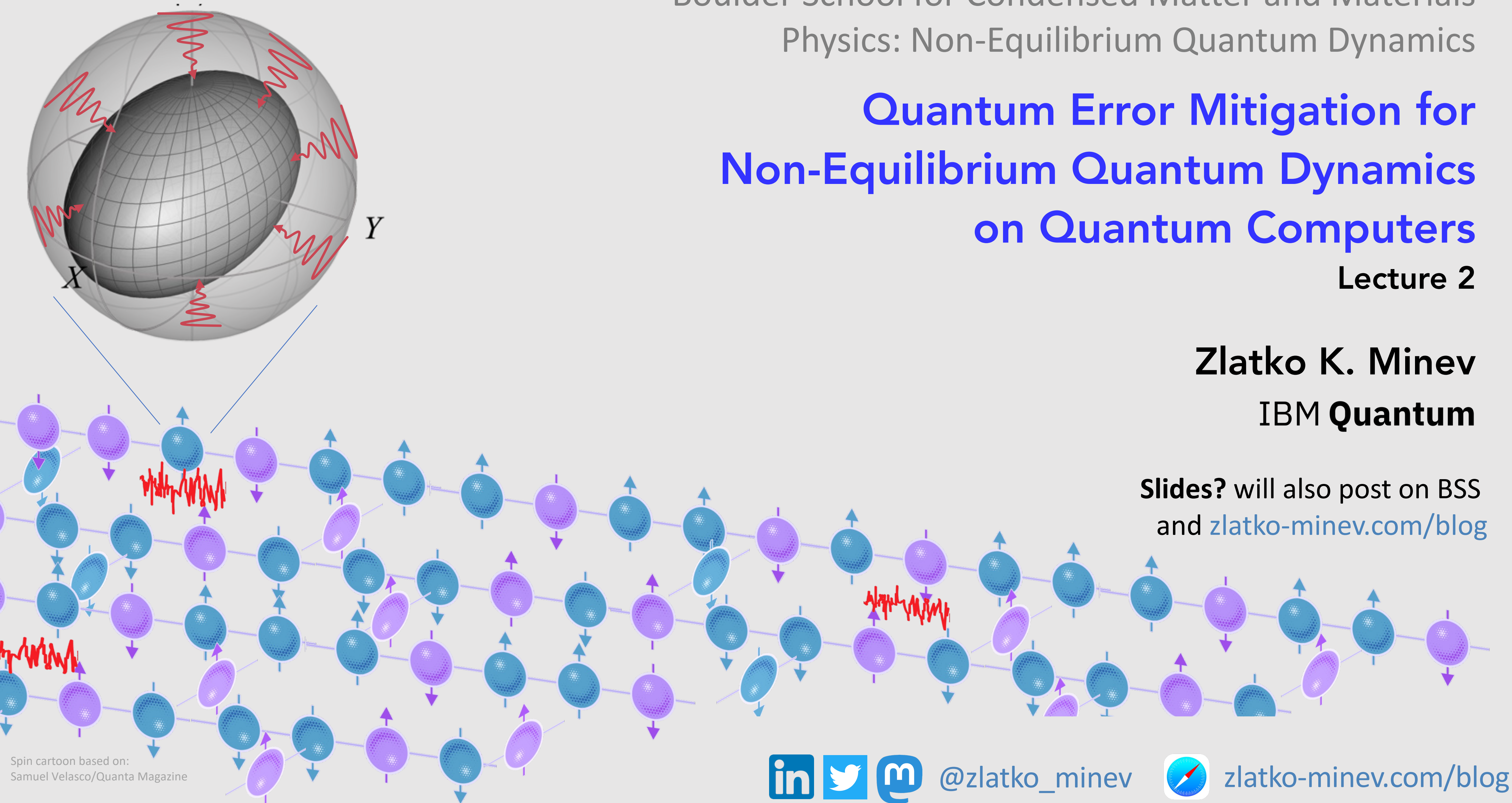
Boulder School for Condensed Matter and Materials Physics: Non-Equilibrium Quantum Dynamics

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?

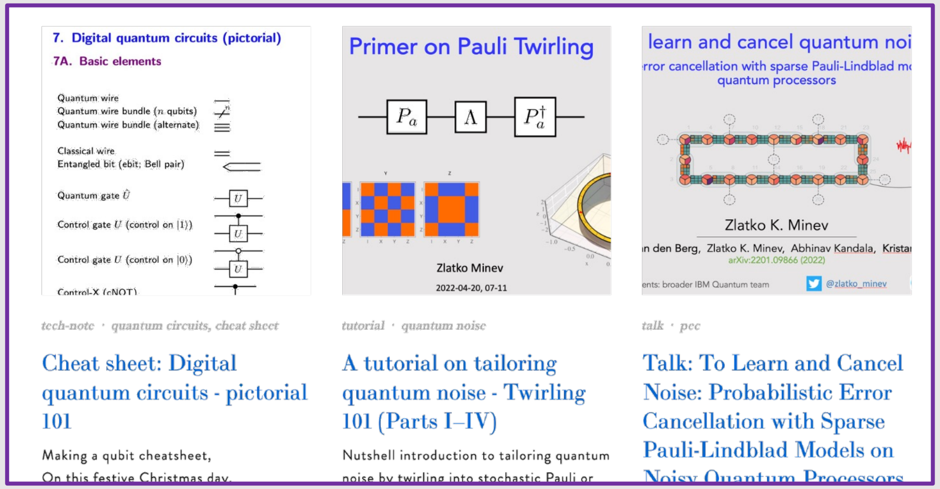
Lecture Slides

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

What is one thing you learned in Lecture 1?

Image: Freepik

Review of Lecture 1

Image: Freepik

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

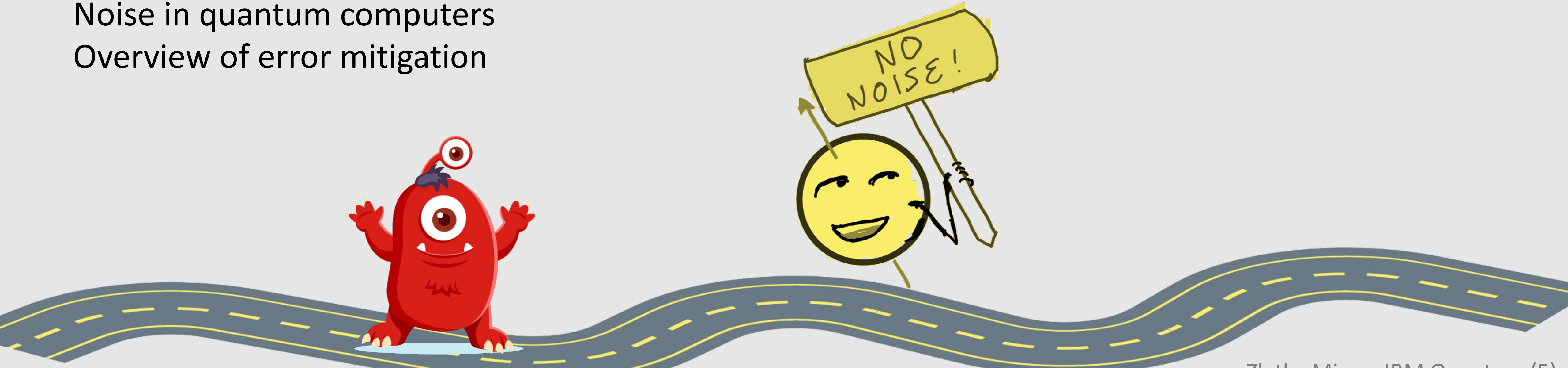
Overview of error mitigation

Mitigation fundamentals

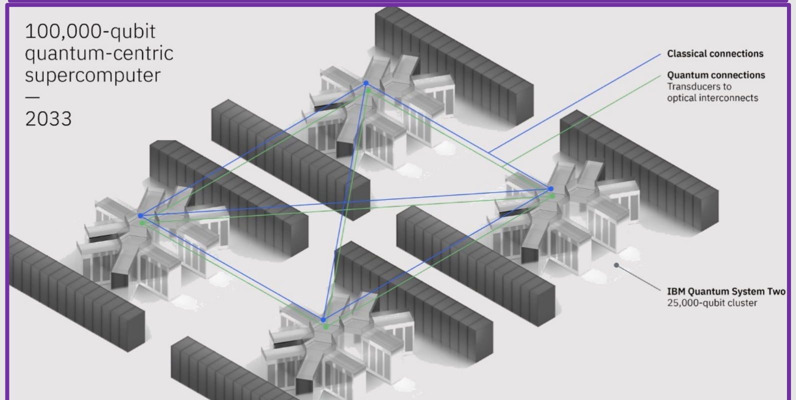
Probabilistic error cancelation (PEC)

Introduction

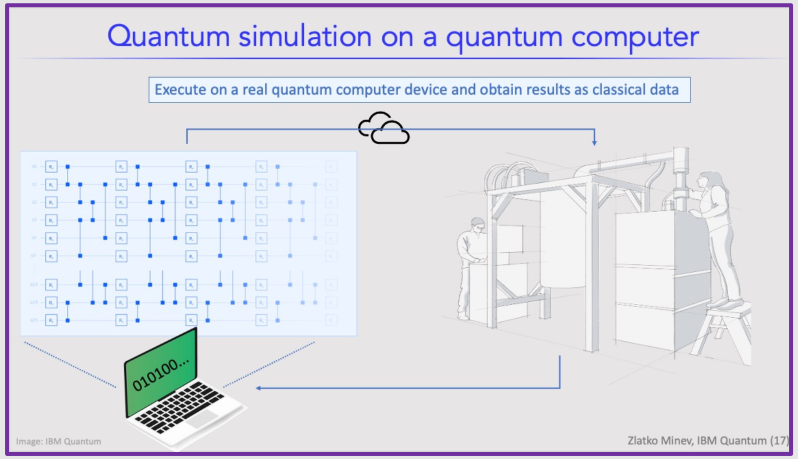
One qubit example

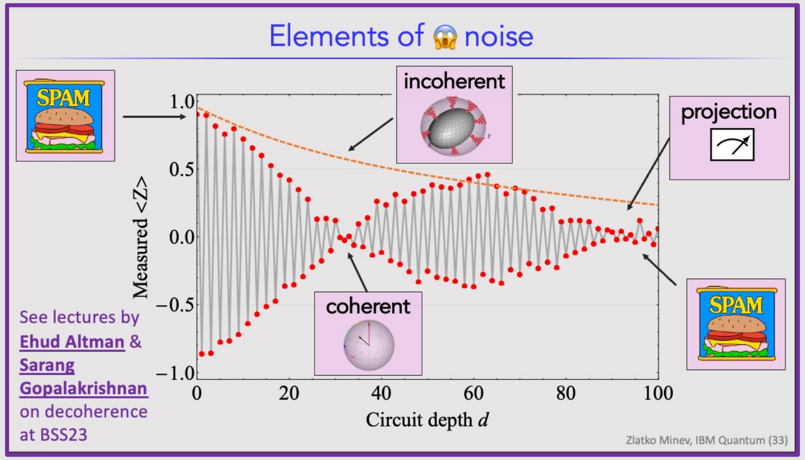


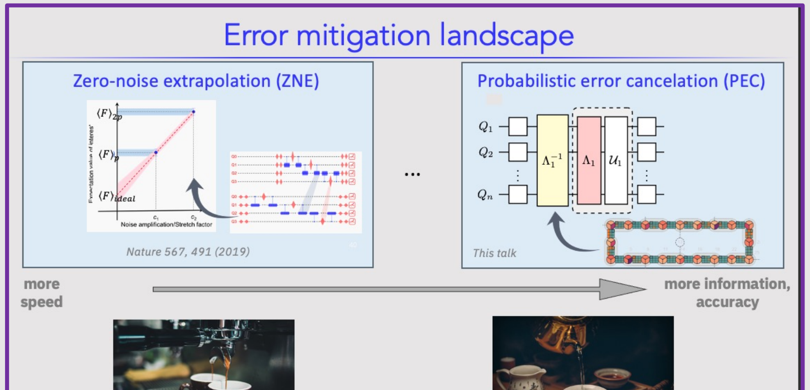
Quantum computers



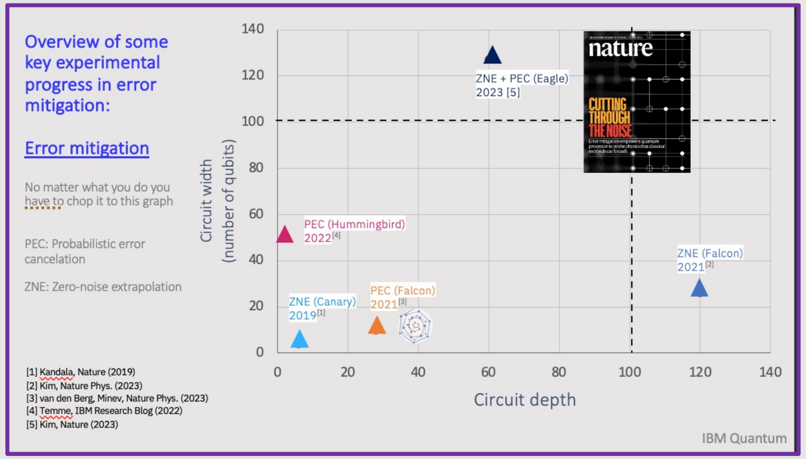
Biggest Problem: Noise

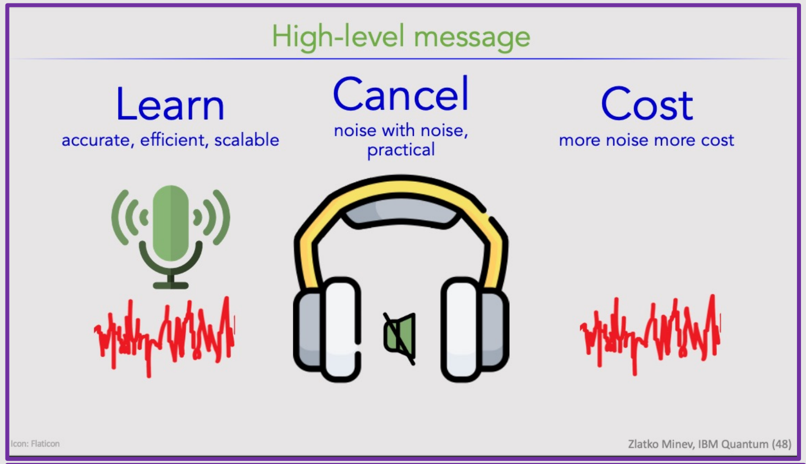


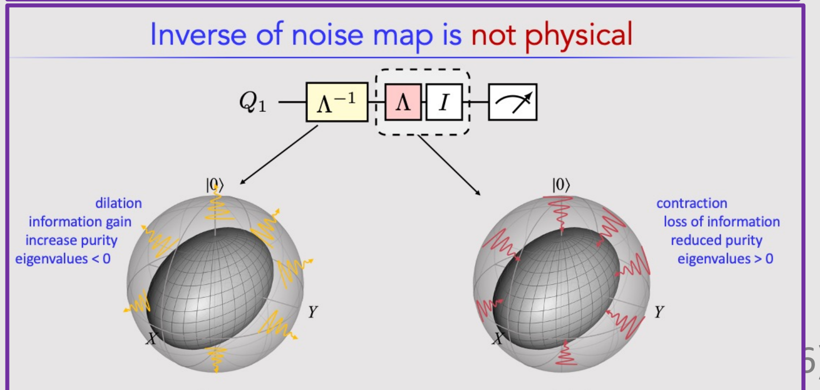




Mitigation







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

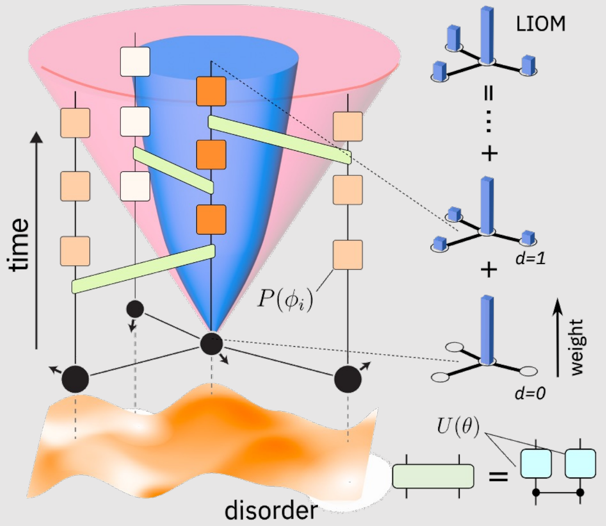
Learning quantum noise
- Challenge
- Overcoming: sparse model

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
…



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

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

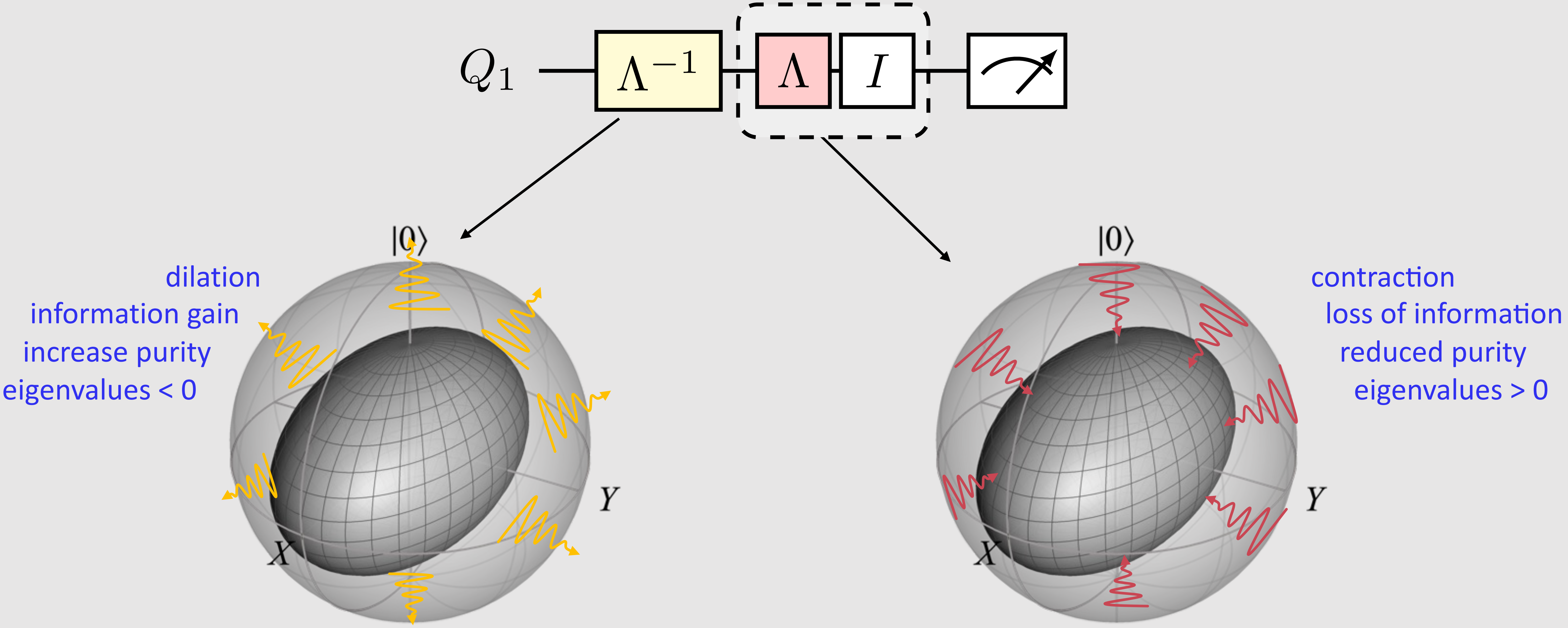
Got Slides?

SCAN ME

Paper: Nature Physics (2023)

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

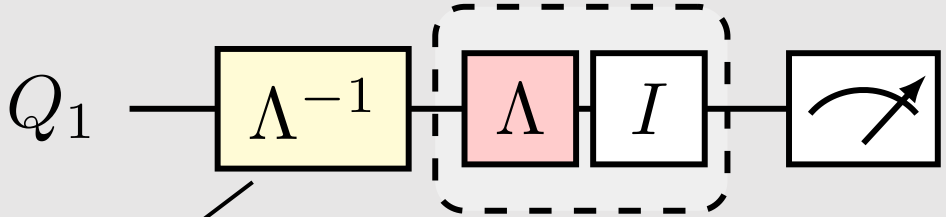
Inverse of noise map is not physical



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

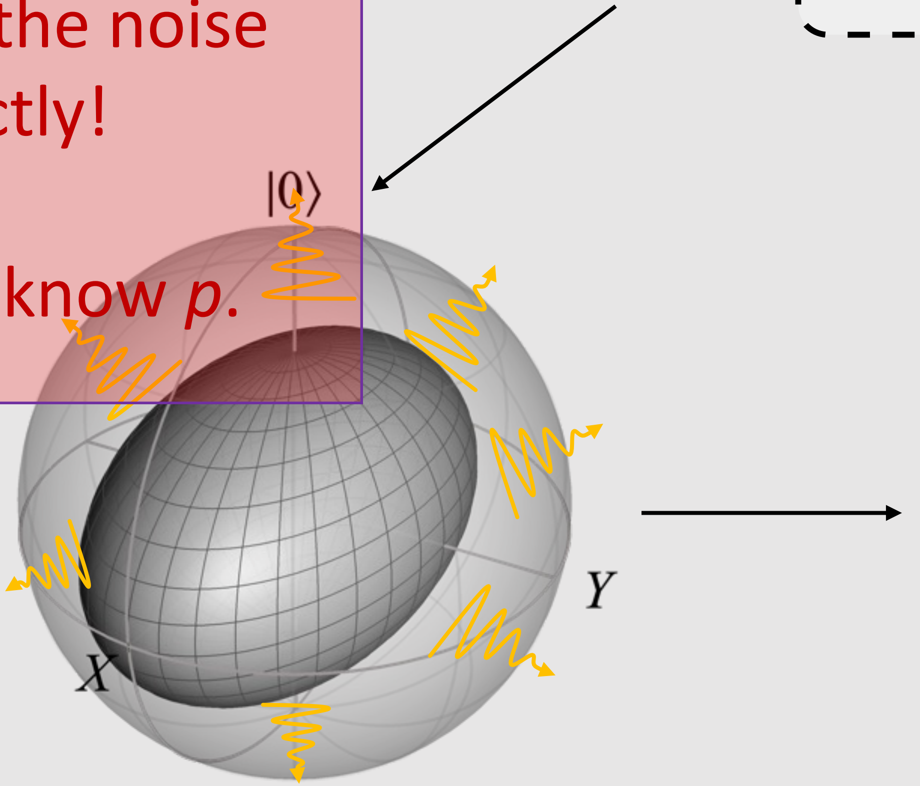
$$\Lambda(\rho)$$

Inverse of noise map is not physical

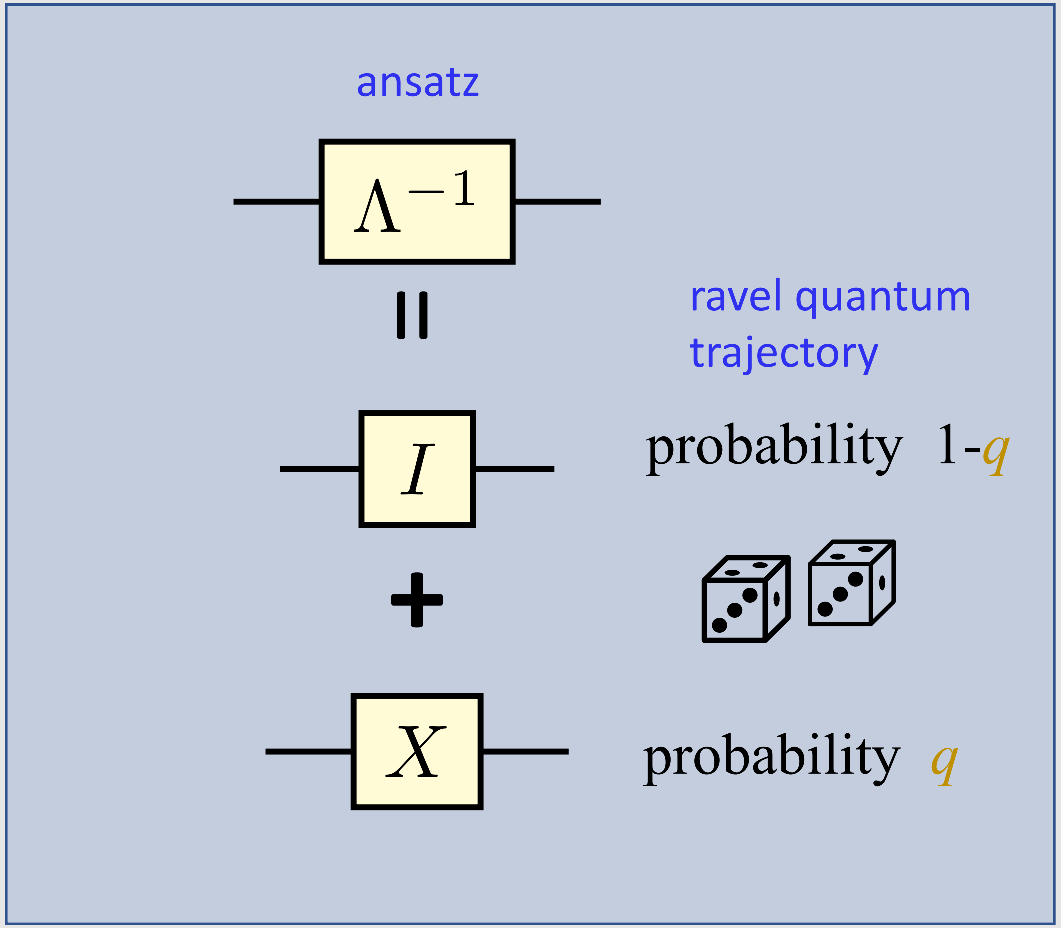


Critically hinges on knowing the noise exactly!

