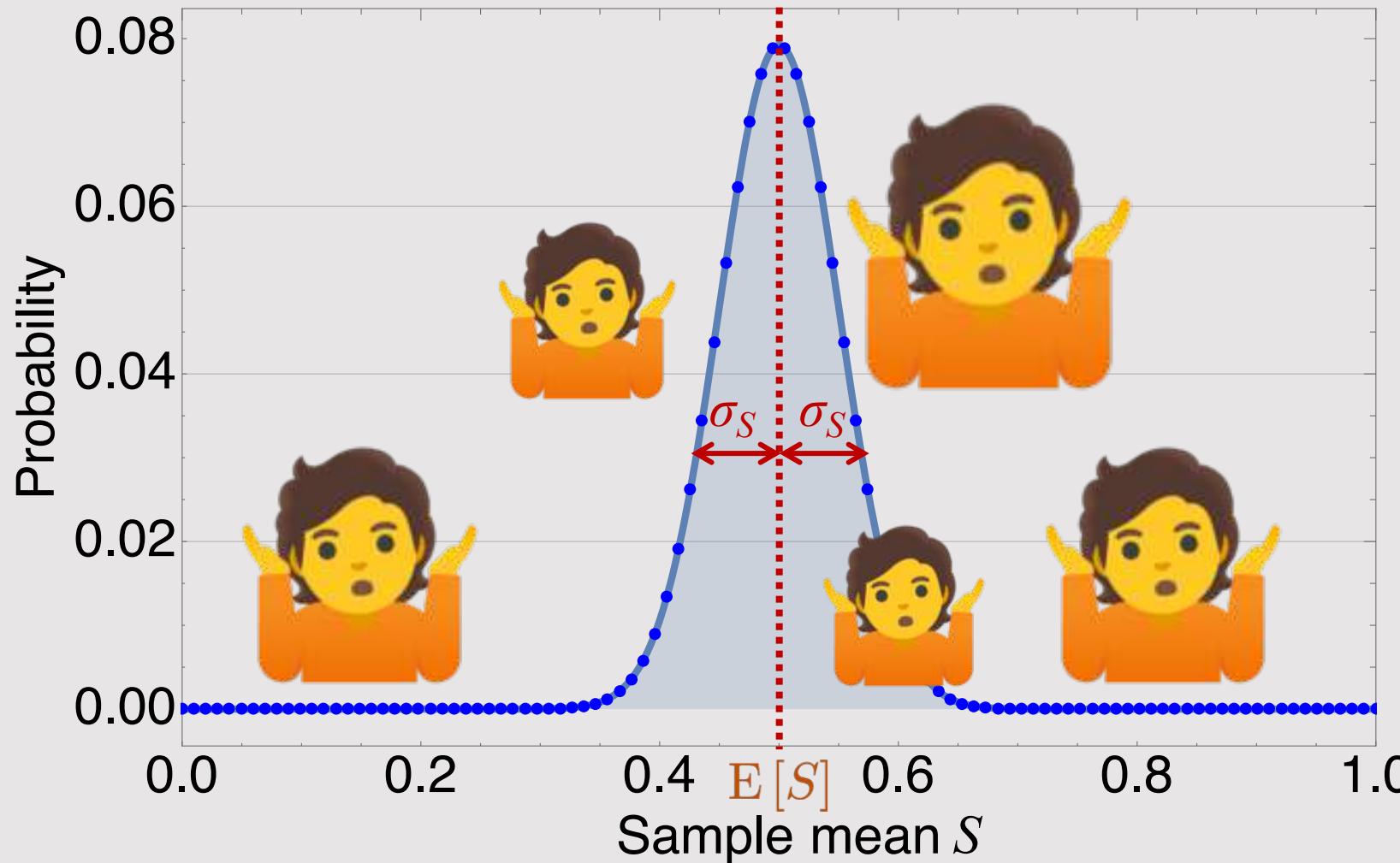
Sampled output distribution

Probability mass function $\mathcal{B}(M = 101, p = 0.5)$



Properties of the output distribution

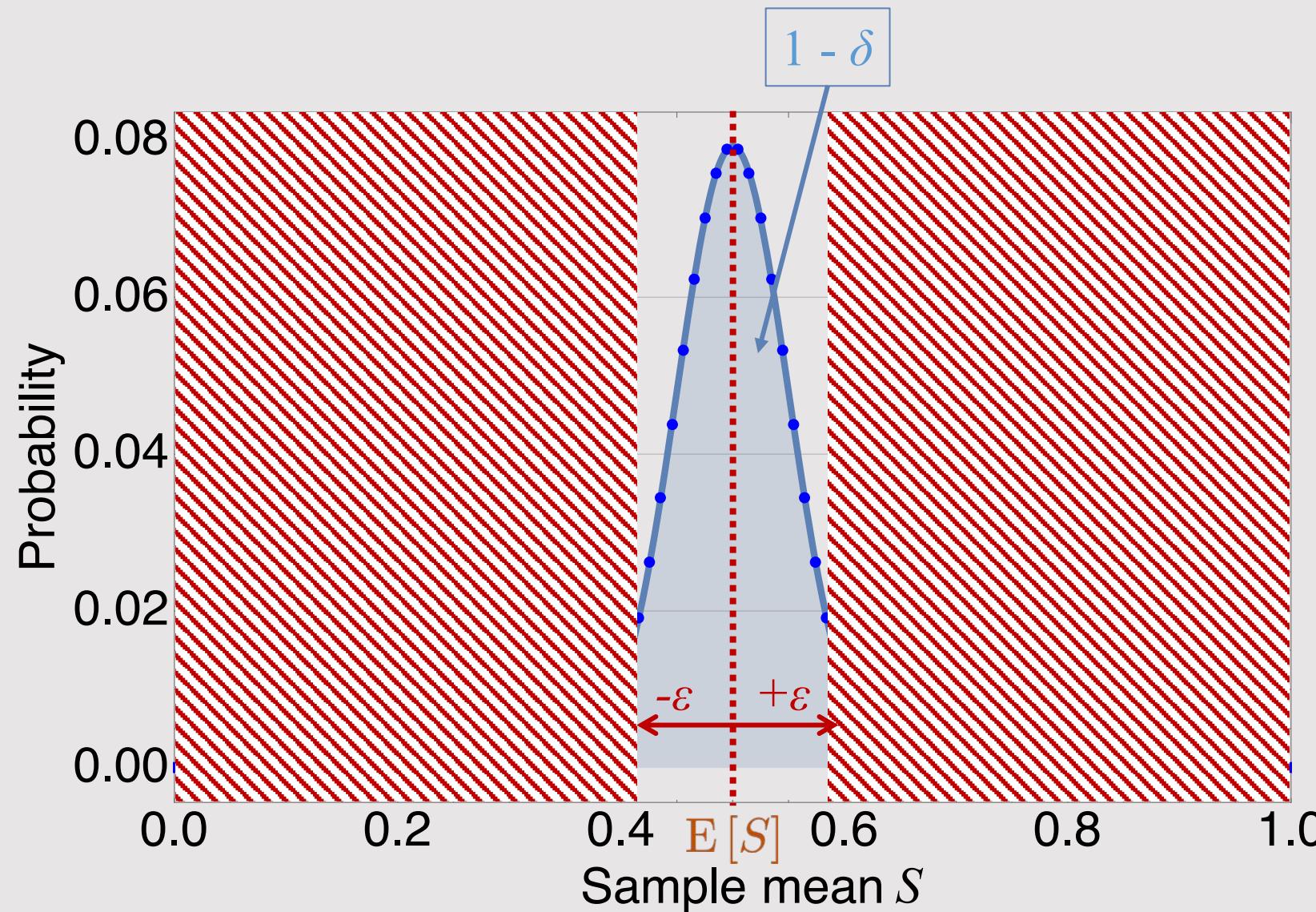
Probability mass function $\mathcal{B}(M = 101, p = 0.5)$



$$E [S] = p$$

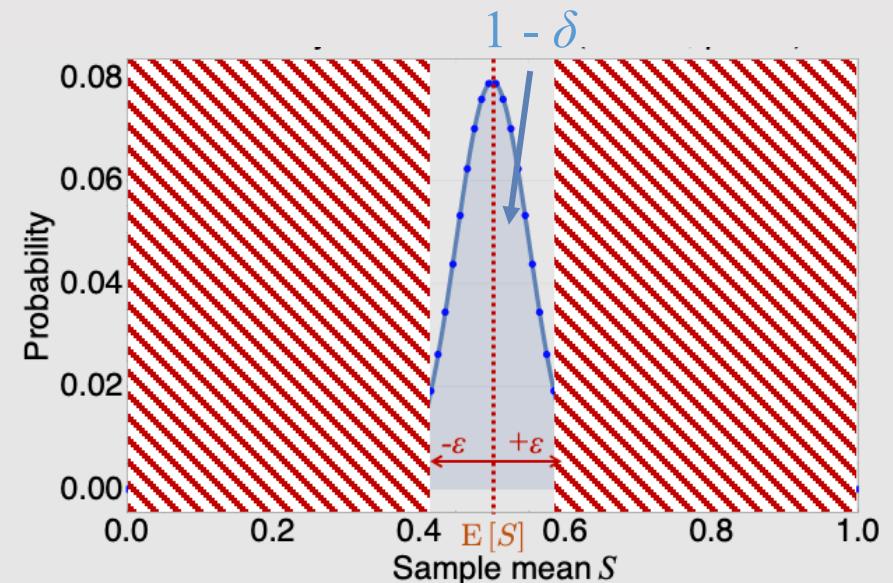
$$\begin{aligned}\sigma_S &= \sqrt{\text{Var}[S]} \\ &= \sqrt{\frac{p(1-p)}{M}}\end{aligned}$$

Concentration Measure for Sampling Expectation Values



Error Bound on Quantum Expectation Values

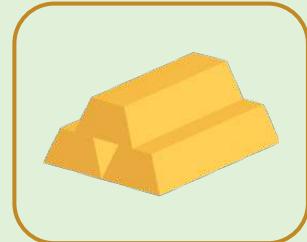
Concentration Measure for Sampling Expectation Values



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 $1 - \delta$.

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

δ specific failure probability for meeting precision ϵ empirically.



Empirical mean & sample properties

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

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

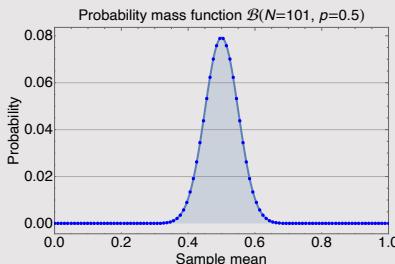
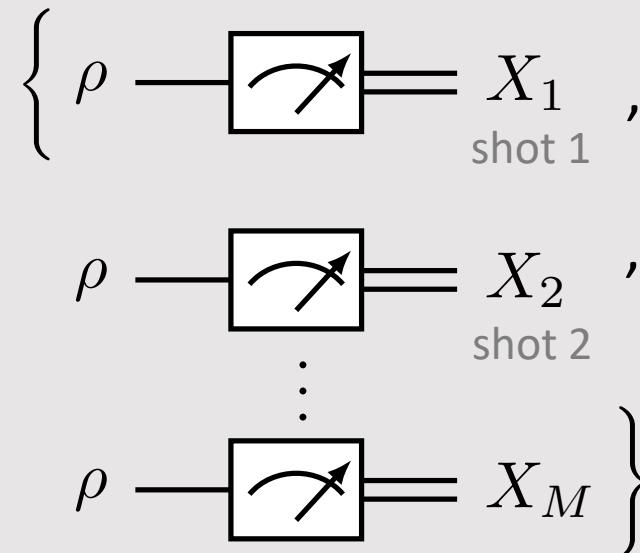
$$M \geq 2\epsilon^{-2} \log(2\delta^{-1}) \quad [M \gtrapprox 4\epsilon^{-2}] .$$

* 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 $\epsilon = \sigma$:

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

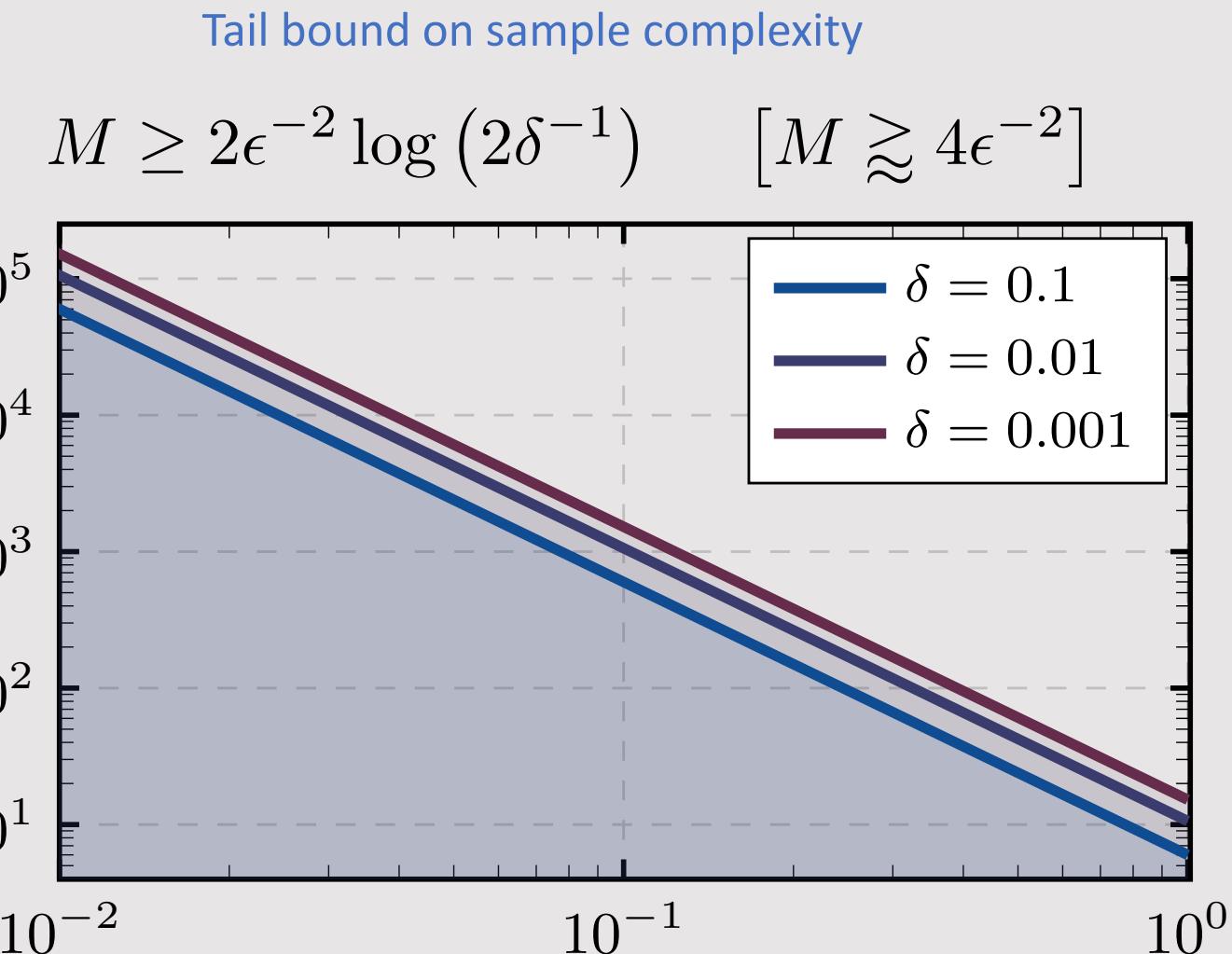
Ideal single qubit measurement with M shots



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.” - Derek

Bound on samples M



Additive precision ϵ

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 $\Pr(\dots \leq \dots)$ and two-sided $\Pr(|\dots| \leq \dots)$. Notation: X is a random variable, $\mu := \mathbb{E}[X]$, $\sigma^2 := \text{Var}[X]$, $S_n := X_1 + \dots + X_n$.

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

¹Markov's inequality bounds the first moment of random variable. Use it when a constant probability bound is sufficient [1, Sec. 3.3].

²Chebyshev 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 $X \in [a, b]$.

³Chernoff 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].

⁴This 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 $\mathbb{E}[X] \pm t$.

