

Figure: Hanneke et al.,
 Nat. Phys. **6**, 13-16 (2009)

Characterization of quantum devices

Steve Flammia



THE UNIVERSITY OF
 SYDNEY

eQUS

ARC CENTRE OF EXCELLENCE FOR
 ENGINEERED QUANTUM SYSTEMS

QIP Tutorial
 Seattle
 14 January 2017

The fault tolerance threshold theorem

You can quantum compute indefinitely and with low overhead, so long as

- 1) your gate error rate is less than ϵ and
- 2) the correlations are sufficiently weak

The fault tolerance threshold theorem

You can quantum compute indefinitely and with low overhead, so long as

- 1) your gate error rate is less than ε and
- 2) the correlations are sufficiently weak

- ✱ ε is pretty small, $\sim 0.1\%$ or less.
- ✱ the overhead is still quite demanding
- ✱ what do “ ε ” and “weak” even mean?

The fault tolerance threshold theorem

You can quantum compute indefinitely and with low overhead, so long as

- 1) your gate error rate is less than ε and
- 2) the correlations are sufficiently weak

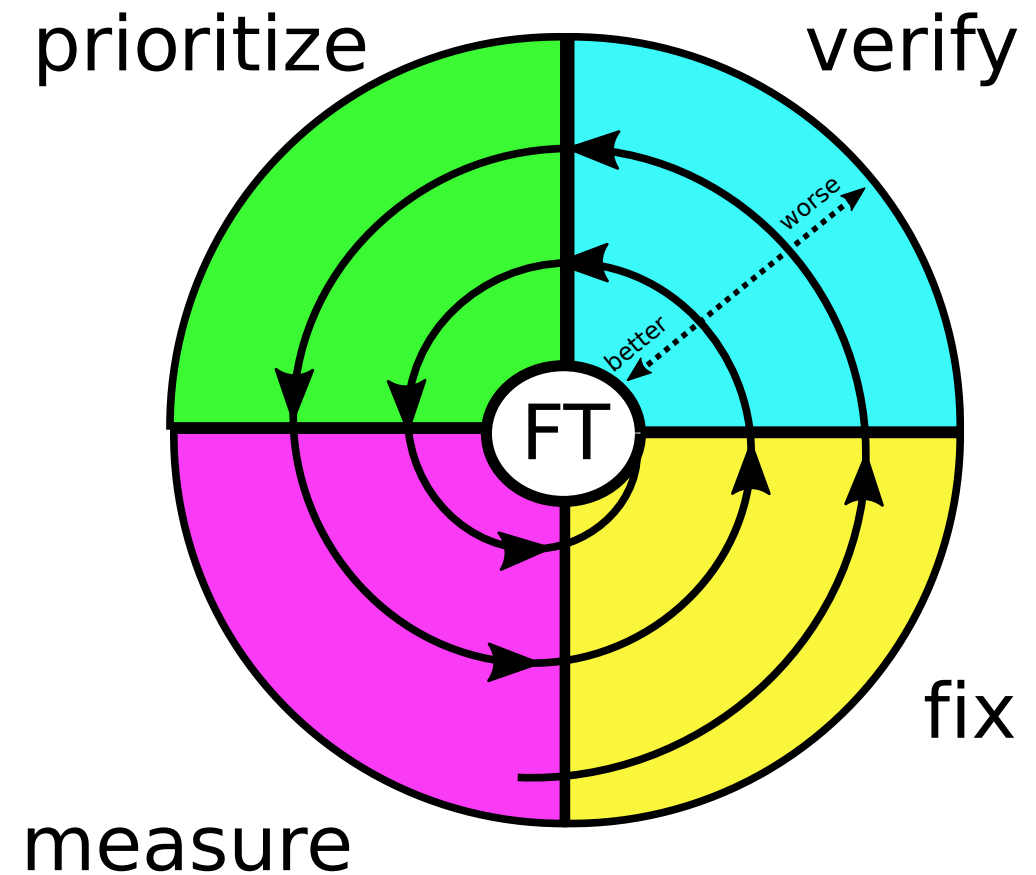
- ✱ ε is pretty small, $\sim 0.1\%$ or less.
- ✱ the overhead is still quite demanding
- ✱ what do “ ε ” and “weak” even mean?

Quantum computers cannot be build “off the shelf”

The engineering cycle

We must iteratively improve devices by estimating sources of smaller and smaller errors

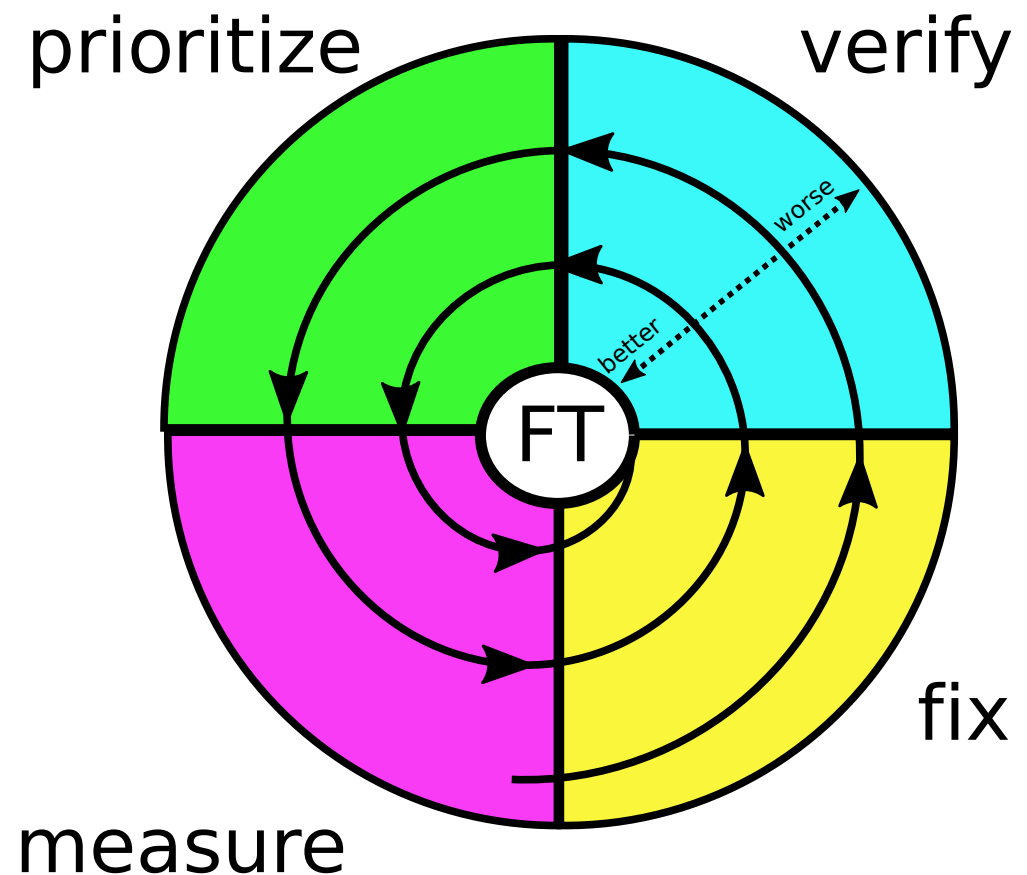
- ✿ prioritize errors
- ✿ measure them accurately
- ✿ fix them
- ✿ verify they've been fixed



The engineering cycle

We must iteratively improve devices by estimating sources of smaller and smaller errors

- ✱ prioritize errors
- ✱ measure them accurately
- ✱ fix them
- ✱ verify they've been fixed



To wind around this cycle, we must:

- ✱ *Design diagnostics* to help identify dominant noise sources
- ✱ *Estimate errors* accurately and with high precision
- ✱ *Quantify the errors* according to FTQC-relevant metrics
- ✱ Do all this with *minimal resources*

Outline (Part I)

- ✱ What are our resources?
 - ✱ complexity dictionary
- ✱ What should we measure?
 - ✱ distance measures for FTQC
 - ✱ reasons we use these measures
- ✱ What are our tools?
 - ✱ *State and process tomography*
 - ✱ standard assumptions and proof of correctness
 - ✱ *Randomized benchmarking*
 - ✱ standard assumptions and proof of correctness

Complexity Dictionary

☀ Measurement Complexity

☀ The number of *distinct measurement settings or state preparations* required

Ex: measure X five times, Y six times, and Z seven times;
this has measurement complexity **3**

☀ Sample Complexity

☀ The number of *independent copies of the system* that you are required to measure

Ex: in the previous example, the sample complexity is $5+6+7 = \mathbf{18}$

Complexity Dictionary

✿ Computational Complexity

- ✿ The *amount of classical computation* required, including the pre- and post-processing of the data

In practice, both space and time complexity are issues, especially the latter for adaptive methods

- ✿ The *amount of quantum computation* required to implement quantum circuits during the protocol

The amount of quantum computing needed to certify a noisy, non-universal device shouldn't require a perfect quantum computer.

Most relevant scaling parameters

- ✱ n the number of qubits
- ✱ d the Hilbert space dimension (usually the whole system, not a single qudit)
- ✱ ε the precision of an estimated error (in some norm)
- ✱ δ the probability of correctness of a randomized algorithm
- ✱ m the length of a sequence of quantum gates (\sim circuit size)

In the majority of situations, the important scaling parameters are

$d, \varepsilon, m.$

This is because n interconverts easily with d , and most methods scale only logarithmically with δ .

Most relevant scaling parameters

- ✱ n the number of qubits
- ✱ d the Hilbert space dimension (usually the whole system, not a single qudit)
- ✱ ϵ the precision of an estimated error (in some norm)
- ✱ δ the probability of correctness of a randomized algorithm
- ✱ m the length of a sequence of quantum gates (\sim circuit size)

Important caveat:

this is data-driven science and *constant factors do matter here!*

Theory results that ignore constants and log factors are important, but *implementations* and *practical heuristics* need to be very competitive to get adopted.

Figures of Merit: States

✱ Fidelity

$$F(\rho, \sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1^2$$

- ✱ Bounded between $[0, 1]$, symmetric
- ✱ If one state is pure, $F(\rho, \psi) = \langle \psi | \rho | \psi \rangle$
- ✱ Note: two conventions exist in the literature: “sqrt(F) vs. F²”

✱ Trace distance

$$T(\rho, \sigma) = \frac{1}{2} \|\rho - \sigma\|_1$$

- ✱ Bounded between $[0, 1]$, symmetric
- ✱ Related to the fidelity via $1 - \sqrt{F} \leq T \leq \sqrt{1 - F}$

- ✱ many other figures of merit are sensible, but we'll stick to these.

Figures of Merit: Channels

✱ “Average Error Rate”

$$r(\mathcal{E}, \mathcal{U}) = 1 - \int d\psi \langle \psi | U^\dagger \mathcal{E}(\psi) U | \psi \rangle$$

✱ Bounded between $[0, d/(d+1)]$

(I know, right?)

✱ It's symmetric

✱ Satisfies $r(\mathcal{E}, \mathcal{U}) = r(\mathcal{E}\mathcal{U}^\dagger, \mathcal{I})$

✱ The process fidelity is $1 - \frac{d+1}{d} r$

✱ *Not* a norm (no triangle inequality)

✱ Physical interpretation: average error over pure states.

It is often convenient to phrase complexity in terms of r , so we'll add it to our list of relevant scaling parameters.

Figures of Merit: Channels

✱ **Diamond distance (completely bounded norm, cb norm)**

$$D(\mathcal{E}, \mathcal{F}) = \sup_{\rho} \frac{1}{2} \|[\mathcal{E} \otimes \mathcal{I} - \mathcal{F} \otimes \mathcal{I}](\rho)\|_1$$

✱ Bounded between $[0, 1]$, symmetric.

✱ Related to the average error rate via¹

$$\frac{d+1}{d} r \leq D \leq d \sqrt{\frac{d+1}{d} r}$$

✱ It's a true norm, and obeys the triangle inequality.

✱ Operational interpretation as the maximum bias of a single-shot measurement distinguishing the two channels

Figures of Merit: why these ones?

Why do people care about the quantity D ?

Chaining: $D(\mathcal{E}_2\mathcal{E}_1, \mathcal{F}_2\mathcal{F}_1) \leq D(\mathcal{E}_1, \mathcal{F}_1) + D(\mathcal{E}_2, \mathcal{F}_2)$

Stability: $D(\mathcal{I} \otimes \mathcal{E}, \mathcal{I} \otimes \mathcal{F}) = D(\mathcal{E}, \mathcal{F})$

These properties are *not* satisfied by r , and they are used crucially in existing proofs of the FT threshold theorem. This makes D appealing to **theorists**.