Need to know *p*.



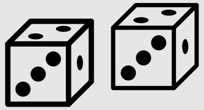
$$\Lambda^{-1}(\rho) = ?$$

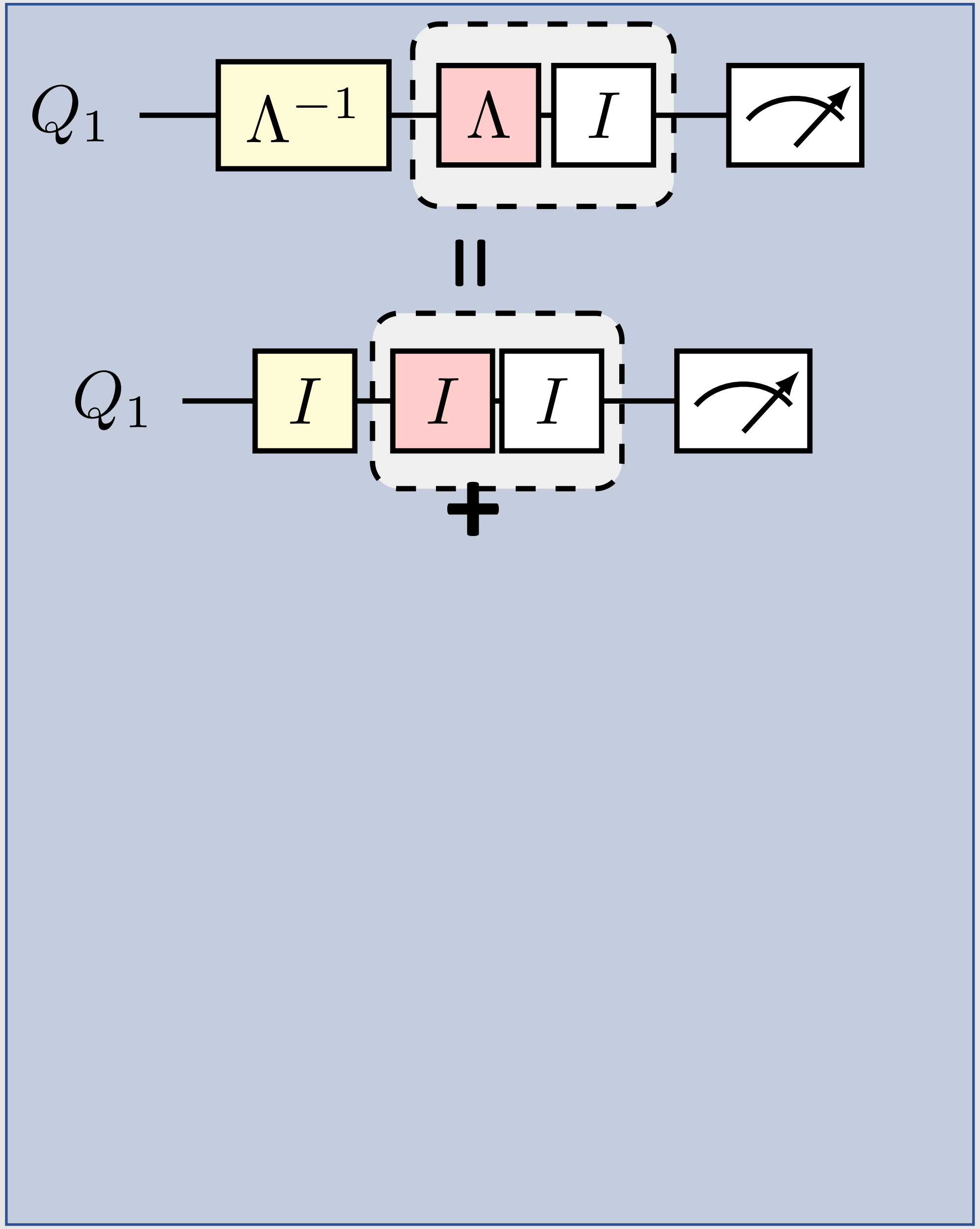


Raveling quantum trajectories to undo noise

probability

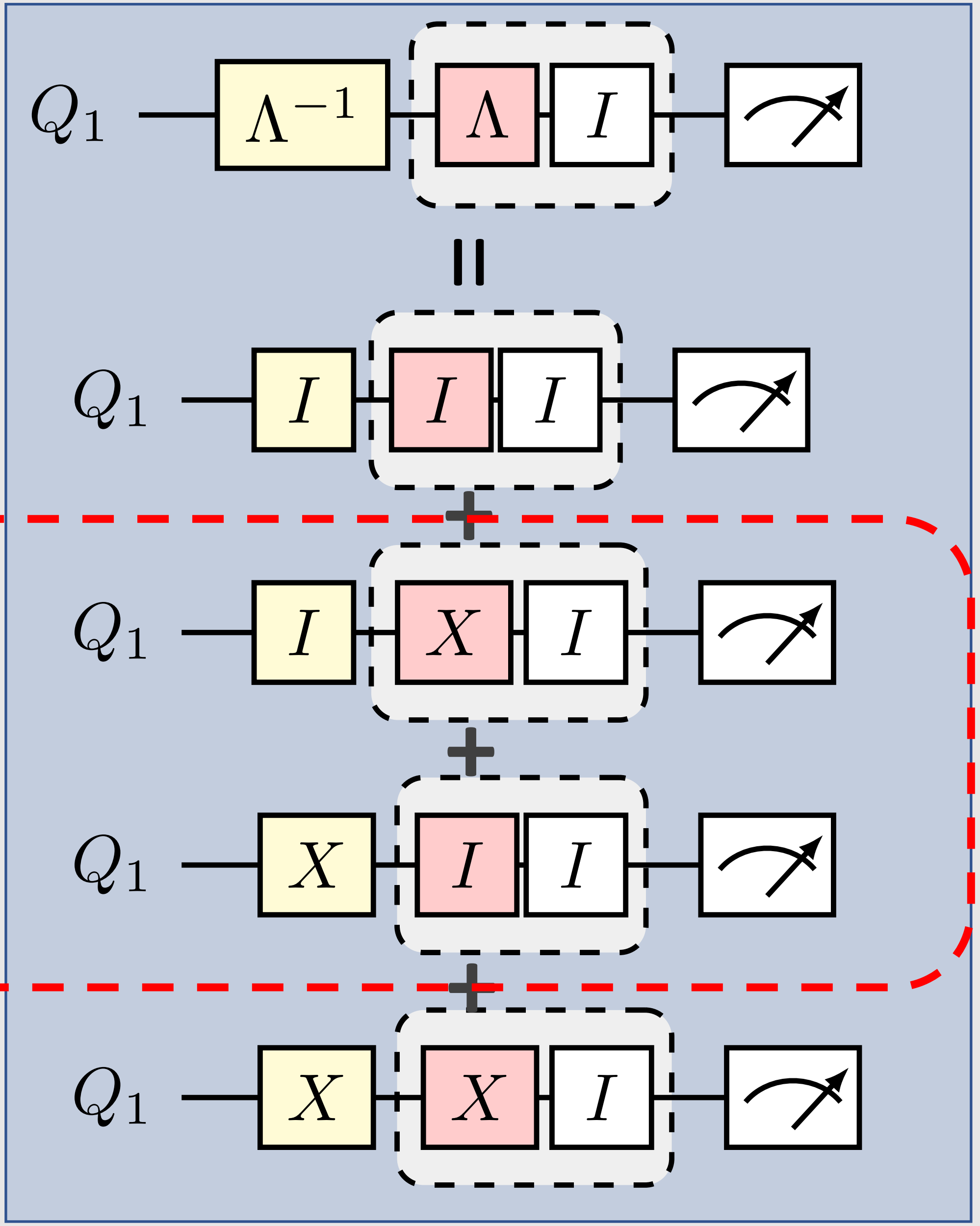
No error (1-q) (1-p)





Raveling quantum trajectories to undo noise

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

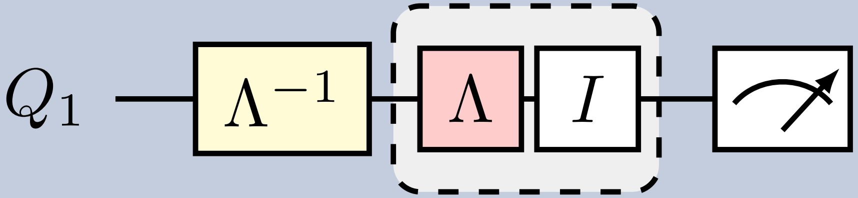


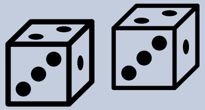
Solution to noise free!

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

Interfere Destructively

Raveling quantum trajectories to undo noise



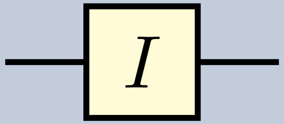


=

probability

sign

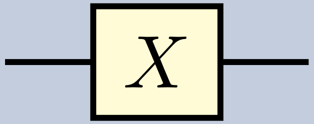
$P_I = |1 - q|/\gamma$



$s_I = \text{sign}(1-q)$

+

$P_X = |q|/\gamma$



$s_X = \text{sign}(q)$

$P_I + P_X = 1$

Solution to noise free!

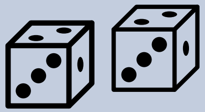
$$q = \frac{-p}{1 - 2p}$$

norm

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

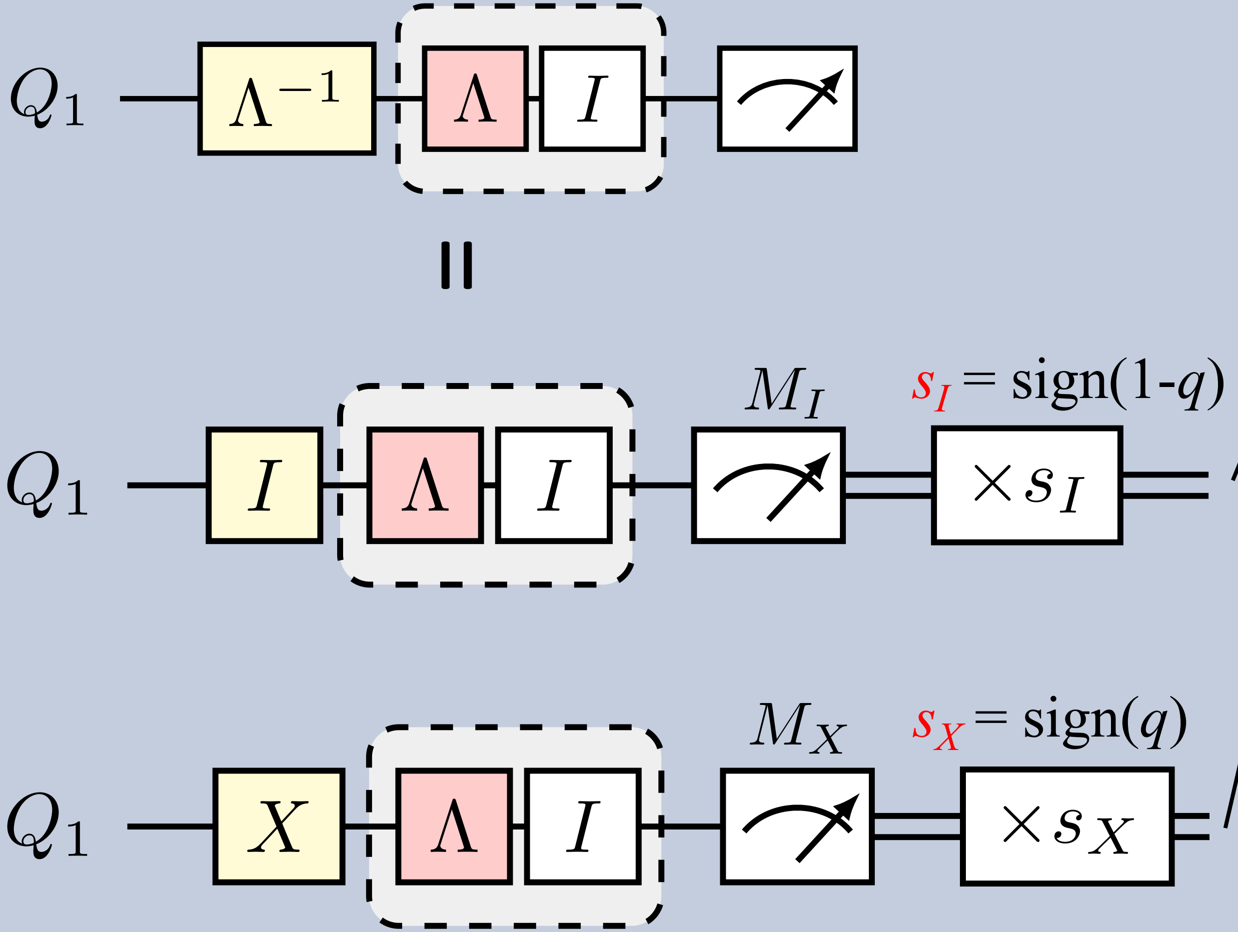
Quasi-probability

Raveling quantum trajectories to undo noise



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

$P_X = |q|/\gamma$



mitigated expectation

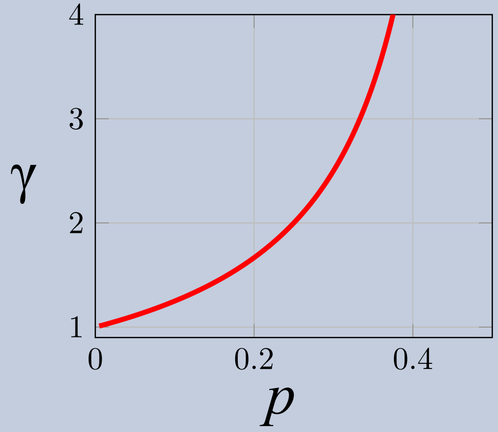
$$\langle M \rangle = \gamma\left(s_I P_I M_I + s_X P_X M_X\right)$$

Gain: Bias-free estimate!

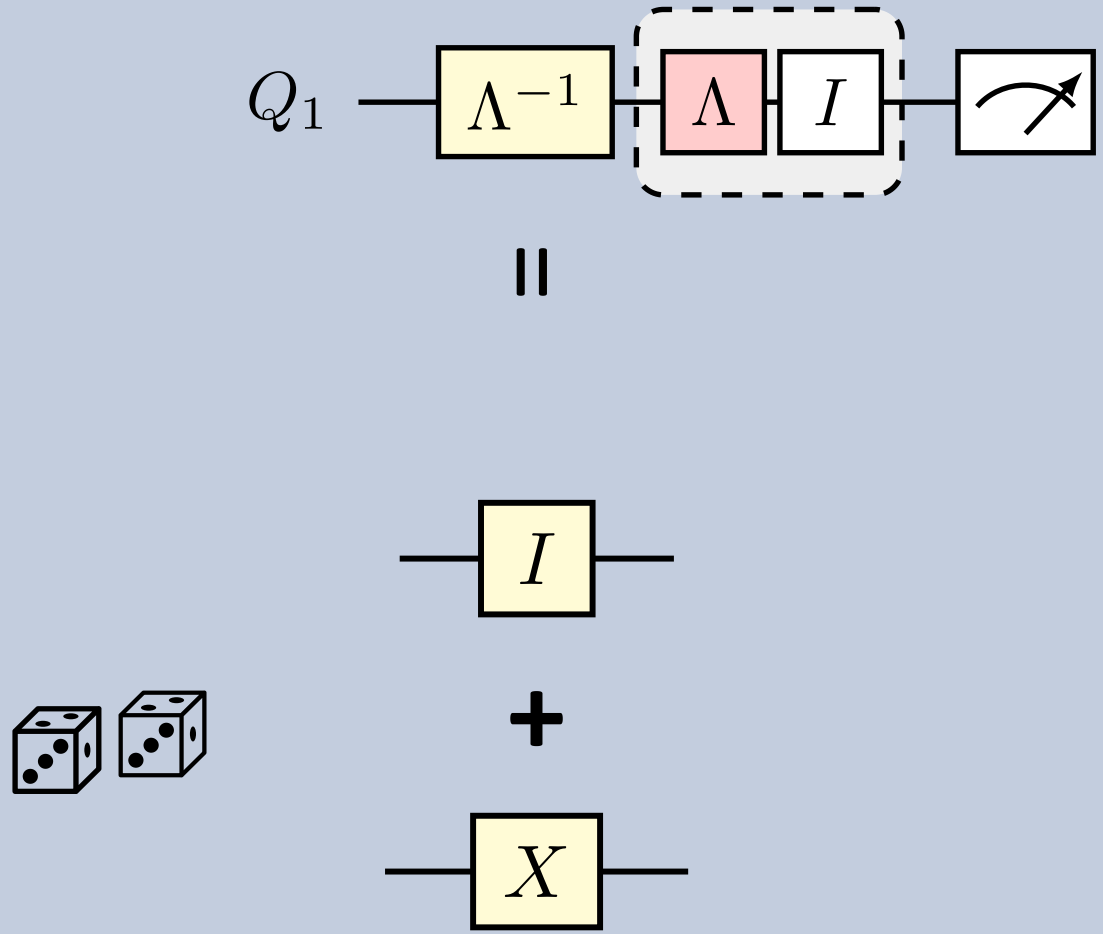
Cost: Variance

Sampling overhead

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



Canceling noise with noise



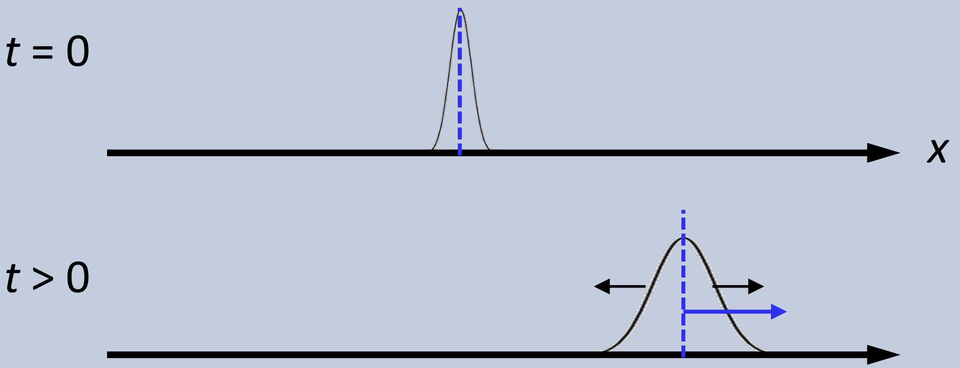
Canceling noise with noise: Drunkard's classical random walk analogy

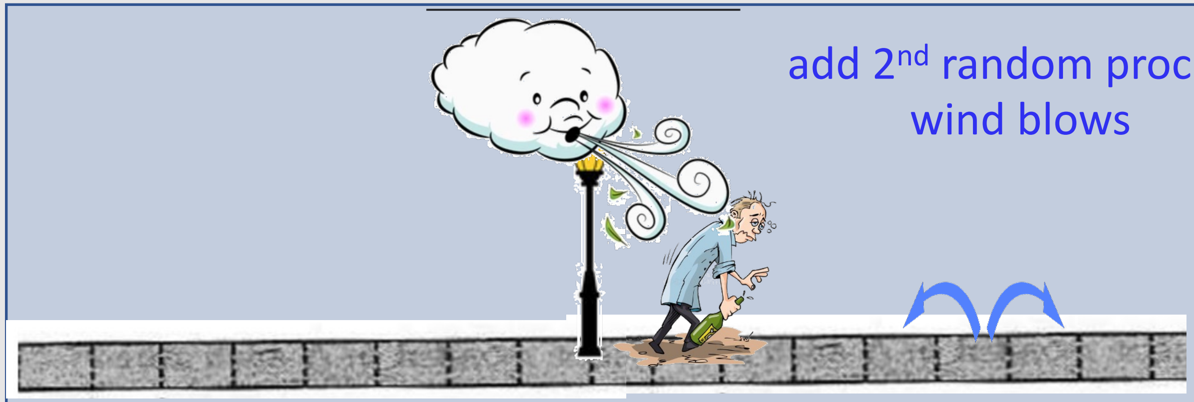
Random step

P(1 step left) = ½-p

P(1 step right) = ½+p

Distribution of random walk





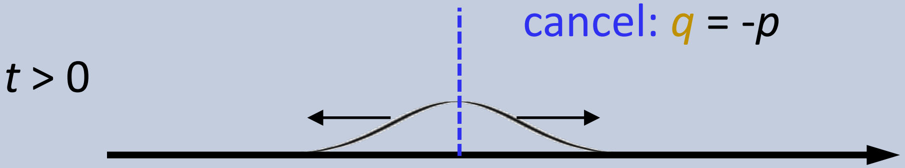
add 2nd random process
wind blows

P(1 step left) = ½-q

P(1 step right) = ½+q

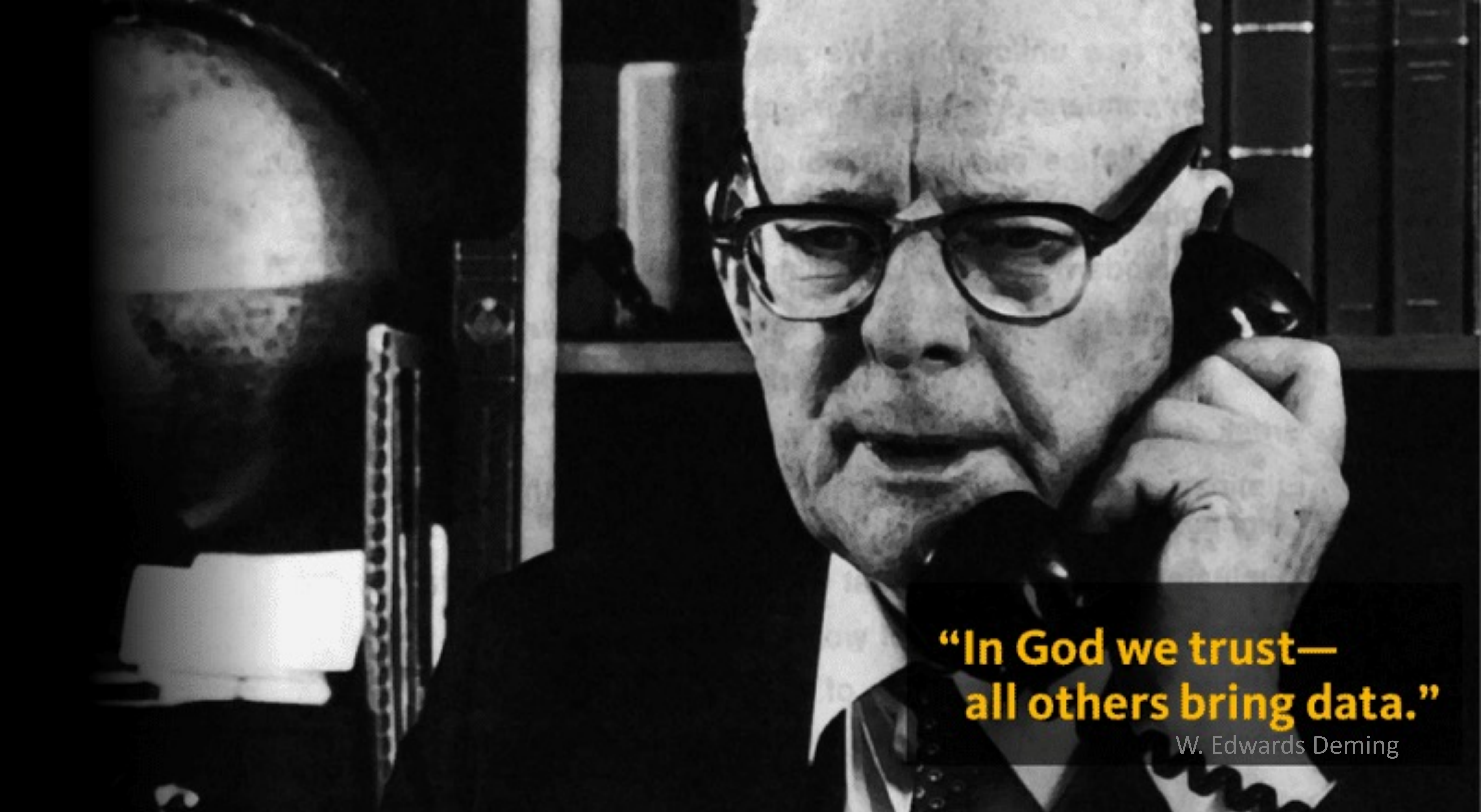
Distribution of random walk with wind

cancel: $q = -p$



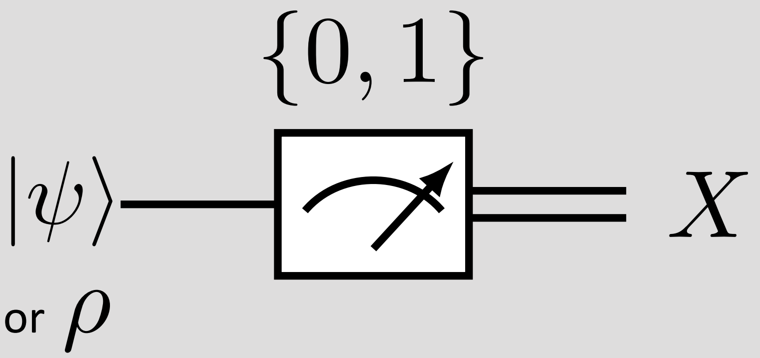
Gain: Bias-free estimate!