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

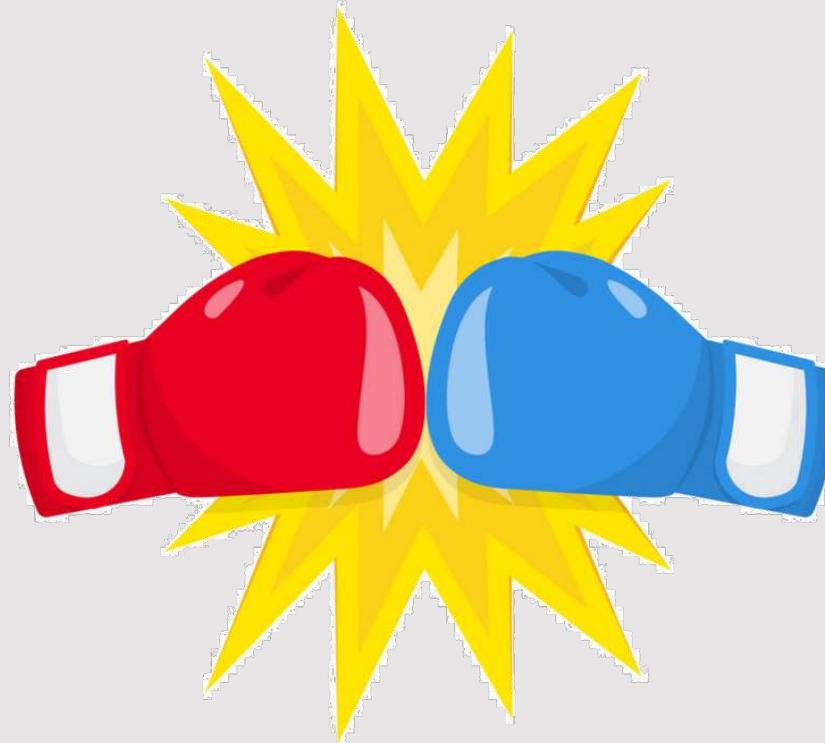


image: makc76



Language of errors and error mitigation

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

@formulas_for_your_comfort

Notation: Super bra ket

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

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

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

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

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

Vectorization isomorphism

$$\begin{aligned} \langle\langle \hat{A} | \hat{B} \rangle\rangle &\leftrightarrow \langle \hat{A}, \hat{B} \rangle \\ &= \text{Tr}(\hat{B} \hat{A}^\dagger) \end{aligned}$$

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

$$\begin{aligned} \langle\langle \hat{A} | \cdot \rangle\rangle &\leftrightarrow \langle \hat{A}, \cdot \rangle \\ &= \text{Tr}(\hat{A}^\dagger \cdot) \end{aligned}$$

$$\begin{aligned} |\hat{A}\rangle\rangle \langle\langle \hat{B} | &\leftrightarrow \hat{A} \langle \hat{B}, \cdot \rangle \\ &\leftrightarrow \hat{A} \text{Tr}(\hat{B}^\dagger \cdot) \end{aligned}$$

Familiar ideas revisited with super notation

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

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

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

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

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

$$a, b \in \Gamma$$

Stochastic Pauli channel

$$\Lambda : L(\mathcal{H}) \rightarrow L(\mathcal{H}) \quad 0 \leq p_a \leq 1 ,$$

$$\Lambda(\rho) = \sum_{a \in \Gamma} p_a P_a \rho P_a , \quad \sum_{a \in \Gamma} p_a = 1 ,$$

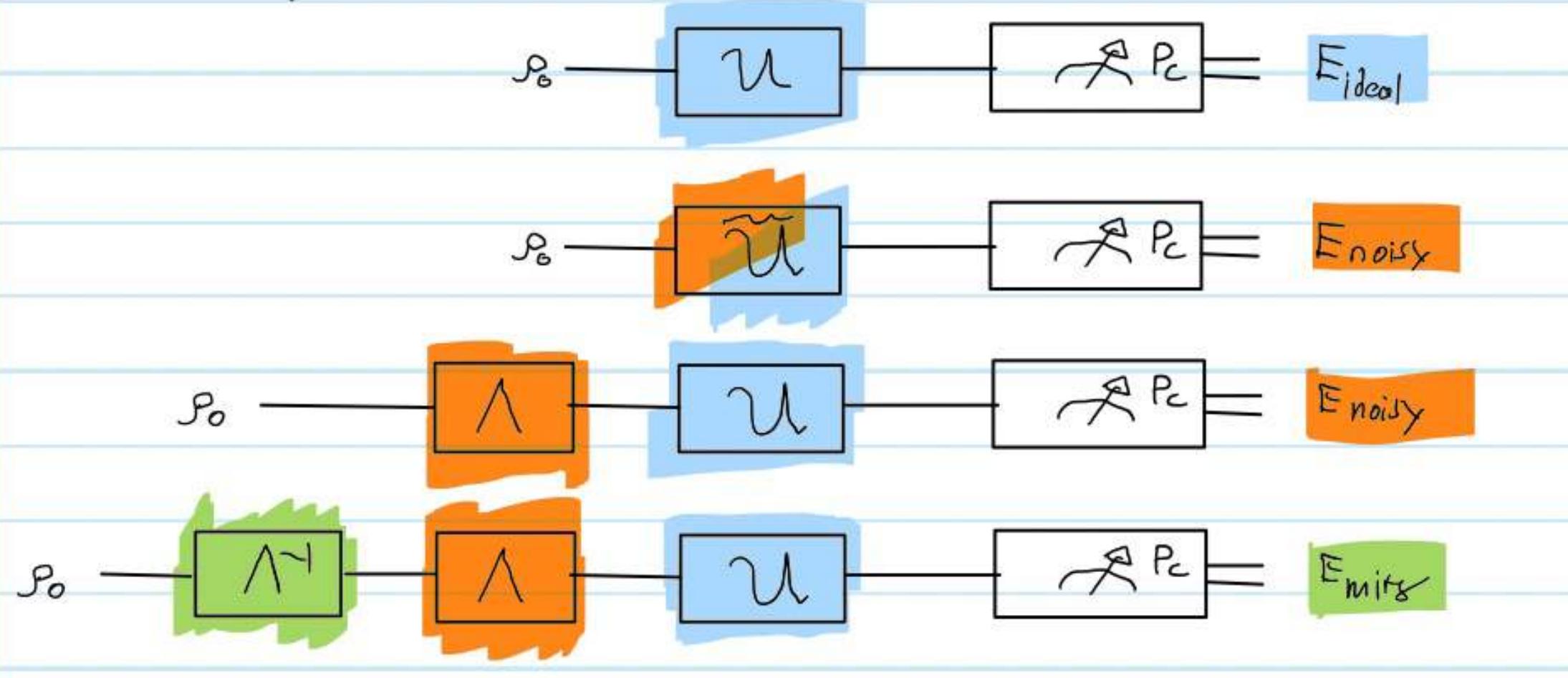
$$\Lambda(\rho) = \frac{1}{\sqrt{|\Gamma|}} \sum_{b \in \Gamma} f_b \operatorname{Tr}(P_b \rho) P_b = \sum_{b \in \Gamma} f_b \frac{|P_b\rangle\langle P_b|}{\langle\langle P_b | P_b \rangle\rangle} ,$$

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

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

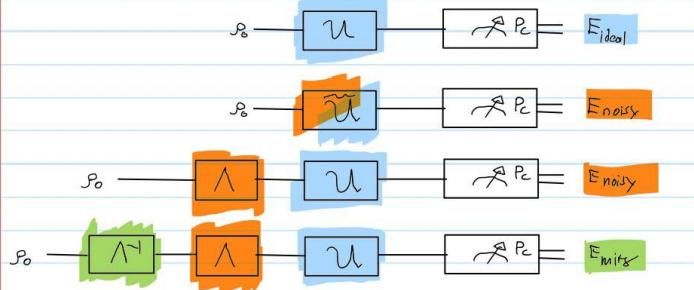
Probabilistic error cancellation: Derivation

From 1 Step



Probabilistic error cancelation: Derivation

From 1 Step



Channel definitions

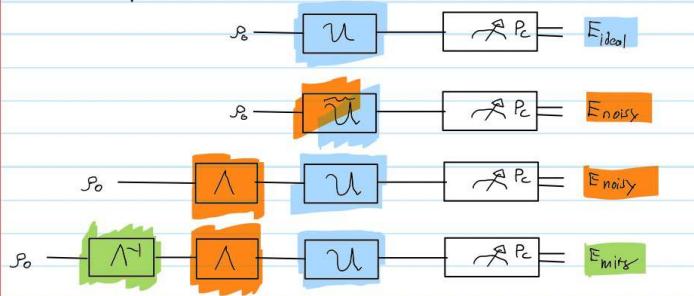
$$U = U \cdot U^\dagger$$

$$\Lambda \approx \sum_a f_a |P_a\rangle \langle P_a|,$$

$$= \sum_b c_b |P_b\rangle \langle P_b|.$$

Probabilistic error cancellation: Derivation

From 1 Step



Channel definitions

$$U = U \cdot U^*$$

$$\langle \hat{P}_c \rangle(\cdot) = \langle P_c \rangle \circ \quad \text{Need}$$

$$\Lambda = \sum_a f_a |P_a\rangle \langle P_a| \quad -1 \leq f_a \leq 1$$

$$= \sum_b c_b |P_b\rangle$$

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

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

$$\vec{c}_b = W \vec{f}_a \quad \vec{f}_a = W c_b$$

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

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

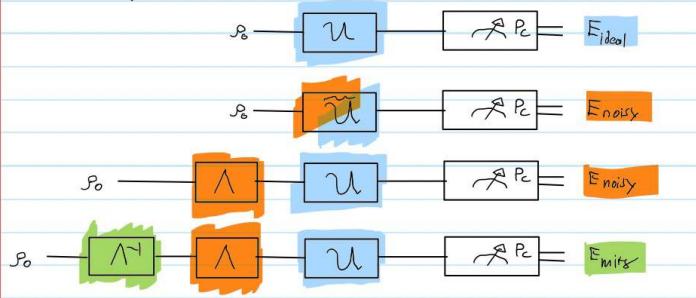
$$= \sum_b c_b^{inv} |P_b\rangle$$

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

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

Probabilistic error cancelation: Derivation

From 1 Step



Channel definitions

$$U = U \cdot U^*$$

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

Need

$$\Lambda \approx \sum_a f_a |P_a\rangle \langle P_a|$$

$$-1 \leq f_a \leq 1$$

$$= \sum_b c_b |P_a\rangle$$

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

$$c_b = \frac{1}{2^n} \sum_a (-1)^{f_a b} f_a$$

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

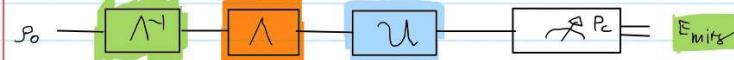
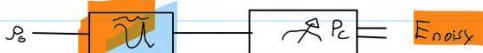
Circuit expectation values

$$E_{\text{ideal}} = \langle \hat{P}_c \rangle_u$$

$$= \langle \langle P_c | U | P_0 \rangle \rangle$$

ideal exp value with noiseless unitary

From 1 Step



Channel definitions

$$U = U \cdot U^\dagger$$

$$\langle \hat{P}_c \rangle(\cdot) = \langle \langle P_c | \cdot | P_0 \rangle \rangle$$

- Need to introduce
 • Super-oper (two-level unitaries)
 • Pauli (classical) $P \in \{I, X, Y, Z\}$
 • Separ-oper (separable)
 • WIF
 • Pauli (classical, separable)
 • SQ (no wif)
 • C (no wif)

$$\Lambda \approx \sum_a f_a |P_a\rangle \langle P_a|$$

$$-1 \leq f_a \leq 1$$

$$= \sum_b c_b |P_b\rangle \langle P_b|$$

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

$$c_b = \frac{1}{2^n} \sum_a (-1)^{f_a b} f_a$$

$$\vec{c}_b = W \vec{f}_a \quad \vec{f}_a = W c_a$$

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

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

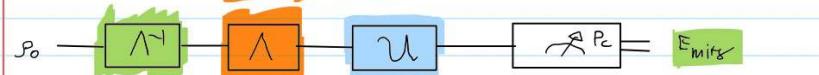
$$= \sum_b c_b^{\text{inv}} |P_b\rangle \langle P_b|$$

$$c_b^{\text{inv}} = \frac{1}{2^n} \sum_a (-1)^{c_a b} f_a$$

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

Circuit expectation values

From 1 Step



Channel definitions

$$U = U \cdot U^\dagger$$

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

- Need to introduce
 - Super-operators
 - Pauli channels
 - Separation of variables
 - WIF
 - Pauli closure requirement
 - SQ no wif
 - C=0,0.5p

$$\Lambda \approx \sum_a f_a |P_a\rangle \langle P_a|$$

$$-1 \leq f_a \leq 1$$

$$= \sum_b c_b |P_b\rangle \langle P_b|$$

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

$$c_b = \frac{1}{2^n} \sum_a (-1)^{f_a b} f_a$$

$$\vec{c}_b = W \vec{f}_a \quad \vec{f}_a = W c_b$$

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

$$\vec{c}_b^{\dagger w} = W \vec{f}_a^\dagger$$

$$c_b^{\dagger w} = \frac{1}{2^n} \sum_a (-1)^{c_a b} f_a$$

$$c_b^{\dagger w} \in \mathbb{R}$$

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

$$= \langle \langle P_c | U | P_0 \rangle \rangle$$

ideal exp value with noiseless unitary

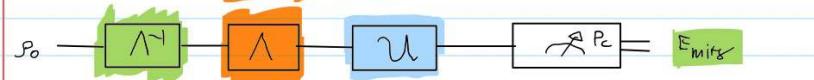
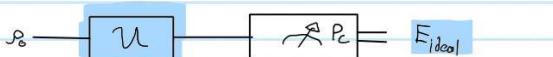
$$E_{\text{noisy}} := \langle \hat{P}_c \rangle_{\tilde{U}}$$

$$= \langle \langle P_c | U \Lambda | P_0 \rangle \rangle$$

noisy-gate expectation value

Circuit expectation values

From 1 Step



Channel definitions

$$U = U \cdot U^\dagger$$

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

- Need to introduce
 - Super-operator basis-unitaries
 - Pauli channels
 - Separate operators
 - WIF
 - Pauli channel representations
 - SQ
 - $\frac{1}{2} \otimes \frac{1}{2}$
 - $\mathbb{C}[0,1]$

$$\Lambda \approx \sum_a f_a |P_a\rangle \langle P_a|$$

$$-1 \leq f_a \leq 1$$

$$= \sum_b c_b |P_b\rangle \langle P_b|$$

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

$$c_b = \frac{1}{2^n} \sum_a (-1)^{f_a b} f_a$$

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

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

$$= \sum_b c_b^{\text{inv}} |P_b\rangle \langle P_b|$$

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

$$c_b^{\text{inv}} = \frac{1}{2^n} \sum_a (-1)^{c_a b} f_a$$

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

$$E_{\text{ideal}} := \langle \hat{P}_c \rangle_U$$

$$= \langle \langle P_c | U | P_0 \rangle \rangle$$

ideal exp value with noiseless unitary

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

$$= \langle \langle P_c | U \Lambda | P_0 \rangle \rangle$$

$$= \langle \langle P_c | \tilde{U} | P_0 \rangle \rangle$$

noisy-gate expectation value

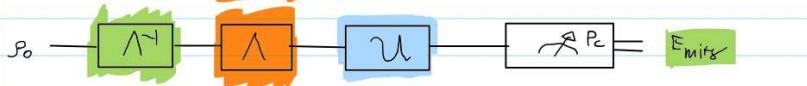
$$E_{\text{mixed}} := \langle \langle P_c | U \Lambda \Lambda^\dagger | P_0 \rangle \rangle$$

$$= \langle \langle P_c | U \Lambda \left(\sum_a f_a^{-1} |P_a\rangle \langle P_a| \right) | P_0 \rangle \rangle$$

$$= \langle \langle P_c | U \Lambda \left(\sum_b c_b^{\text{inv}} P_b \right) | P_0 \rangle \rangle$$

Circuit expectation values

From 1 Step



Channel definitions

$$U = U \cdot U^\dagger$$

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

- Need to introduce
 - Super-operators
 - Separable bipartite variables
 - Pauli channels
 - Separable variables
 - CNOT
 - Pauli channels
 - Separable
 - SQM
 - CNOT

$$\Lambda \approx \sum_a f_a |P_a\rangle \langle P_a|$$

$$-1 \leq f_a \leq 1$$

$$= \sum_b c_b |P_b\rangle \langle P_b|$$

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

$$C_b = \frac{1}{2^n} \sum_a (-1)^{f_a b} f_a$$

$$\vec{C}_b = W^{\dagger} \vec{f}_a \quad \vec{f}_a = W C_b$$

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

$$\vec{C}_b^{\text{inv}} = W \vec{f}_a^{\dagger}$$

$$= \sum_b c_b^{\text{inv}} |P_b\rangle \langle P_b|$$

$$c_b^{\text{inv}} = \frac{1}{2^n} \sum_a (-1)^{f_a b} f_a$$

$$C_b^{\text{inv}} \in \mathbb{R}$$

$$E_{\text{ideal}} := \langle \hat{P}_c \rangle_{\text{u}}$$

$$= \langle \langle P_c | U | P_0 \rangle \rangle$$

ideal exp value with noiseless unitary

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

$$= \langle \langle P_c | U \Lambda | P_0 \rangle \rangle$$

$$= \langle \langle P_c | \tilde{U} | P_0 \rangle \rangle$$

noisy-gate expectation value

$$E_{\text{mixed}} := \langle \langle P_c | U \Lambda \Lambda^\dagger | P_0 \rangle \rangle$$

$$= \langle \langle P_c | U \Lambda \left(\sum_a f_a^{-1} |P_a\rangle \langle P_a| \right) | P_0 \rangle \rangle$$