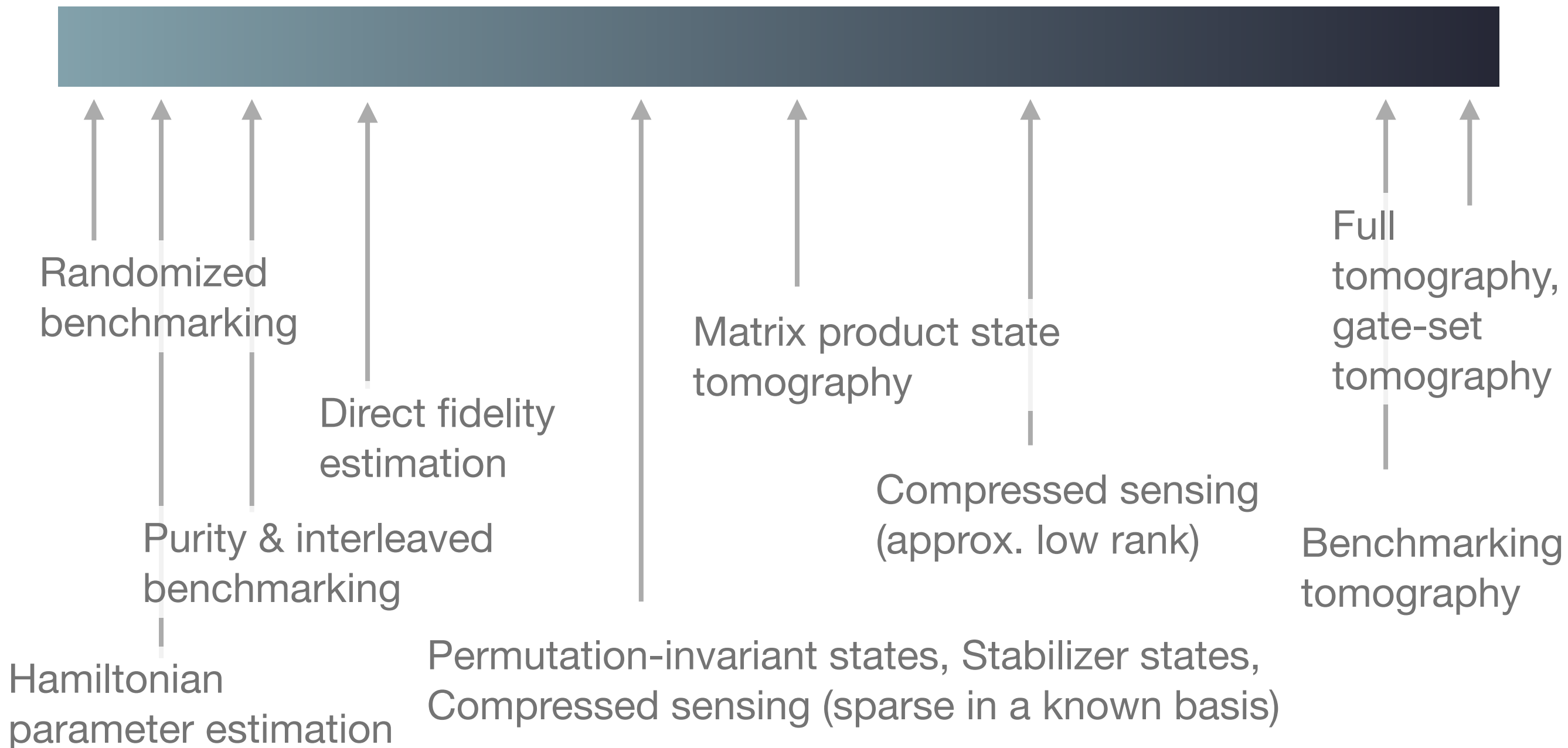
Why do people care about the quantity r ?

The average error rate r is easy to measure! Randomized benchmarking can estimate r accurately in a wide variety of circumstances. This makes r appealing to **experimentalists**.

Strategies for Characterizing Noise

Low complexity,
Less information

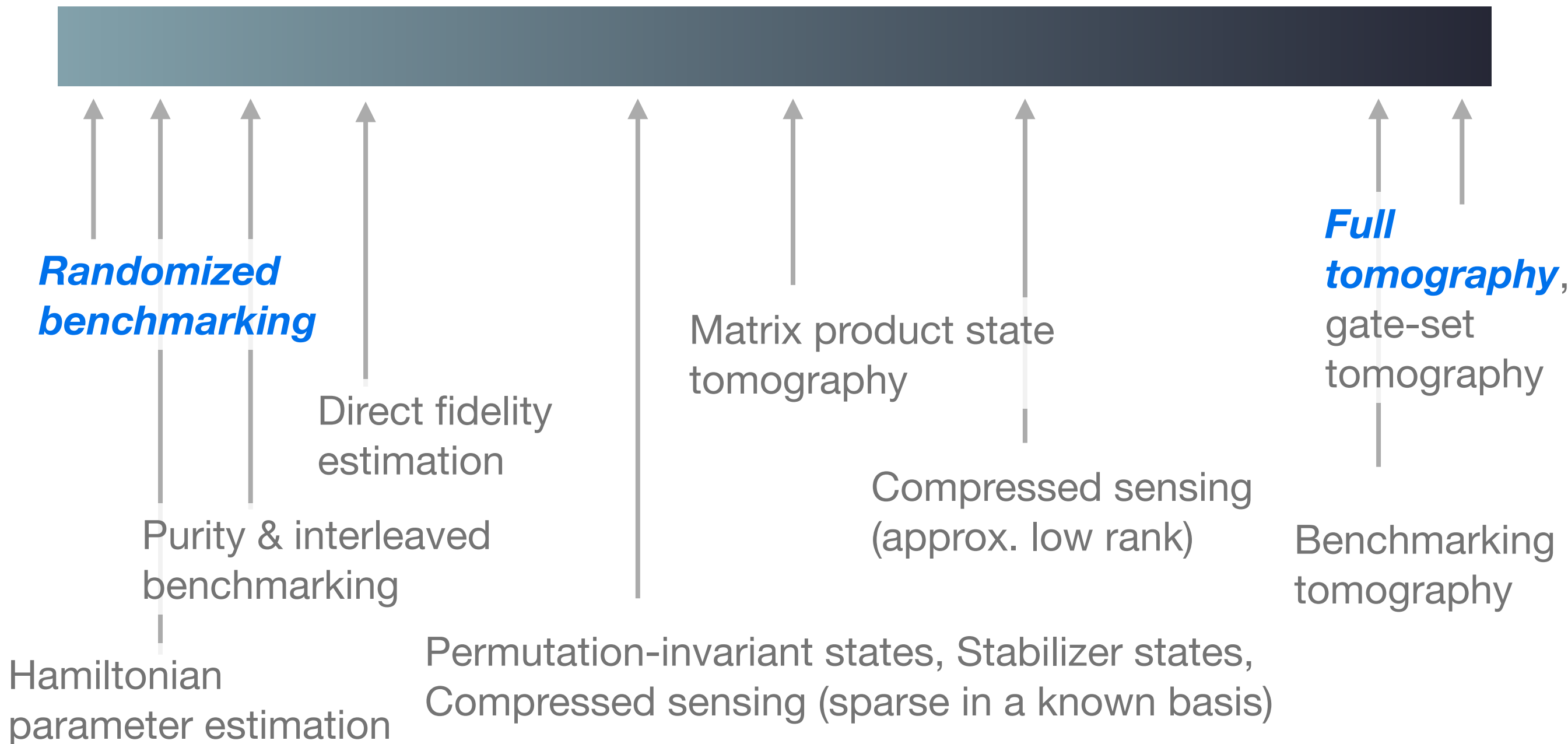
High complexity,
More information



Strategies for Characterizing Noise

Low complexity,
Less information

High complexity,
More information



Diagrammatic notation

- Tensor network notation

$$R^\rho_{\sigma\mu\nu} \implies \text{Diagram of a blue circle labeled } R \text{ with four legs}$$

- Contraction

$$\text{Diagram of two blue circles connected by two arcs} := \sum_{i,j} \text{Diagram of a blue circle with legs } i, j \text{ connected to another blue circle with legs } i, j$$

- Grouping and splitting

$$\text{Diagram of two blue circles with multiple legs and arcs} = \text{Diagram of two blue circles with double lines} = \text{Diagram of two blue circles with a thick black line between them}$$

Conventional	Einstein	TNN
$\langle \vec{x}, \vec{y} \rangle$	$x_\alpha y^\alpha$	
$M\vec{v}$	$M^\alpha_\beta v^\beta$	
AB	$A^\alpha_\beta B^\beta_\gamma$	
$\text{Tr}(X)$	X^α_α	

State Tomography

✱ What is state tomography?

$$\text{Tr}(E \cdot) \text{ ————— } \rho$$

- ✱ Prepare a given (unknown) state ρ
- ✱ Measure a POVM with effects E .
- ✱ These experiments define a probability distribution:

$$p(E|\rho) := \text{Tr}(E\rho)$$

- ✱ Now sample repeatedly to get an estimate of p .

State Tomography

✻ What is state tomography?

$$\text{Tr}(E \cdot) \text{ ————— } \rho$$

We can formalize this with the notion of a *sampling operator*.

$$\mathcal{A} : \mathbb{H}^{d \times d} \rightarrow \mathbb{R}^k \quad [\mathcal{A}(\rho)]_j = \text{Tr}(E_j \rho)$$

The j^{th} element is just the probability of the j^{th} outcome.

Thus, the input and output are linearly related.

State Tomography

✱ What is state tomography?

$$\text{Tr}(E \cdot) \text{ ————— } \rho$$

Now suppose that, because of experimental noise, we must write the data of empirical estimates y as

$$y = \mathcal{A}(\rho) + z \quad z \sim \mathcal{N}(0, \sigma)$$

Suppose further that the sampling operator is full rank. Then the minimum variance unbiased estimator is the least squares estimate:

$$\hat{\rho} = (\mathcal{A}^T \mathcal{A})^{-1} \mathcal{A}^T y$$

Assumptions of State Tomography

This estimator $\hat{\rho} = (\mathcal{A}^T \mathcal{A})^{-1} \mathcal{A}^T y$ gives us a provably optimal estimate of the true state, so we're done, right?

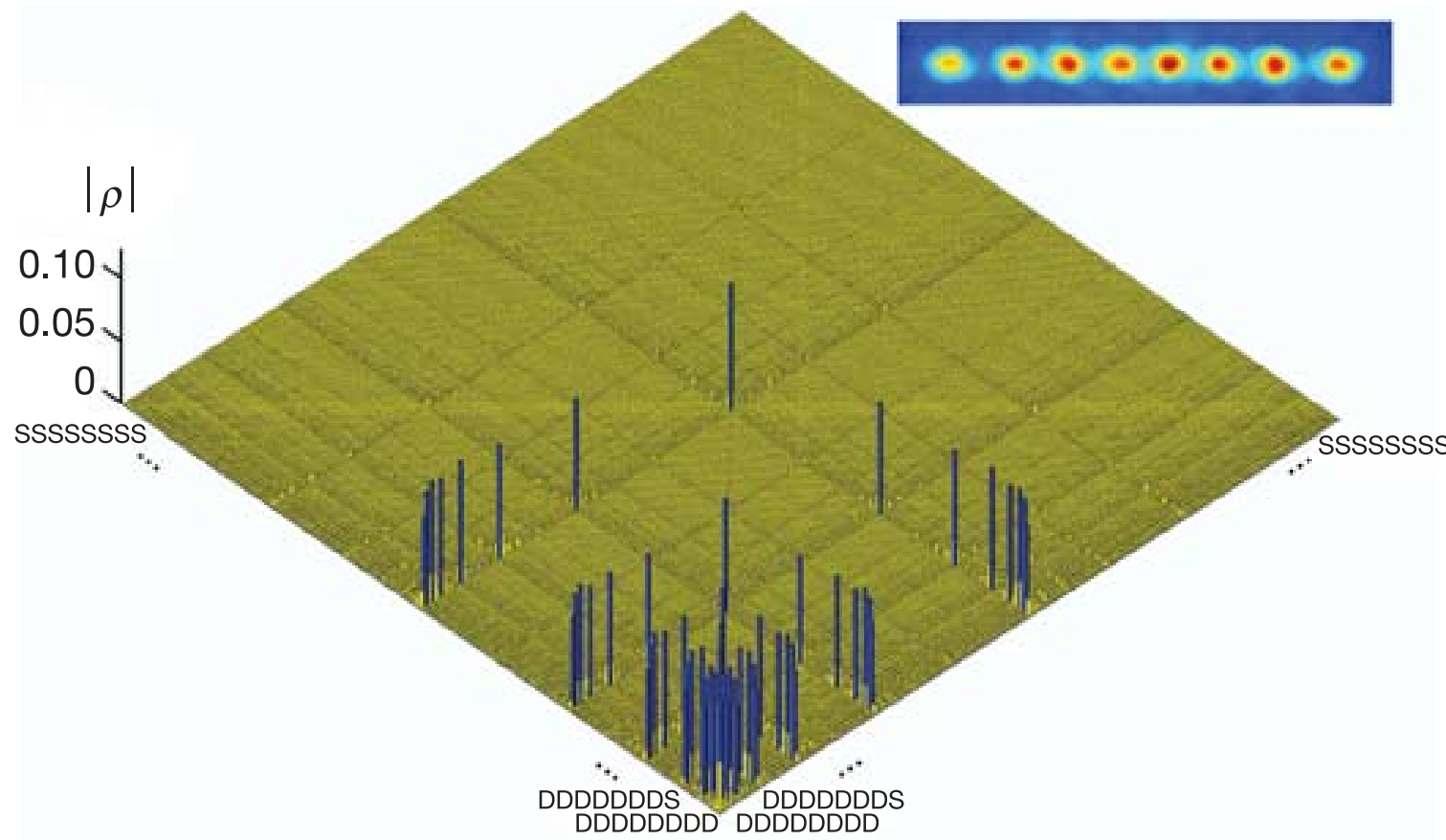
$$\text{Tr}[E\mathcal{E}(\rho)] \quad \text{Tr}(E \cdot) \text{ --- } \boxed{\mathcal{E}} \text{ --- } \rho$$

- ✱ Noise means we only reconstruct (noise + state)
 - ✱ **State Preparation And Measurement errors (SPAM)** when doing tomography of quantum channels
- ✱ Noise is assumed iid (exchangeable prior), so drift is not tolerated
- ✱ It is *extremely* resource intensive! Complexity is a key bottleneck
- ✱ The reconstructed state (channel) might not even be positive (CP)
- ✱ Quantifying the uncertainty of the reconstruction is challenging

Tomography in Practice

Despite the formidable complexity of doing tomography, experiments have succeeded in doing tomography in a variety of physical systems.

- The record for the largest implementation of “standard” tomography is in an **8 qubit** ion trap experiment done in Innsbruck by the Blatt group
- This experiment used a **maximum likelihood estimator** (instead of linear inversion), defined as $\hat{\rho} = \arg \max_{\rho} \Pr(y|\rho, \mathcal{A})$
- The post-processing alone took **two weeks** to get the estimate + error bars!



Tomography in Practice

Open-source software implementations of quantum state tomography, reflecting different strategies for estimation:

- Tomographer (Faist & Renner 2016)

C++ / command-line • <https://tomographer.github.io>

Uses Metropolis–Hastings algorithm to compute Bayesian region estimates, then expands to find confidence regions.