Cost: Variance



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

Summary: Qubit measured in the computational basis



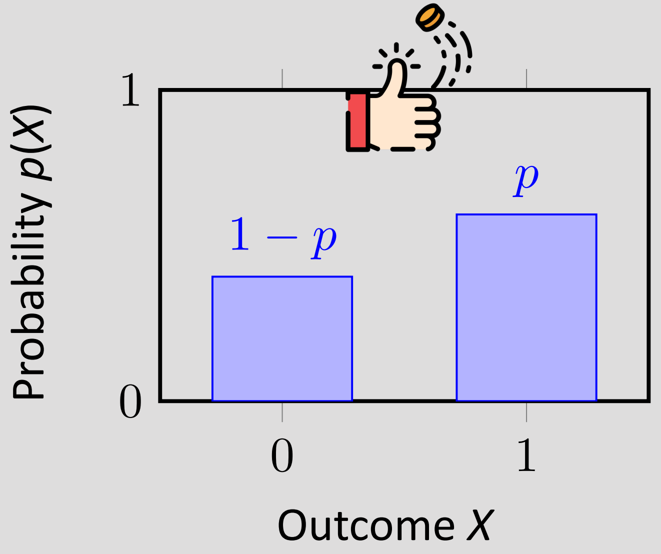
Set of possible outcomes

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

$$X \in \Sigma$$

Measurement operators

x	$\hat{\mu}(x)$
$0:$	$\hat{\mu}(0) = \|0\rangle \langle 0\|$
$1:$	$\hat{\mu}(0) = \|1\rangle \langle 1\|$



Probability to measure outcome

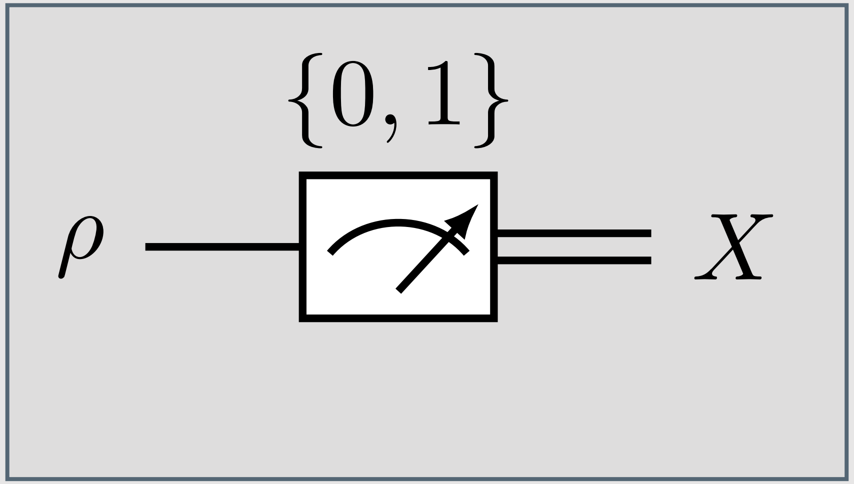
$$\begin{cases} X = 0: & p(X=0) = \mathrm{Tr}\left(\hat{\mu}(0)^\dagger \rho\right) = \mathrm{Tr}\left(|0\rangle\langle 0| \rho\right) = \frac{1}{2}\left(1 + \langle Z \rangle\right) \\ X = 1: & p(X=1) = \mathrm{Tr}\left(\hat{\mu}(1)^\dagger \rho\right) = \mathrm{Tr}\left(|1\rangle\langle 1| \rho\right)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

Bernoulli distribution. Single shot outcome follows a Bernoulli distribution:

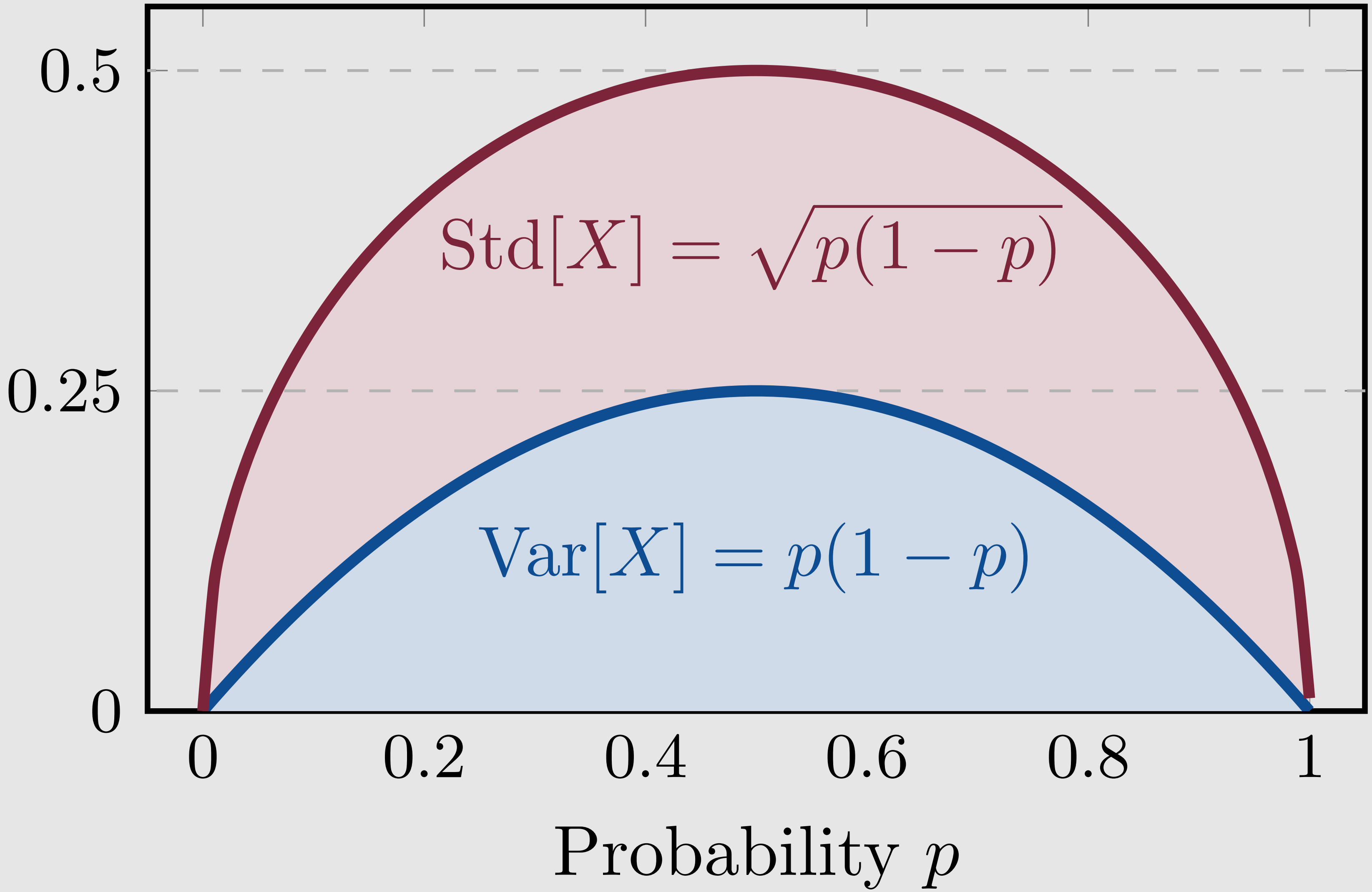
$$X \sim \mathrm{Bernoulli}\,(p) \qquad p := \mathrm{Tr}\left(|1\rangle\langle 1| \rho\right) \in [0, 1]$$

$$\mathrm{E}[X] = p \qquad \mathrm{Var}[X] = p(1-p)$$

Quantum projection noise for a single shot

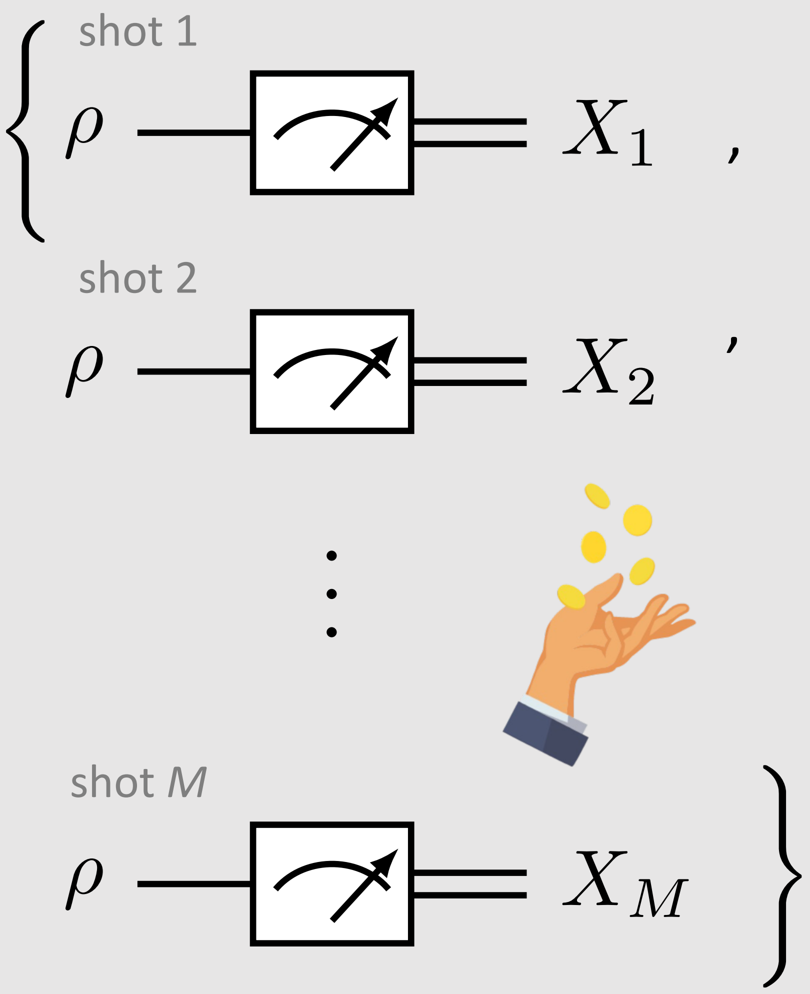


Variance of the random classical variable vs. probability to obtain 1



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

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

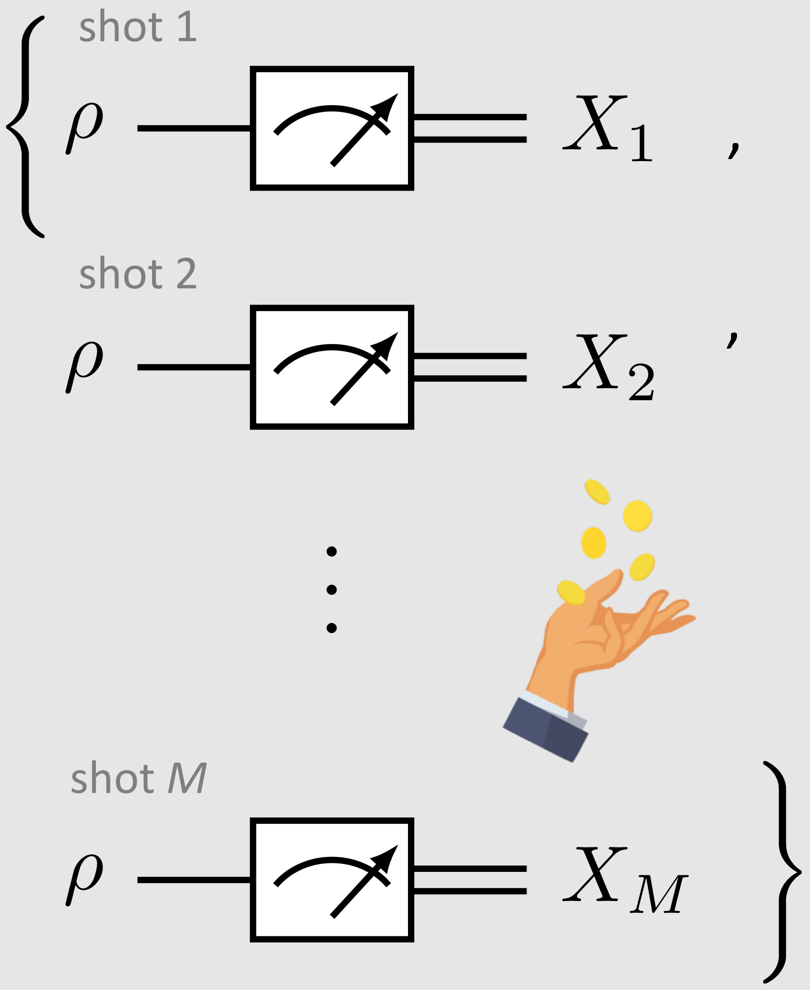
$$X_1, X_2, \ldots, X_M \sim \Pr\left[X = x\right] = \left\langle \hat{\mu}\left(x\right) \right\rangle$$

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

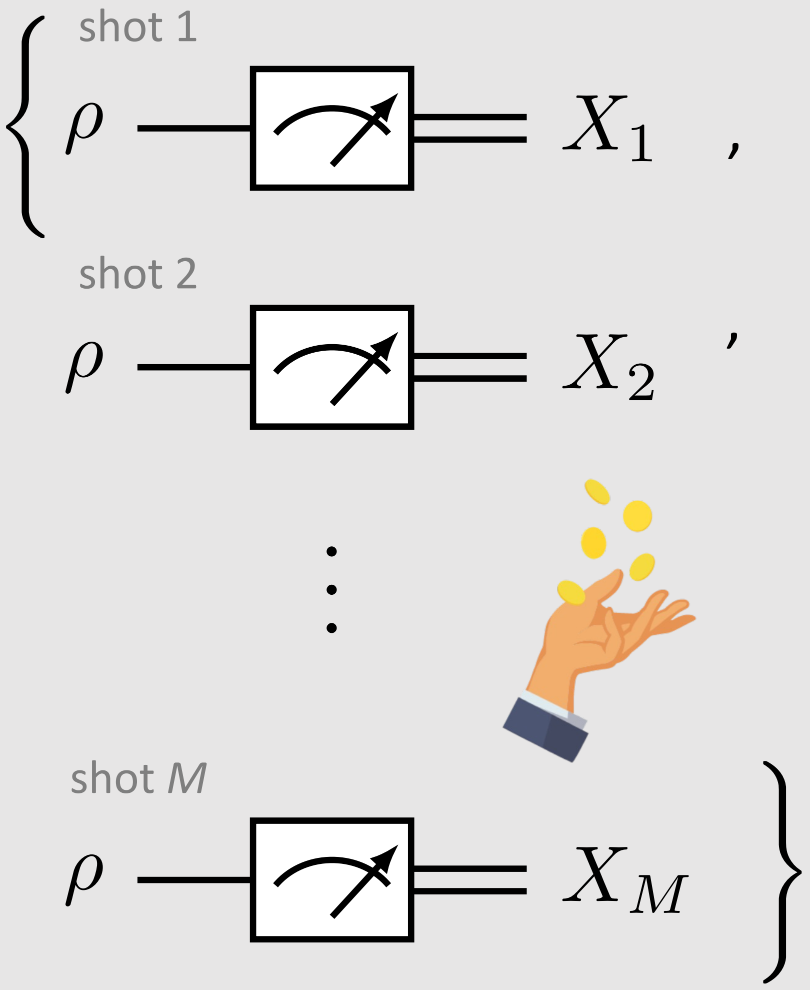
$$= \left\langle \hat{M} \right\rangle$$

relation to quantum operator we derived earlier (unbiased estimator)

$$= p$$

relation to probability to measure 1

How noisy is our estimate of the empirical mean?



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

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

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

$$= \frac{1}{M}\mathbb{V}[X]$$

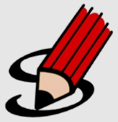
$$= \frac{p(1-p)}{M}$$

$$\sigma_S = \sqrt{\mathbb{V}[S]}$$

$$= \sqrt{\frac{\mathbb{V}[X]}{M}}$$

$$= \sqrt{\frac{p(1-p)}{M}}$$

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)

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

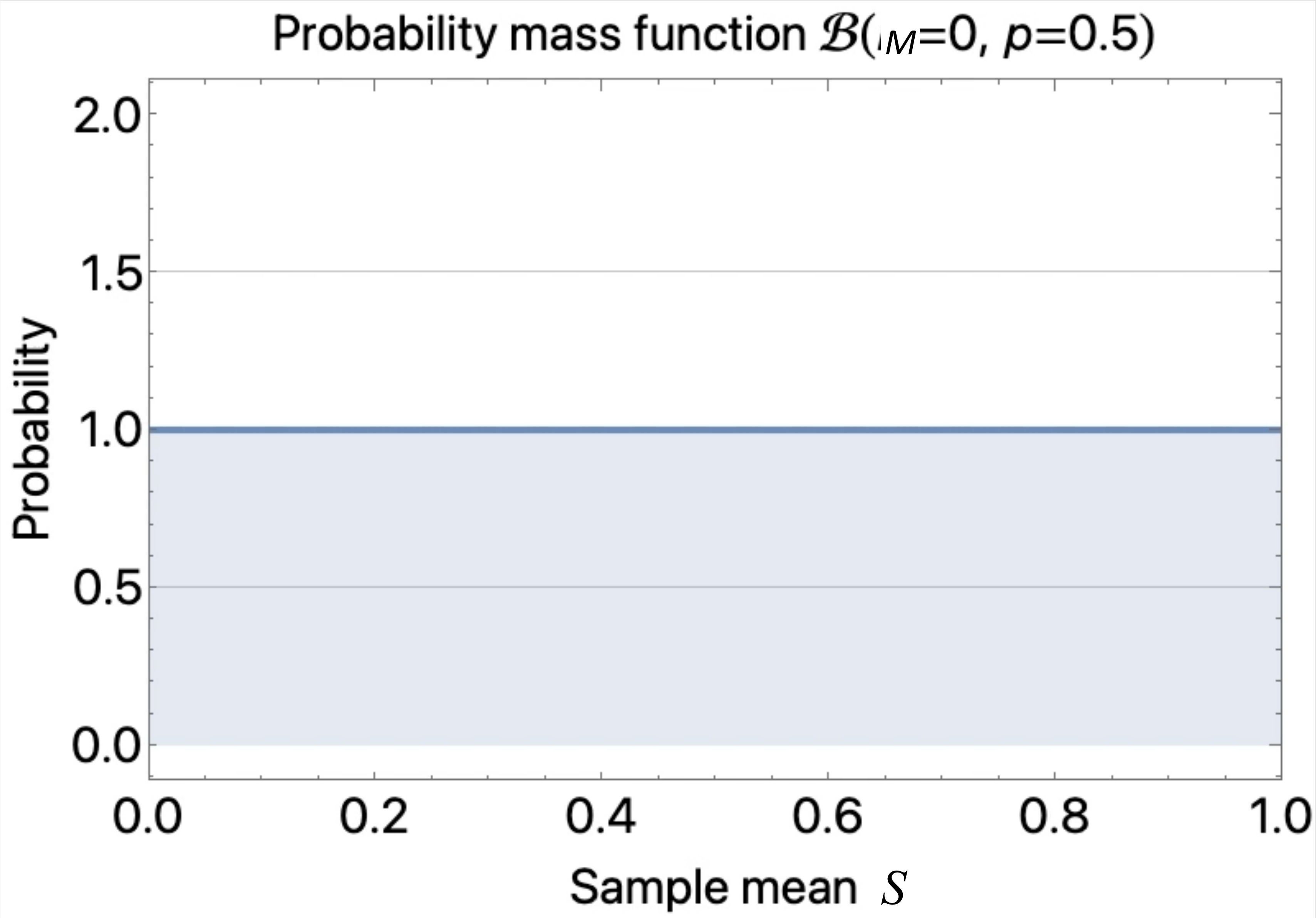
The variance is reduced by the number of sampler we take!

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

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

Hand Image by iconicbestiary on Freepik

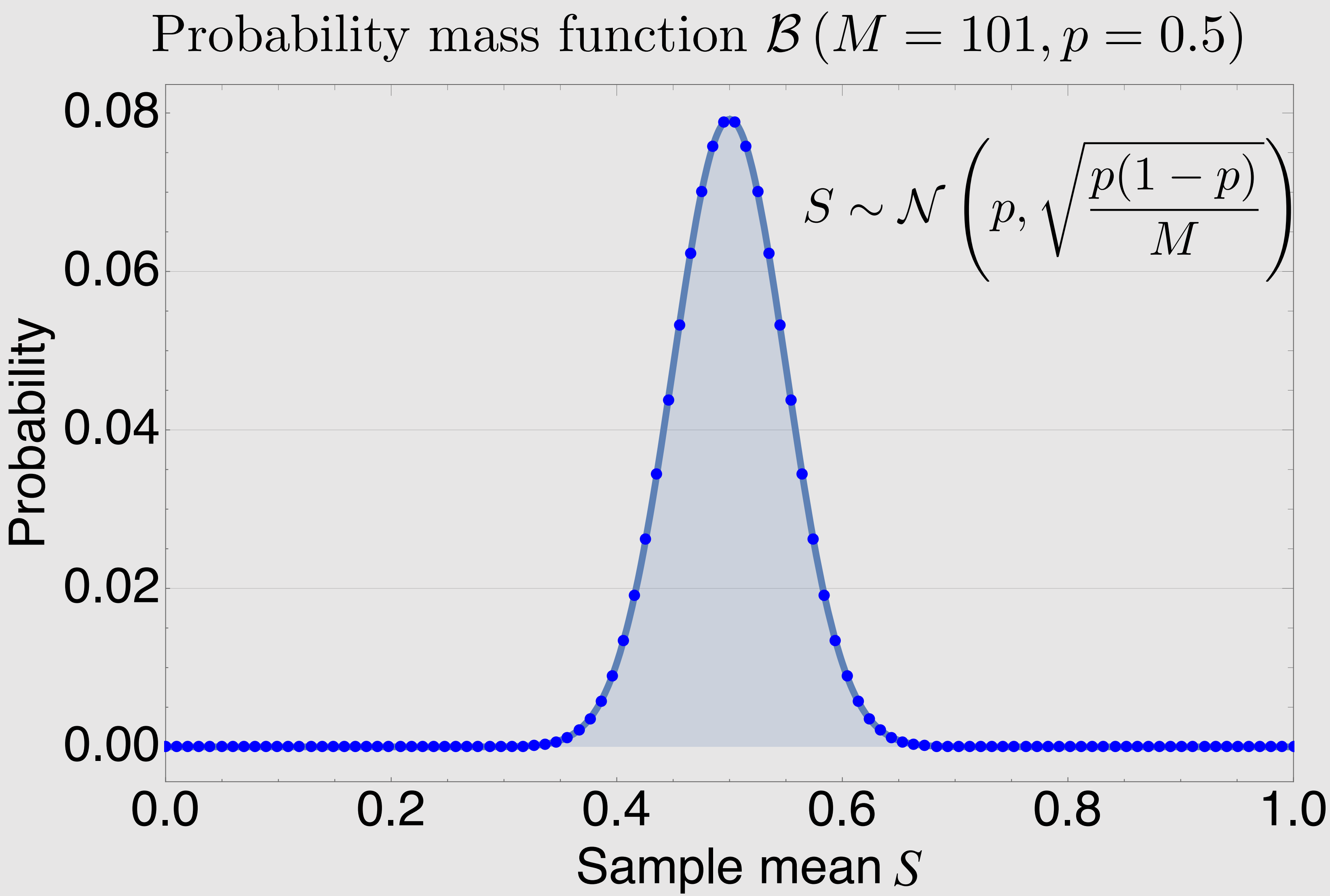
Animation of convergence of shots expectation value and mean