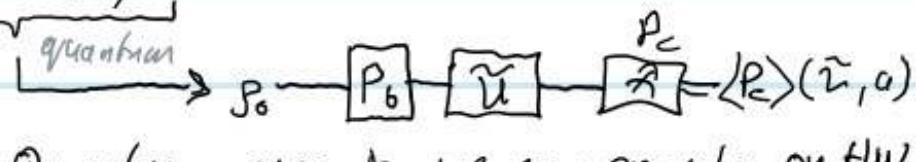
$$= \langle \langle P_c | U \Lambda \left(\sum_b c_b^{\text{inv}} P_b \right) | P_0 \rangle \rangle$$

$$= \sum_b c_b^{\text{inv}} \langle \langle P_c | U \Lambda P_b | P_0 \rangle \rangle$$

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

$$= \sum_b c_b^{\text{inv}} \underbrace{\langle P_c \rangle(\tilde{U}, b)}$$

Classical Post process



$$E_{\text{mitg}} := \langle\langle P_c | U_1 \Lambda^\dagger | \rho_0 \rangle\rangle$$

$$= \langle\langle P_c | U_1 \left(\sum_a f_a^{-1} |P_a\rangle \langle P_a| \right) | \rho_0 \rangle\rangle$$

$$= \langle\langle P_c | U_1 \left(\sum_b c_b^{\text{inv}} P_b \right) | \rho_0 \rangle\rangle$$

$$= \sum_b c_b^{\text{inv}} \langle\langle P_c | U_1 P_b | \rho_0 \rangle\rangle$$

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

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



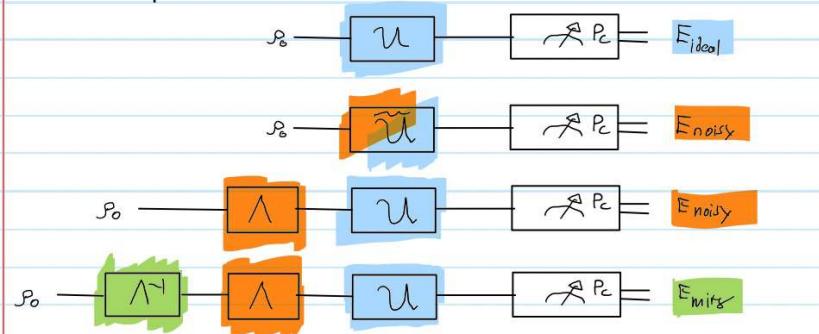
Quantum circuit we can execute on H/W
and find exp. value from.

∴ 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:

From 1 Step



Channel definitions

$$U = U \cdot U^\dagger$$

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

- Need to introduce
 - Super-oper (two-level unitaries)
 - Pauli channels
 - Separ.-op. unitaries
 - WIF
 - Pauli channel representations
 - SQ
 - t=0
 - C=0.03p

$$\Lambda \approx \sum_a f_a |P_a\rangle \langle P_a|$$

$$-1 \leq f_a \leq 1$$

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

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

$$\vec{c}_b = W \vec{f}_a \quad \vec{f}_a = W c_b$$

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

$$\vec{c}_b^{\text{inv}} = W \vec{f}_a^\dagger$$

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

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

Quasi probability distribution

C_6^{inv} can be outside $[0, 1]$

$\sum_i C_6^{\text{inv}} = \gamma \geq 1$ generally for Λ not unitary

e.g. 6A + 1Y chapter

$$\begin{aligned}\Lambda &= (1-p)[\cdot I + pX \cdot X] \\ \Lambda^{-1} &= ((1-p)^2 + p^2)I + \frac{-p}{1-p}X \cdot X\end{aligned}$$

$$C_{\Sigma}^{\text{inv}} = 1 + \frac{p}{1-p}$$

$b = (0, 0)$
Chao vector

$$C_X^{\text{inv}} = -\frac{p}{1-p}$$

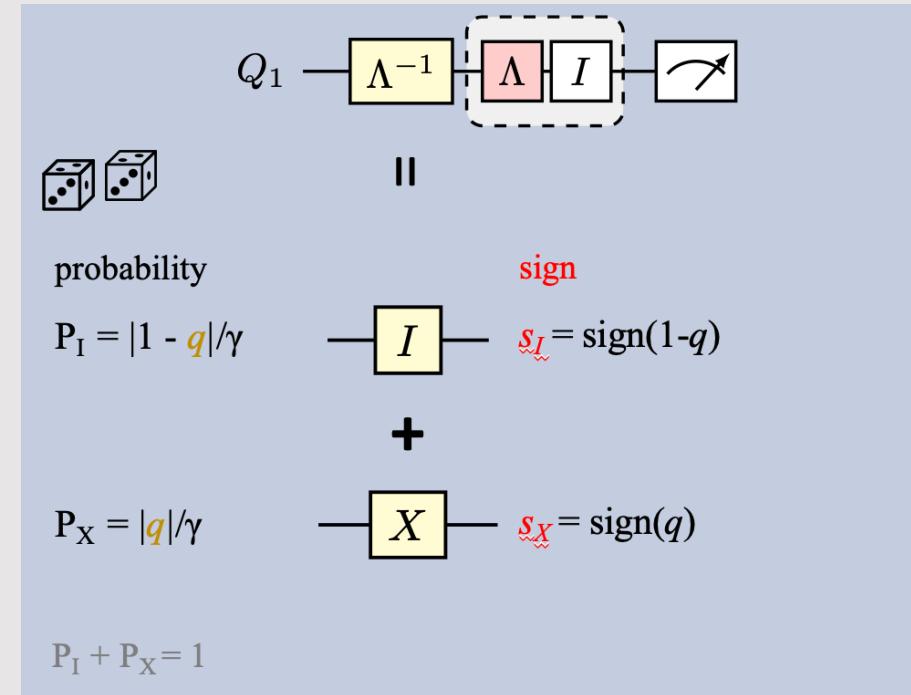
$b = (1, 0)$

Turn into probabilities

$$C_G^{\text{inv}} = \text{sgn}(C_6^{\text{inv}}) \frac{|C_6^{\text{inv}}|}{\gamma} \gamma$$

Sign $\in \{-1, +1\}$ $|C_6^{\text{inv}}|$ γ
 Prob $\in [0, 1]$ Scale

$$\begin{aligned}\tilde{C}_6^{\text{inv}} &:= \frac{|C_6^{\text{inv}}|}{\|C_6^{\text{inv}}\|_1} \xrightarrow{\text{defn of } \|\cdot\|_1} \|C_6\|_1 = \sum_6 |C_6^{\text{inv}}| \leftrightarrow L_1 \text{ norm} \\ &= \frac{|C_6^{\text{inv}}|}{\gamma}\end{aligned}$$



Emitigated

$$E_{\text{mitig}} = \sum_b C_b^{\text{inv}} \langle \hat{P}_c \rangle (\tilde{u}, b)$$

$$= \sum_b \underbrace{\text{sgn}(C_b^{\text{inv}})}_{\gamma} \frac{|C_b^{\text{inv}}|}{\gamma} \langle \hat{P}_c \rangle (\tilde{u}, b)$$

$$= \gamma \sum_b \underbrace{\text{sgn}(C_b^{\text{inv}})}_{\text{scale}} \underbrace{\overline{C}_b^{\text{inv}}}_{\text{classical part}} \underbrace{\langle \hat{P}_c \rangle (\tilde{u}, b)}_{\text{Value QC circuit}}$$

scale classical part -
processing. Value QC circuit
can run & find value on HW

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.

C_b^{inv} come outside $\{\mathbb{0}, \mathbb{1}\}$

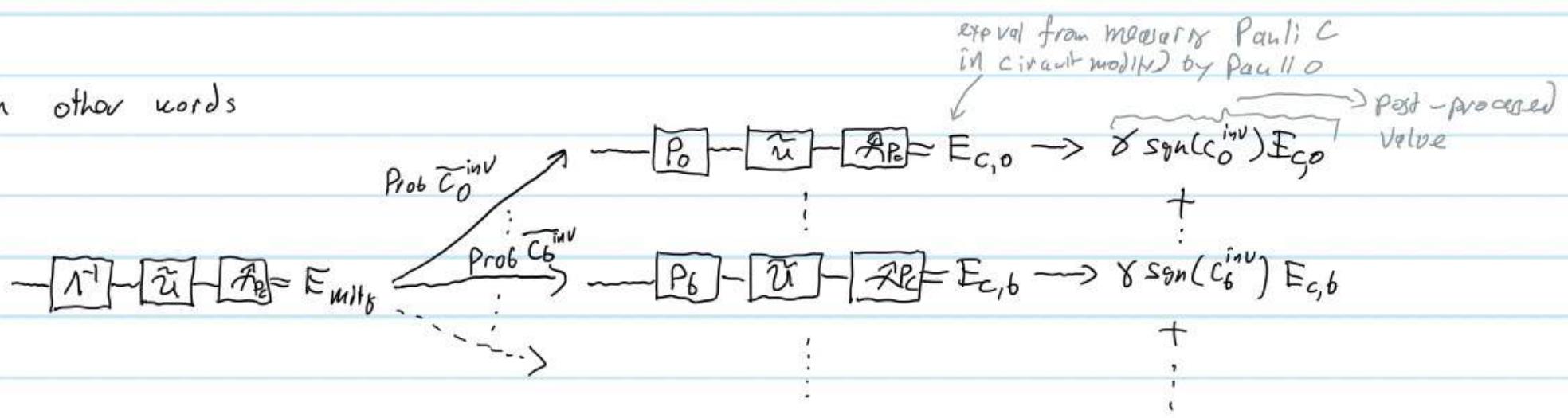
$$\begin{aligned} \notag C_b^{\text{inv}} &\approx \gamma \geq 1 \quad \text{generally for } \Lambda \text{ not unitary} \\ \notag \text{eg bit flip channel} \quad \Lambda &= (1-p)I + pX \cdot X \\ \notag \Lambda' &= (1-p)I + \frac{-p}{1-p} X \cdot X \\ \notag C_{\Sigma}^{\text{inv}} &= 1 + \frac{p}{1-p} \quad C_X^{\text{inv}} = -\frac{p}{1-p} \\ \notag \stackrel{\text{p}}{\stackrel{\text{b}}{\stackrel{\text{b}=(0,0)}{\text{b}=(1,0)}}} \text{Chase vector} & \stackrel{\text{p}}{\stackrel{\text{b}}{\stackrel{\text{b}=(1,0)}{\text{b}=(1,0)}}} \end{aligned}$$

Turn into probabilities

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

$$\begin{aligned} \overline{C}_b^{\text{inv}} &:= \frac{|C_b^{\text{inv}}|}{\|C_b^{\text{inv}}\|_1} \rightarrow \|C_b^{\text{inv}}\|_1 = \sum_b |C_b^{\text{inv}}| \rightarrow L_1 \text{ norm} \\ &= \frac{|C_b^{\text{inv}}|}{\gamma} \end{aligned}$$

In other words



Quantum Probability Distribution

c_b^{inv} can be outside $[0, 1]$ generally for N not unitary

$\frac{d}{dt} c_b^{\text{inv}} = \gamma \approx 1$ for $t \in [0, 1]$

eg bit flip error $\Lambda = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

$\Lambda^t = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

$c_b^{\text{inv}} = \frac{1-p}{1-p} = \frac{p}{1-p}$

Turn into probability

$C_b^{\text{inv}} = \frac{\text{sgn}(c_b^{\text{inv}})}{|c_b^{\text{inv}}|} \gamma$

$\tilde{c}_b^{\text{inv}} = \frac{|c_b^{\text{inv}}| \text{sgn}(c_b^{\text{inv}})}{|c_b^{\text{inv}}|} = |c_b^{\text{inv}}| = \frac{q}{1-q}$

$E_{\text{mitg}} = \sum_b \tilde{c}_b^{\text{inv}} \langle \tilde{b} \rangle (\tilde{u}, b)$

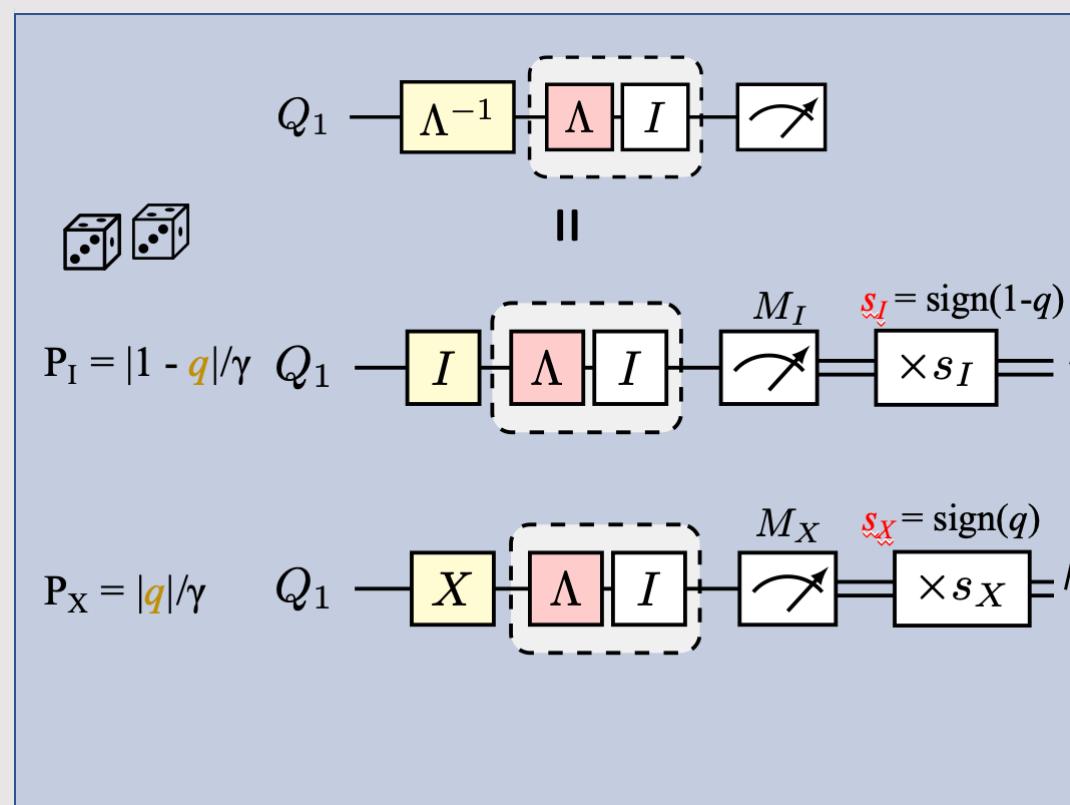
$$E_{\text{mitg}} = \sum_b \frac{\text{sgn}(c_b^{\text{inv}})}{|c_b^{\text{inv}}|} \gamma \langle \tilde{b} \rangle (\tilde{u}, b)$$

$$= \gamma \sum_b \text{sgn}(c_b^{\text{inv}}) \tilde{c}_b^{\text{inv}} \langle \tilde{b} \rangle (\tilde{u}, b)$$

$$\text{scale classical part} = \sqrt{1-q} \text{ D.C. error}$$

post-processing

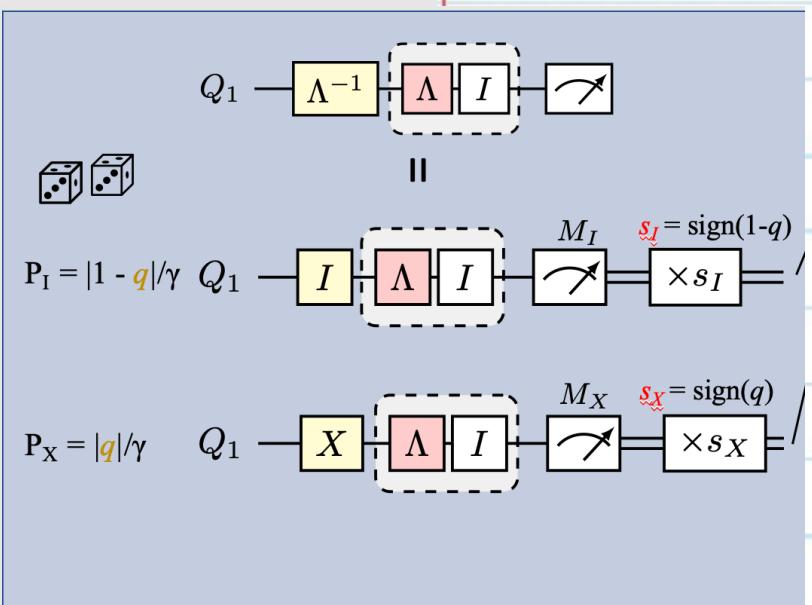
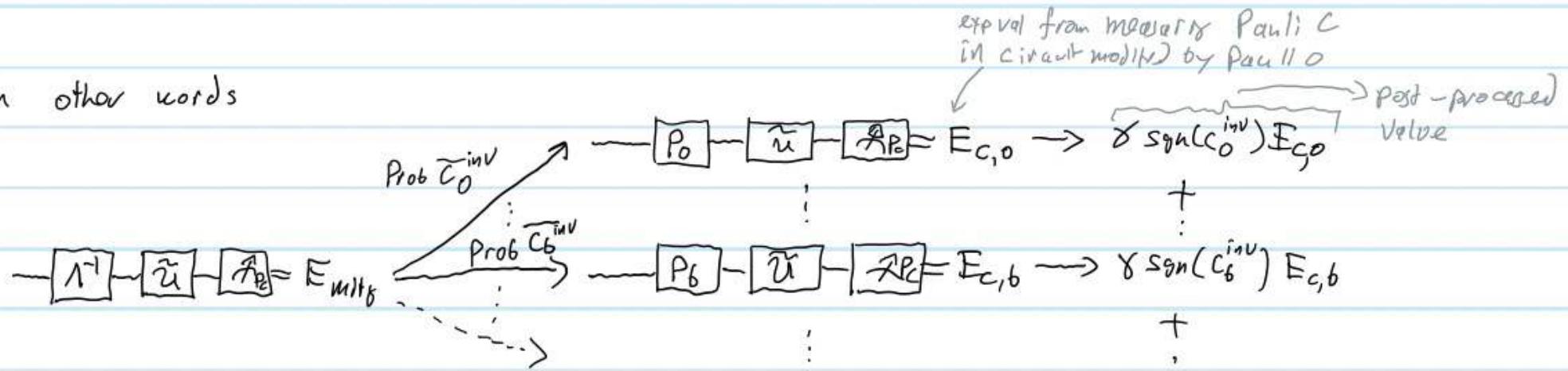
In this form, the decomposition of the error-mitigated expectation value is simply a sum over expectation values of Pauli-gate modified circuits, whose value can be obtained from direct quantum computer execution, weighted by a probability, c_b bar inverse, and the sign. The elements that perform the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