- QInfer (Granade *et al.* 2016)

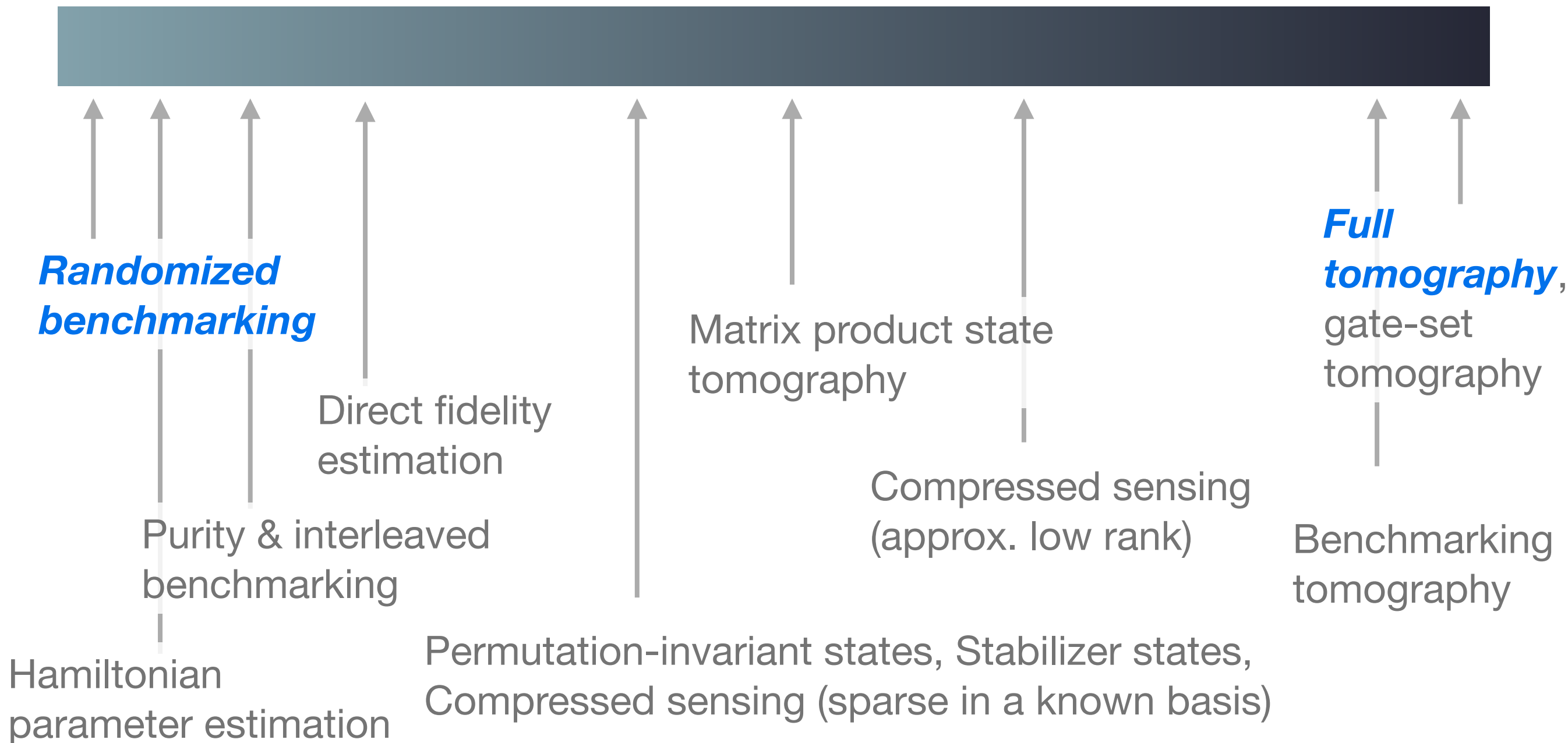
Python • qinfer.org

Uses particle filtering to approximate posterior distribution and to report credible regions.

Strategies for Characterizing Noise

Low complexity,
Less information

High complexity,
More information



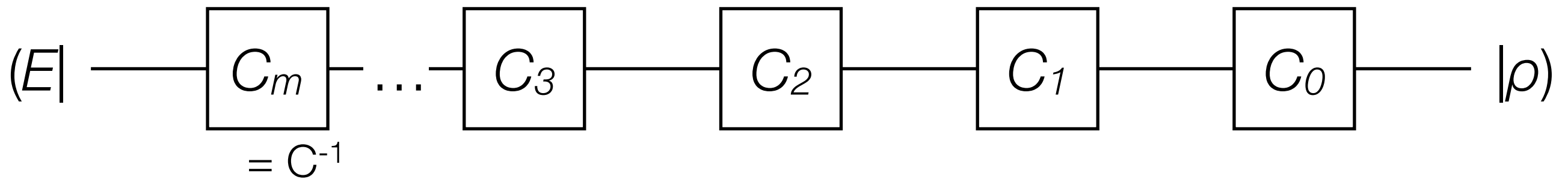
Tomography versus RB

Tomography was limited by two main factors:

- SPAM errors, leading to low accuracy in the estimate
- High complexity, making practical implementation difficult

Randomized Benchmarking (RB) is a method that tries to solve both of these problems, but at the cost that it provides much less information (though hopefully it is *relevant* information)

Randomized Benchmarking

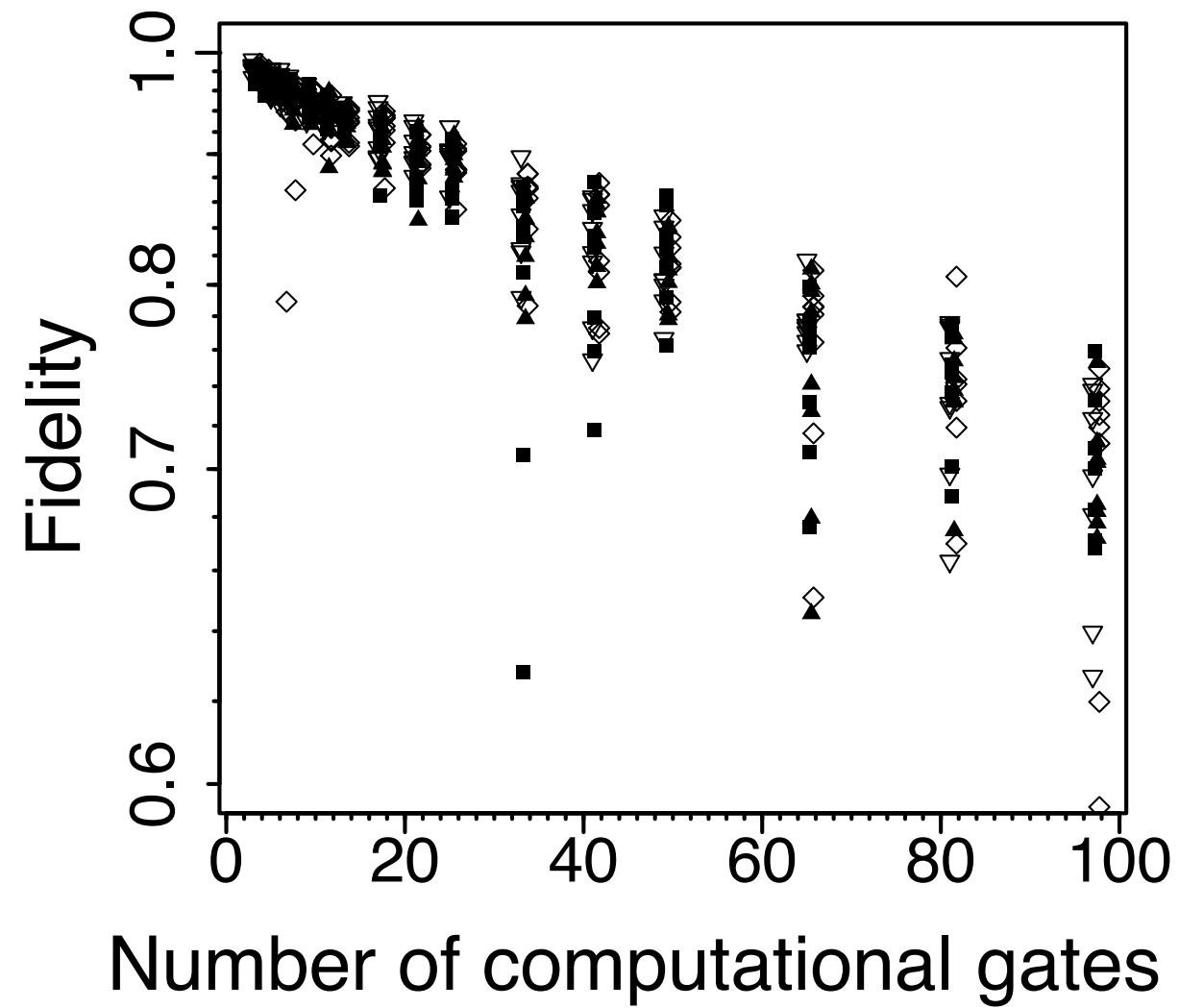


- ✿ Choose a random set s of m Clifford gates
- ✿ Prepare the initial state in the computational basis
- ✿ Apply the Clifford sequence, and add the inverse gate at the end of the sequence
- ✿ Measure in the computational basis

Repeat to estimate $F_{m,s} = \Pr(E|s,\rho)$

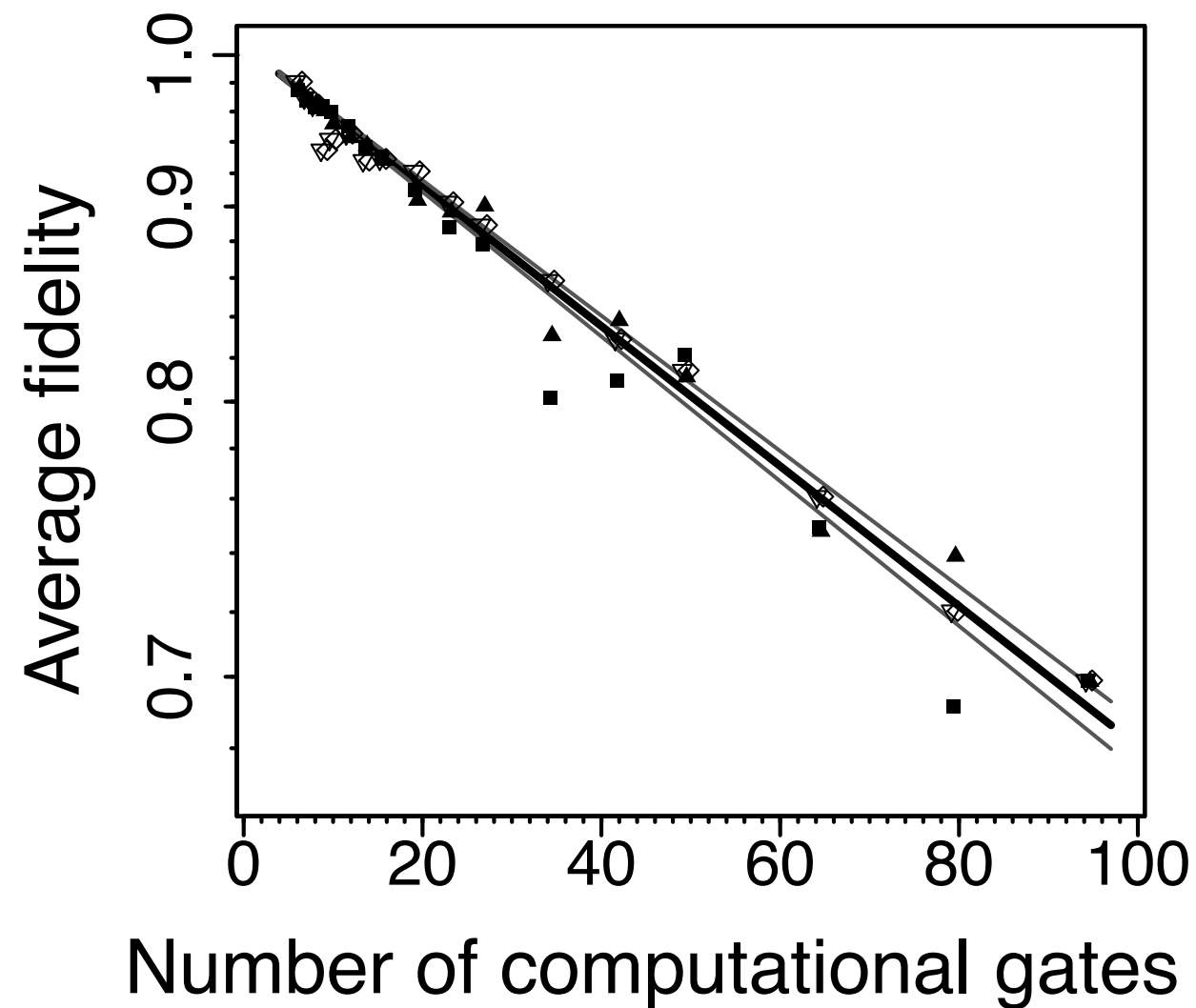
Simple procedure tests performance of large quantum circuits

Randomized Benchmarking



Randomized Benchmarking

If the noise is **time-** and **gate-independent**, we get:



“0th order model”

Fit to the model

$$\bar{F}_m = A + B f^m$$

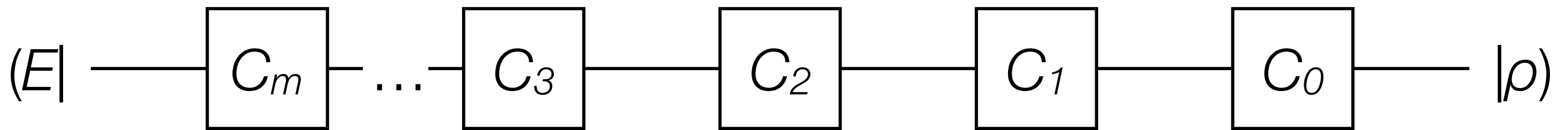
Note this is **not** a linearizable model!