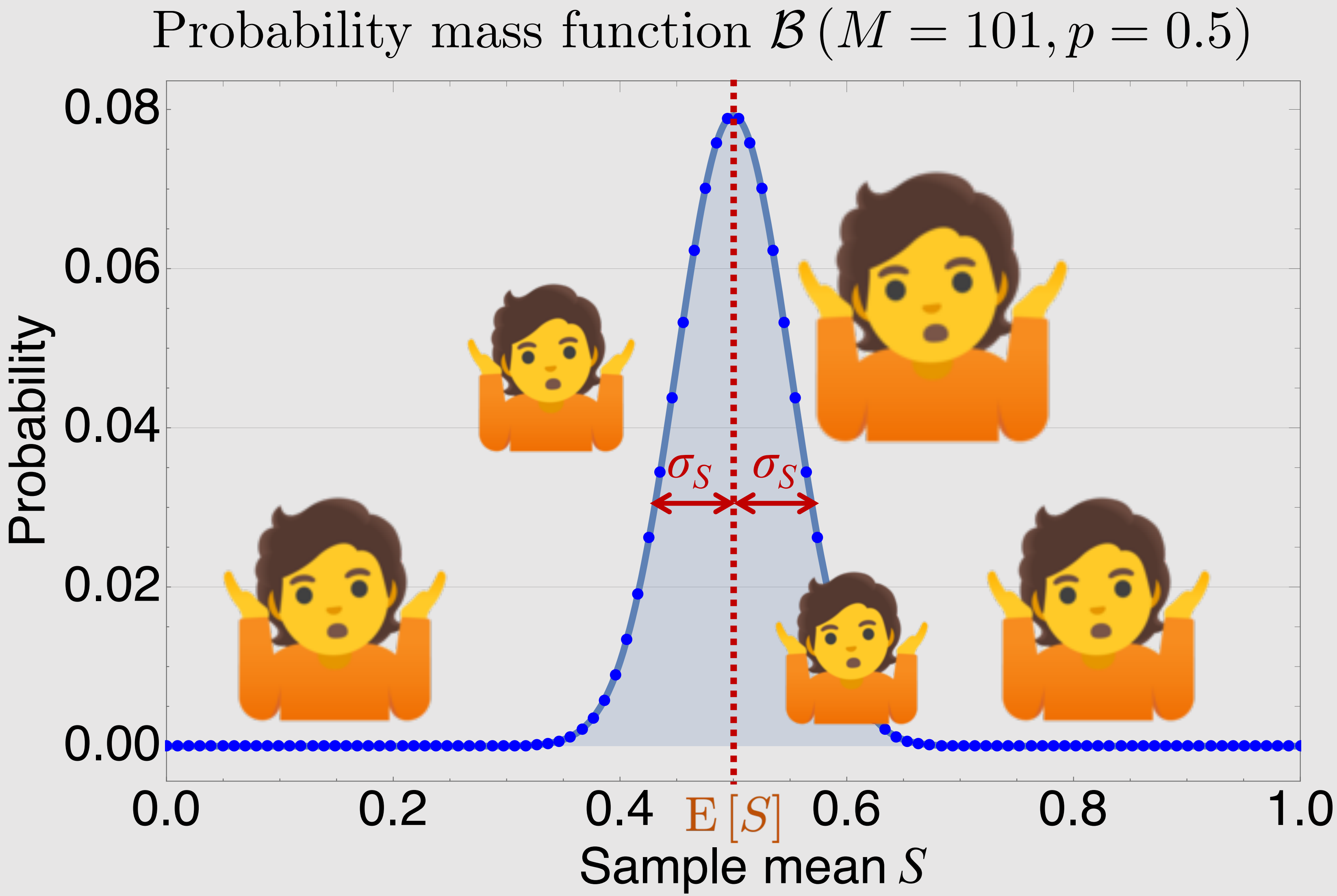
Binomial distribution

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

Sampled output distribution



Properties of the output distribution

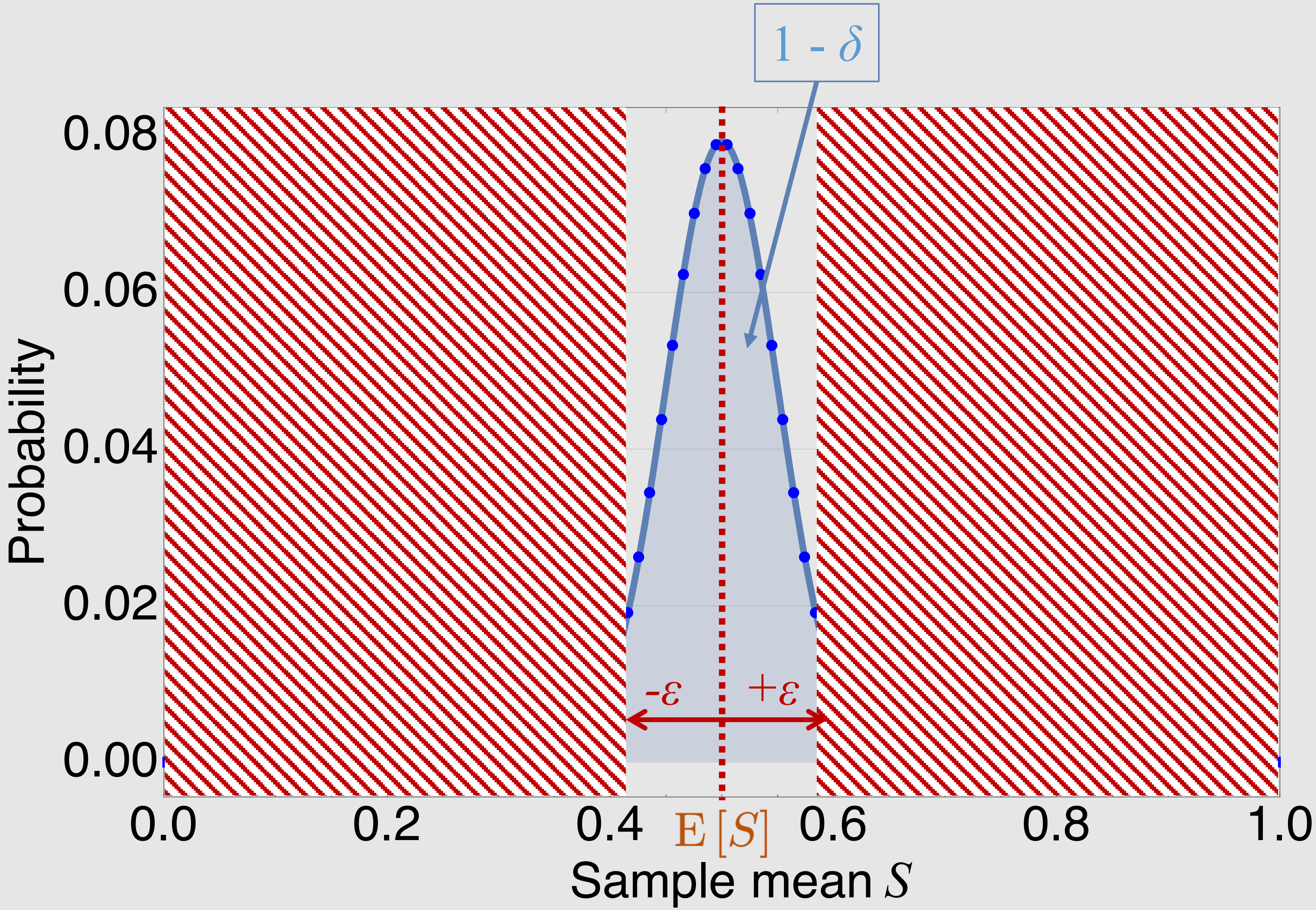


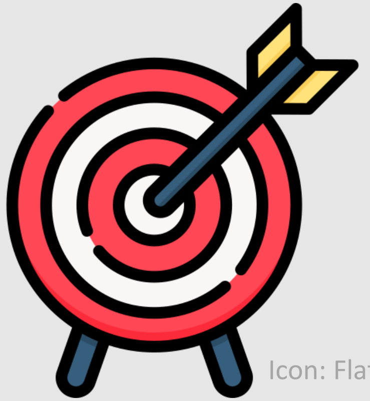
$$\mathrm{E}\,[S] = p$$

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

person: facebook

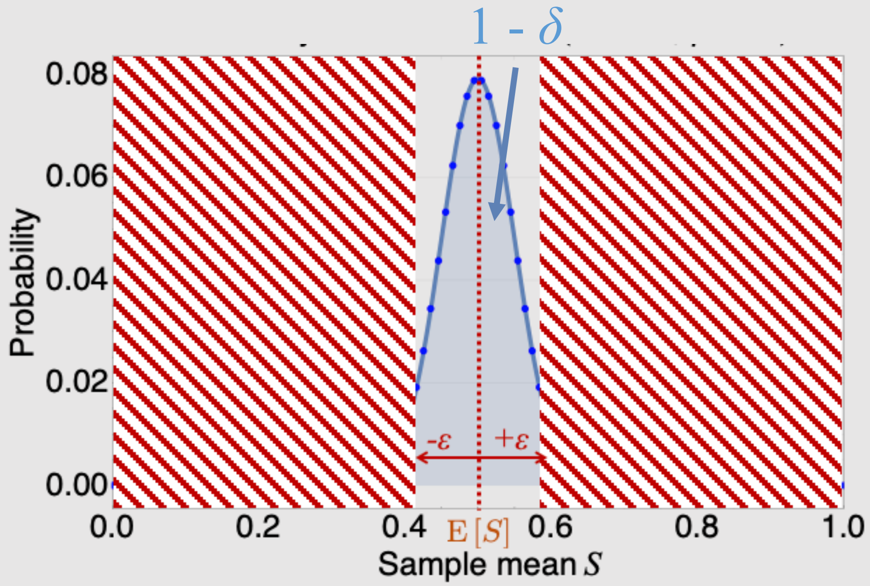
Concentration Measure for Sampling Expectation Values





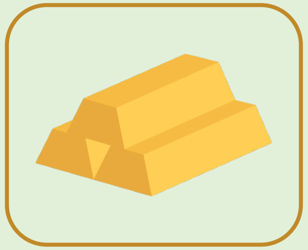
Icon: Flaticon

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)]

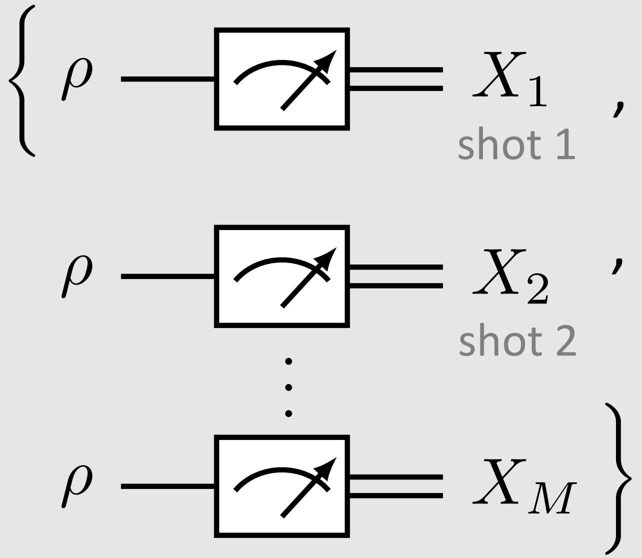
$$M \geq 2\epsilon^{-2}\log\left(2\delta^{-1}\right) \qquad \left[M \gtrapprox 4\epsilon^{-2}\right] \; .$$

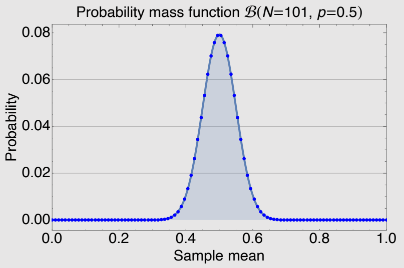
* Can find even tighter bound here owing to smaller [0,1] range

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

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

Ideal single qubit measurement with M shots



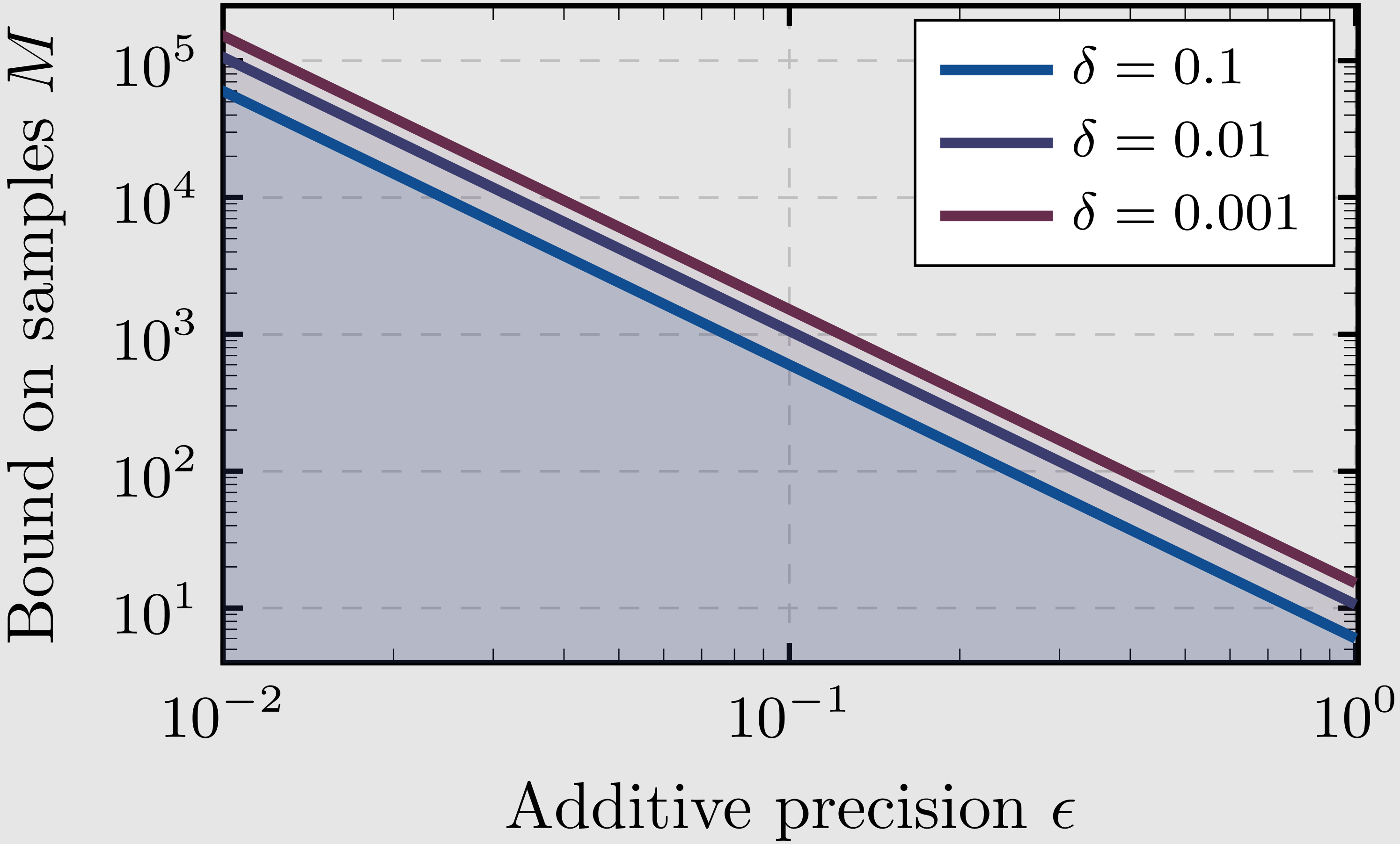


Observe that the **probability** δ is much cheaper than the **precision** ε.
Observe, n is not part of the equation.

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

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 $\Pr(|\cdots| \leq \cdots)$. 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	$X \geq 0$	$\Pr[X \geq a] \leq \frac{\mathbb{E}[X]}{a}$ $\forall a > 0$	$\Pr[X \geq kE[X]] \leq \frac{1}{k}$ $k > 1$	[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)	$\Pr[X - \mathbb{E}[X] \geq a]) \leq \frac{\sigma^2}{\sigma^2 + a^2}$	$\forall a > 0,\ \sigma^2 = \mathrm{Var}[X]$	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 $S_n = X_1 + \cdots + X_n$ $X_i \in [a_i, b_i]$ $\forall i$	$\Pr[S_n - \mathbb{E}[S_n] \geq t] \leq \exp\left(\frac{-2t^2 n^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$		[1, Sec. 3.5]
(two-sided)[4]	(same as above)		$\Pr[\lvert S_n - \mathbb{E}[S_n]\rvert > t] \leq 2\exp\left(\frac{-2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$	$\forall t \in (0, \frac{1}{2})$	[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	$\Pr\left[\left\lvert \frac{S_n}{n} - \mu \right\rvert \geq \epsilon\right] \leq 2\exp(-2n\epsilon^2)$	$\forall \epsilon > 0$	[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 $\mathbb{E}[\bar{X}] \pm t$.

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

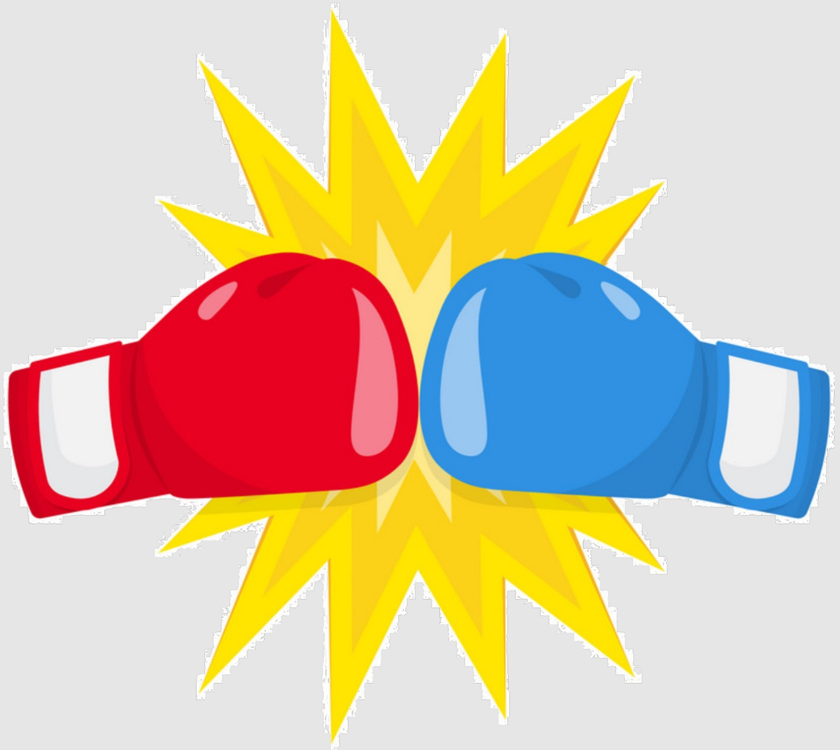


image: makc76

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

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

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

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

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

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

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

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

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

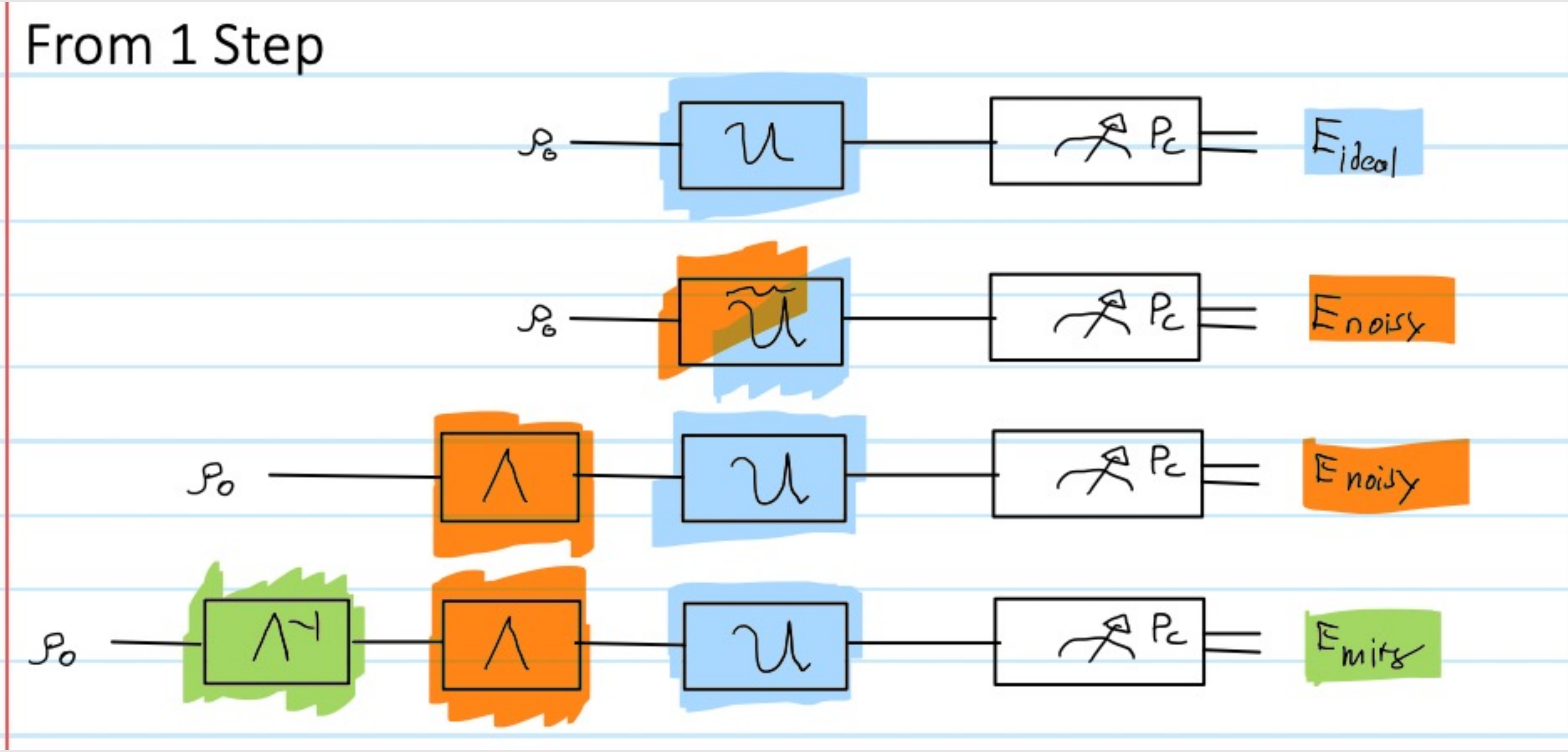
$$\sum_{a\in\Gamma} p_a = 1 \,,$$

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

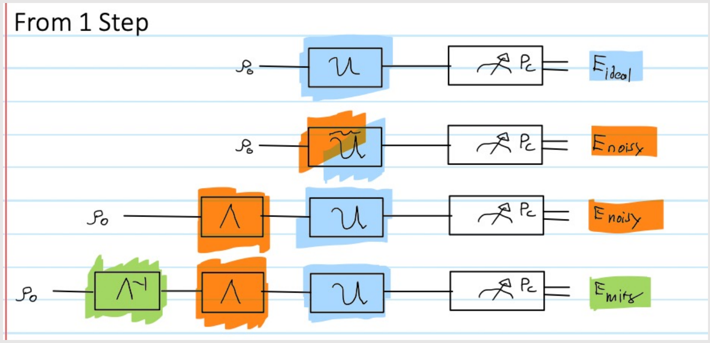
$$\Lambda(P_b) = f_b P_b \,, \quad \forall a \in \Gamma \,,$$

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

Probabilistic error cancelation: Derivation



Probabilistic error cancelation: Derivation



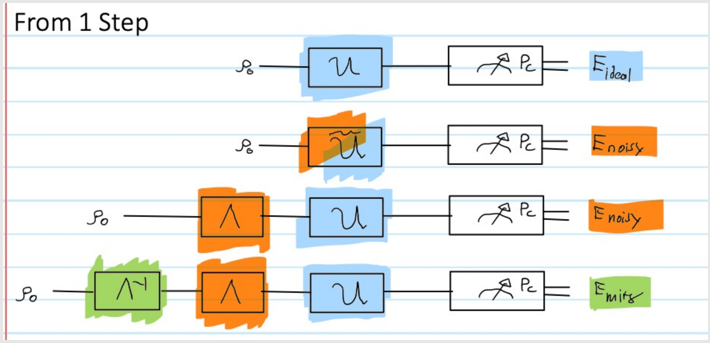
Channel definitions

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

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

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

Probabilistic error cancelation: Derivation



Channel definitions

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

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

Need

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

$-1 \le f_a \le 1$

$= \sum_b c_b P_b \cdot$

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

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

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

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

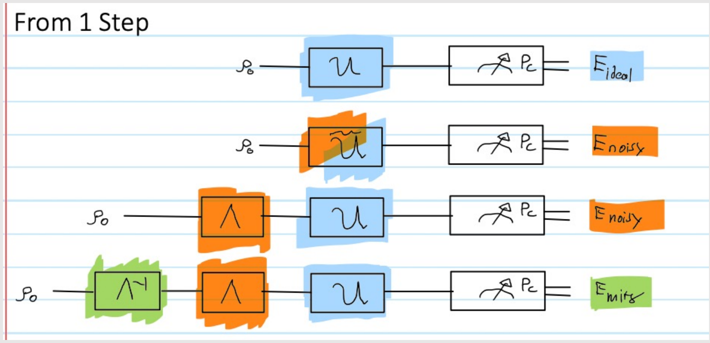
$= \sum_b c_b^{inv} P_b \cdot$

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

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

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

Probabilistic error cancelation: Derivation



Channel definitions

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

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

Need

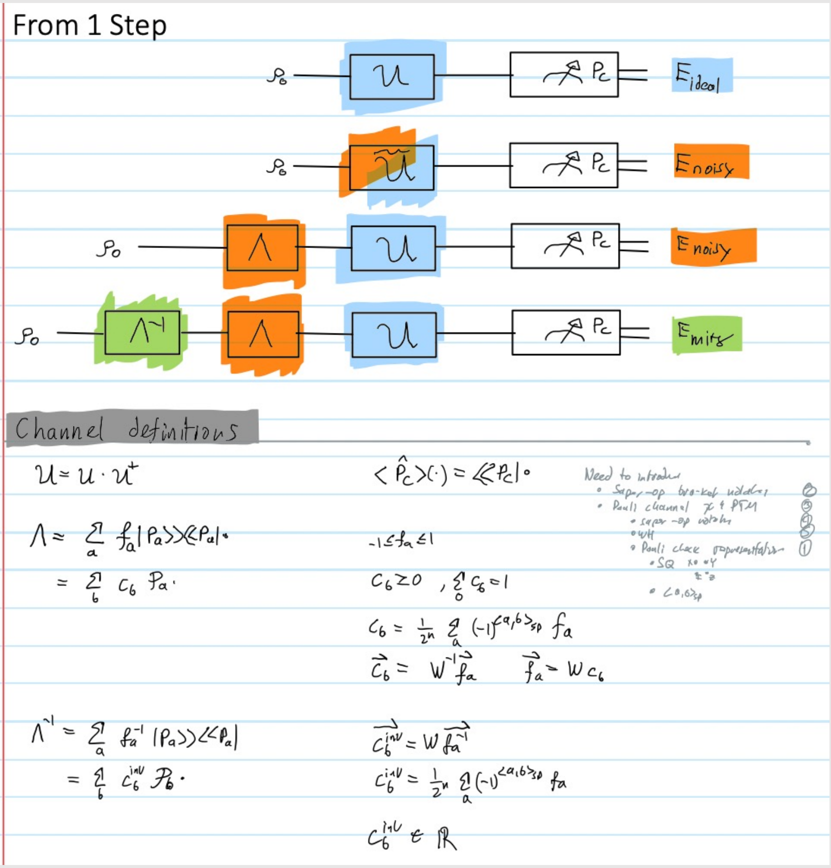
$$\Lambda = \sum_a f_a |P_a\rangle\rangle\langle\langle P_a| \cdot \qquad -1 \le f_a \le 1$$

$$= \sum_b c_b \mathcal{P}_a \cdot \qquad c_b \ge 0, \ \sum_0 c_b = 1$$

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

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

Circuit expectation values



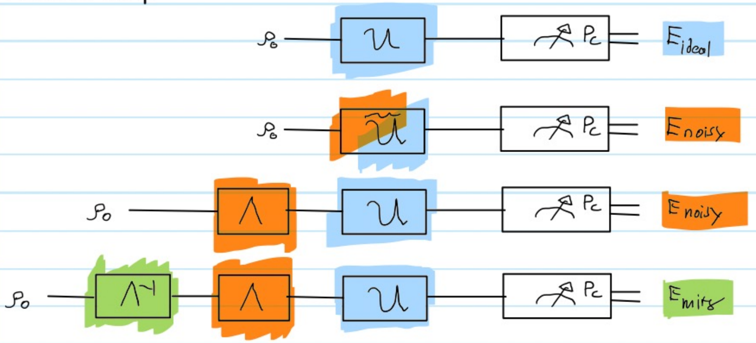
$$E_{ideal} := \langle \hat{P}_c \rangle_{\mathcal{U}}$$