Estimator

post processing.

In other words



$$E_{c,\text{mitg}} = \sum_b \gamma \text{sgn}(C_b^{\text{inv}}) E_{c,b}$$

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

$$\left\{ \underbrace{\tilde{C}_b^{inv}}_{\text{prob}} : -\underbrace{[P_b \rightarrow \cancel{Y} \rightarrow \cancel{RP_b} = Y \in \{-1, +1\}]}_{\text{Pauli observable}} \rightarrow X = \text{sgn}(C_b^{inv}) Y \right\}$$

single-shot outcome
ie r-vals

Let's say we sample M instances, randomly sample assign value for b and obtain one-shot values on QC, ie one random instance of $Y=1$ or $Y=-1$, which we can pass-probs.

The results are thus for classcar random variables

$$\{x_1, x_2, \dots, x_m\} \quad \text{or} \quad \{x_m : m=1, \dots, M\}$$

Probabilistic error cancelation: Derivation

where each $x_m \in \{1, -1\}$ and is distributed to modulator.

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:

$$E_M := \gamma \frac{1}{M} \sum_{m=1}^M x_m = \frac{1}{M} \sum_{m=1}^M \gamma \text{sgn}(c_{b_m}^{\text{inv}}) \underbrace{Y_{b_m}}_{\substack{\text{rand outcome of un-taken for } b_m \text{ circ} \\ \text{only current}}} (\underbrace{U_1}_{\substack{\text{Pauli chosen for } m\text{-th shot}}}, b_m)$$

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 \rightarrow

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

Unbiased estimator

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

$$= \mathbb{E}[\delta X_m] \quad \text{no } X_m \text{ is different}$$

$$= \mathbb{E}\left[\delta \operatorname{sgn}(c_{b_m}^{\text{inv}}) Y_{b_m}(\tilde{u}, b_m)\right] \quad \text{where rand var is } b_m \text{ now, not just } m, \text{ so}$$

$$\underset{b \in \mathcal{Y}}{\approx} \mathbb{E}\left[\delta \operatorname{sgn}(c_b^{\text{inv}}) Y_b(\tilde{u}, b)\right] \quad \text{Prob}[b] = \bar{C}_b^{\text{inv}}$$

drop notation

$$= \sum_b \mathbb{E}\left[\delta \operatorname{sgn}(c_b^{\text{inv}}) Y_b\right] \text{Prob}[b]$$

$$= \sum_b \underbrace{\delta \operatorname{sgn}(c_b^{\text{inv}})}_{\text{post-process of outcome}} \underbrace{\bar{C}_b^{\text{inv}}}_{\text{sample prob}} \underbrace{\mathbb{E}[Y_b]}_{\text{rand outcome}} \longrightarrow$$

$$= \sum_b \delta \operatorname{sgn}(c_b^{\text{inv}}) \frac{c_b^{\text{inv}}}{\delta} \langle\langle p_c | u \cap p_b | p_0 \rangle\rangle$$

note: $\langle\langle p_c | (\tilde{u}, b) \rangle\rangle = \langle\langle p_c | u \cap p_b | p_0 \rangle\rangle$

$$= \mathbb{E}[Y_b]$$

\mathbb{P} 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 :

$$\mathbb{E}[f(b) Y_b] = f(b) \langle\langle p_c | (\tilde{u}, b) \rangle\rangle$$

Probabilistic error cancelation: Derivation

$$= \langle\langle p_c | u \wedge \left(\sum_b c_b^{\text{inv}} p_b \right) | p \rangle\rangle$$

the value y_b but only on the label b :

$$\mathbb{E}[f(b) | y_b] = f(b) \langle \hat{p}_c | \tilde{u}_{,b} \rangle$$

$$= \langle\langle p_c | u | \lambda^{-1} | p \rangle\rangle$$

$$= \langle\langle p_c | u | p \rangle\rangle$$

$$= \langle \hat{p}_c \rangle (u_{\text{ideal}})$$

without noise

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



(Optional step) Variance

Variance of E_M

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

x_m iid
can drop subscript m
and emphasize b and value X

$$= \frac{\sigma^2}{M} \mathbb{V}[X(\tilde{u}, b)]$$

note the sum variance is just
rescaled by σ^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

Prob. error cancelation (PEC)

Circuit cutting (knitting) of gates

Circuit cutting of wires

Classical sim. algorithms (QP)

Channel \mathcal{C}

noise inverse

non-local gate

large unitary

unitary

Quantum Error Mitigation for Non-Equilibrium Quantum Dynamics

Lecture 1

Big picture

Why quantum computers?

Status and outlook

Why error mitigation?

Noise in quantum computers
Overview of error mitigation

Mitigation fundamentals

Probabilistic error cancellation (PEC)

Introduction

One qubit example

General derivation

Next lectures

Learning noise

State-of-art PEC experiments

Key techniques: Twirling

T-REX mitigation

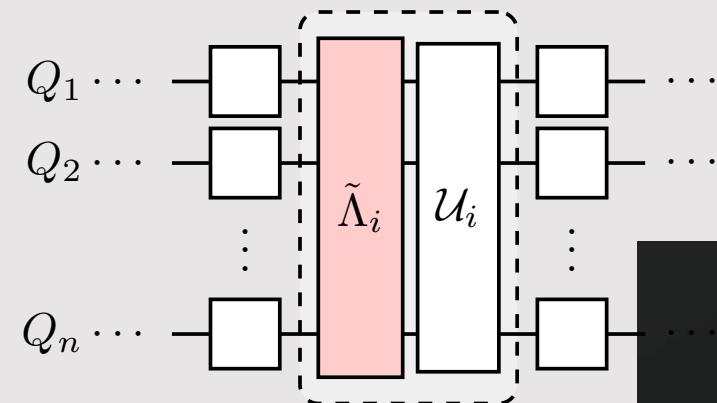
...

State-of-art experiments at the 100Q+, depth 50+: uncovering local integrals of motion

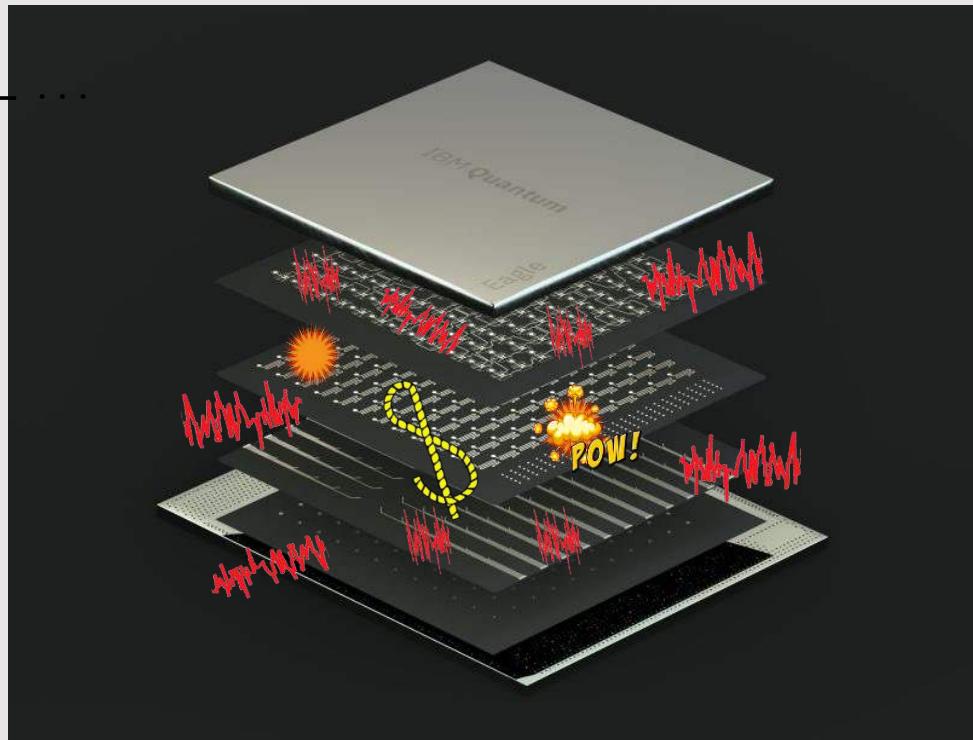
...



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
...

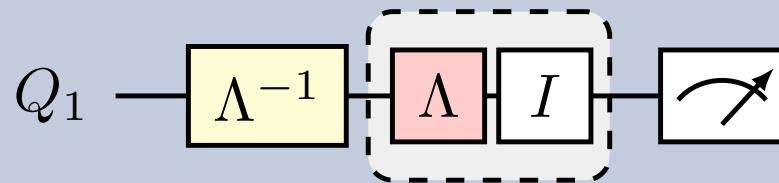


Control errors
Photon shot noise
1/f charge noise
1/f flux noise
Nonequilibrium quasiparticles
Leakage
Cosmic rays
...

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

Practical challenges

Small scale

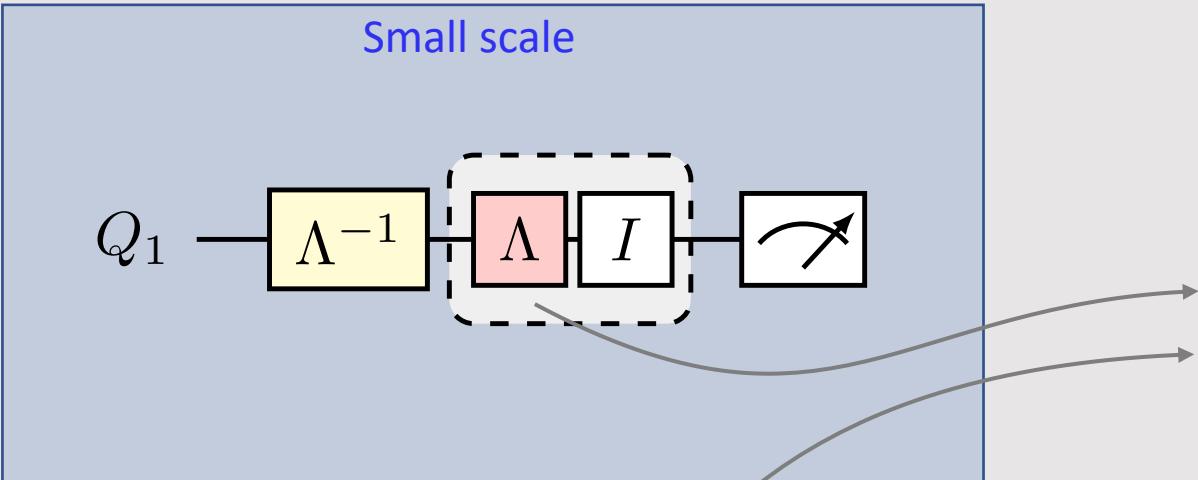


Critically hinges on knowing the full noise near perfectly

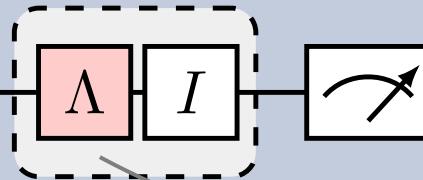
Despite the method's theoretical appeal (1-10), practical challenges have limited its demonstration to the one and two-qubit level (2, 3)

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

 Q_1

Small scale

 Λ^{-1}  Λ I 

2 qubits

10 qubits

50 qubits

noise param values $10^{-2} - 10^{-5}$ additive error sampling cost ($>10^2 - 10^{10}$)

255 parameters

 10^{12} parameters 10^{60} parameters

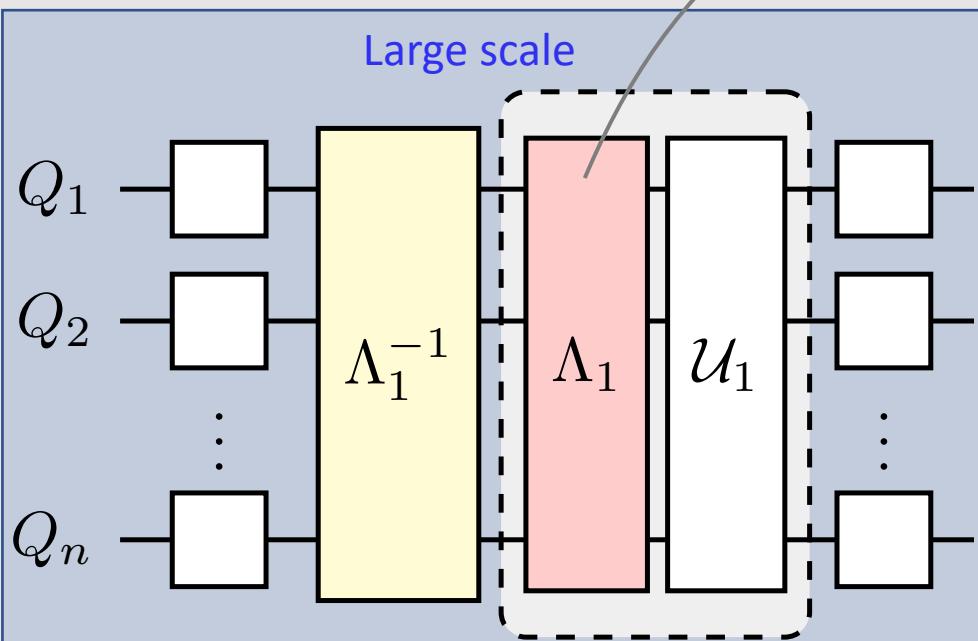
Challenges

learning complexity

- efficient
- scalable
- accurate
- compact, tractable representation

noise in full device

- cross-talk
- correlated errors
- parallel gates

 Q_1 Q_2 \vdots Q_n

Large scale

 Λ_1^{-1} Λ_1 \mathcal{U}_1

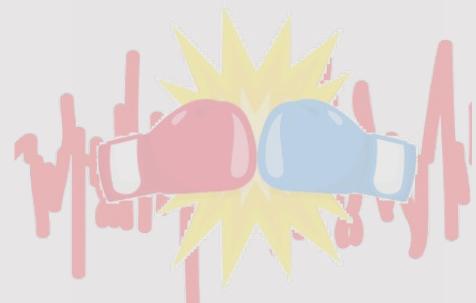
Outline



Idea

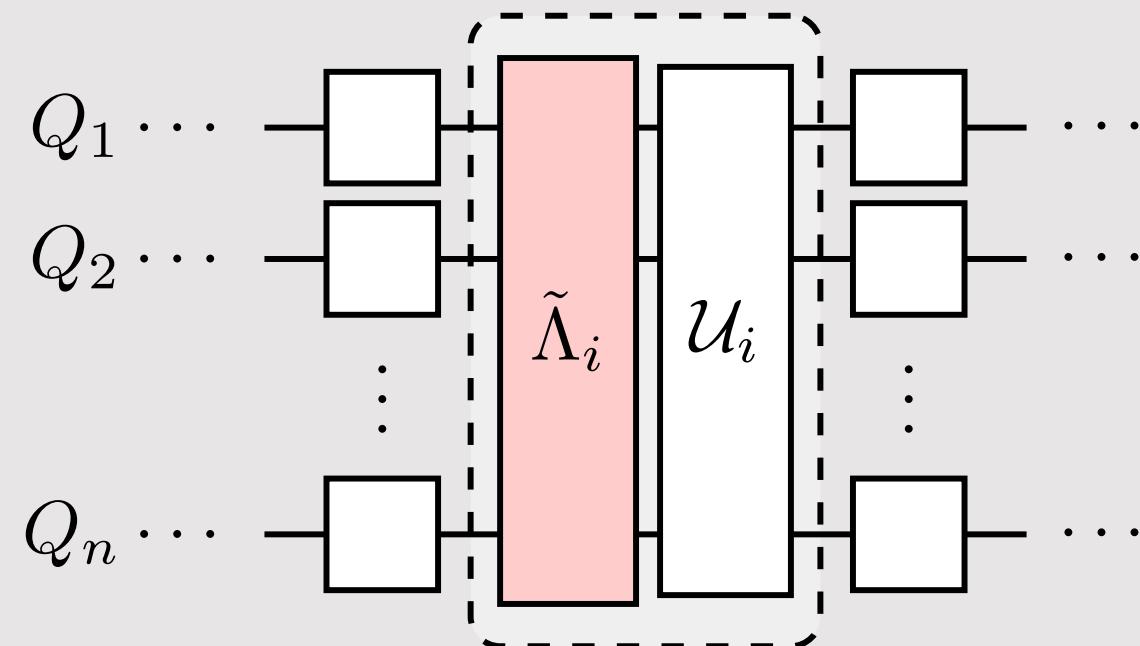


Learn



Cancel
(realization)