$$f = \frac{d\mathcal{F}_{\text{avg}}(\Lambda) - 1}{d - 1}$$

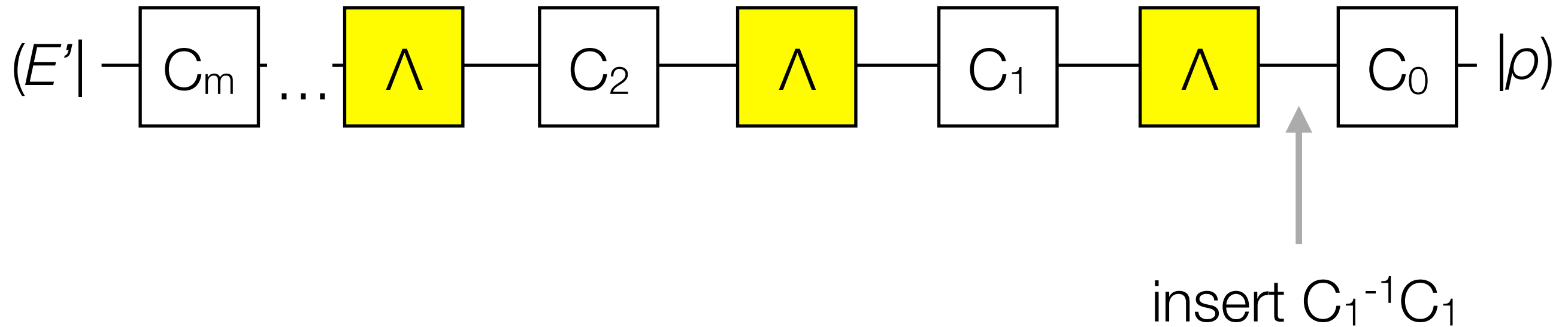
$$\mathcal{E}_{\text{physical}} = \Lambda_{\text{noise}} \mathcal{U}_{\text{ideal}}$$

$$\mathcal{F}_{\text{avg}}(\Lambda) = \int d\psi \text{Tr}[\psi \Lambda(\psi)]$$

Randomized Benchmarking

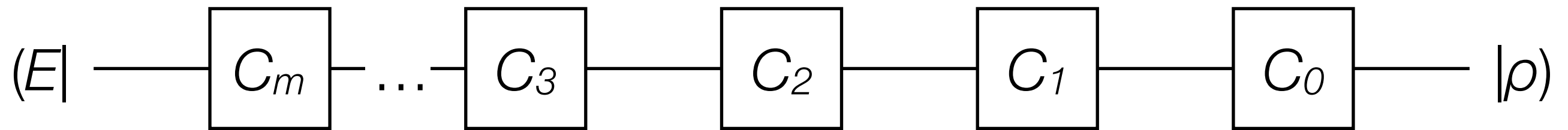


Factor each Clifford into ideal + noise

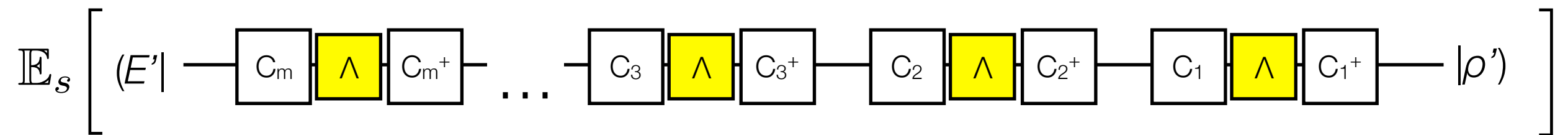


Repeat this insertion everywhere, and absorb extra gates into the endpoints as SPAM

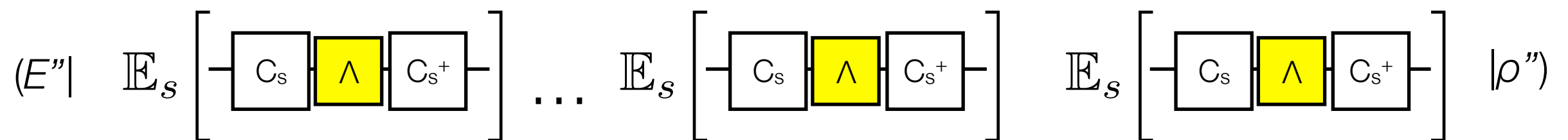
Randomized Benchmarking



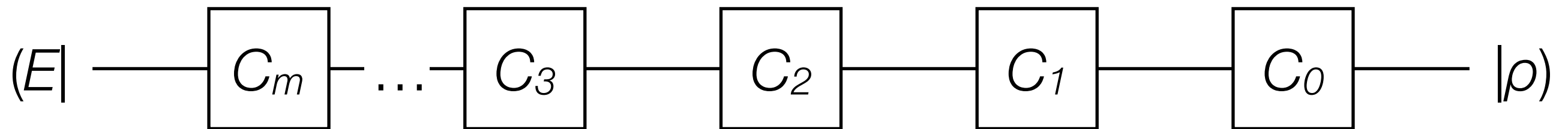
Look at the expected value of the probabilities over sequences



By independence, this factorizes into a product of average channels
(Care must be taken on the boundaries)



Randomized Benchmarking



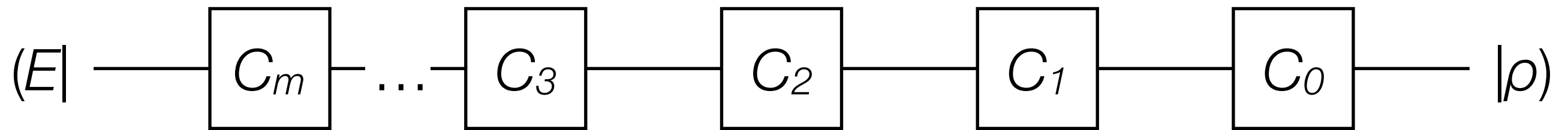
Each term in the product is a group average

$$\mathbb{E}_S \left[\text{---} \boxed{C_s} \text{---} \boxed{\Lambda} \text{---} \boxed{C_s^\dagger} \text{---} \right] = \frac{1}{|C|} \sum_j \left[C_j \Lambda C_j^{-1} \right] = \int dU \mathcal{U} \Lambda \mathcal{U}^\dagger = \bar{\Lambda}$$

The last line follows because the Clifford group is a 2-design

Now using Schur's lemma, we find $\bar{\Lambda}(\rho) = p \rho + (1 - p) \frac{\mathbb{1}}{d}$

Randomized Benchmarking



The averaged data looks like a product of depolarizing channels

$$\bar{F}_m = A + B f^m = (E| \text{---} \boxed{\bar{\Lambda}^m} \text{---} |\rho) \quad \bar{\Lambda}(\rho) = p \rho + (1 - p) \frac{\mathbb{1}}{d}$$

Moreover, the constant p is related to the average gate infidelity:

$$p = 1 - \frac{d}{d-1} r(\bar{\Lambda})$$

Thus, this fit gives us a straightforward way to estimate r .

This is the **average error of the average Clifford gate**.

Key Assumptions for Randomized Benchmarking

As with tomography, RB also depends on several key assumptions:

- **Gate-Independence:** The noise incurred by each gate cannot depend on which *Clifford* gate we performed.
- **Markovianity:** The noise processes can be described by a CPTP map acting on the system of interest.
- **Time-independence:** The noise processes should not drift during the course of the experiment, or the reported answer will also be an average over this dependence.
- **2-design:** The group should be a 2-design, but not universal (no T-gates).

Relaxing and certifying these assumptions is an area of active work.

Randomized Benchmarking in Practice

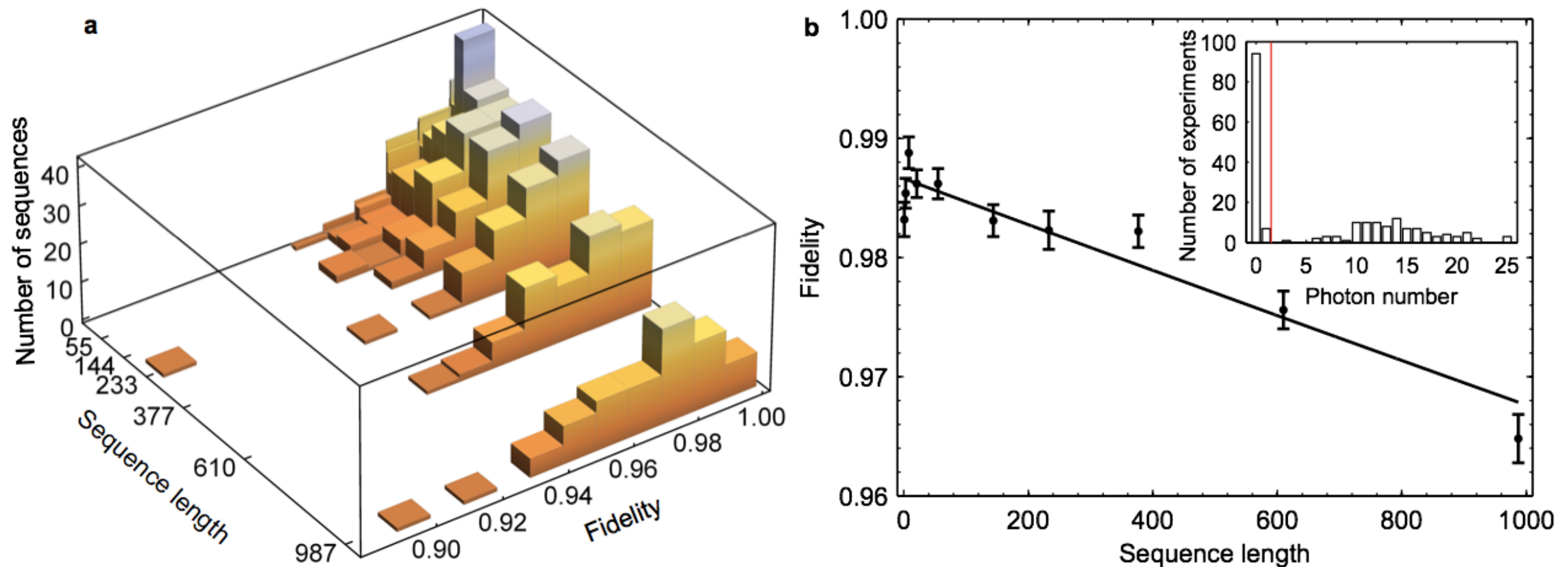


FIG. 2: (Color online) Results of the single-qubit benchmarking experiments. (a) Histogram of sequences of a given length with a given fidelity. Fidelity is discretized to 0.01 precision because 100 experiments were performed for each sequence. (b) Mean fidelity for each sequence length with error bars. The black trace is a least-squares fit to Eq. (1) yielding an EPG of $2.0(2) \times 10^{-5}$. (Inset) Summed histogram of bright and dark calibration experiments with a red line indicating the detection threshold.

Randomized Benchmarking in Practice

How many random sequences and repetitions are necessary?