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

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
- Pauli channel χ + PTM
- super-op notation
- W-H
- Pauli class representation
- $\langle a,b \rangle_{sp}$

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

$= \sum_b c_b \mathcal{P}_a \cdot$ $\quad c_b \ge 0, \; \sum_b c_b = 1$

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

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

$\Lambda^{-1} = \sum_a f_a^{-1} |P_a\rangle\rangle\langle\langle P_a|$ $\quad \vec{c}_b^{\,inv} = W \vec{f_a^{-1}}$

$= \sum_b c_b^{inv} \mathcal{P}_b \cdot$ $\quad c_b^{inv} = \frac{1}{2^n} \sum_a (-1)^{\langle a,b \rangle_{sp}} f_a$

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

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

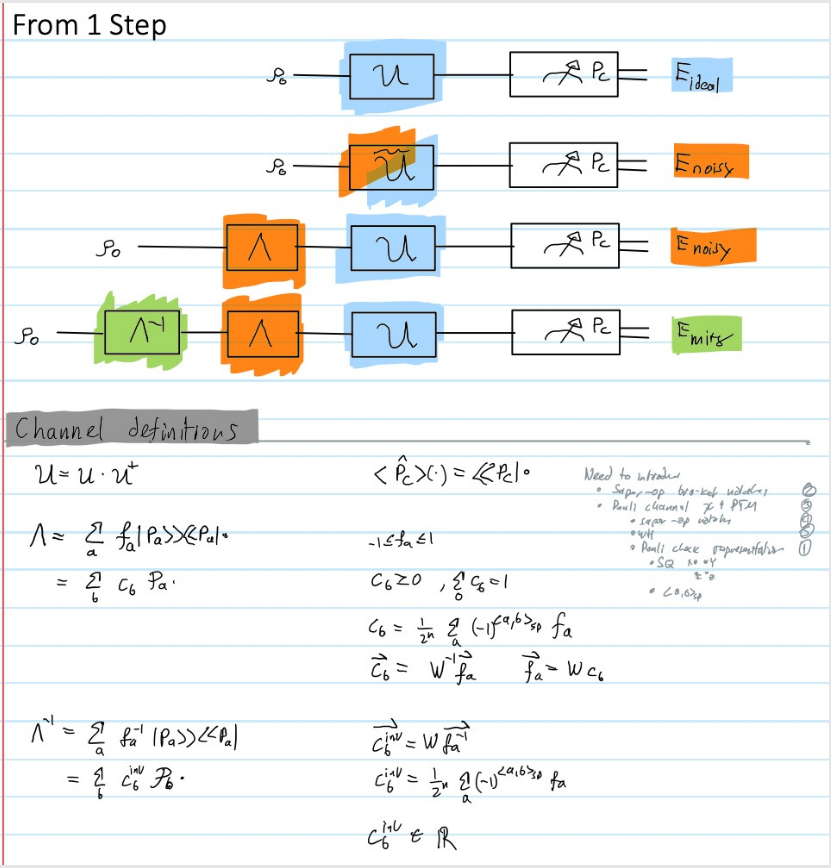
$= \langle\langle P_c | \mathcal{U} | \rho_0 \rangle\rangle$ — ideal exp value with noiseless unitary

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

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

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

Circuit expectation values



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

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

ideal exp value with noiseless unitary

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

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

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

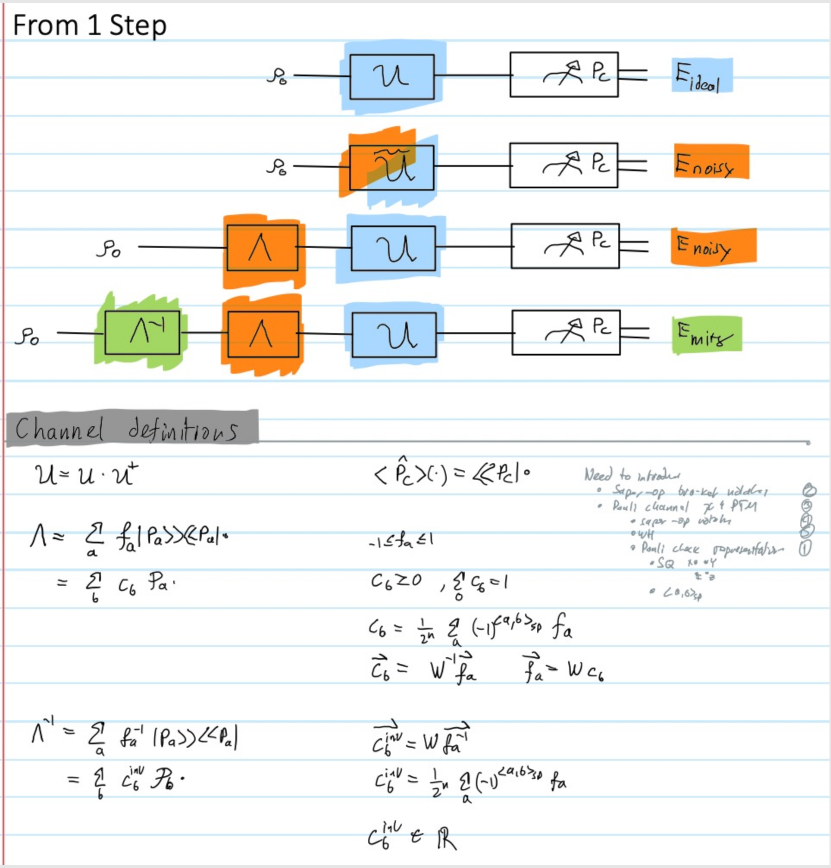
noisy-gate expectation value

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

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

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

Circuit expectation values



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

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

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

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

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

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

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

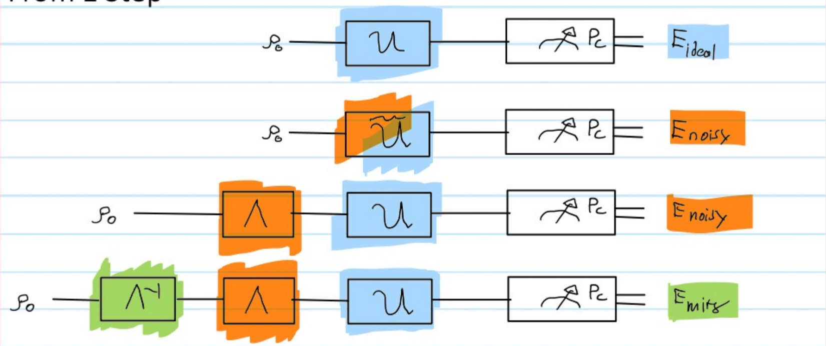
sum of trajectories with weight $c_b^{inv} \in \mathbb{R}$

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

classical post process

quantum

From 1 Step



Channel definitions

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

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

$\Lambda = \sum_a f_a |P_a\rangle\rangle\langle\langle P_a| \cdot$
$= \sum_b c_b \mathcal{P}_a \cdot$

$-1 \le f_a \le 1$

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

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

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

$\Lambda^{-1} = \sum_a f_a^{-1} |P_a\rangle\rangle\langle\langle P_a|$
$= \sum_b c_b^{inv} \mathcal{P}_b \cdot$

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

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

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

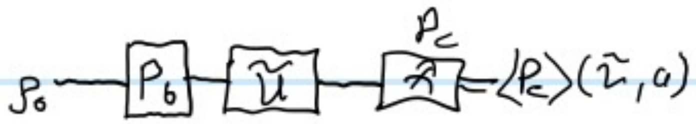
Need to introduce
- Super-op bra-ket notation (2)
- Pauli channel χ + PTM (3)
  - super-op notation (4)
  - WH (5)
  - Pauli clock representation (1)
    - SQ X Y Z
  - $\langle a,b \rangle_{sp}$

$$E_{mitg} := \langle\langle P_c | \mathcal{U} \Lambda \Lambda^{-1} | \rho_0 \rangle\rangle$$
$$= \langle\langle P_c | \mathcal{U} \Lambda \Big(\sum_a f_a^{-1} |P_a\rangle\rangle\langle\langle P_a| \Big) |\rho_0\rangle\rangle$$
$$= \langle\langle P_c | \mathcal{U} \Lambda \Big(\sum_b c_b^{inv} \mathcal{P}_b \Big) |\rho_0\rangle\rangle$$
$$= \sum_b c_b^{inv} \langle\langle P_c | \mathcal{U} \Lambda \mathcal{P}_b | \rho_0 \rangle\rangle$$

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

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

classical post process — quantum



Quantum circuit we can execute on HW 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... cant 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 $\Lambda = (1-p)\, I\cdot I + p\, X\cdot X$

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

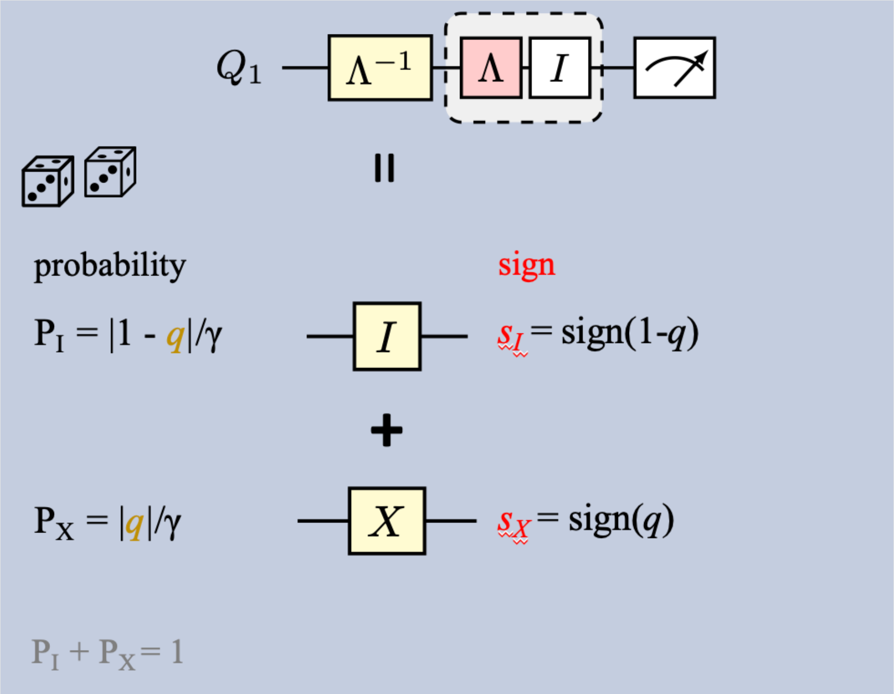
$c_I^{inv} = 1 + \frac{p}{1-2p}$ ($b=(0,0)$ choose vector) $\quad c_X^{inv} = \frac{-p}{1-2p}$ ($b=(1,0)$)

Turn into probabilities

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

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

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



E mitigated

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

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

$$= \gamma \sum_b \mathrm{sgn}(C_b^{inv}) \bar{C}_b^{inv} \langle \hat{P}_c \rangle (\hat{\mathcal{U}}, b)$$

scale — classical post-processing — 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 $\Lambda = (1-p) I \cdot I + p X \cdot X$

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

$C_I^{inv} = 1 + \frac{p}{1-2p}$ ($b = (0,0)$ choice vector) $\quad C_X^{inv} = \frac{-p}{1-2p}$ ($b = (1,0)$)

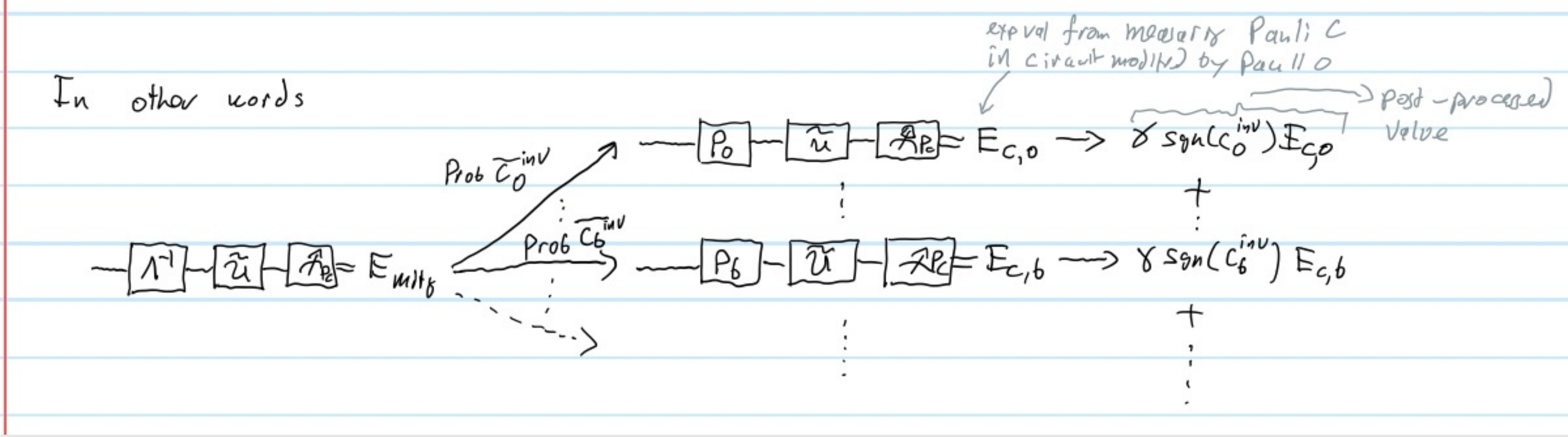
Turn into probabilities

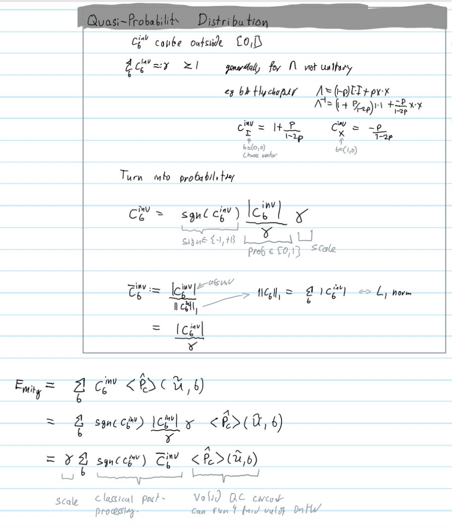
$$C_b^{inv} = \underbrace{\mathrm{sgn}(C_b^{inv})}_{\text{sign} \in \{-1,+1\}} \underbrace{\frac{|C_b^{inv}|}{\gamma}}_{\text{prob} \in [0,1]} \underbrace{\gamma}_{\text{scale}}$$

$$\bar{C}_b^{inv} := \frac{|C_b^{inv}|}{\|C_b^{inv}\|_1} \quad \text{(abs val)} \qquad \|C_b\|_1 = \sum_b |C_b^{inv}| \Leftrightarrow L_1 \text{ norm}$$

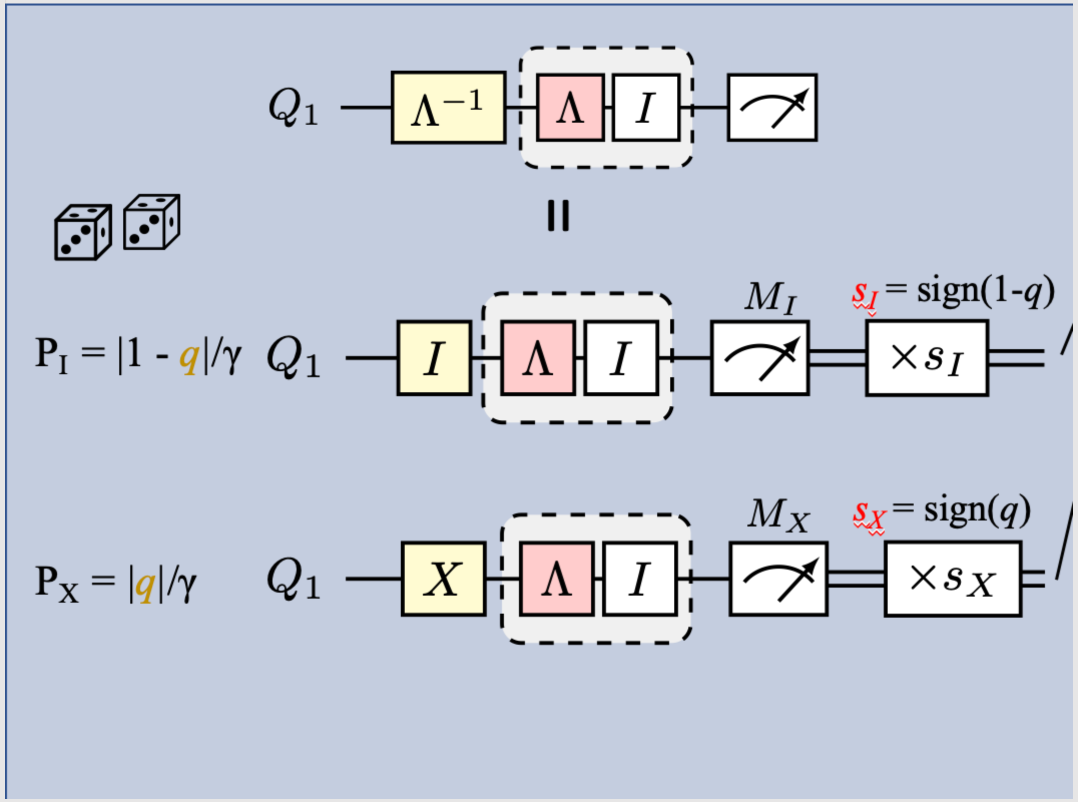
$$= \frac{|C_b^{inv}|}{\gamma}$$

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.





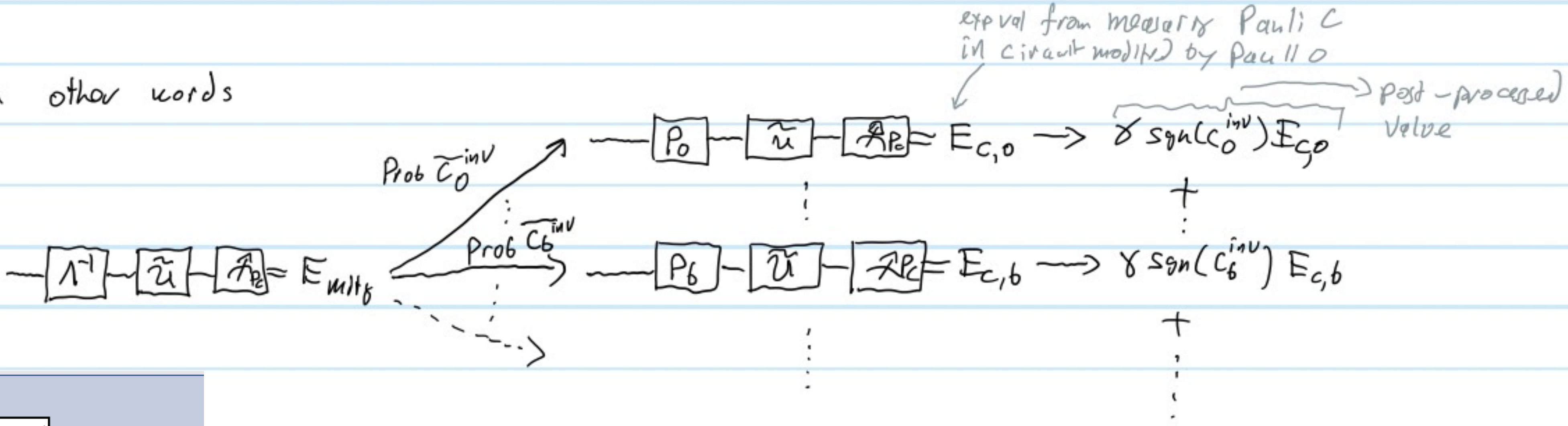
In this form, the decomposition of the error error mitigated expectation value is simply a sum over expectation values of Pauli-gate modified circuits, whose value can be obtained from direct quantum computer execution, weighted by a probability, c bar b inverse, and the sign.
The elements that perform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



Estimator

post processing:

In other words

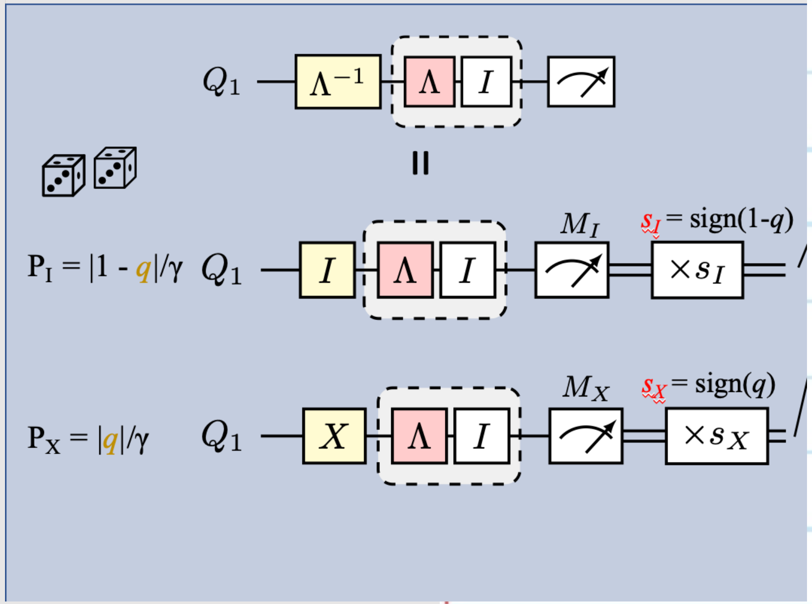


$$E_{c,mit} = \sum_b \gamma\, sgn(C_b^{inv})\, E_{c,b}$$

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

$$\left\{ \bar{C}_b^{inv} : -[P_b]-[\tilde{\mathcal{U}}]-[\bar{\mathcal{X}}P_c]= Y \in \{-1,+1\} \longrightarrow X = \text{sgn}(C_b^{inv})\, Y \right\}$$

$\bar{C}_b^{inv}$: prob

Y: Pauli observable single-shot outcome, ie e-vals.

X: post processed value X, $X \in \{+1,-1\}$

Let's say we sample M instances, randomly sampling a single value for b and obtaining 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

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

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 $\Lambda = (1-p)I\cdot I + pX\cdot X$

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

$C_I^{inv} = 1 + \frac{p}{1-2p}$ ($b=(0,0)$ choice vector) $\qquad C_X^{inv} = \frac{-p}{1-2p}$ ($b=(1,0)$)

Turn into probabilities

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

$$\bar{C}_b^{inv} := \frac{|C_b^{inv}|}{\|C_b^{inv}\|_1} = \frac{|C_b^{inv}|}{\gamma}$$

(absolute value) $\|C_b\|_1 = \sum_b |C_b^{inv}| \Leftrightarrow L_1$ norm

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

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.

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:

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

rand outcome of m-th shot for σth circ.

noisy circuit

Pauli chosen for m-th shot

There are now 2 random processes:

σ_m : which pauli σ we pick for shot m

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

Unbiased Estimator of the Ideal, noise-free circuit expectation

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

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

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

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

drop notation

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

$$= \sum_b \gamma \, \mathrm{sgn}(c_b^{inv}) \, \bar{c}_b^{inv} \, \mathbb{E}_Y[Y_b]$$

post-process of outcome; sample prob; rand outcome

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

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.

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

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

Probabilistic error cancelation: Derivation

the value Y_σ but only on the label σ:

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

γ

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

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

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

$$= \langle \hat{P}_c \rangle (\mathcal{U}_{ideal})$$

without noise!