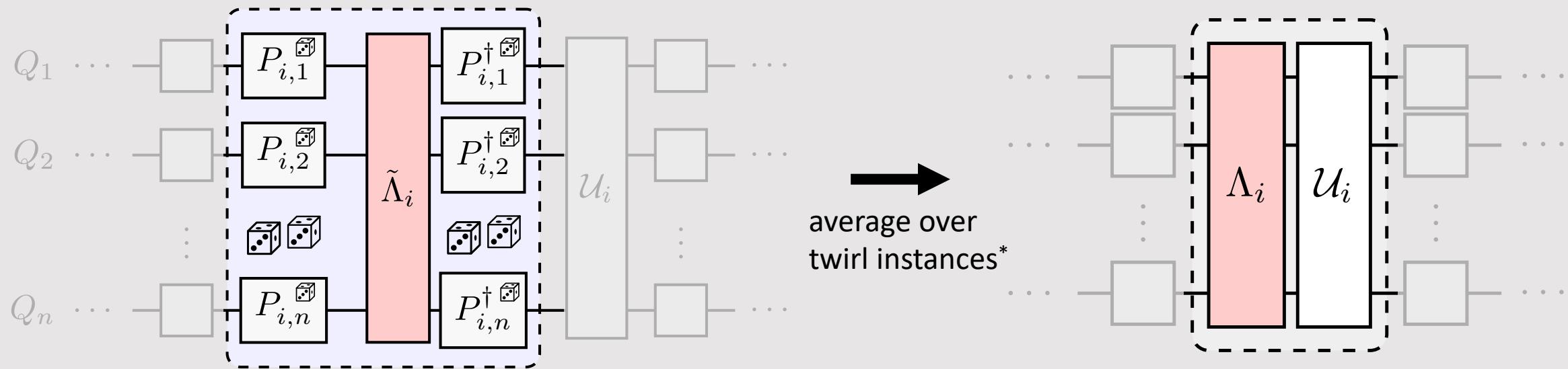
Step 1: Simplify the noise



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

Step 1: Simplify the noise: twirl

twirl reduces to noise $4^n \times 4^n$ 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. ...



Stochastic Pauli channel

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

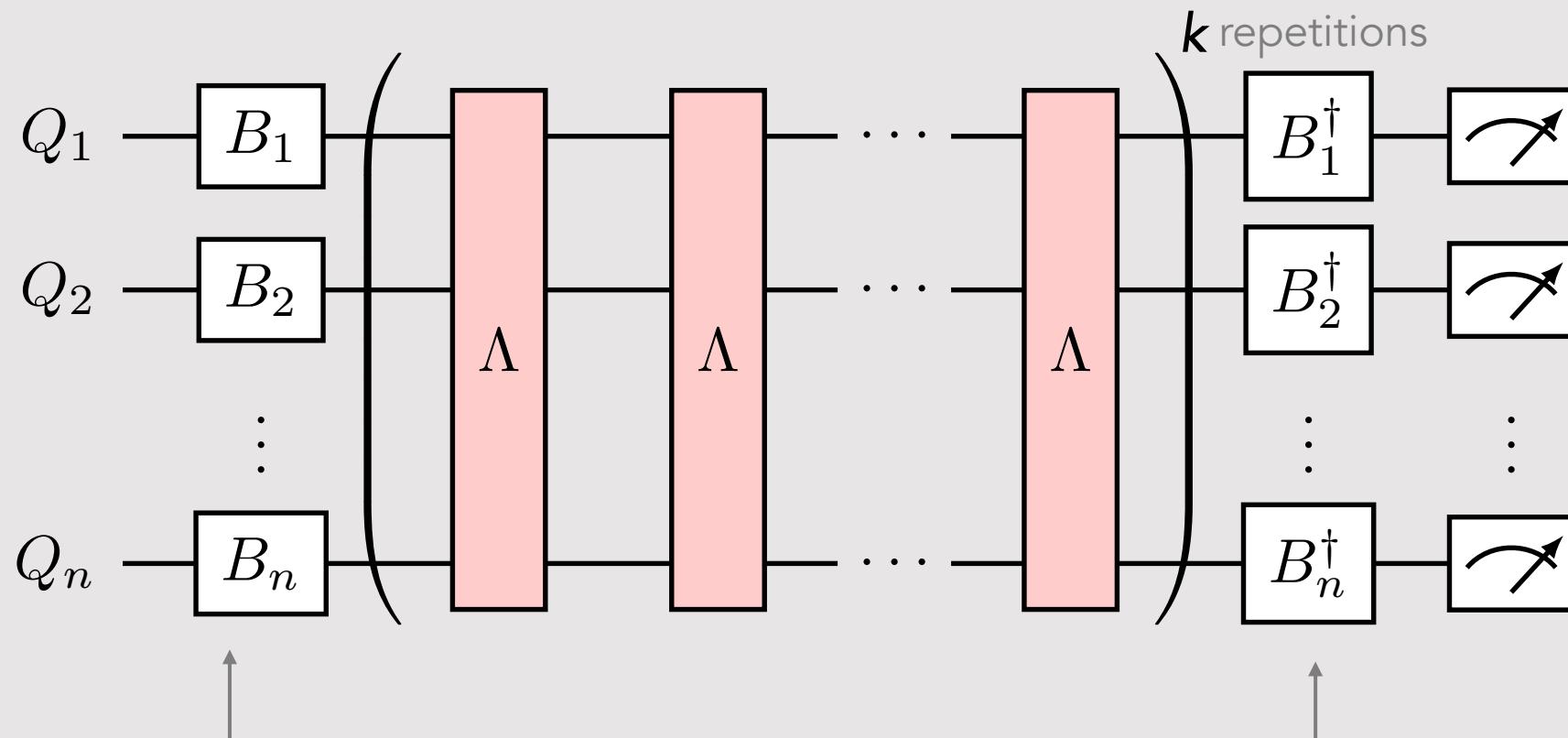
$$\Lambda(P_a) = f_a P_a$$

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

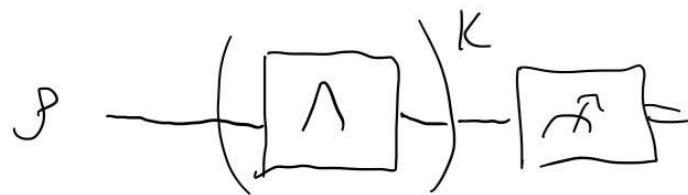
Akin to:

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



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

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

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:

$$\rho = \frac{1}{2^n} \sum_6^1 \rho_6 \hat{P}_6$$

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

$$\rho' = \frac{1}{2^n} \sum_6^1 \rho'_6 \hat{P}'_6$$

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.

Posterior state: Action of the channel on the input state

Pauli decomposition of posterior state

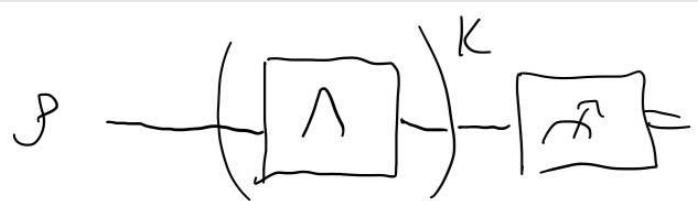
Since the channel is linear

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

$$\Lambda(\rho) = \Lambda\left(\frac{1}{2^n} \sum_6^1 \rho_6 \hat{P}_6\right)$$

$$= \frac{1}{2^n} \sum_6^1 \rho_6 \Lambda(\hat{P}_6)$$

Just need to know action of ρ_6 .



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

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

Or written another way:

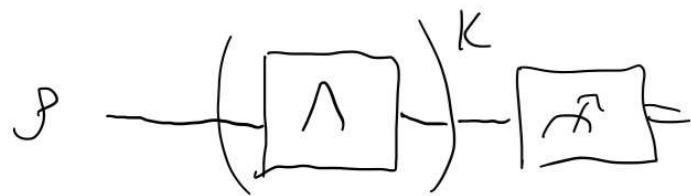
$$\begin{aligned} P_c' &= \sum_b p_b \frac{1}{2^n} \text{Tr}(P_c^\dagger \Lambda(\hat{P}_b)) \\ &= \sum_b p_b \frac{\langle\langle P_c | \Lambda | P_b \rangle\rangle}{\langle\langle P_c | P_c \rangle\rangle} \end{aligned}$$

$$\begin{aligned} \Lambda(\rho) &= \Lambda\left(\sum_b p_b \Lambda(\hat{P}_b)\right) \\ &= \frac{1}{2^n} \sum_b p_b \Lambda(\hat{P}_b) \quad \text{Just need to know each } \Lambda(\hat{P}_b) \\ &= P' \\ &= \frac{1}{2^n} \sum_c P_c' \hat{P}_c \quad \text{since } \text{Tr}(\hat{P}_a^\dagger \hat{P}_b) = S_{ab} 2^n \\ &\quad \text{orthogonal} \end{aligned}$$

Solve for P_c' by equate & orthogonal use:

$$\begin{aligned} \text{Tr}\left(P_c^\dagger \sum_b p_b \Lambda(\hat{P}_b)\right) &= \text{Tr}\left(P_c^\dagger \sum_{c'} p_{c'}' \hat{P}_{c'}\right) \\ \sum_b p_b \text{Tr}(P_c^\dagger \Lambda(\hat{P}_b)) &= \sum_{c'} p_{c'}' \text{Tr}(\hat{P}_c^\dagger \hat{P}_{c'}) \\ &= \sum_{c'} p_{c'}' \delta_{cc'} 2^n \\ \frac{1}{2^n} \sum_b p_b \text{Tr}[P_c^\dagger \Lambda(\hat{P}_b)] &= P_c' \quad \checkmark \end{aligned}$$

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):

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

For a single step

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

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

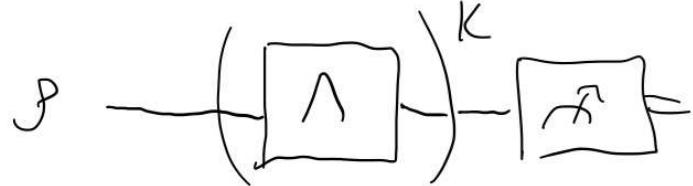
$$= \boxed{\rho_c} \quad \text{now sub in}$$

$$= \sum_b f_b \underbrace{\langle\langle P_c | \Lambda | P_b \rangle\rangle}_{2^n}$$

$$\rightarrow \text{for a Pauli channel } \langle\langle P_c | \Lambda | P_b \rangle\rangle = f_c S_{bc} 2^n$$

$$= \boxed{f_c \rho_c} \leftarrow \begin{array}{l} \text{initial state coefficient} \\ \text{fidelity } c \text{ of } \Lambda \end{array}$$

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

$$\Rightarrow \rho_{c=2} = \text{Tr}(\hat{Z}\rho) = +1$$

For repeating the channel K times

$$\rho_0 \xrightarrow{\left(\begin{array}{c|c} \Lambda & \\ \hline & \end{array}\right)^K} \left(\begin{array}{c|c} \hat{P}_c & \\ \hline & \end{array}\right)$$

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

$$\Rightarrow \rho'_c = \sum_b f_b \underbrace{\langle \hat{P}_c | \Lambda^K | \rho_b \rangle}_{2^n}$$