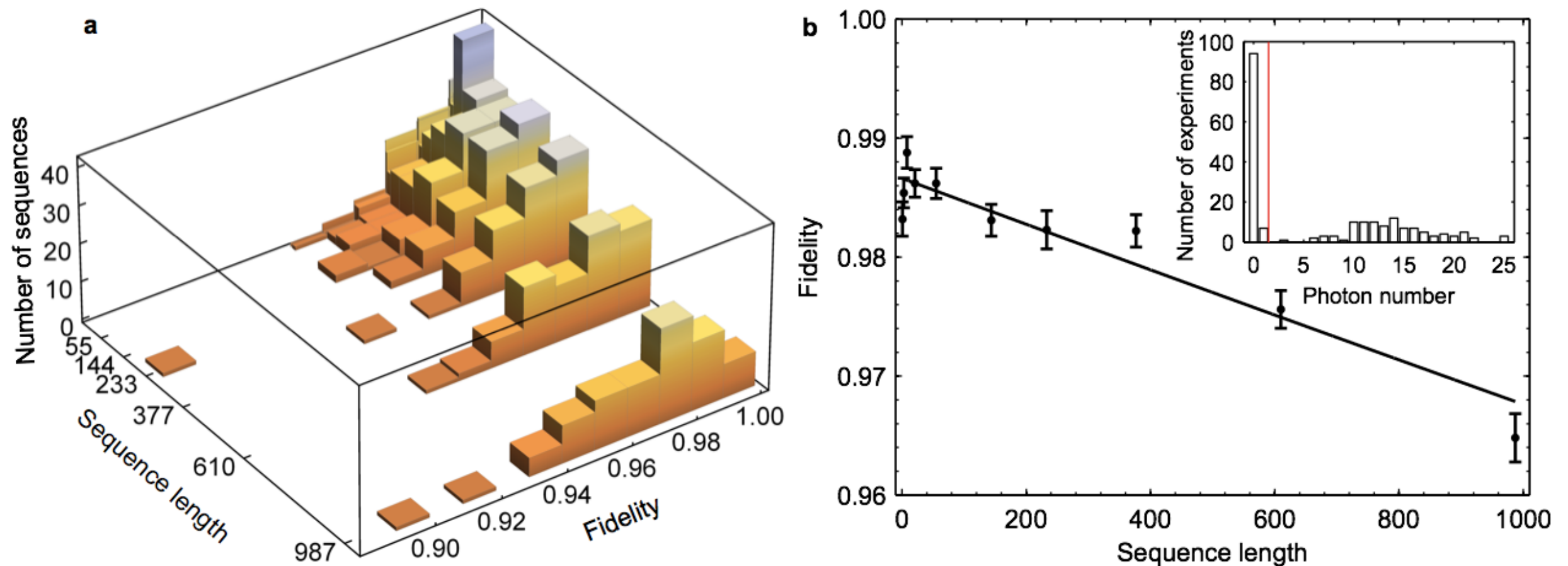
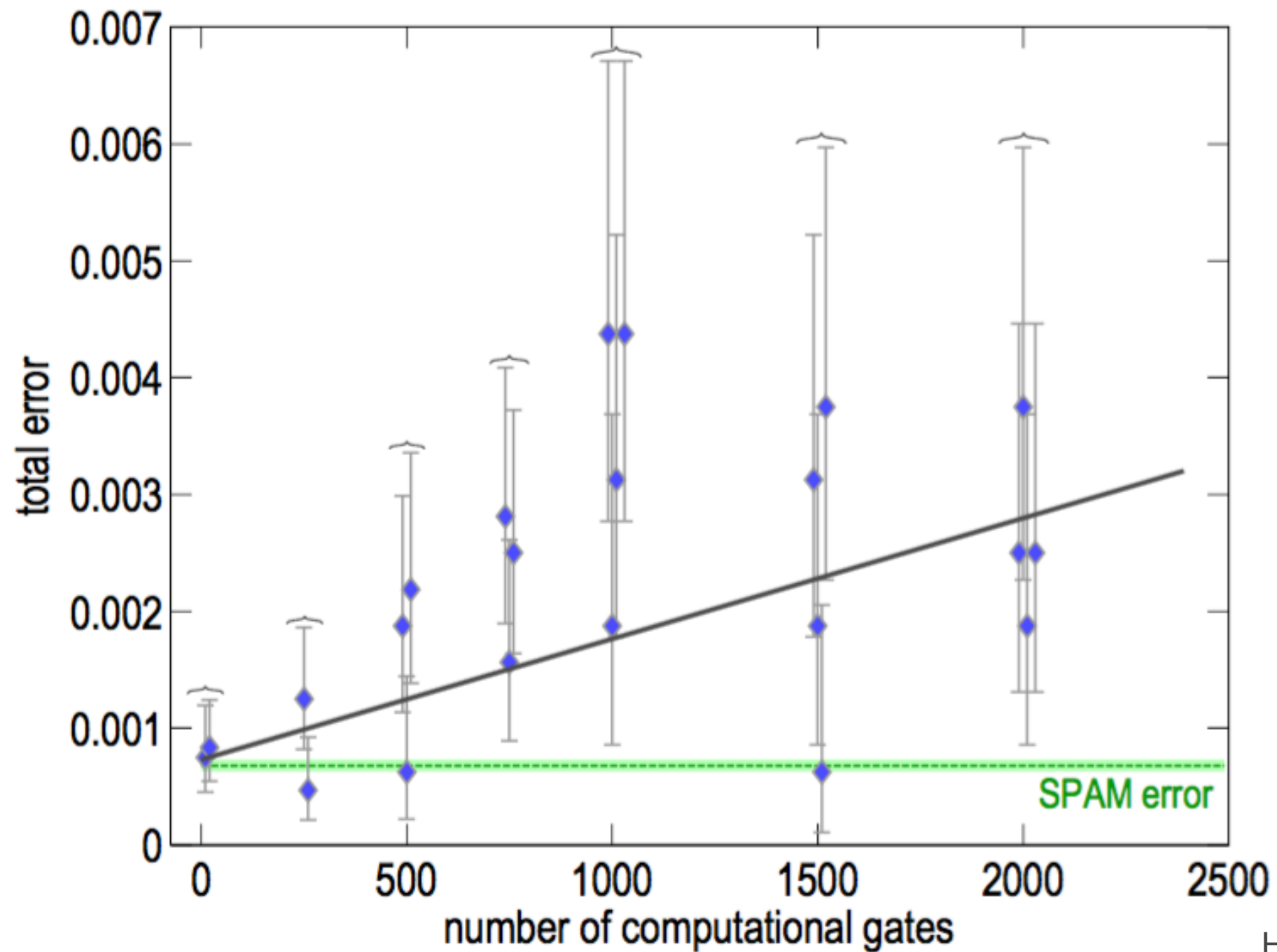


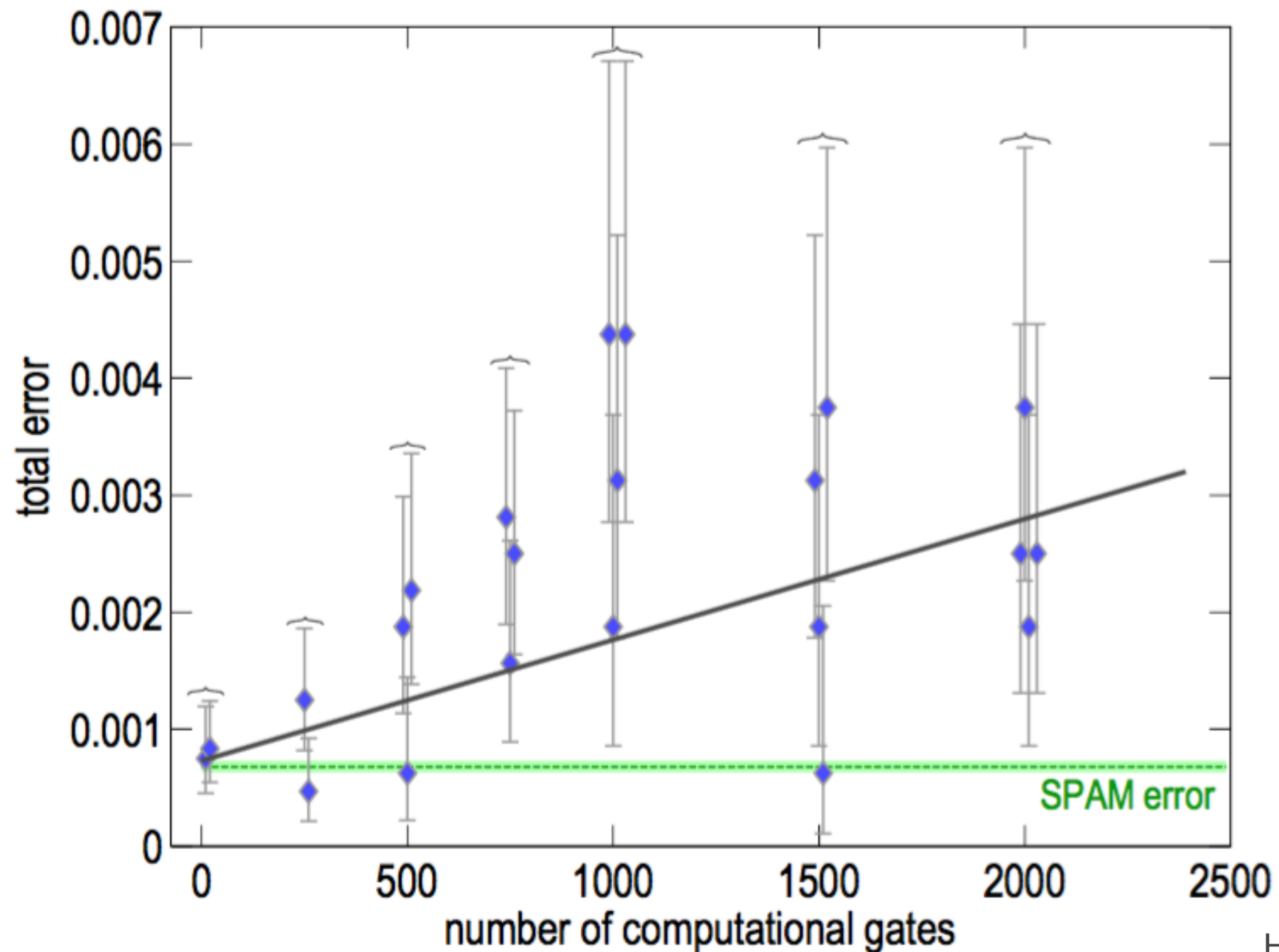
FIG. 2: (Color online) Results of the single-qubit benchmarking experiments. (a) Histogram of sequences of a given length with a given fidelity. Fidelity is discretized to 0.01 precision because 100 experiments were performed for each sequence. (b) Mean fidelity for each sequence length with error bars. The black trace is a least-squares fit to Eq. (1) yielding an EPG of $2.0(2) \times 10^{-5}$. (Inset) Summed histogram of bright and dark calibration experiments with a red line indicating the detection threshold.

Randomized Benchmarking in Practice



Randomized Benchmarking in Practice

How many random sequences and repetitions are necessary?



Randomized Benchmarking in Practice

Software solutions for randomized benchmarking estimation in practice:

- `nlinfit` (MATLAB) / `curve_fit` (SciPy)
Estimates least-squares fit for RB, but may report non-physical SPAM parameters (e.g.; $A = 1.5$ and $B = -0.5$).
- QInfer (Granade, Ferrie, & Cory 15)
Python • qinfer.org
Provides region estimates, Cramér-Rao bounds for RB experiments. Uses one shot per sequence to concentrate inference on relevant parameters.
- `pyGSTi` (Nielsen *et al.*)
Python • pygsti.info
Estimates RB as a special case of the more general but more expensive gate set tomography (GST) procedure.

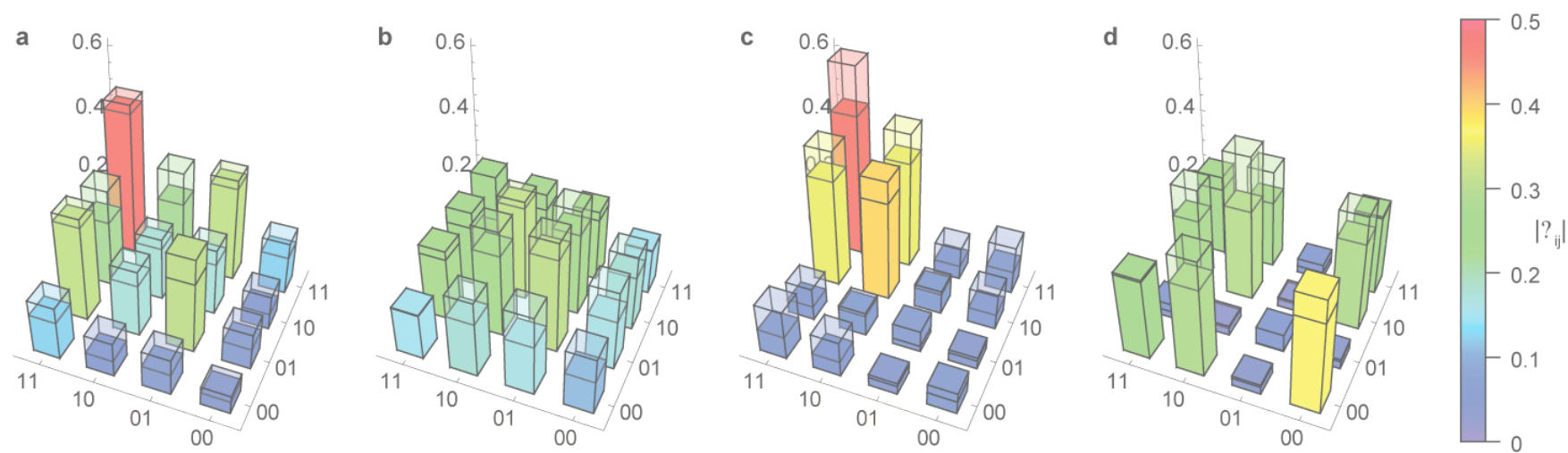


Figure: Hanneke et al.,
Nat. Phys. **6**, 13-16 (2009)

Characterization of quantum devices (part II)

Steve Flammia



THE UNIVERSITY OF
SYDNEY

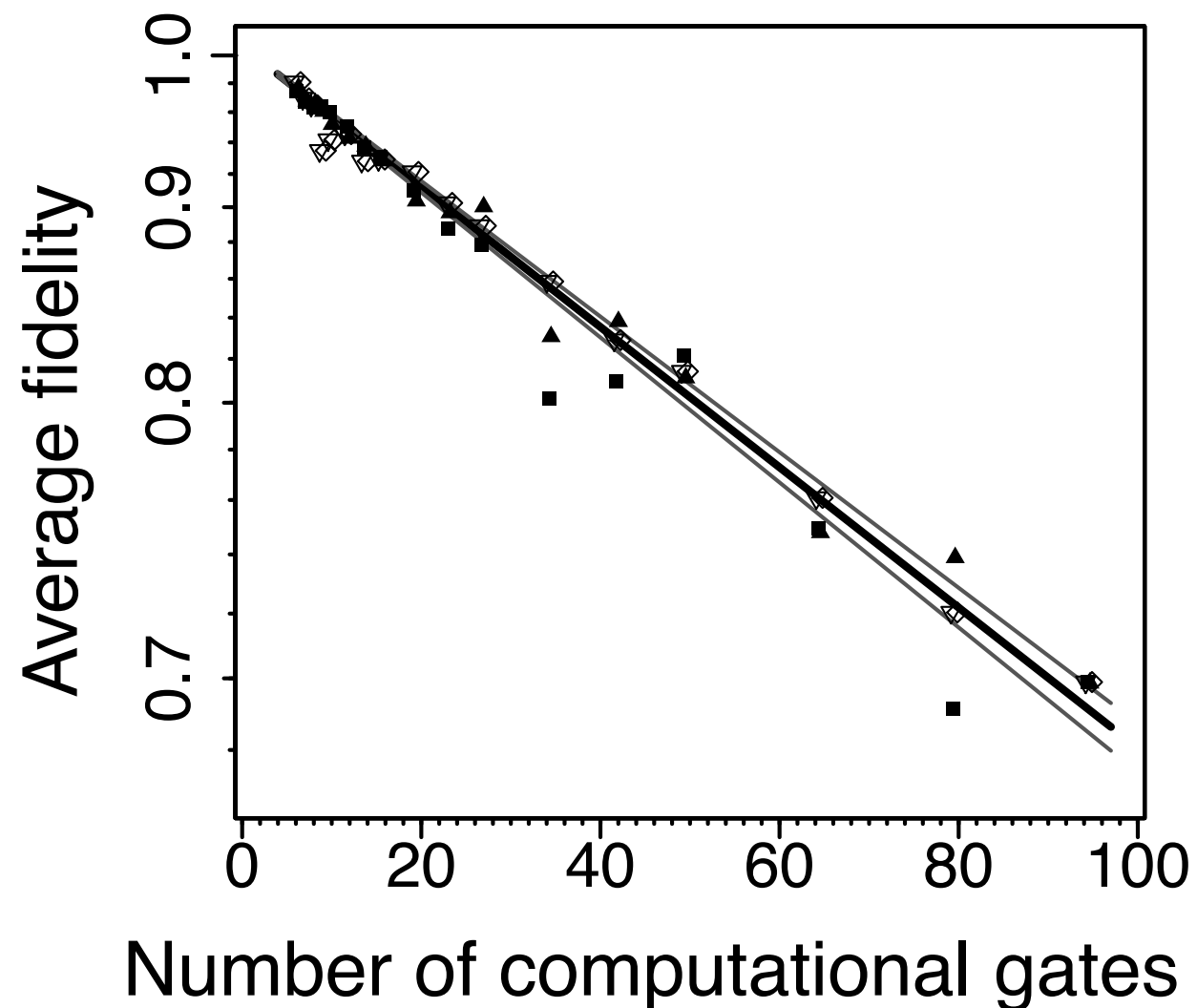
eQUS

ARC CENTRE OF EXCELLENCE FOR
ENGINEERED QUANTUM SYSTEMS

QIP Tutorial
Seattle
14 January 2017

Randomized Benchmarking

We saw how RB allows us to estimate the average error rate of our average Clifford noise, as long as certain natural assumptions are met.