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

(Optional step) Variance

Variance of E_M

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

with respect to

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

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

General derivation

Next lectures

Learning noise

State-of-art PEC experiments

Key techniques: Twirling

T-REX mitigation

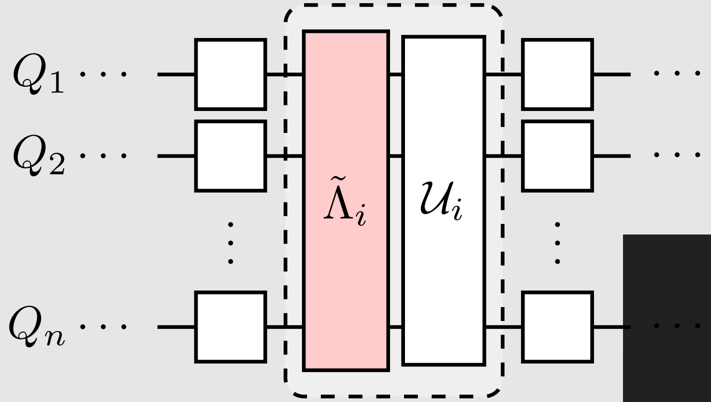
…

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

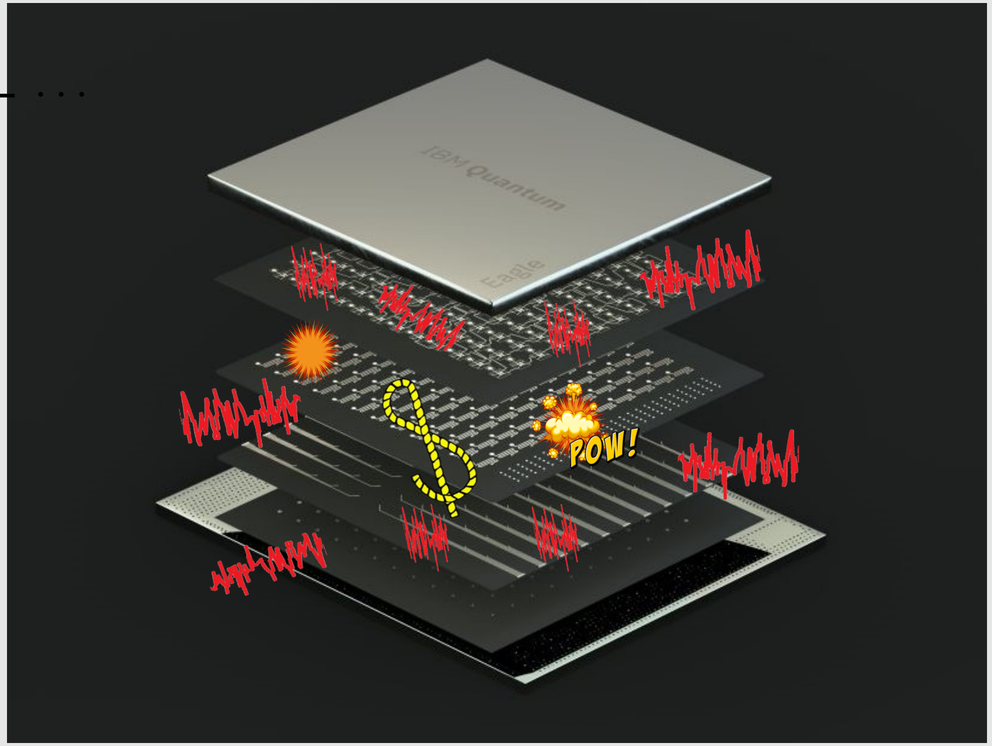
…

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

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



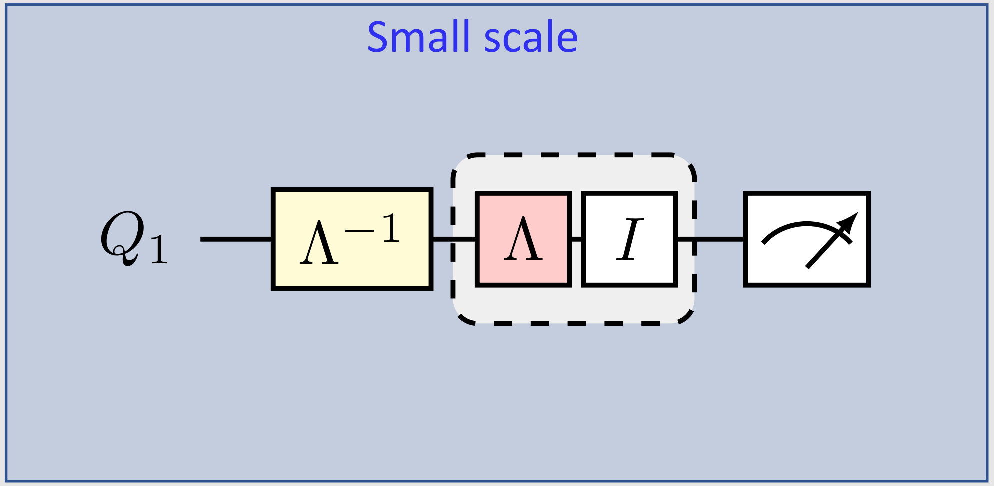
Energy relaxation T_1
Dephasing T_2
Coherent errors ZZ
Classical crosstalk
Quantum crosstalk
State preparation error
Measurement correlated errors
…



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

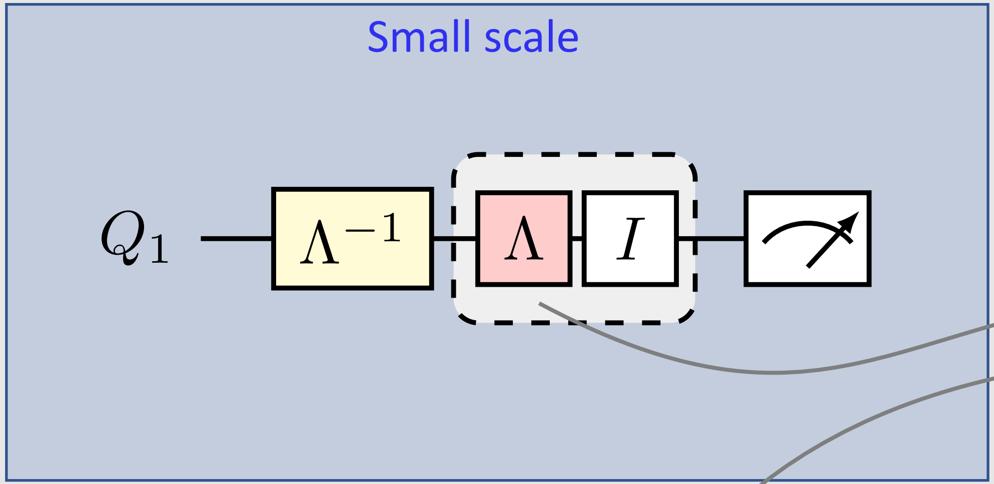


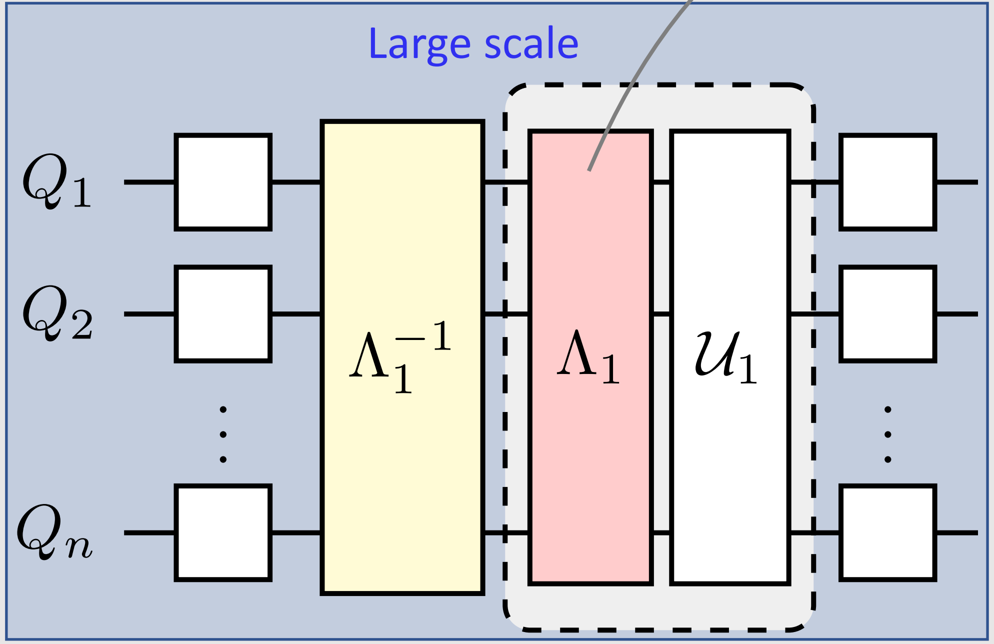
Critically hinges on knowing the full noise near perfectly

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

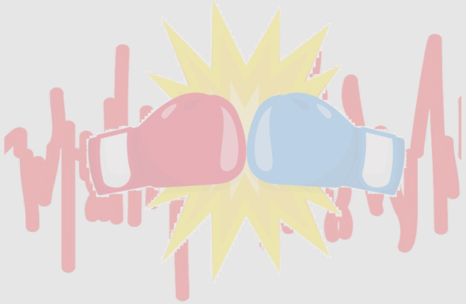
noise in full device

- cross-talk
- correlated errors
- parallel gates

Outline

Idea

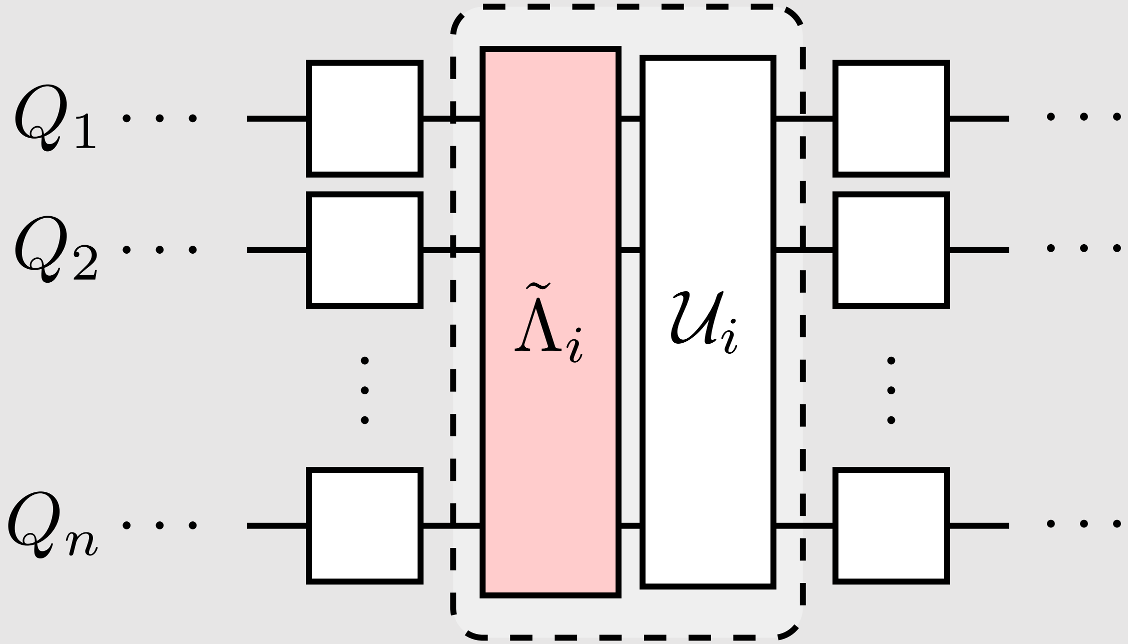
Learn



Cancel
(realization)

Icon: Flaticon

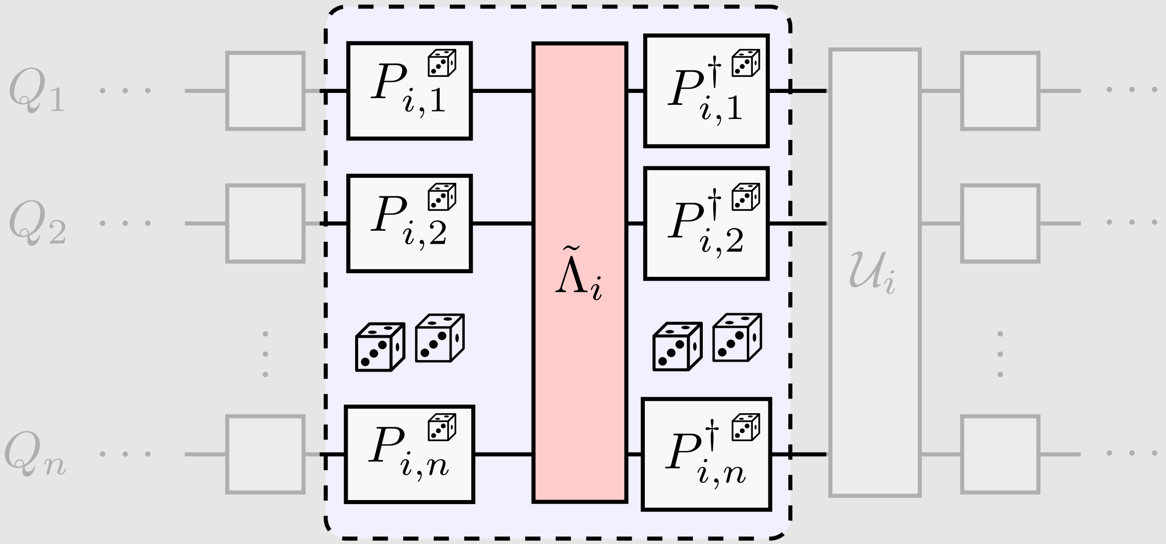
Step 1: Simplify the noise



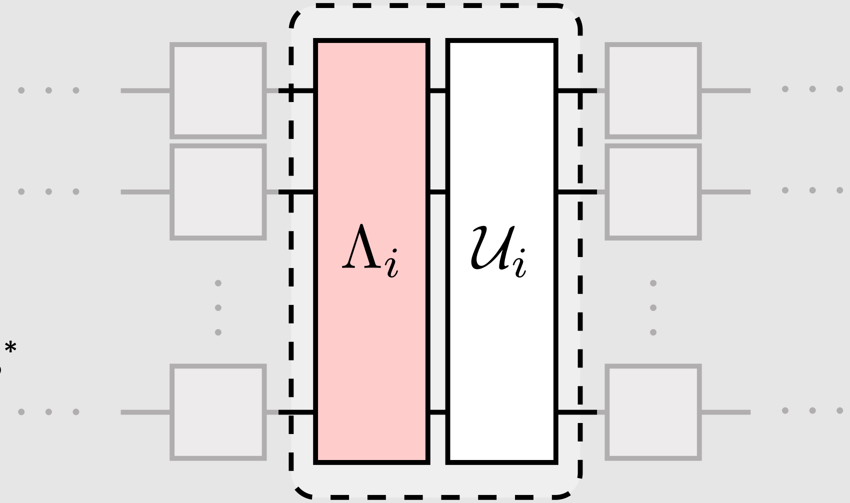
noise that includes cross-talk errors, etc.
characterized by some 4^n x 4^n matrix

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



average over twirl instances*



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

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

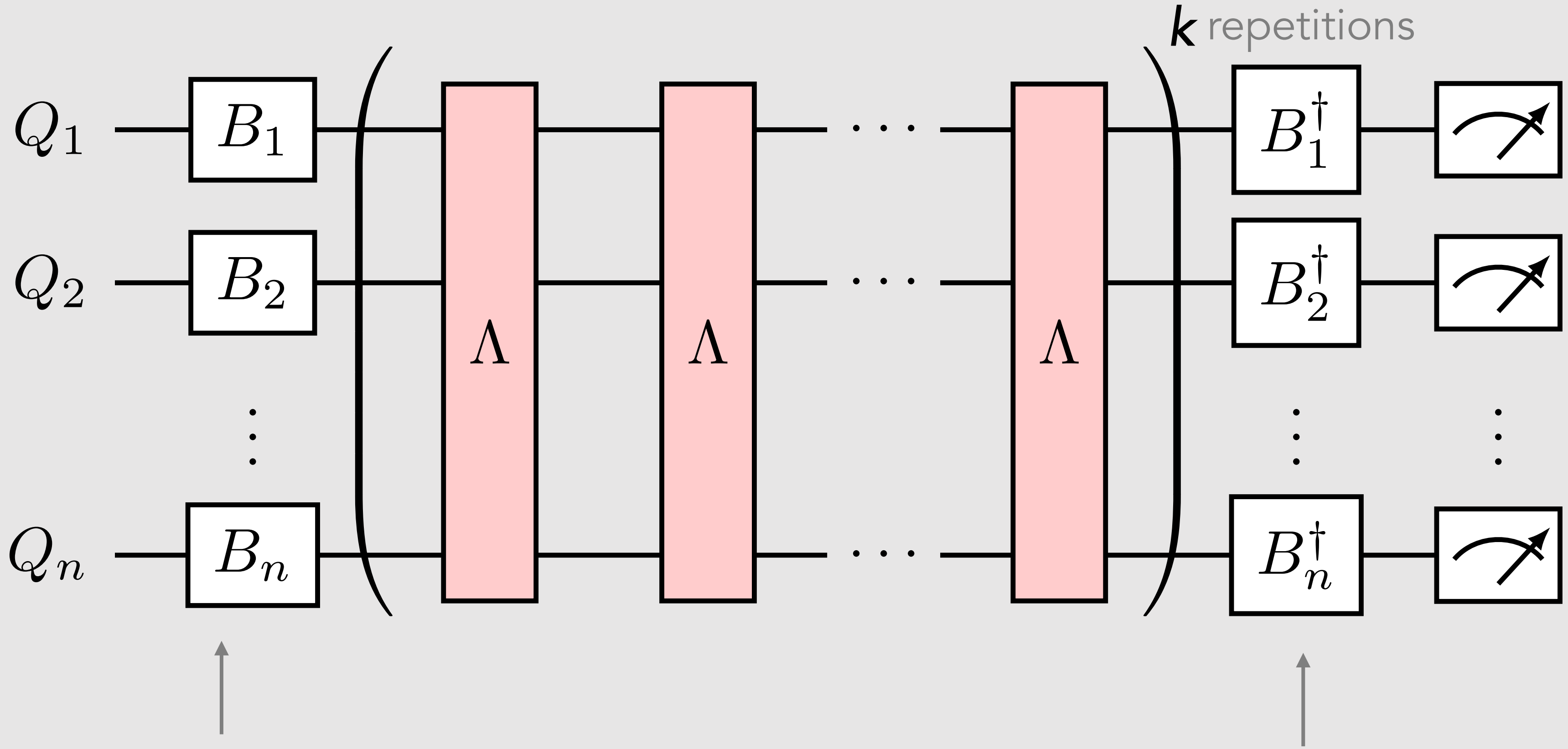
$$\Lambda(P_a) = f_a P_a$$

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

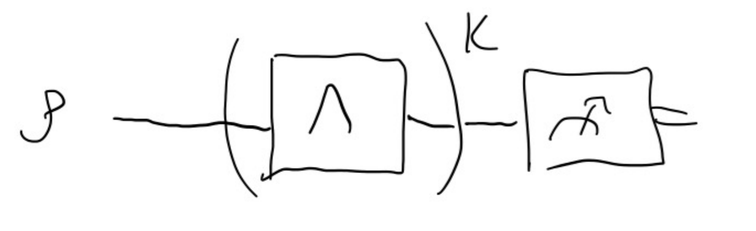
Akin to:

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:

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

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

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

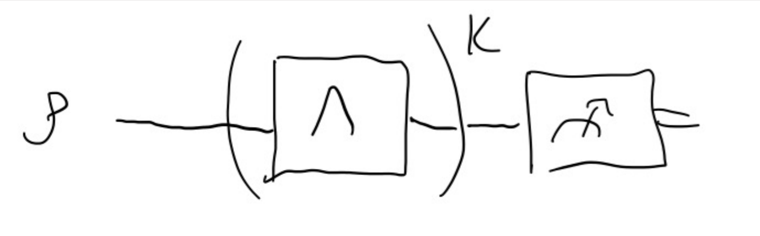
Pauli decomposition of posterior state

Since the channel is linear $$\Lambda(a\hat{A} + b\hat{B}) = a\Lambda(\hat{A}) + b\Lambda(\hat{B})$$

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

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

Just need to know action on P_b



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

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

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

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

just need to know action on P_b

$$= \rho'$$

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

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

orthogonal

Solve for ρ'_c by equate & orthogonal use:

$$Tr\left(P_c^\dagger \frac{1}{2^n}\sum_b \rho_b \Lambda(\hat{P}_b)\right) = Tr\left(P_c^\dagger \frac{1}{2^n}\sum_{c'} \rho'_{c'} \hat{P}_{c'}\right)$$

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

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

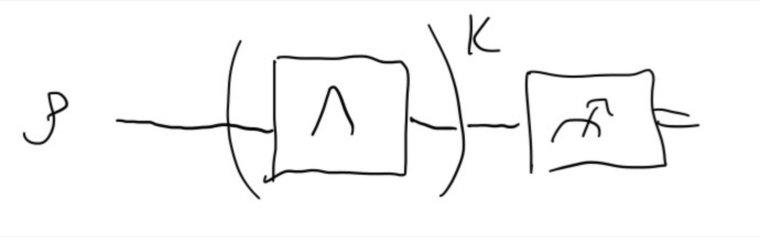
$$\frac{1}{2^n}\sum_b \rho_b Tr\left[P_c^\dagger \Lambda(\hat{P}_b)\right] = \rho'_c \quad ✓$$

or written another way:

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

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

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):

For a single step

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

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

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

$$= \rho_c$$

now sub in

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

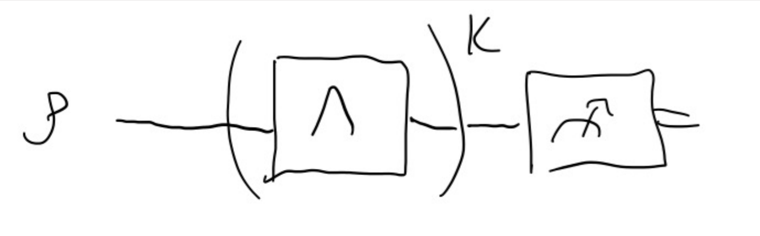
$\rightarrow$ for a pauli channel $\langle\langle P_c | \Lambda | P_b \rangle\rangle = f_c \, \delta_{bc} \, 2^n$

$$= f_c \, \rho_c$$

ρ_c ← initial state coefficient

f_c ← fidelity c of Λ

Notes



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

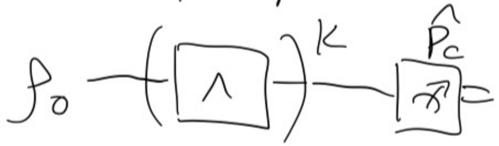
$$\Lambda(P_a) = f_a P_a$$

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

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

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

For repeating the channel K times



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

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

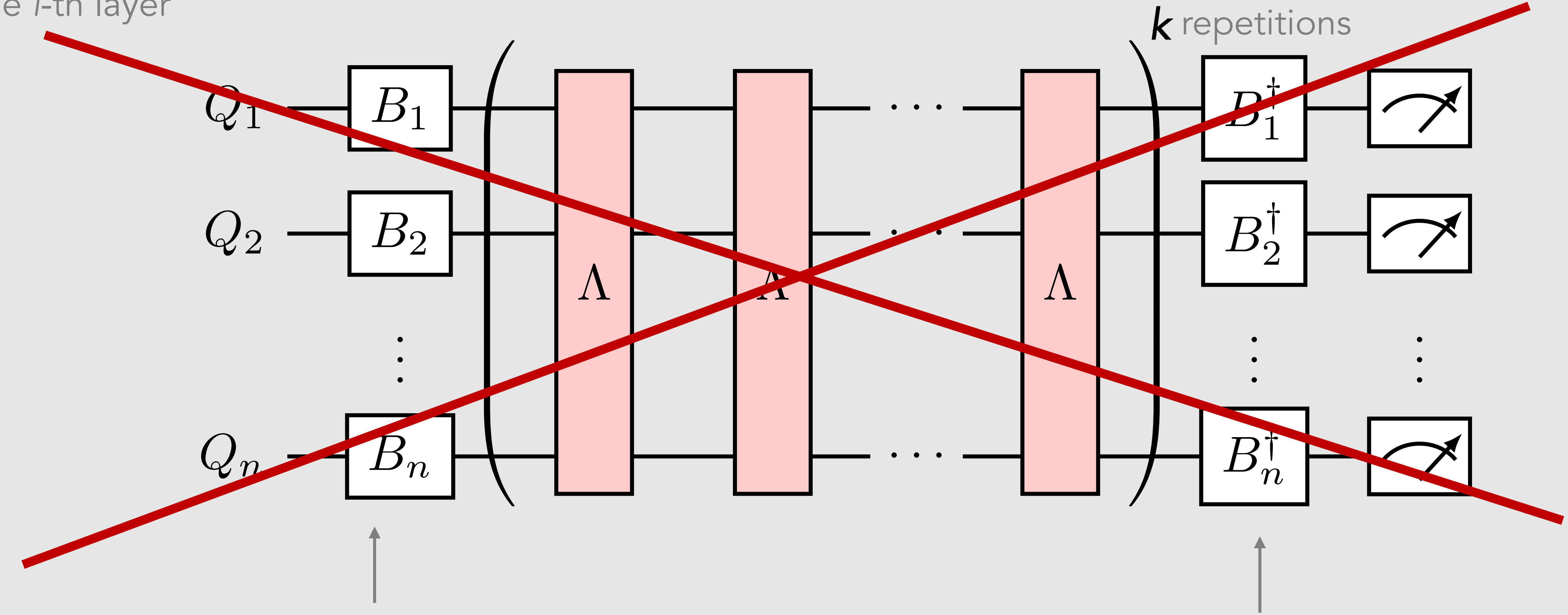
$= \sum_b f_b^K \rho_b$

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

Can include all diagonal SPAM

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

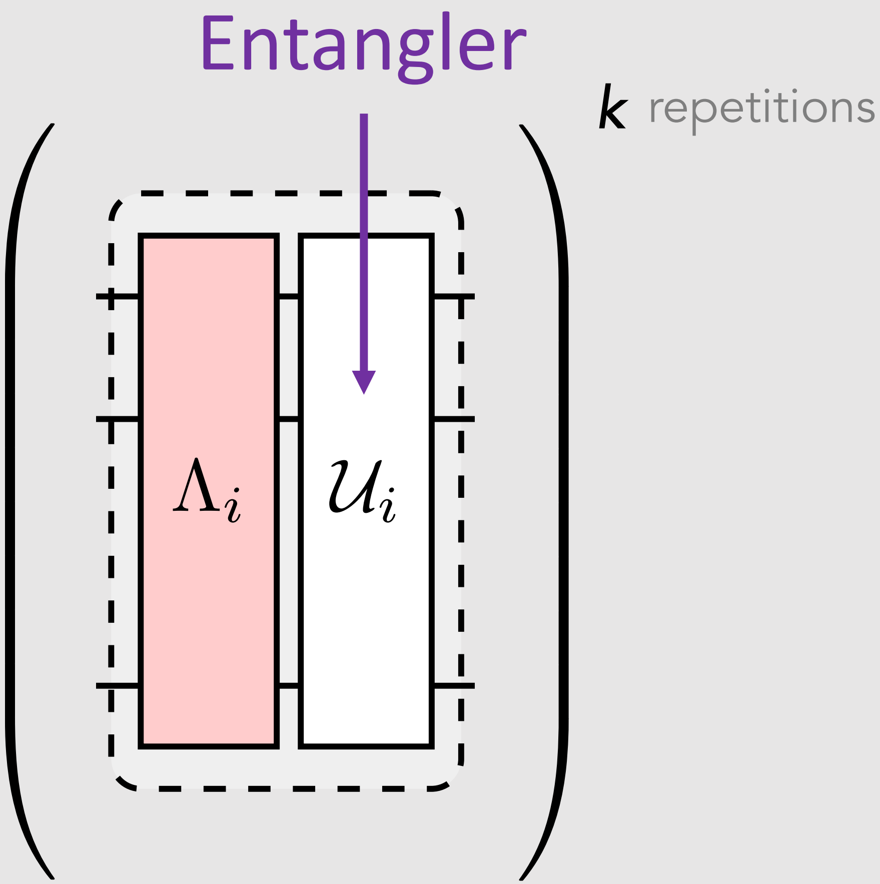
Akin to:

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)

Step 2: Ideally, amplify the noise and learn



Ideally wish

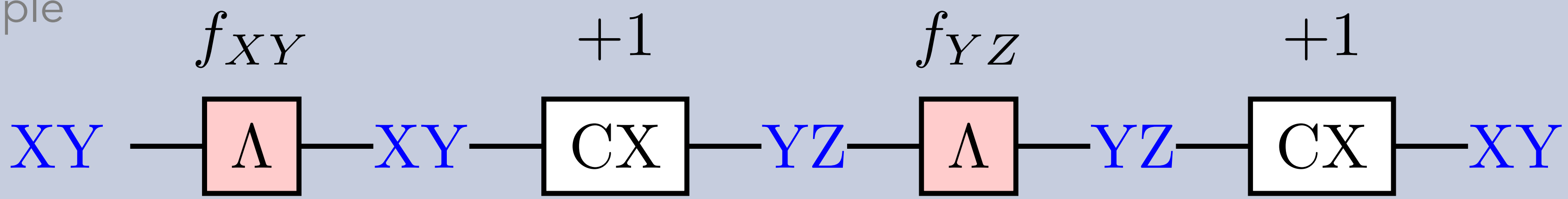
$$\Lambda_i^k(P_a) = f_{ia}^k P_a$$

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

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

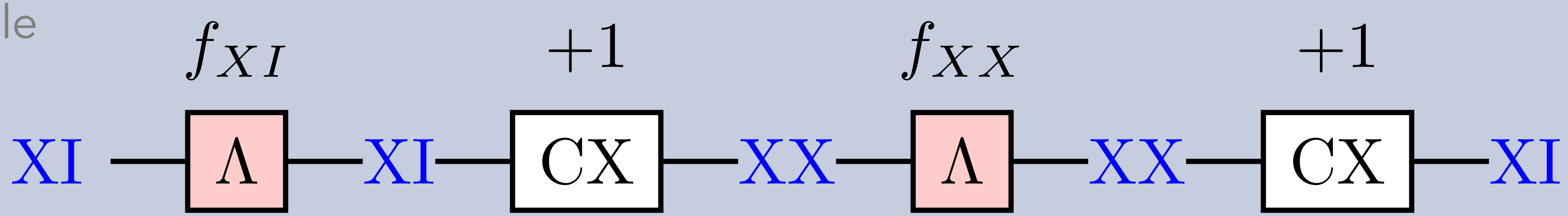
2Q example



action of CX here can be undo by SQ gates

$f_{XY} f_{YZ} XY$

2Q example



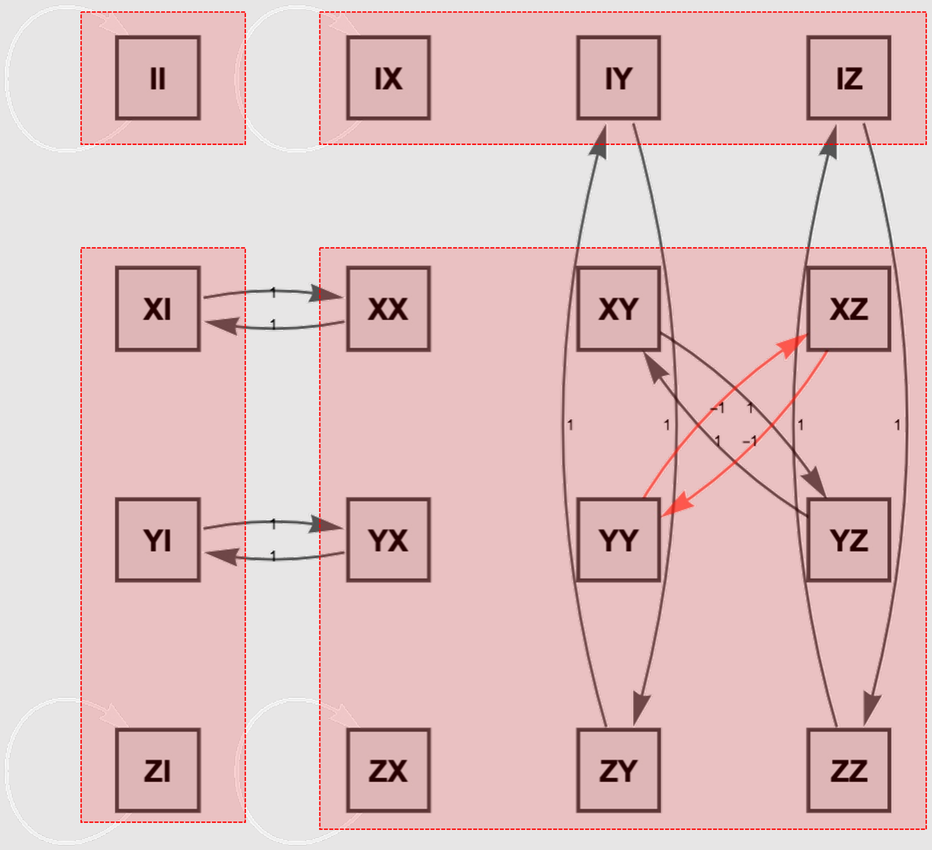
action of CX here can **not** be undone by SQ gates

$f_{XI} f_{XX} XI$

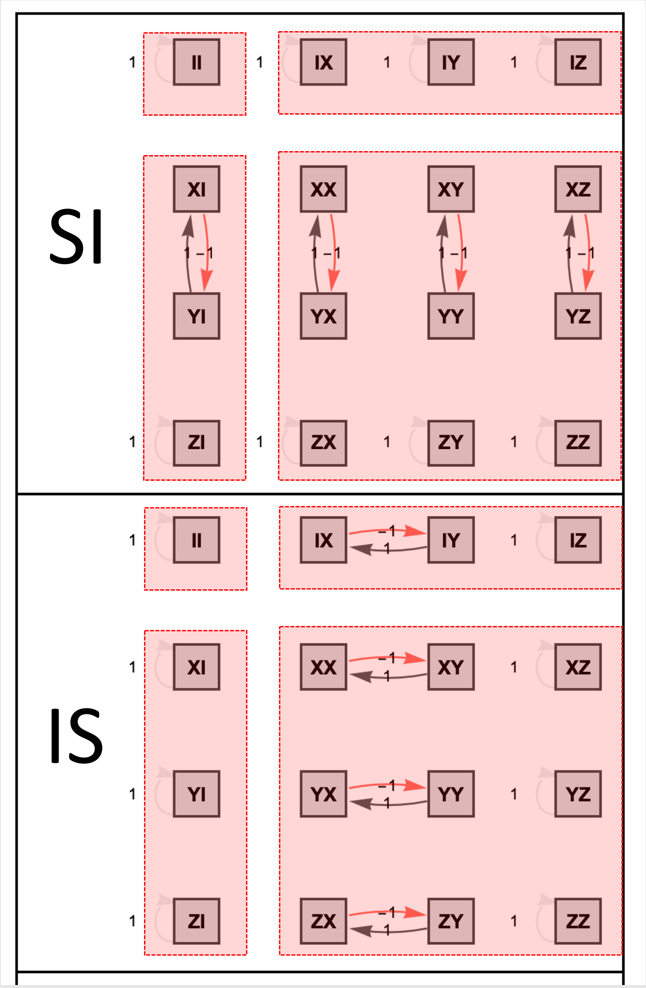
fundamental degeneracy – can not undo some non-local – need entangling operation

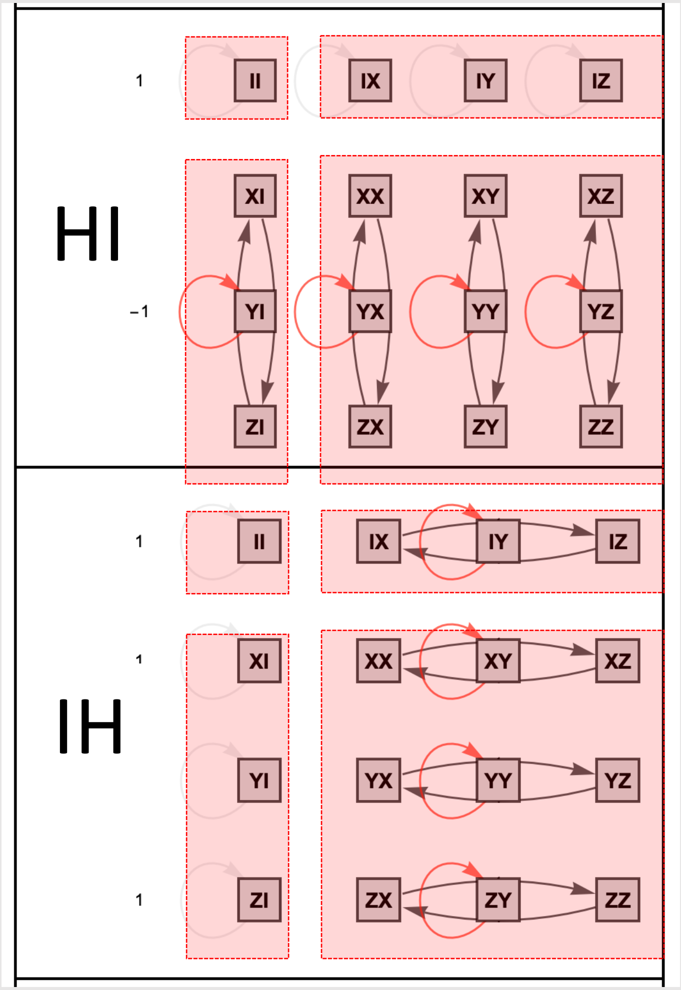
How to gates move state Paulis around?

cNOT

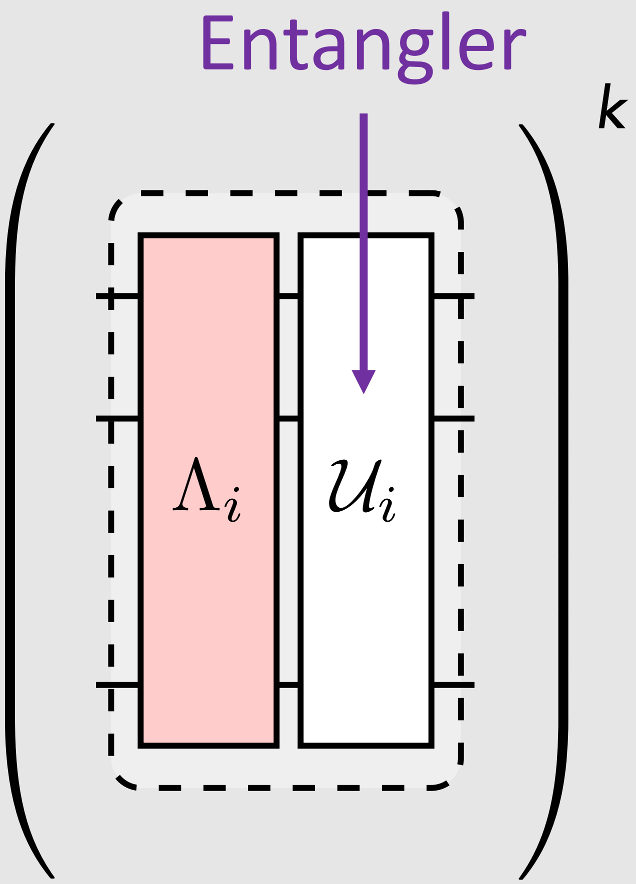


Example single qubit gates





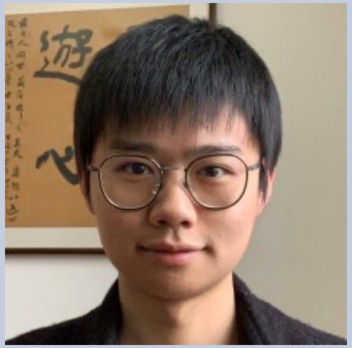
Not so simple to learn noise of entangling gates

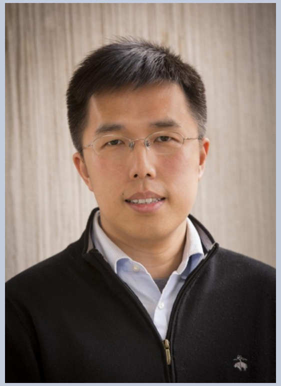


Ideally wish

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

Fundamental no-go theorem on learning

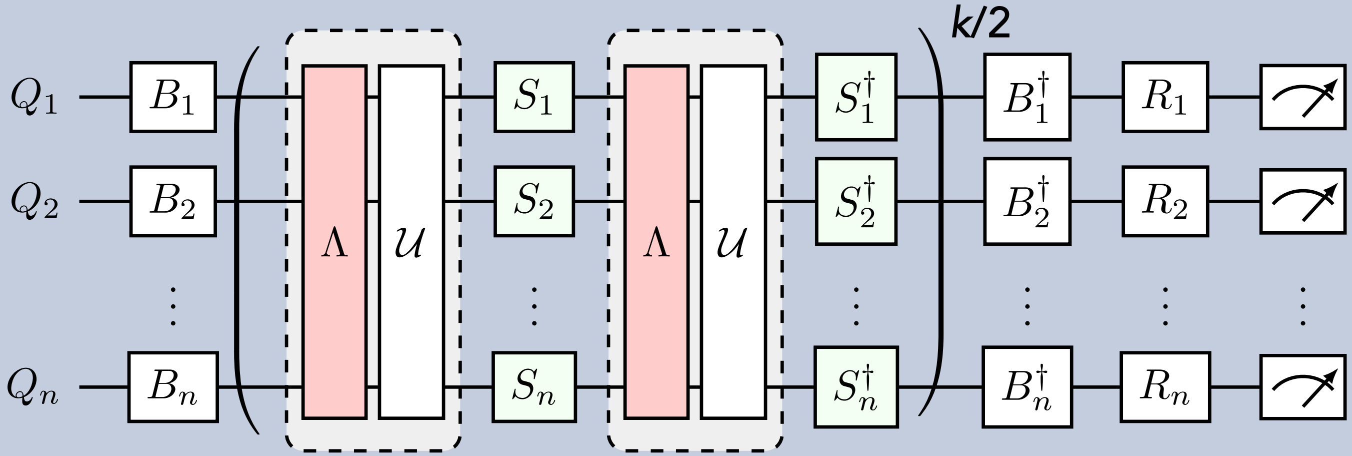




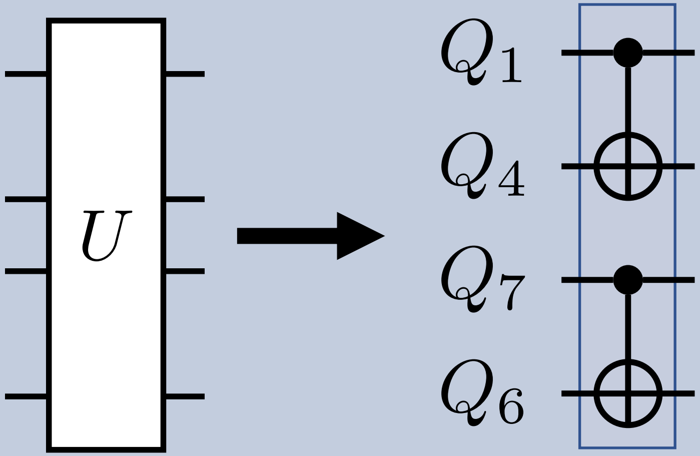
For general and in-depth
Senrui Chen, Y. Liu, M. Otten, A. Seif, B. Fefferman, L. Jiang
arXiv:2206.06362 (2022)
or supplement of our paper for qubit version and work by
S. Flammia, S Benjamin, and teams.