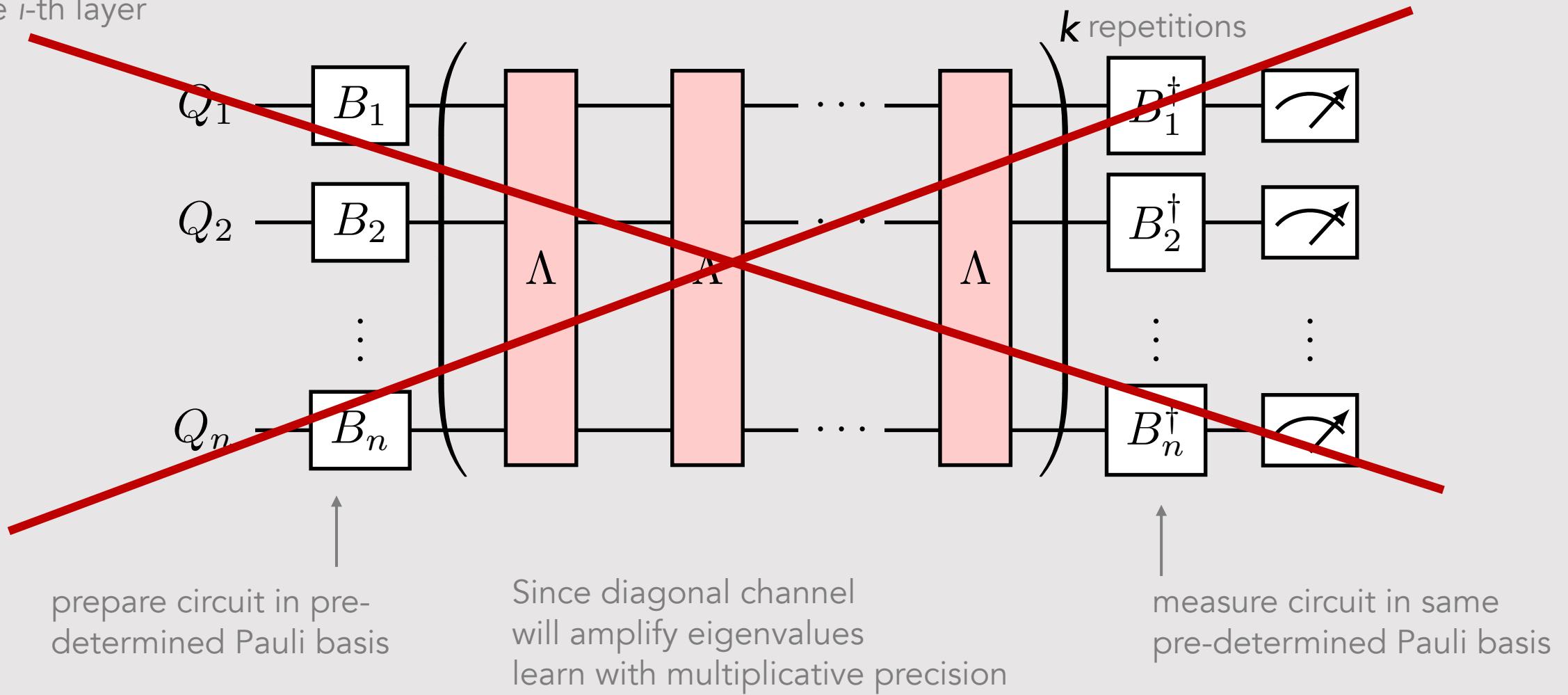
$$= \sum_b f_b^K \rho_b$$

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

Can include all diagonal SAW

Step 2 wish: amplify noise

for the i -th layer



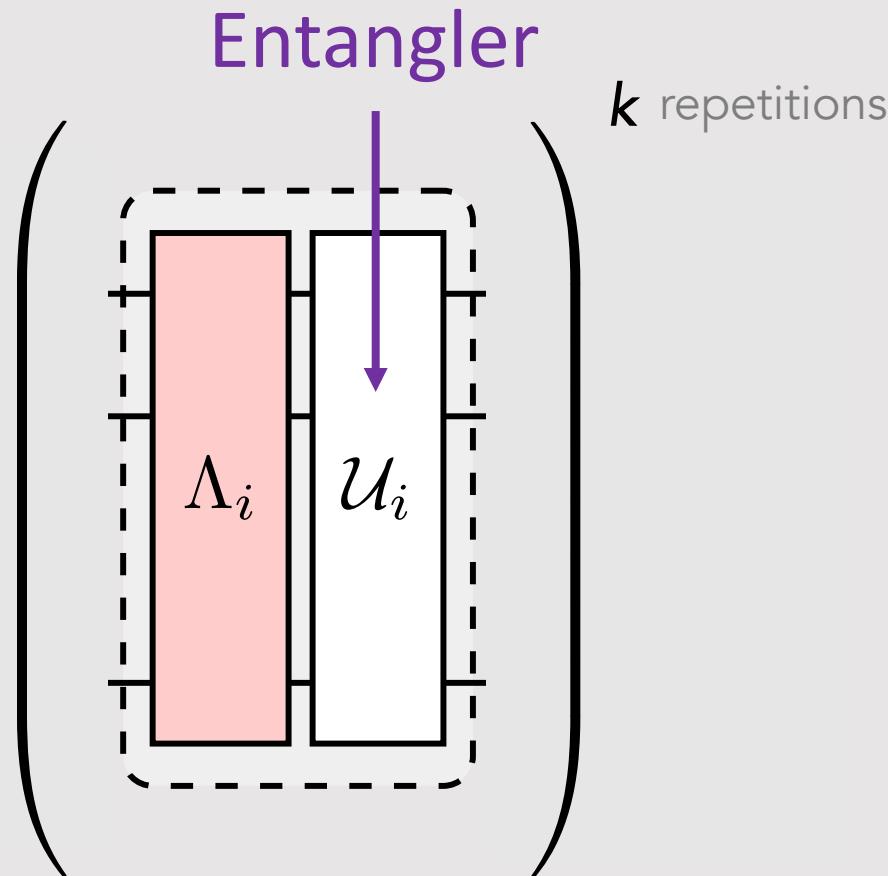
Akin to:

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), ...

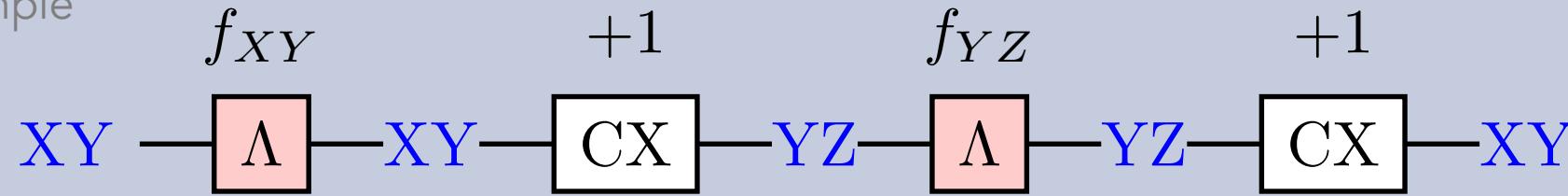
Erhard *et al.*, arXiv:1902.08543; Ferracin *et al.*, arXiv:2201.10672, ...

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

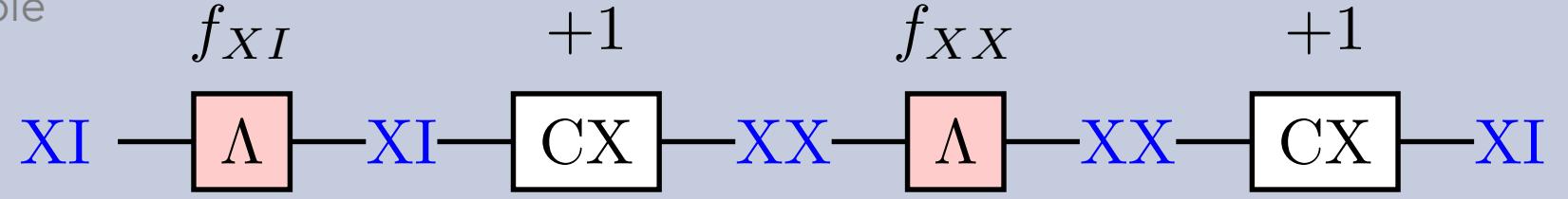
$$\Lambda(P_a) = f_a P_a$$

2Q example



action of CX here can be undo by SQ gates

2Q example

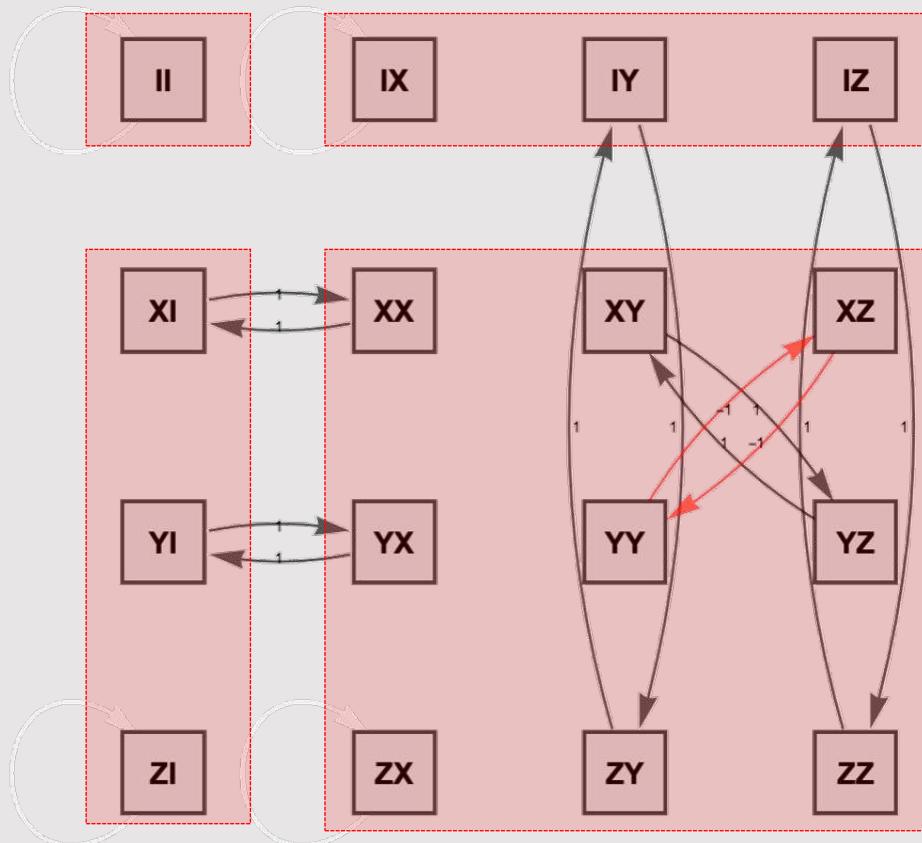


action of CX here can not be undone by SQ gates

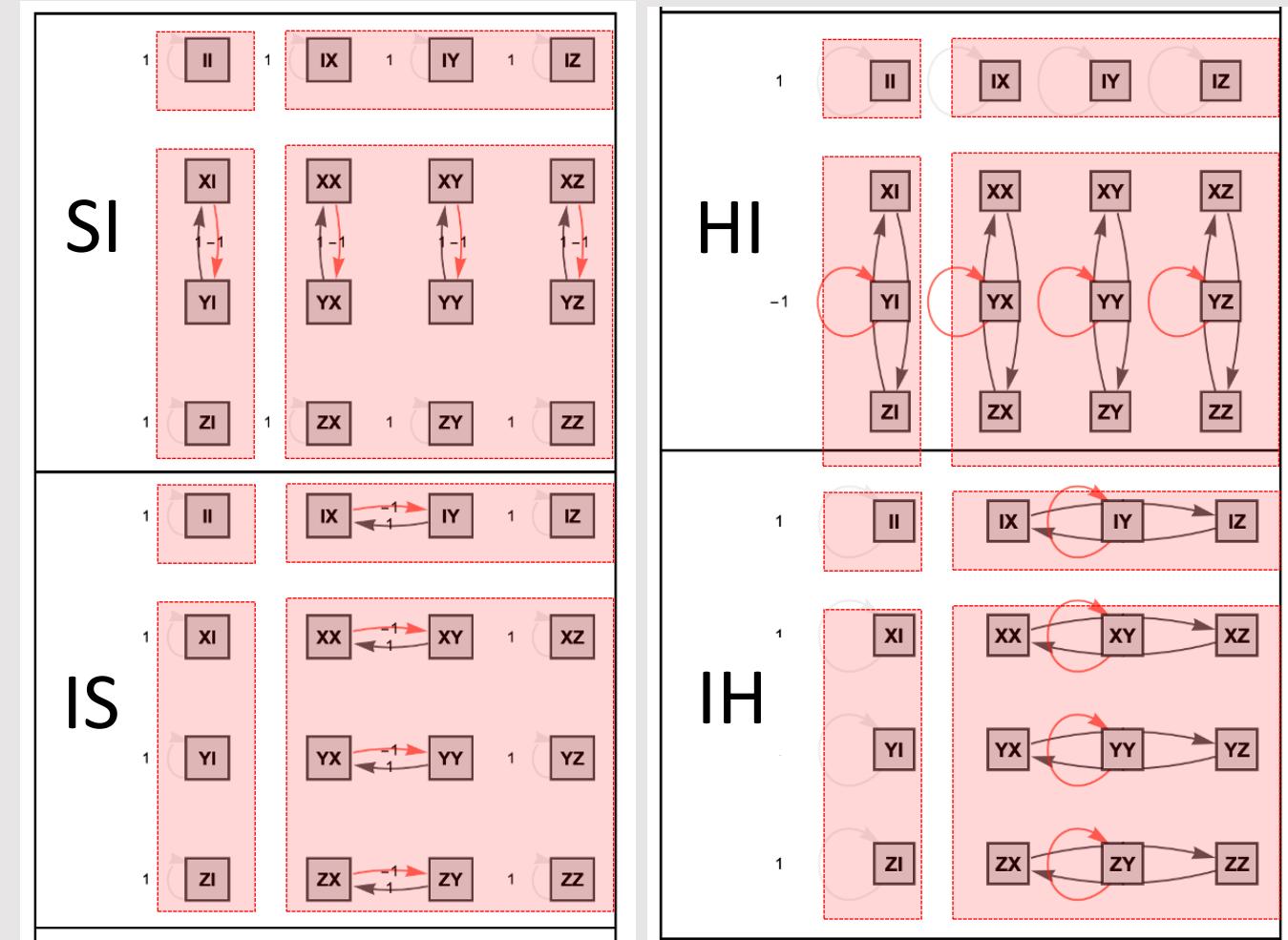
fundamental degeneracy – can not undo some non-local – need entangling operation

How to gates move state Paulis around?

cNOT

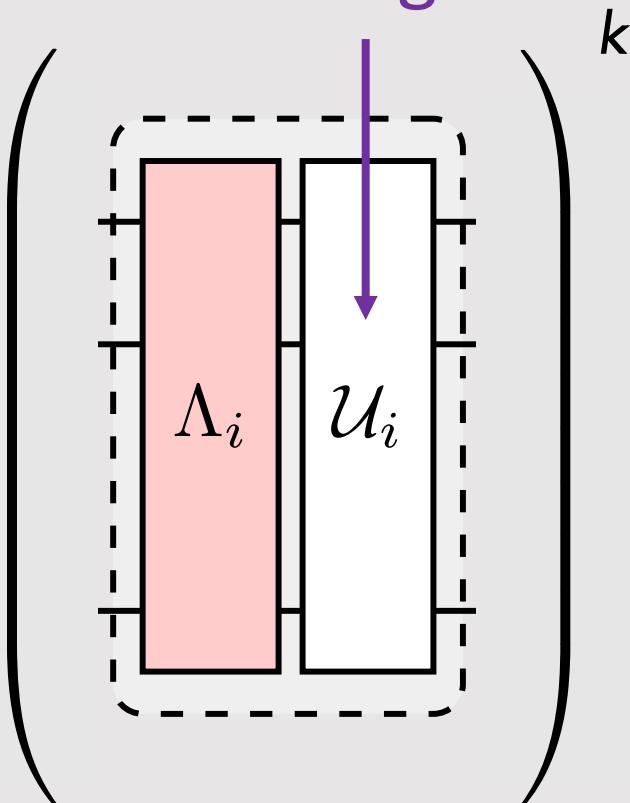


Example single qubit gates



Not so simple to learn noise of entangling gates

Entangler



Ideally wish

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

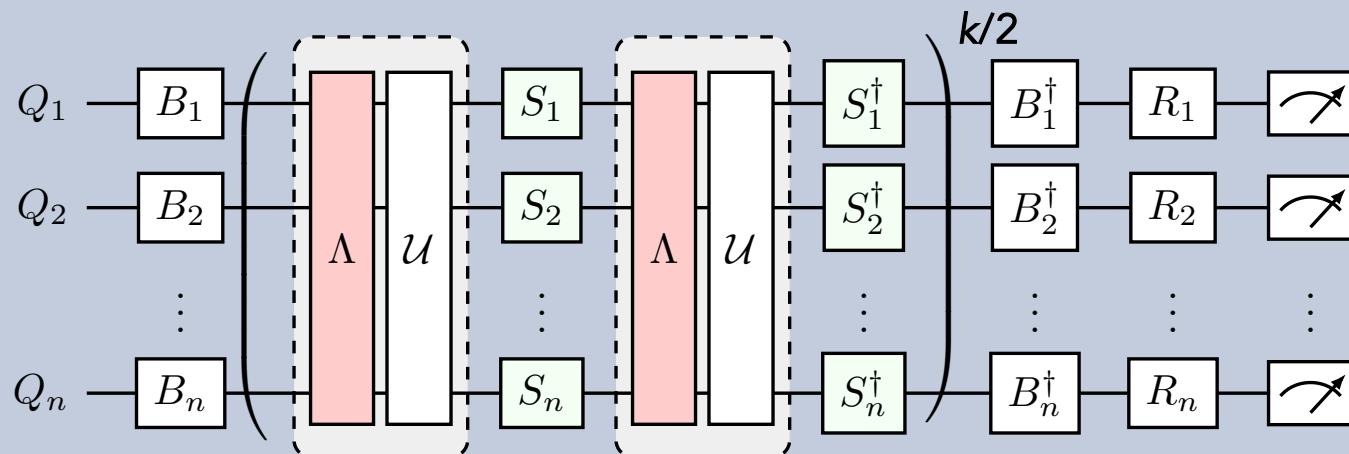
Fundamental no-go theorem on learning



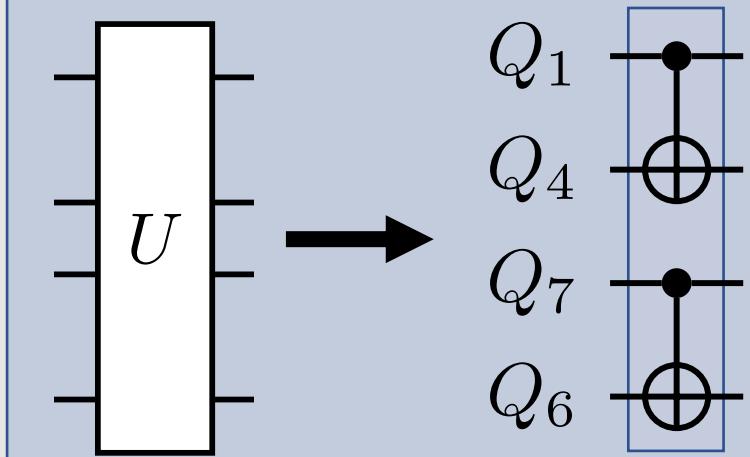
For general and in-depth
Senrui Chen, Y. Liu, M. Otten, A. Seif, B. Fefferman, L. Jiang
arXiv:2206.06362 (2022)
or supplement of our paper for qubit version and work by
S. Flammia, S Benjamin, and teams.