- How expensive is benchmarking?
- Can it be used for estimating FTQC thresholds?
- What happens when we break the RB assumptions?
- Can it be extended to learn more about our noise?

RB with confidence

- ✿ Random Clifford circuits have small variance;
 - plug this into standard arguments to get guidance for how to choose # of sequences

- ✿ General upper bound

$$\sigma_m^2 \leq O(mr)$$

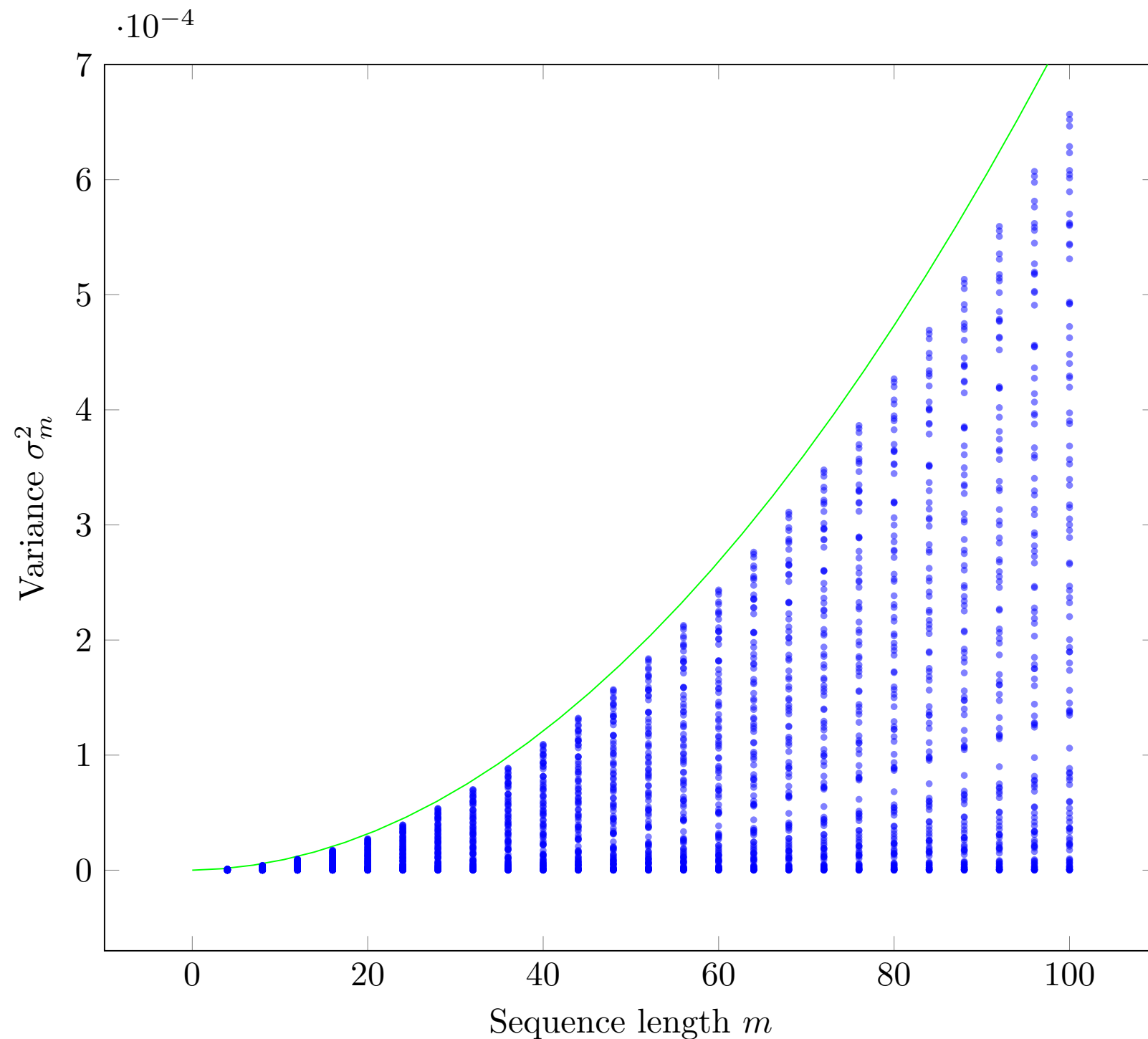
- ✿ For qubits, this can be improved to

$$\sigma_m^2 \leq O(m^2 r^2)$$

- ✿ When the noise is diagonal in the Pauli basis, this improves even further to

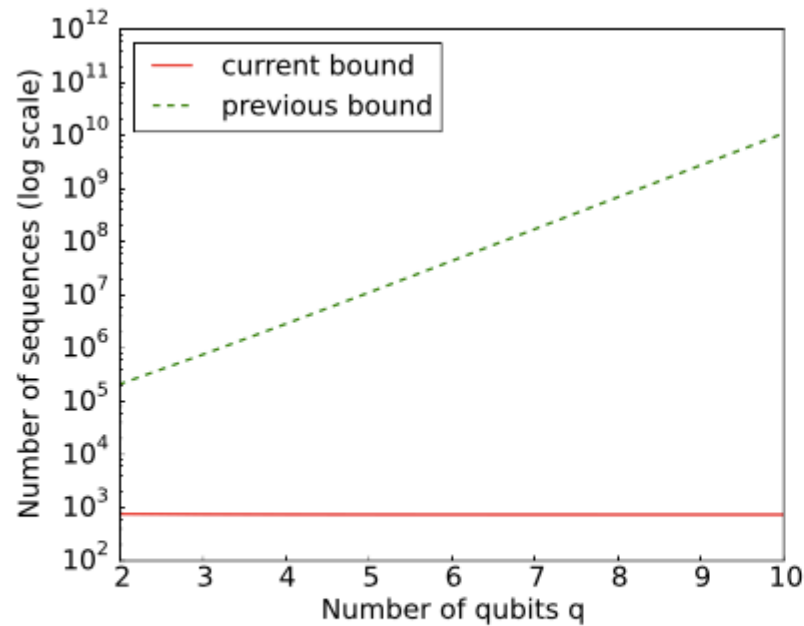
$$\sigma_m^2 \leq O(mr^2)$$

Tight bounds for qubits

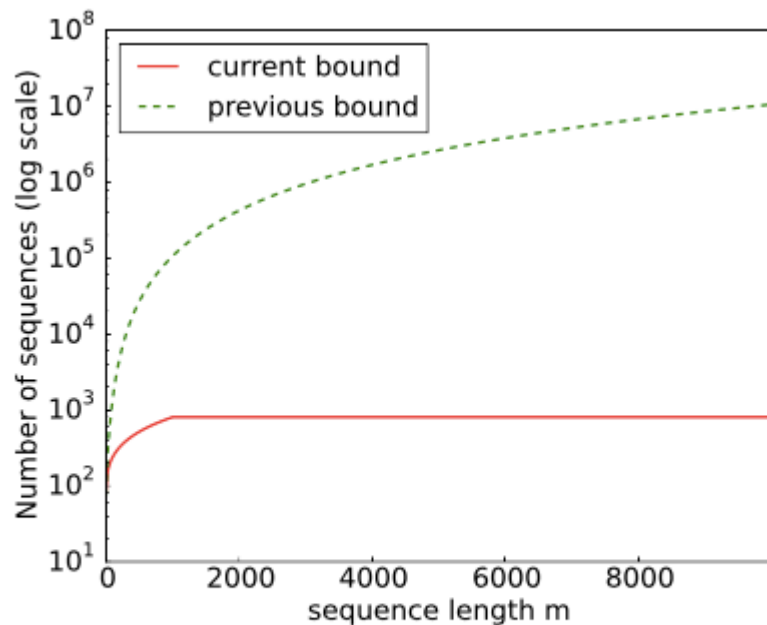


This result leads to estimates on the order of **< 100** sequences.

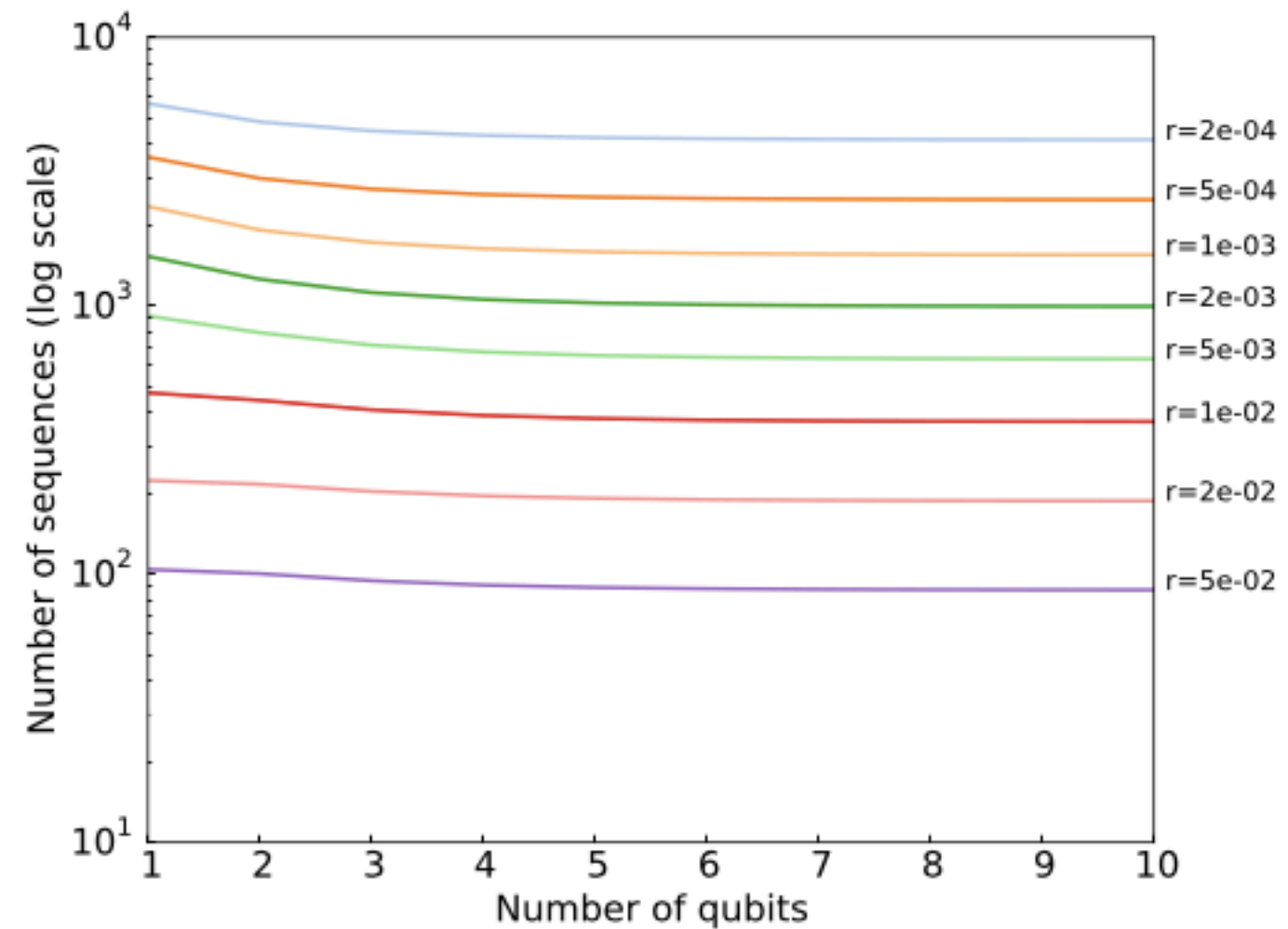
Tight bounds for **multiple** qubits



(a) Plot of the scaling with system size for eq. (15) and previous bounds [40].



(b) Plot showing scaling with sequence length for new (eqs. (10) and (11)) and previous [40] bounds



Accelerated RB

The variance at each sequence length in RB depends on two terms:

- **Within-sequence variance** ($\mathbb{E}_s [\text{Var}(p|s)]$):
Completely described by binomial sampling.
- **Inter-sequence variance** ($\text{Var}_s [\mathbb{E}(p|s)]$):
Described by RB complexity analysis.

If sequences are reused many times, then inter-sequence variance dominates.

Accelerated RB

If each sequence is used *once* before drawing new one, binomial sampling exactly describes RB performance.