Solution: Custom protocol + weak assumption

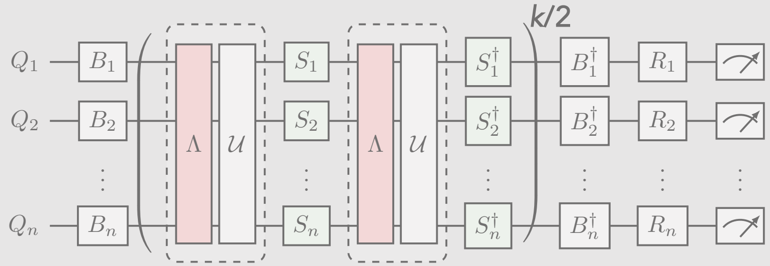
Learning circuits

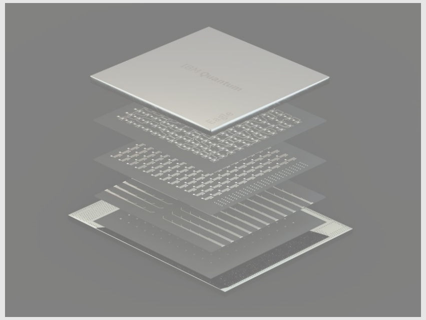


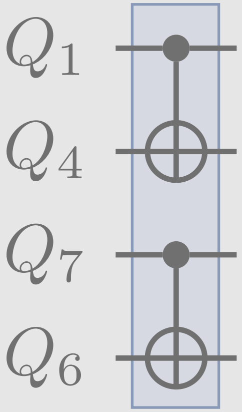
Example



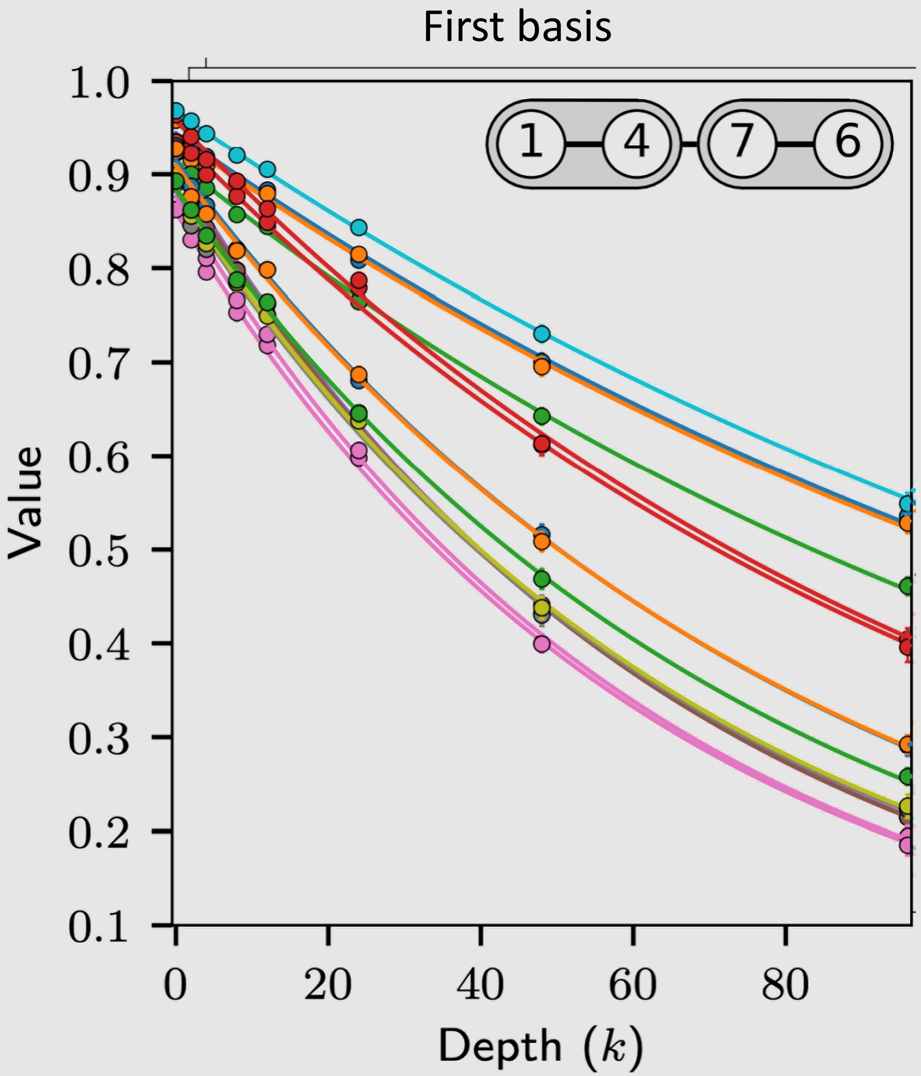
Learning the noise: raw data



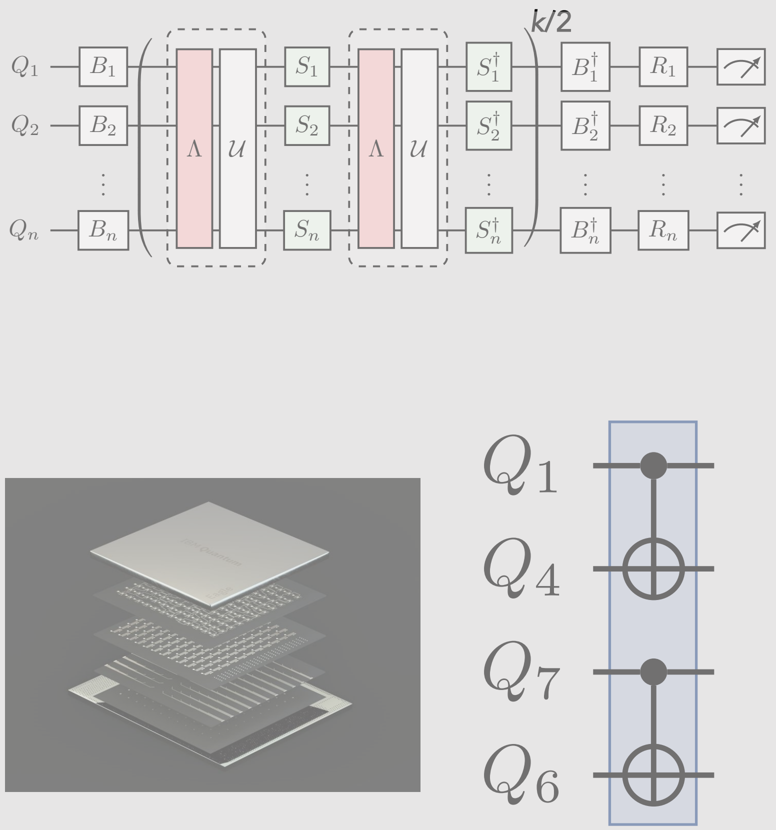




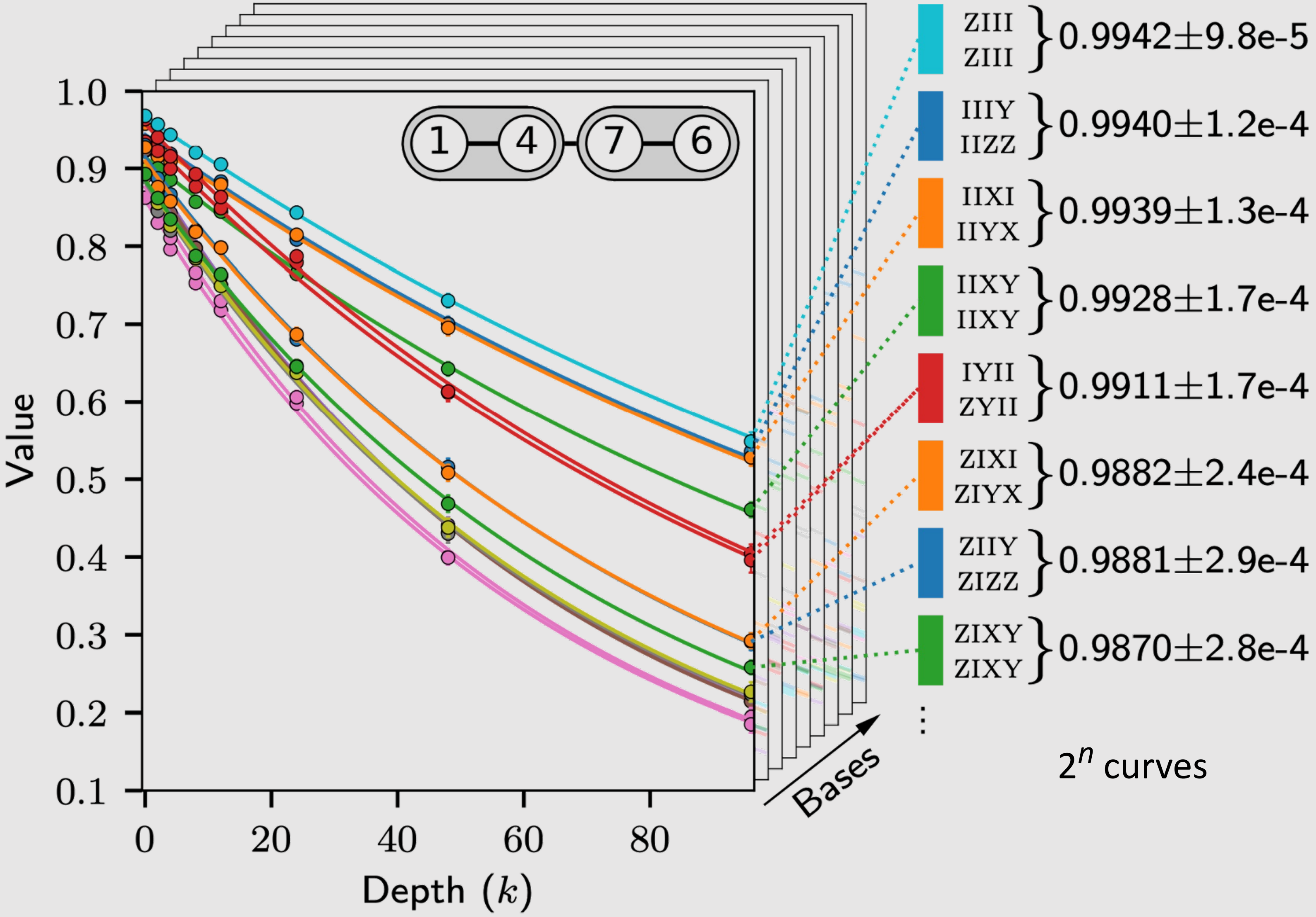
9 bases, 100 randomizations, 216 shots



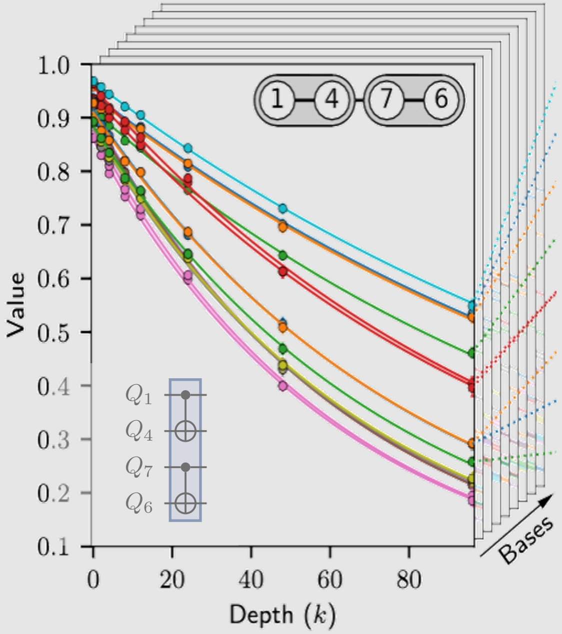
Learning the noise: raw data



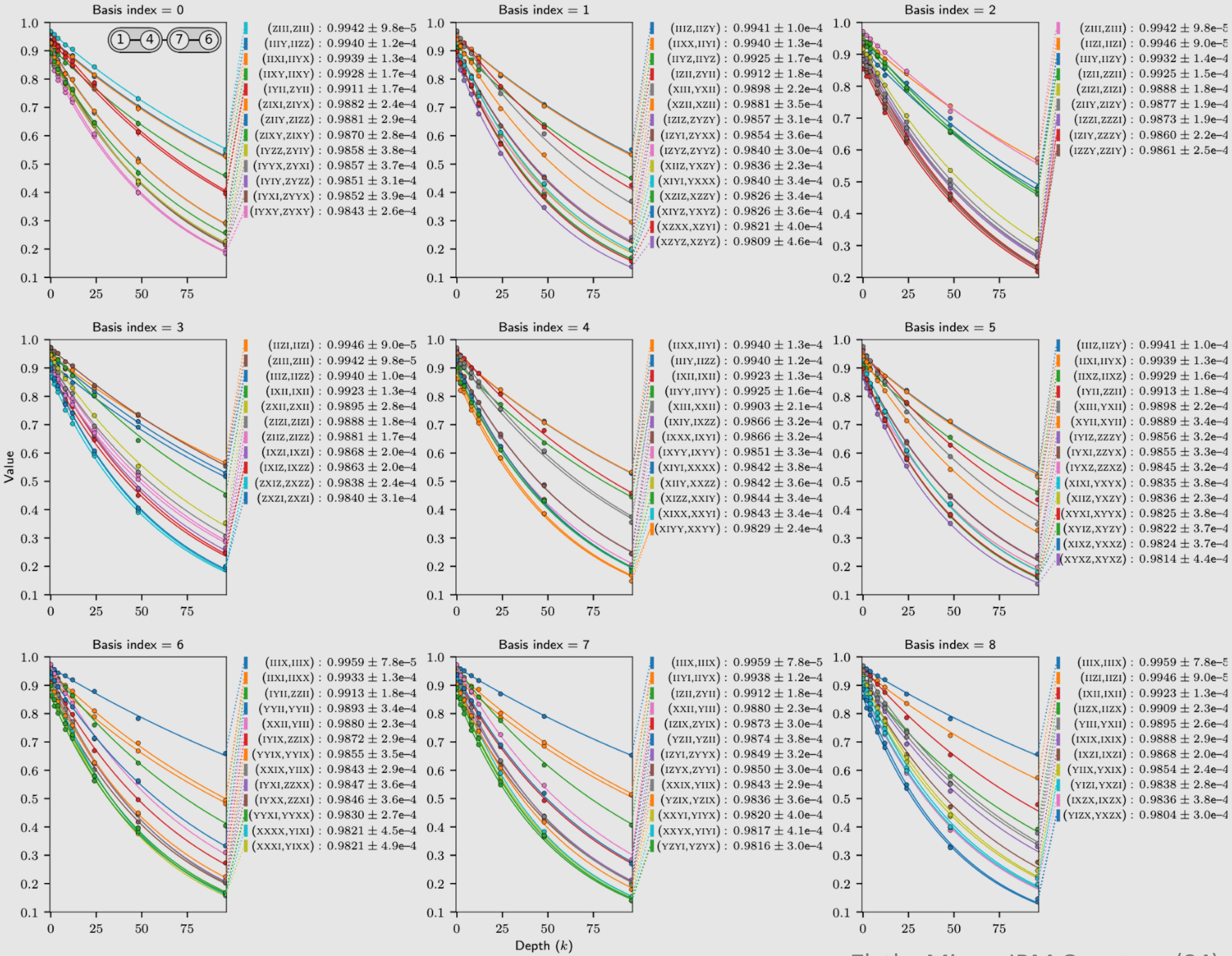
9 bases, 100 randomizations, 216 shots



Raw data

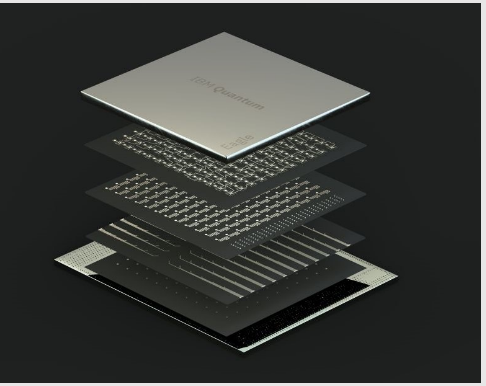


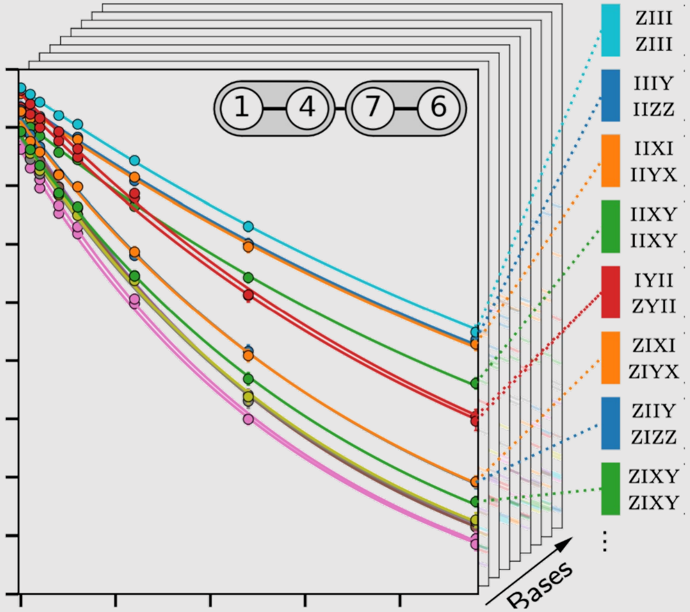
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Value
0
20
40
60
80
Depth (k)
1 4 7 6
Q_1
Q_4
Q_7
Q_6
Bases



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
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
Value
Depth (k)

Reconstructing quantum channel from measurement data

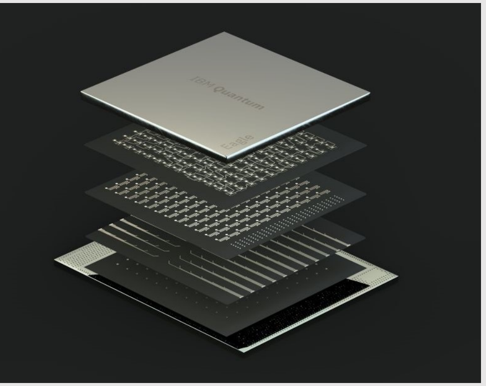


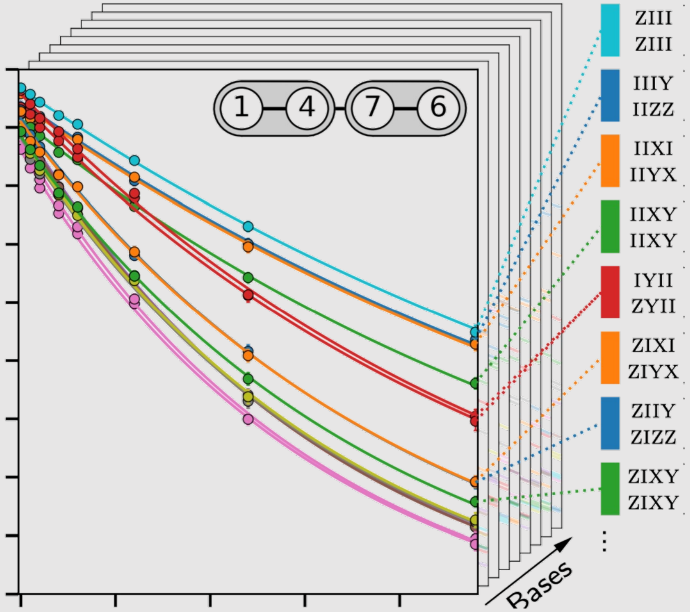


$$\Lambda(\rho) = \sum_{a=0}^{4^n-1} c_a P_a \rho P_a^\dagger$$

Still 4^n

Sparse Pauli-Lindblad model

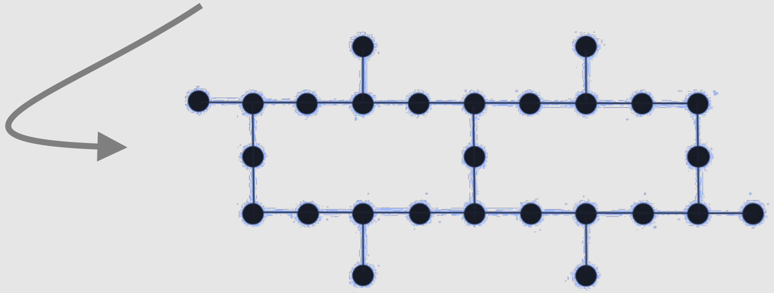




$$\Lambda(\rho) = \sum_{a=0}^{4^n-1} c_a P_a \rho P_a^\dagger$$

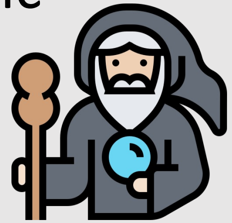
$$\Lambda(\rho) = \exp[\mathcal{L}](\rho)$$

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



Highlight: Ewout van den Berg

Magic



icon: Eucalyp

Notes

Dissipator for a given Pauli

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

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

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

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

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

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

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

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

Each sub channel

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

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

$$= \exp(\gamma \hat{\vec{P}}_a)\exp(\gamma \hat{\vec{I}})\exp\left(-\frac{1}{2}\gamma^2 [\hat{\vec{P}}_a, \hat{\vec{I}}]\right) \cdots \quad (\to 0)$$

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

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

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

Each sub channel

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

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

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

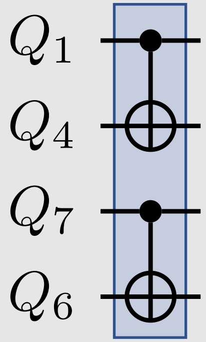
$\hat{P}_a^{\,n-\text{odd}} = \hat{I}$

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

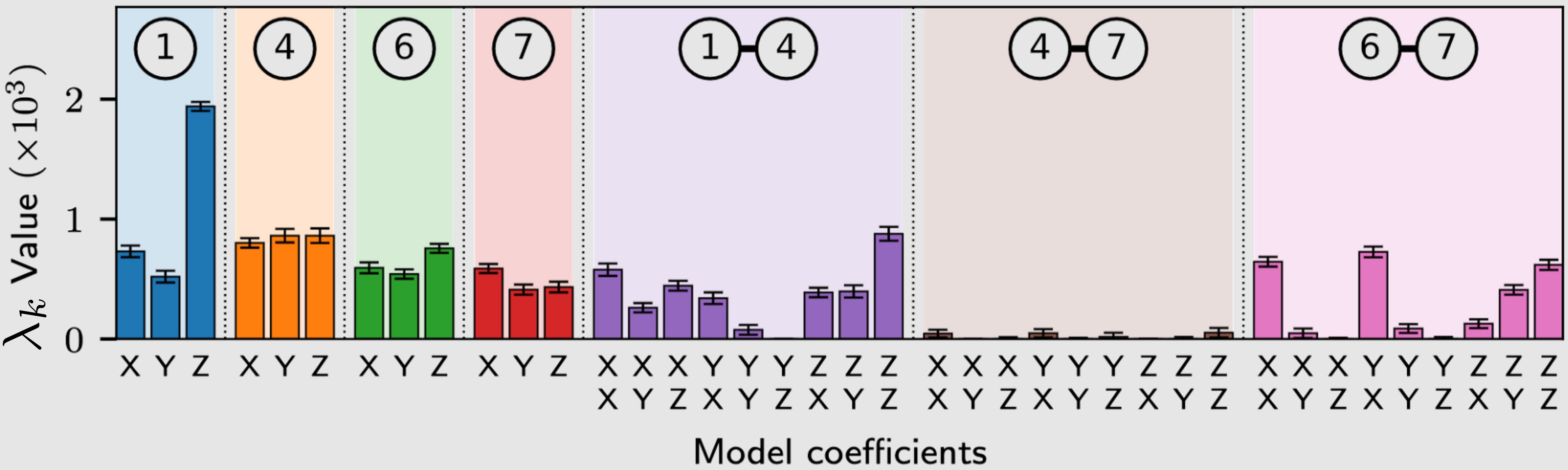
$$= \cosh(\gamma)\, \hat{I} + \sinh(\gamma)\, \hat{P}_a$$

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

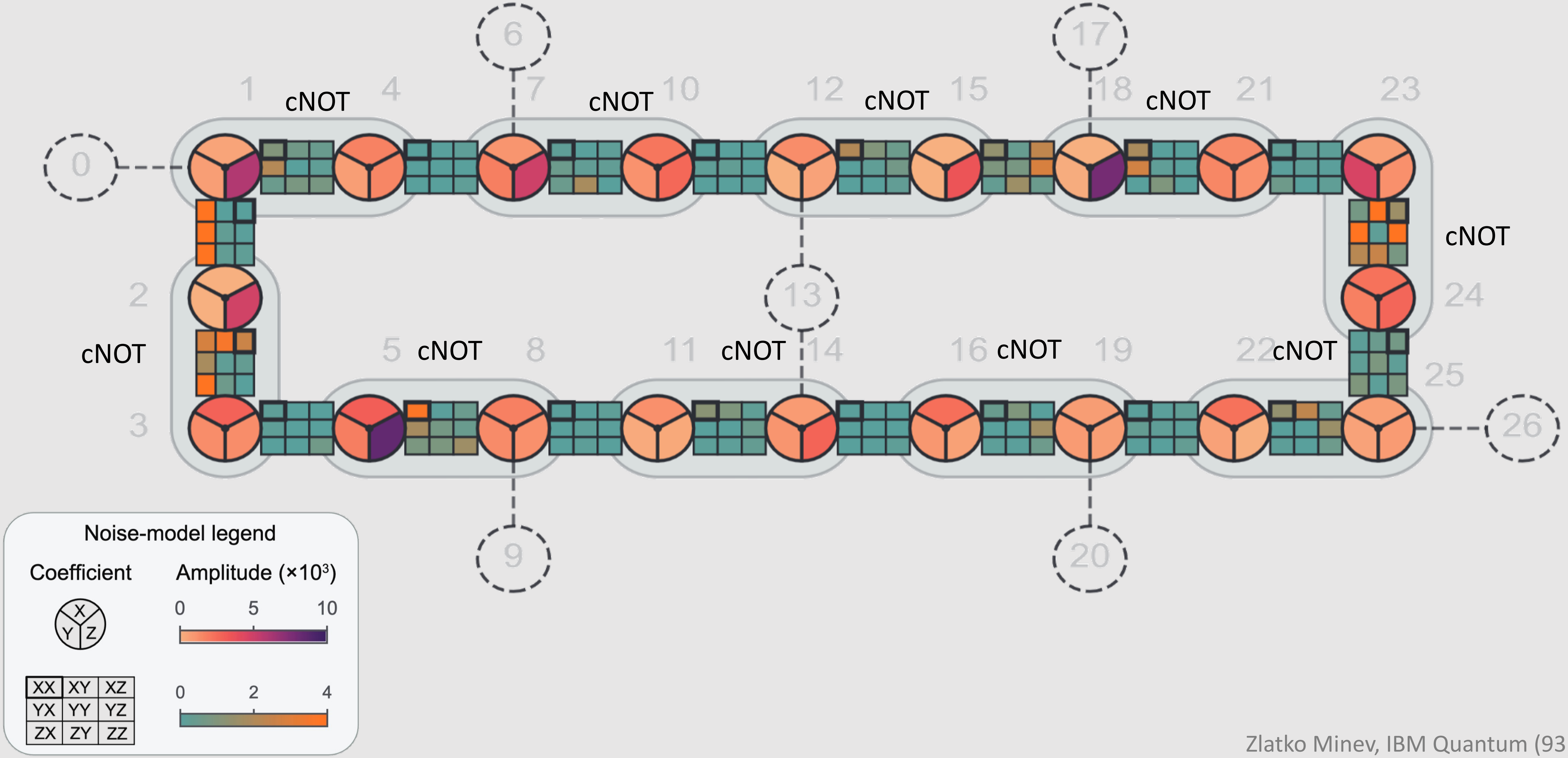
Sparse Lindblad tomogram



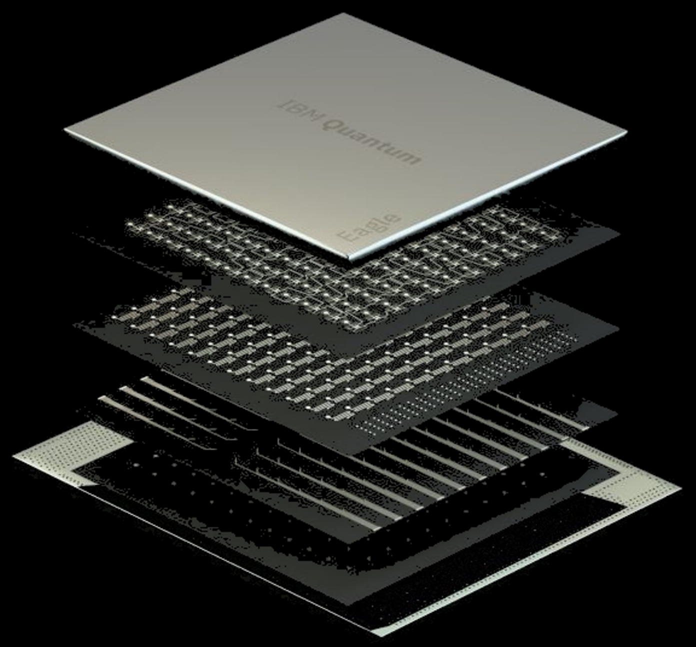
$$\mathcal{L}(\rho) = \sum_{k \in \mathcal{K}} \lambda_k \left(P_k \rho P_k^\dagger - \rho \right)$$

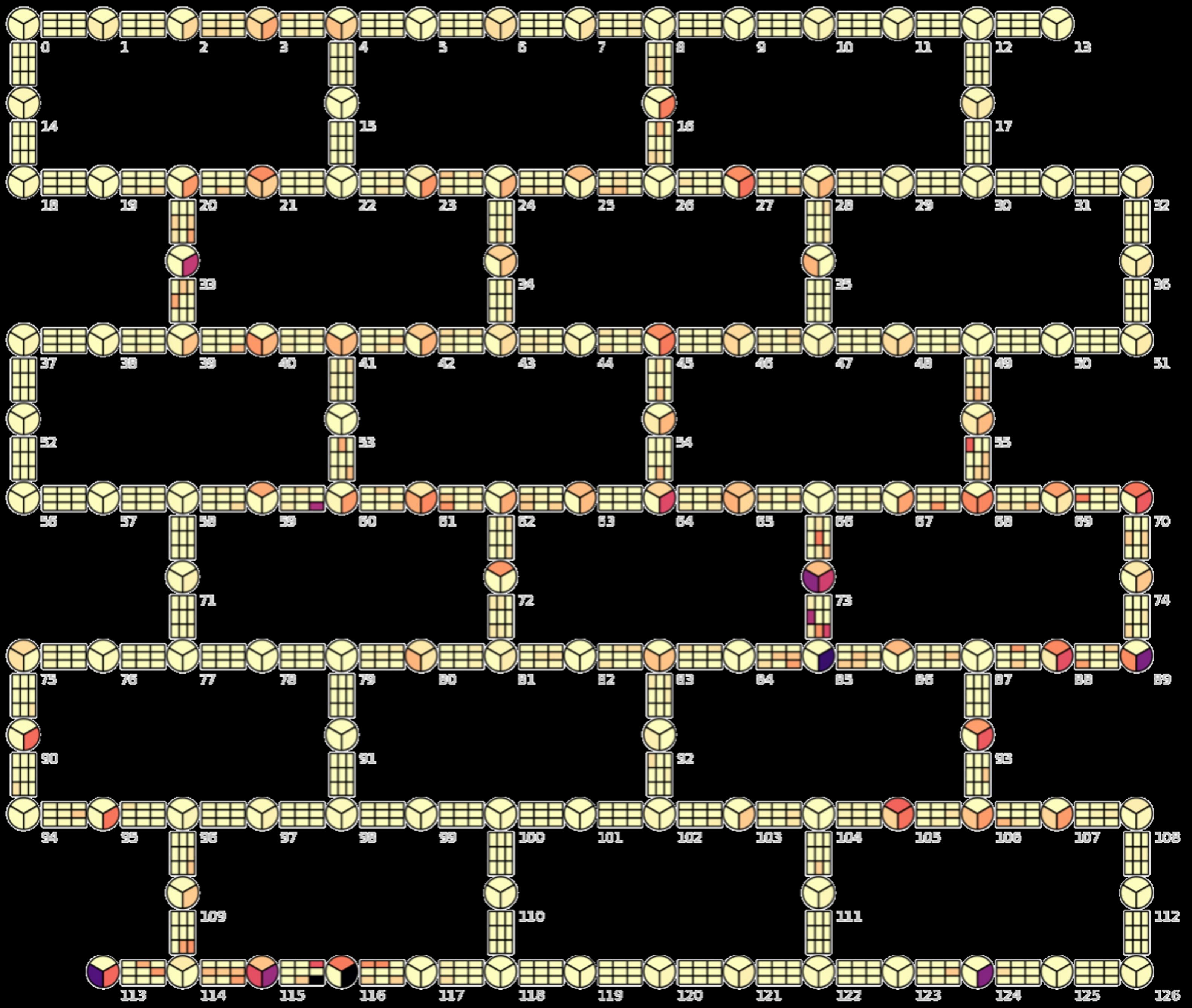


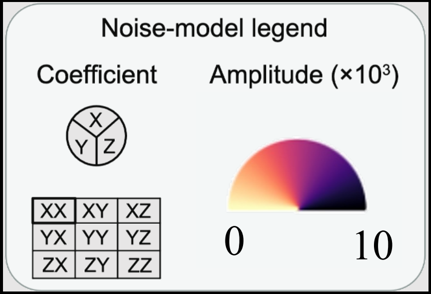
Noise tomogram for 20Q Ising-ring Trotter layer



Noise tomogram for 127Q Trotter layer







Same number of learning circuits as for 4Q

Zlatko Minev, IBM Quantum