Solution: Custom protocol + weak assumption

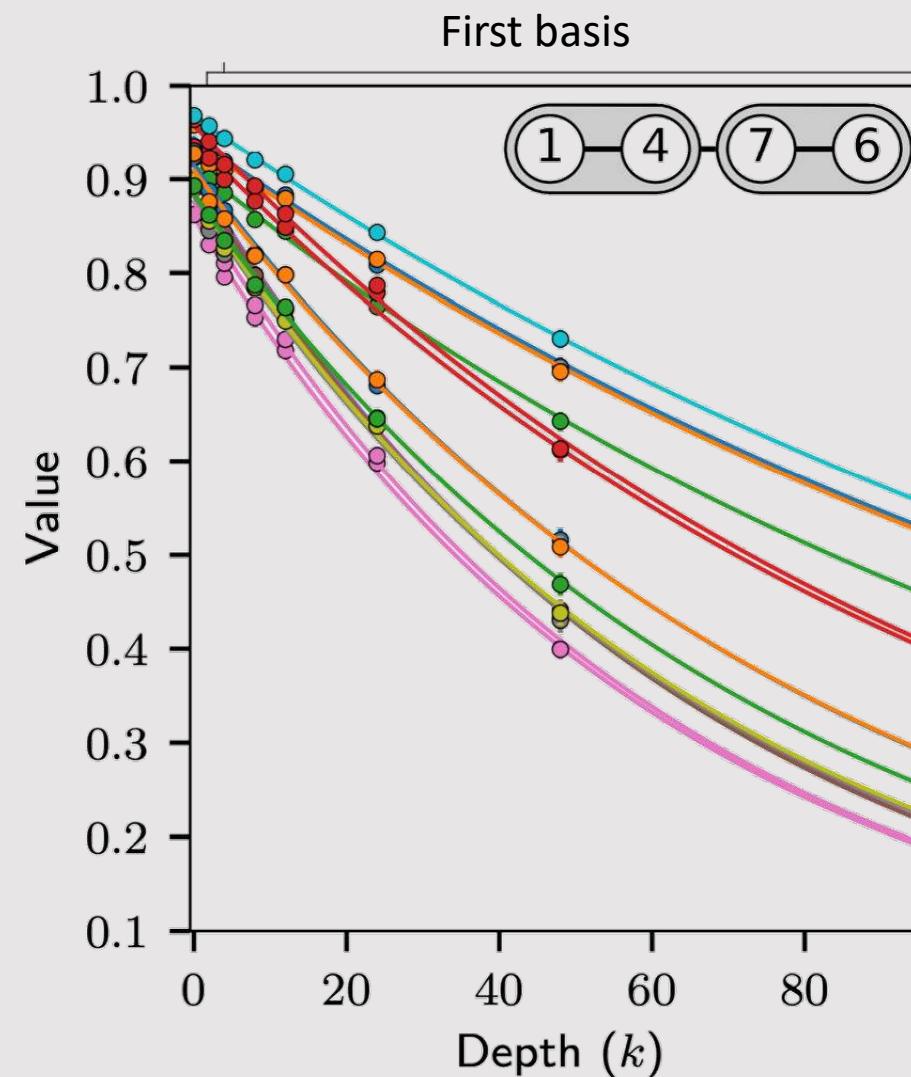
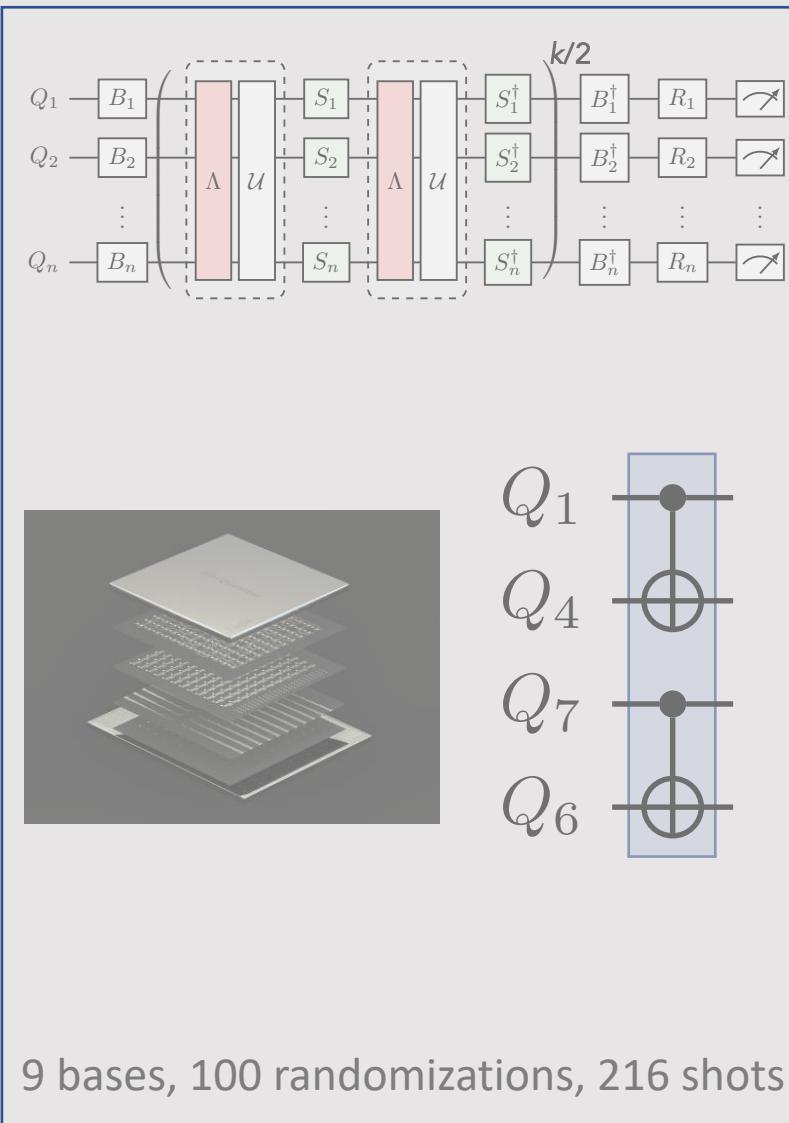
Learning circuits



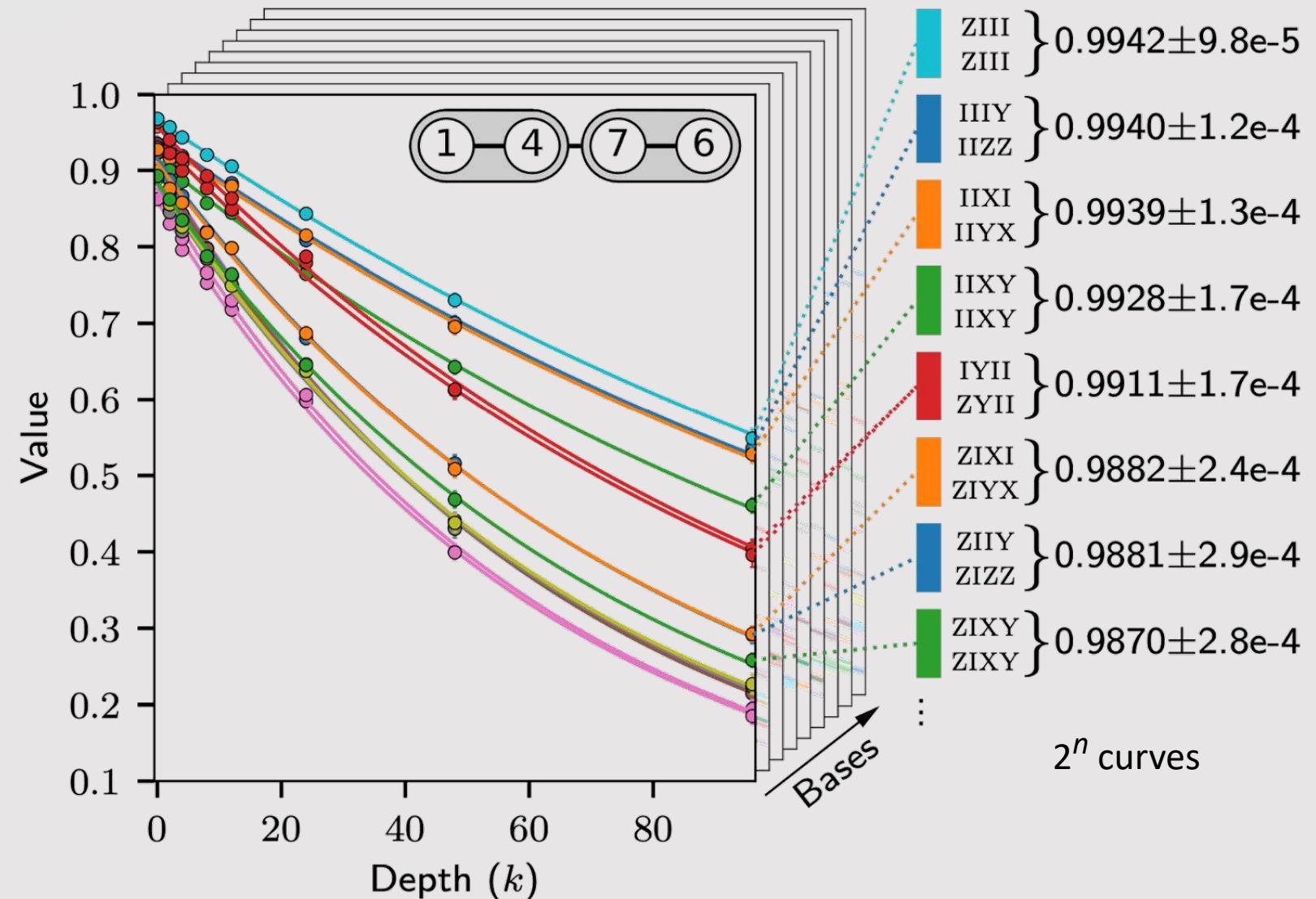
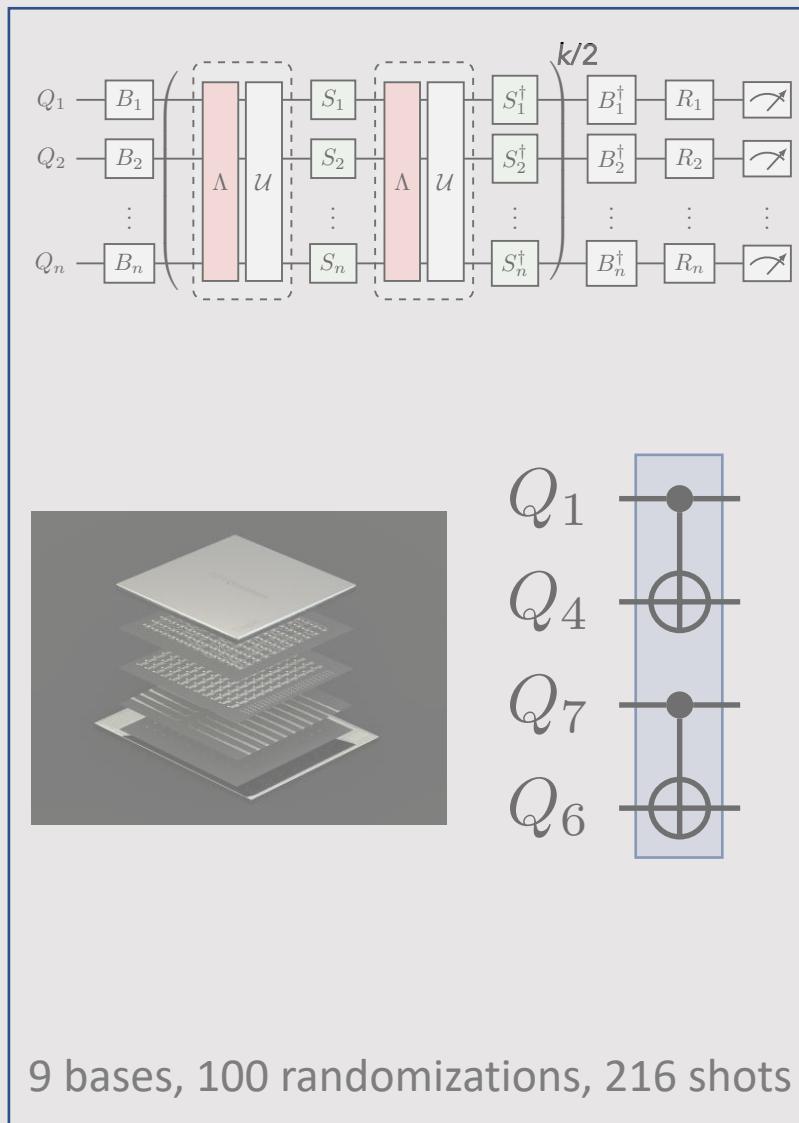
Example



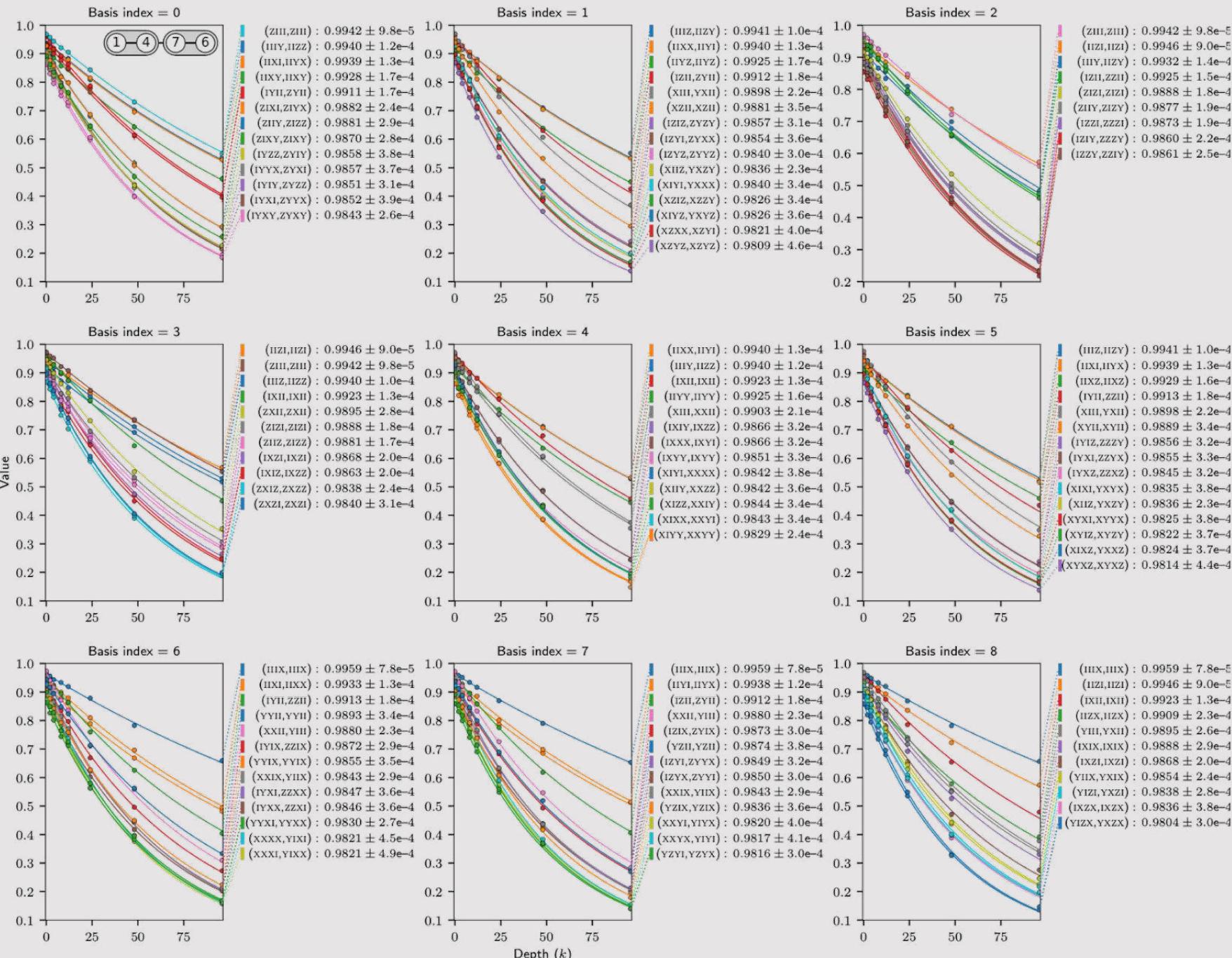
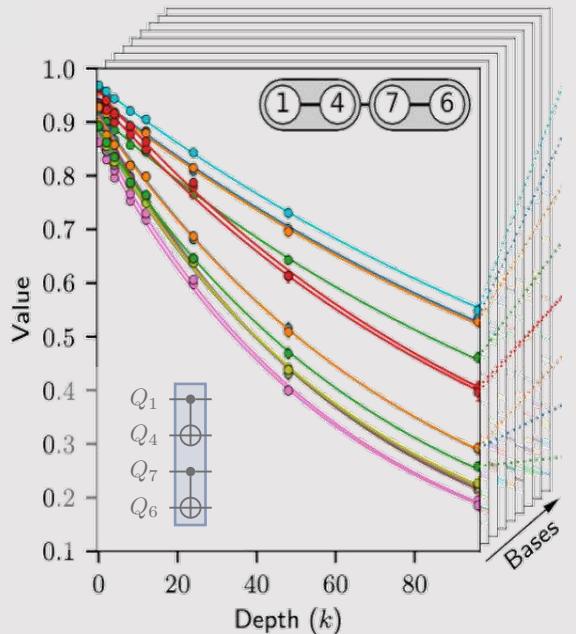
Learning the noise: raw data