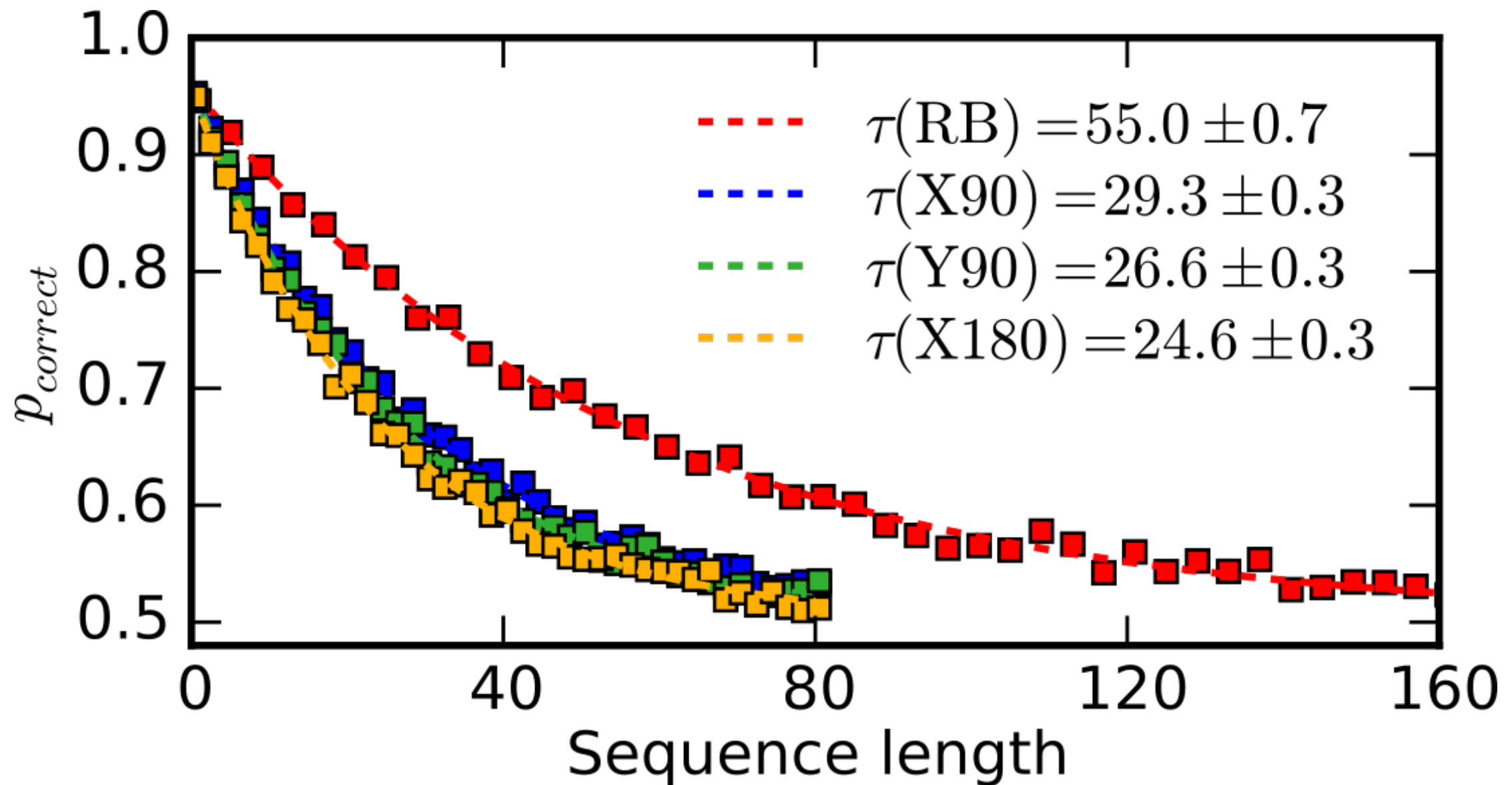
This is **optimal** for learning p !

inter-sequence variance is sensitive to coherence of noise and other effects, rather than p itself.

Bonus: Cramér-Rao bound is easy to evaluate in accelerated (single-shot) limit.

- Optimal sequence length ($\approx 1 / r$).
- Formal test on achievable precision for single-shot RB experiments.

Randomized Benchmarking in Practice



Single-shot limit is achievable!

Comparison to the FTQC threshold

Fault-tolerant thresholds are proven in terms of D , but we only estimate r in RB experiments.

Problem: these are only loosely related!

$$\frac{d+1}{d}r \leq D \leq d\sqrt{\frac{d+1}{d}r}$$

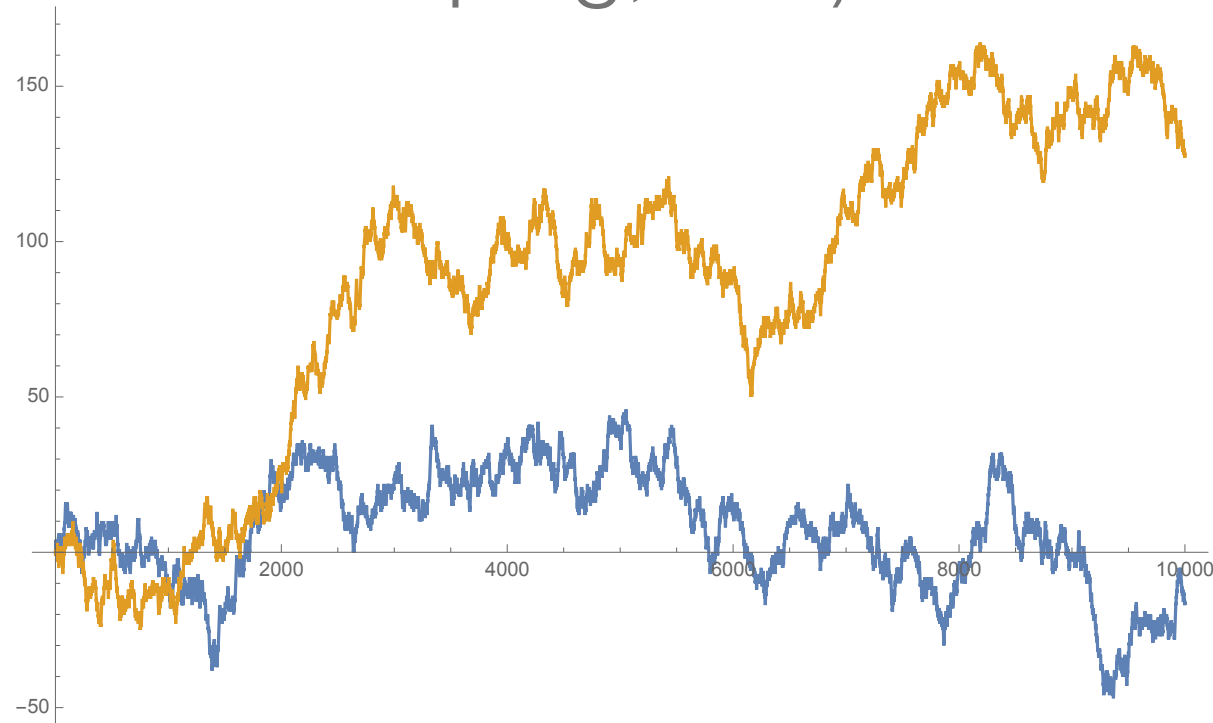
Which types of noise are most detrimental for **fault-tolerance**?

In particular, **Pauli noise**, where most thresholds are estimated, saturates the *lower bound*, which could be orders of magnitude off from D !

Coherent noise

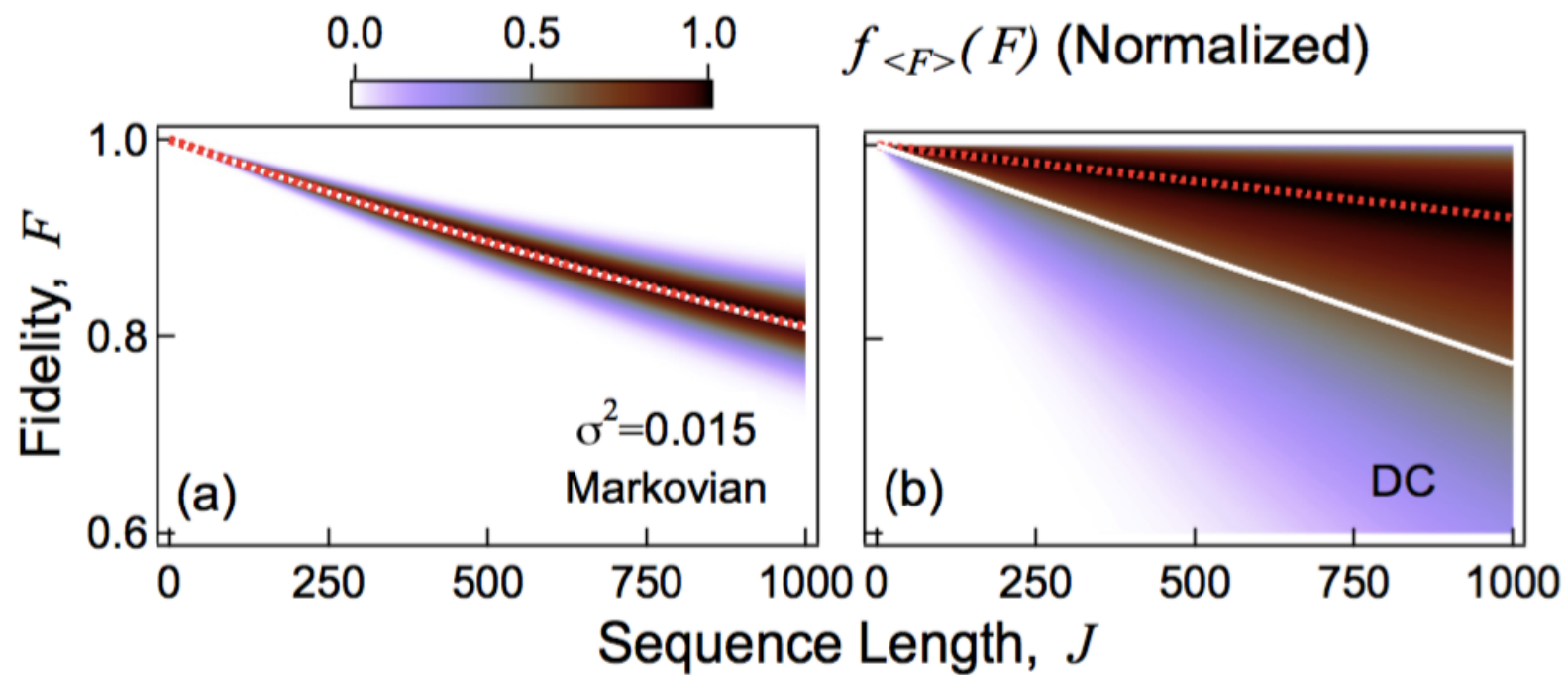
Benchmarking measures the **average error rate**.

For the same fixed average error, this quantity does not care if the noise is **coherent** (unitary) or **incoherent** (e.g. dephasing, amplitude damping, etc.)

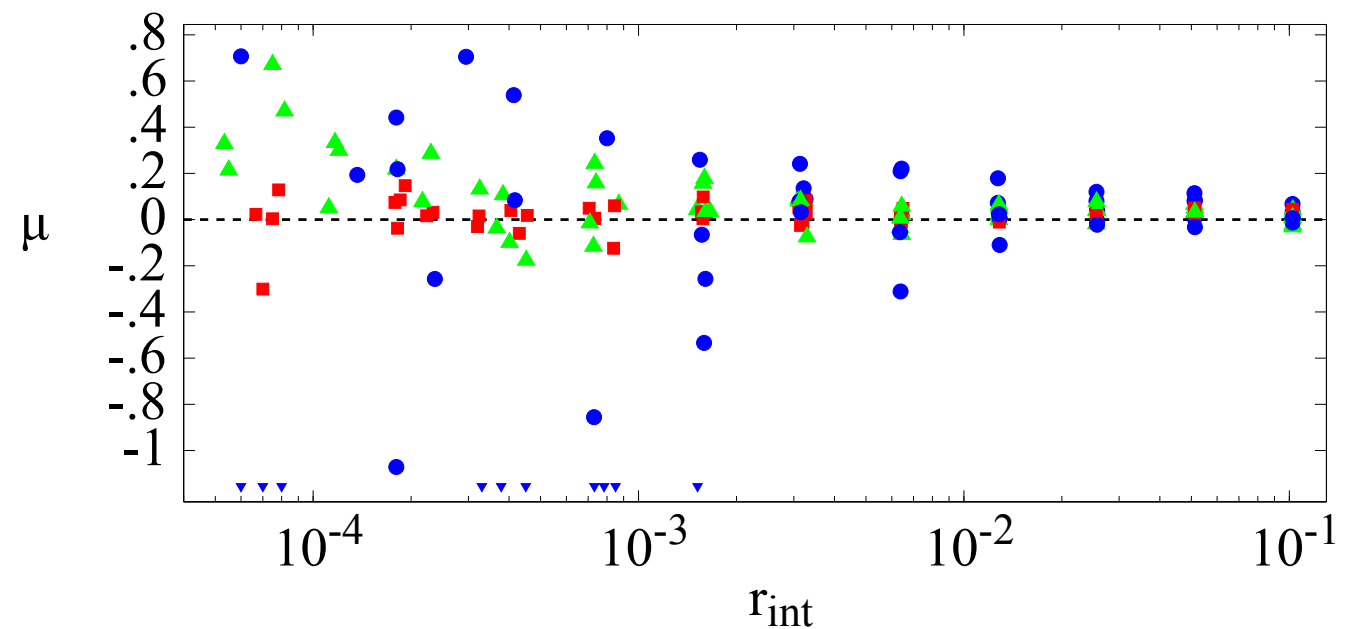
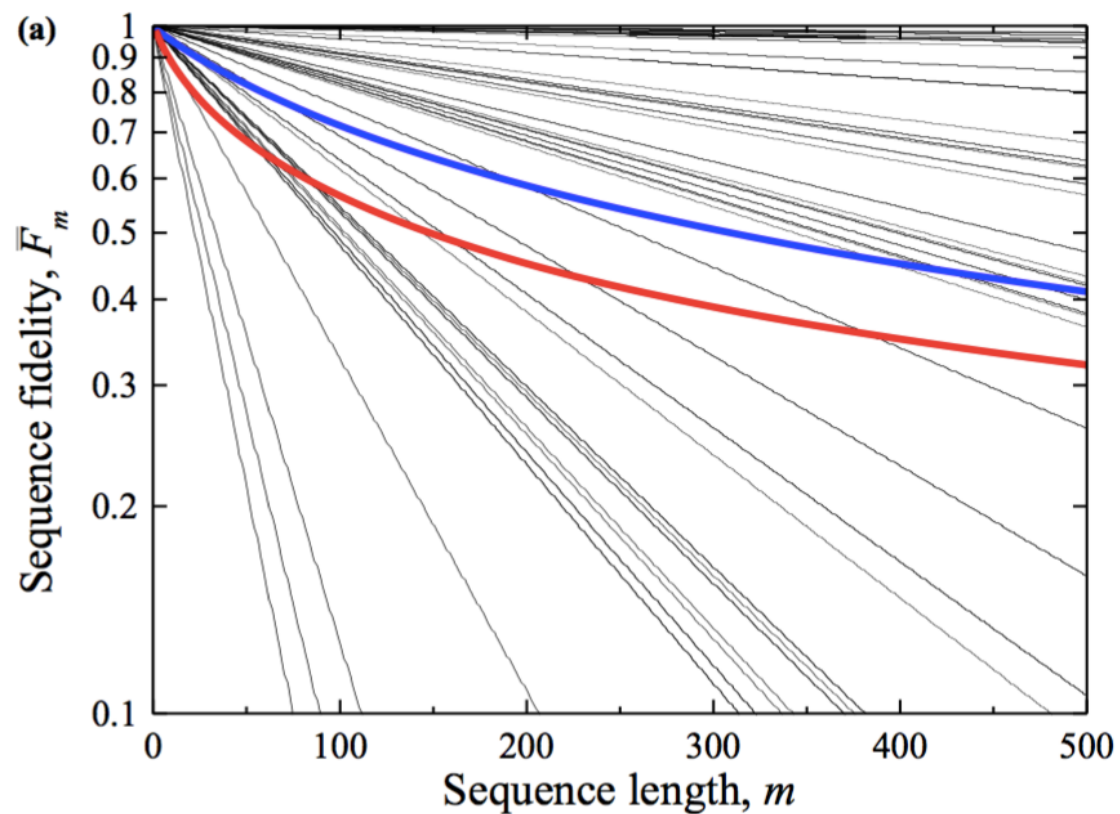