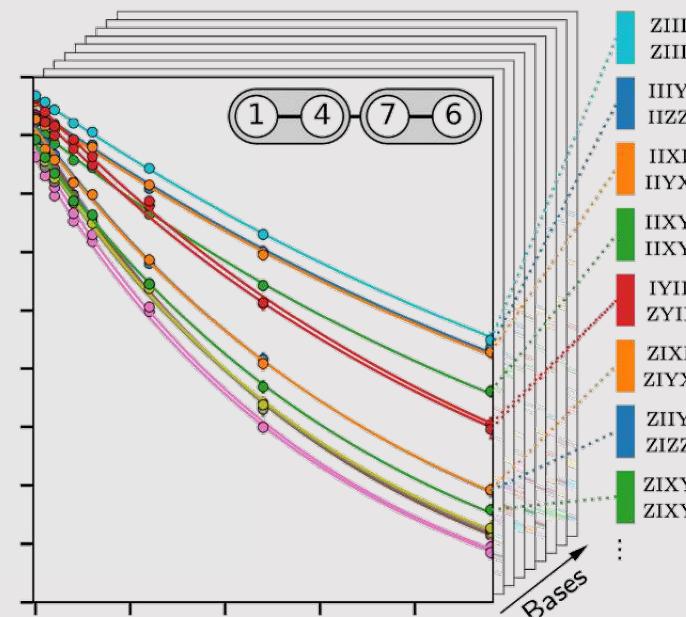
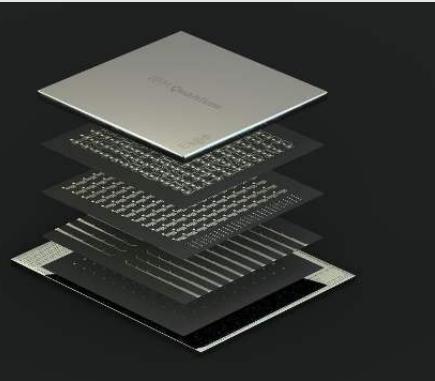
Learning the noise: raw data



Raw data



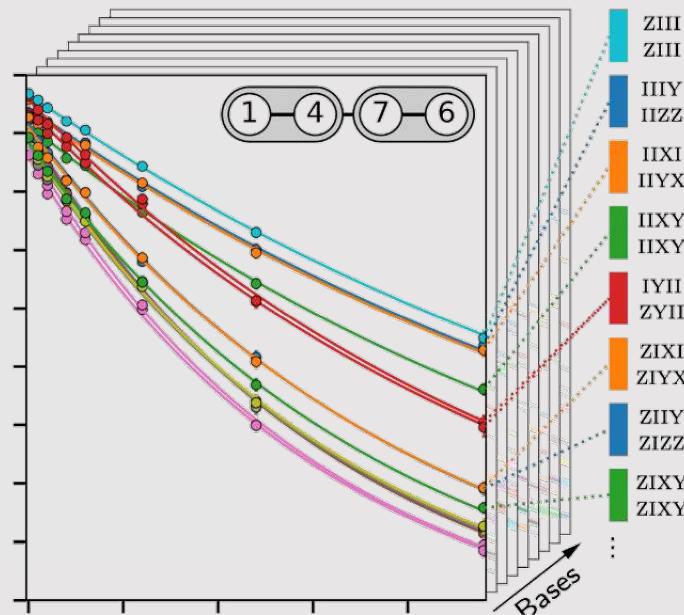
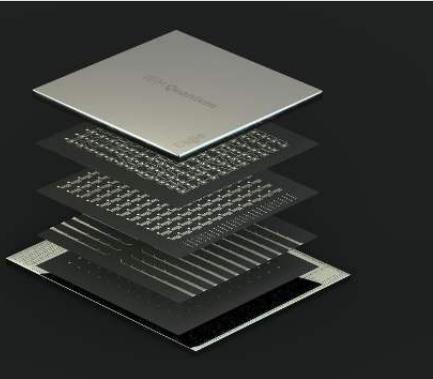
Reconstructing quantum channel from measurement data



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

Still 4^n

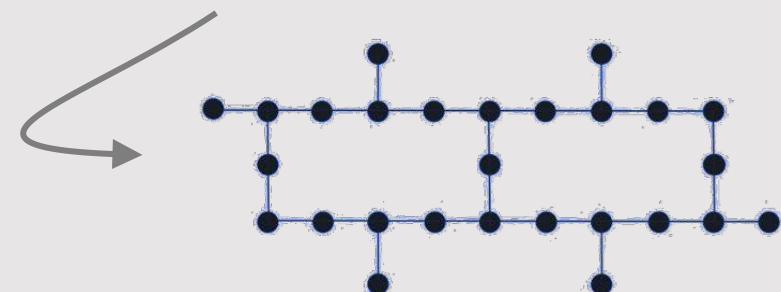
Sparse Pauli-Lindblad model



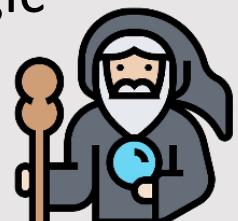
$$\Lambda(\rho) = \sum_{a=0}^{4^n - 1} c_a P_a \rho P_a^\dagger$$

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

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



Magic



icon: Eucalyp

Highlight: Ewout van den Berg

Zlatko Minev, IBM Quantum (86)

Notes

Dissipator for a given Pauli

$$\begin{aligned} \hat{\mathcal{D}}[\rho]_g &= P \hat{\rho} P^+ - \frac{1}{2}(P P^+ + P^+ P) \\ &= P \hat{\rho} P^+ - P \\ &= (P \cdot P - I) \rho \end{aligned}$$

$$\begin{aligned} \hat{\mathcal{L}} &= \sum_a \gamma_a \left(\hat{\rho}_a^\dagger - \frac{1}{I} \right) \\ &= \sum_a \gamma_a \hat{\mathcal{D}}[\hat{\rho}_a] \end{aligned}$$

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

$$= \prod_a \Lambda[\rho_a]$$

$$\text{but } [\hat{\mathcal{D}}[\hat{\rho}_a], \hat{\mathcal{D}}[\rho_b]] = 0$$

Each sub channel

$$\begin{aligned} \Lambda[\sum P_a] &= \exp(\gamma D\{\vec{P}\}) \\ &= \exp(\gamma(\vec{P}_a + \vec{I})) \\ &= \exp(\gamma \vec{P}_a) \exp(\gamma \vec{I}) \exp(-\frac{1}{2} \gamma^2 \sum \vec{P}_a \cdot \vec{I}) \dots \\ &= \exp(\gamma \vec{P}_a) \exp(\gamma \vec{I}) \\ &\hookrightarrow \sum_n \frac{\gamma^n \vec{I}^n}{n!} = \sum_n \frac{\gamma^n}{n!} \vec{I}^n \\ &= \exp(\gamma) \vec{I} = \exp(\gamma) \vec{I} \cdot \vec{I} \end{aligned}$$

Each sub channel

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

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

note

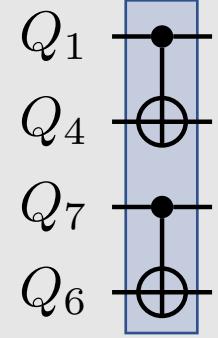
$$\begin{aligned}\hat{P}_a^{n-\text{even}} &\approx \hat{P}_a \\ \hat{P}_a^{n-\text{odd}} &= \hat{I}\end{aligned}$$

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

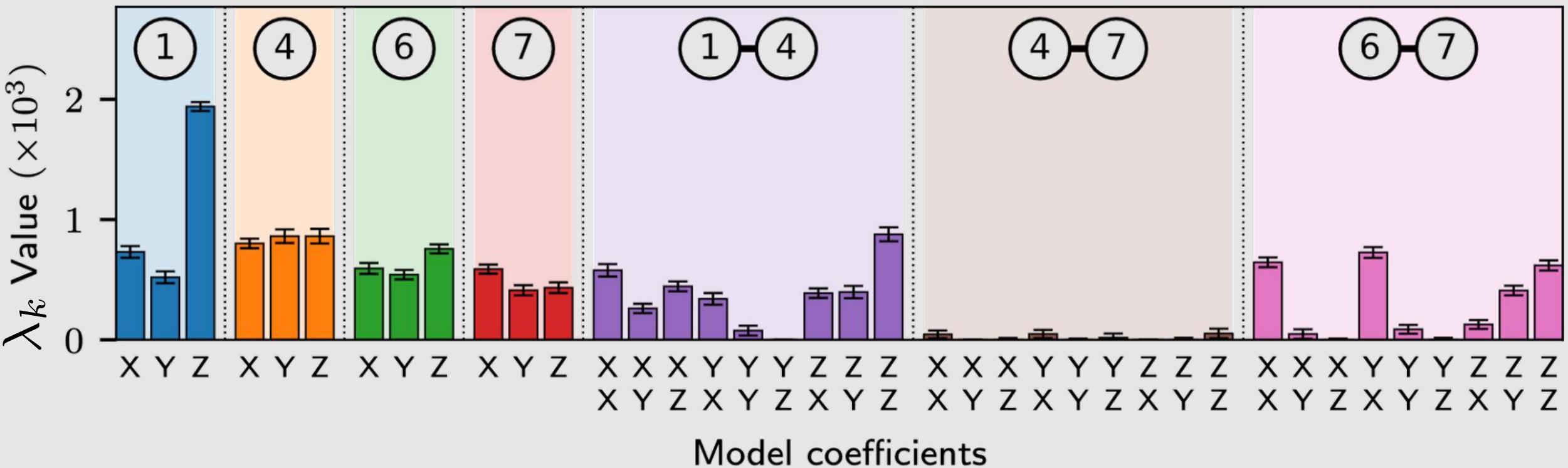
$$= \cancel{\cosh(\gamma)} \hat{I} + \cancel{\sinh(\gamma)} \hat{P}_a$$

$$\lambda [\Sigma P_a] = \cosh(\gamma_a) \hat{I} + \sinh(\gamma_a) \hat{P}_{cl}$$

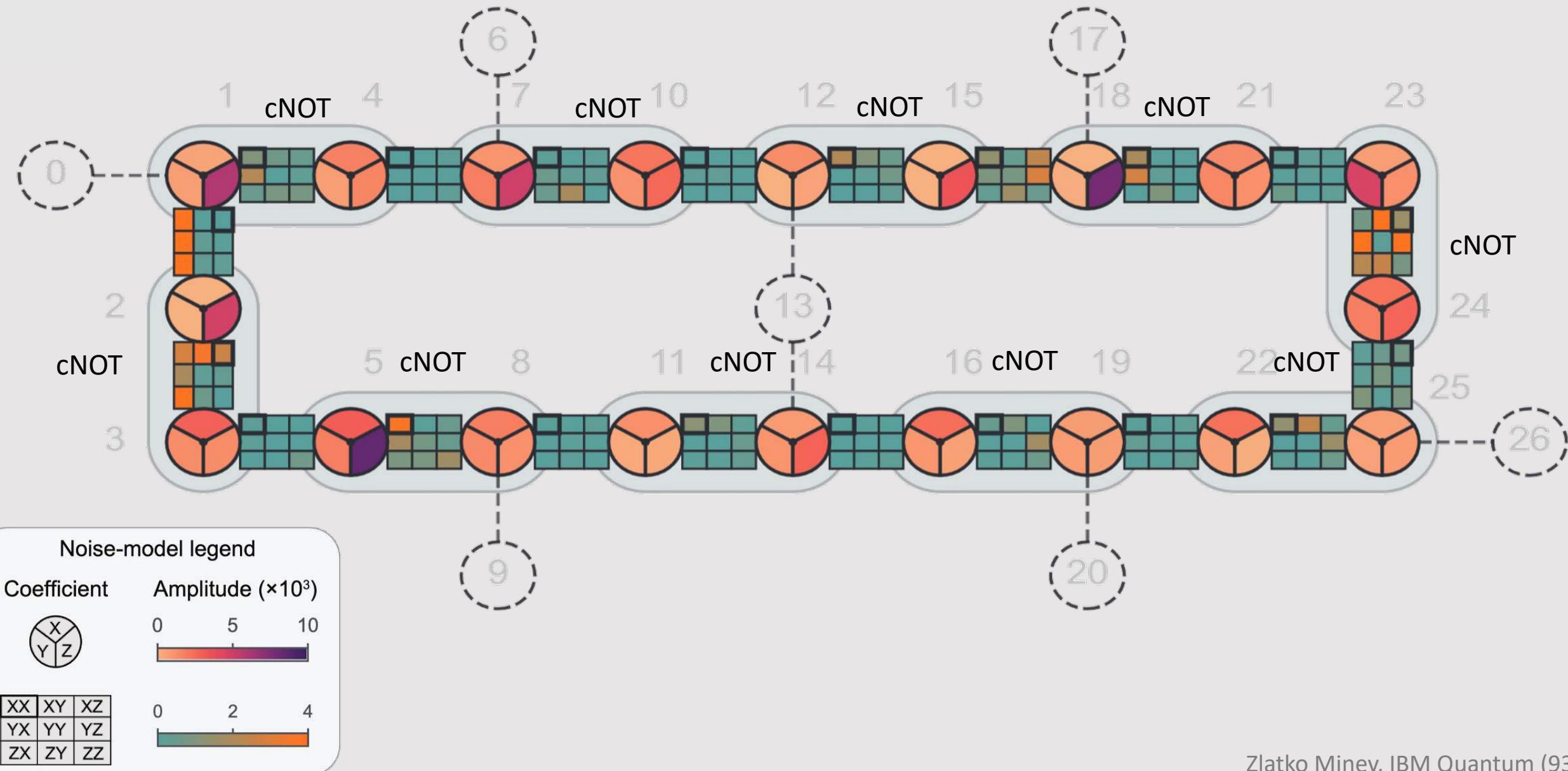
Sparse Lindblad tomogram



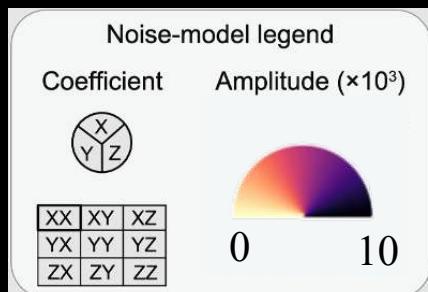
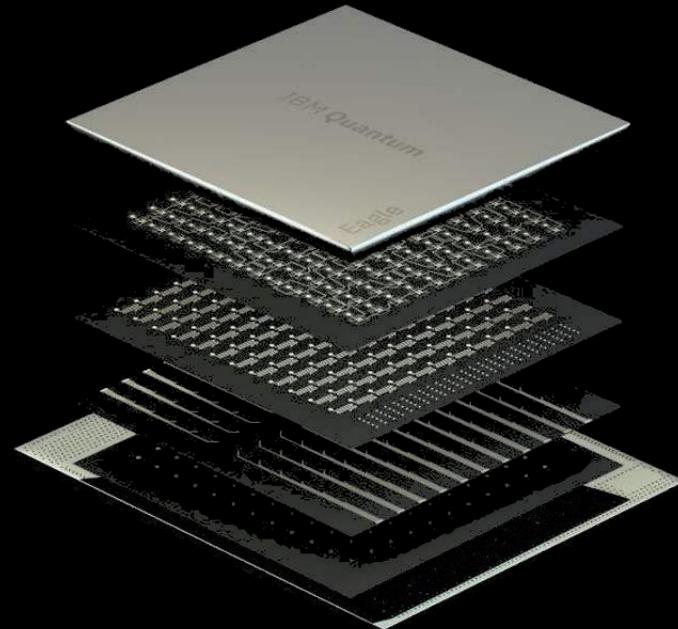
$$\mathcal{L}(\rho) = \sum_{k \in \mathcal{K}} \lambda_k \left(P_k \rho P_k^\dagger - \rho \right)$$



Noise tomogram for 20Q Ising-ring Trotter layer



Noise tomogram for 127Q Trotter layer



Same number of learning circuits as for 4Q