These two noise types have radically different effects on the **worst-case error**, that is relevant for fault-tolerance thresholds

Coherent errors and gate dependent noise



The average and worst-case error can be radically different!

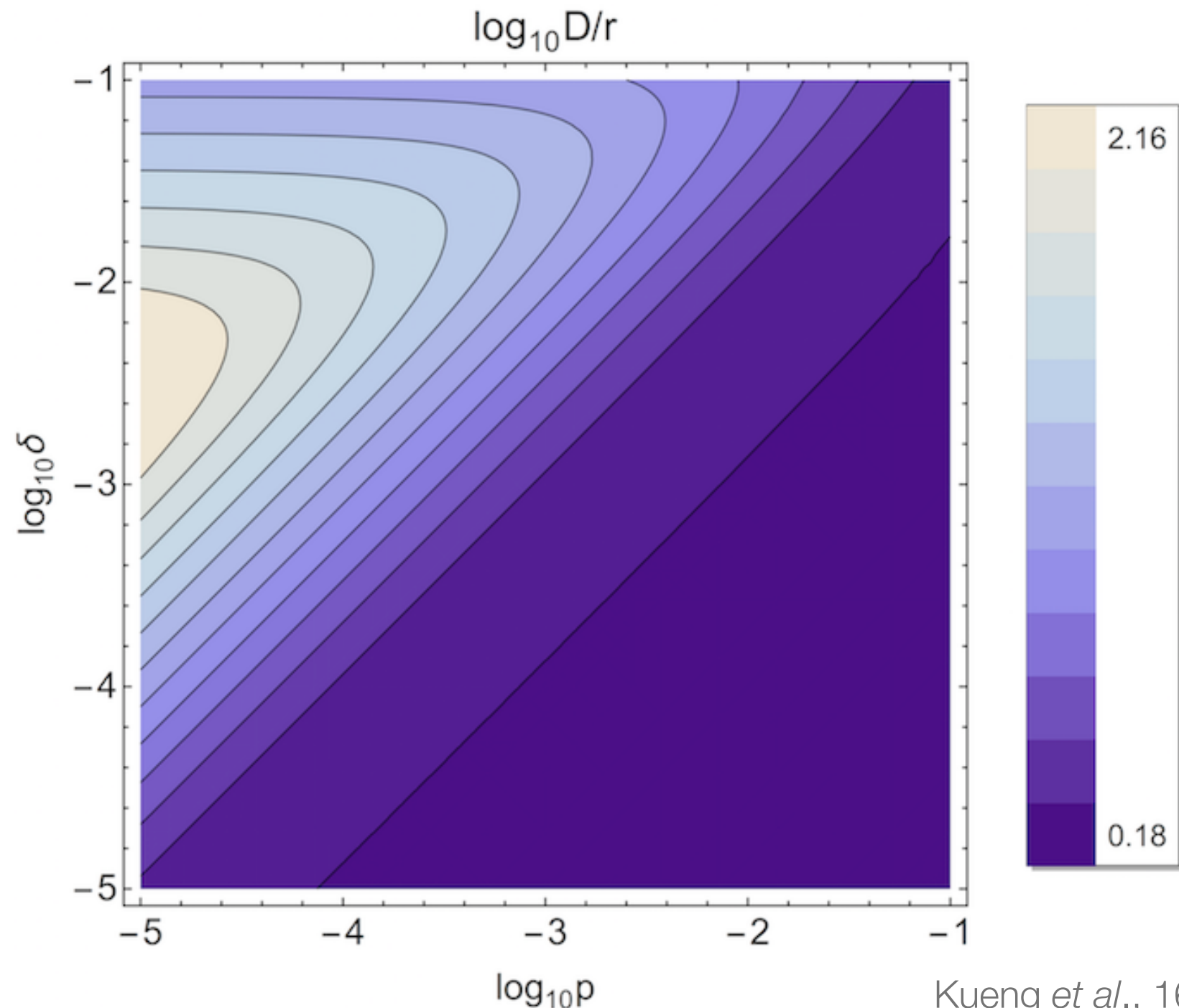


Coherent errors are the culprit

$$\mathcal{E}(\rho) = p\sigma_z e^{-i\delta\sigma_z} \rho e^{i\delta\sigma_z} \sigma_z + (1-p)e^{-i\delta\sigma_z} \rho e^{i\delta\sigma_z}$$

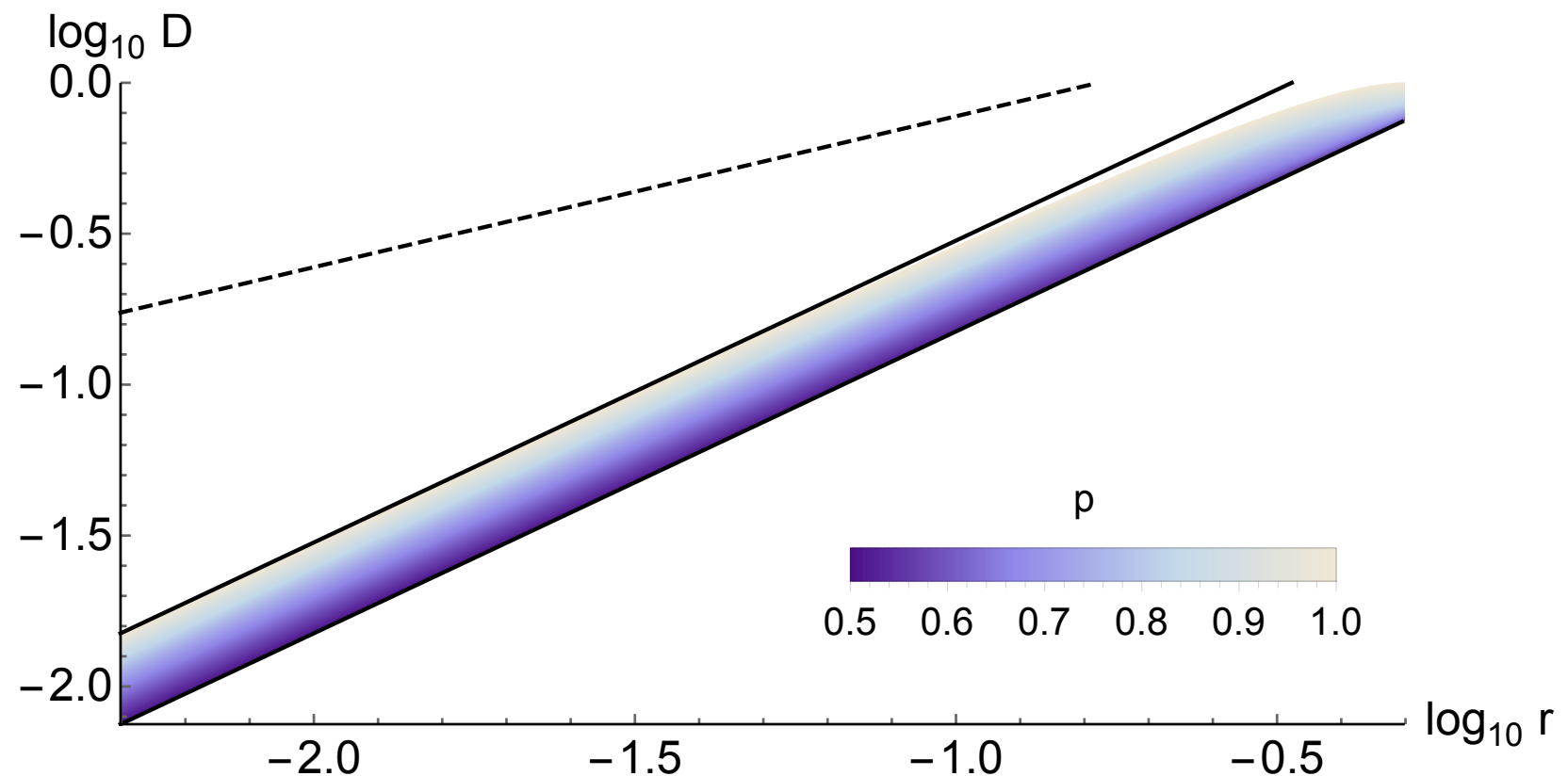
This noise interpolates between low-frequency (unitary) and high-frequency (stochastic) noise.

Unless unitary rotation angles are at least as small as stochastic error, they **dominate** the noise scaling.



Amplitude damping is more benign

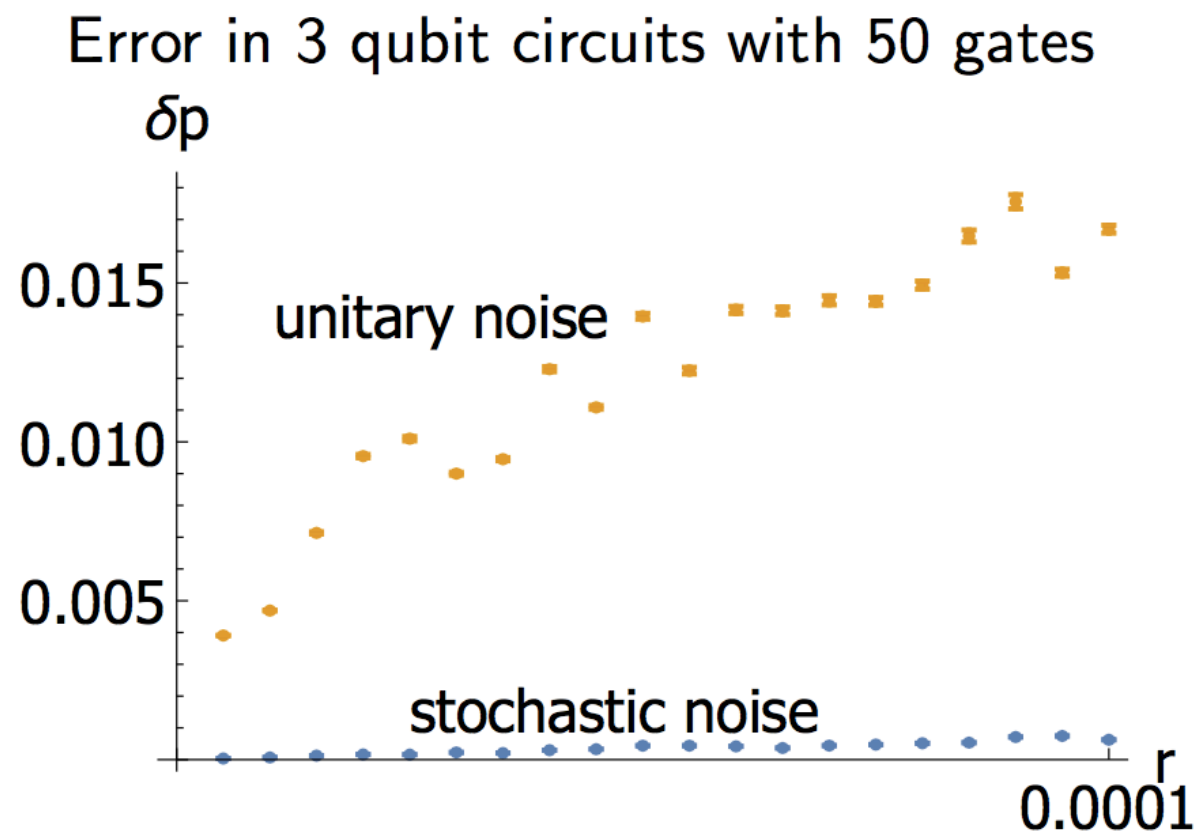
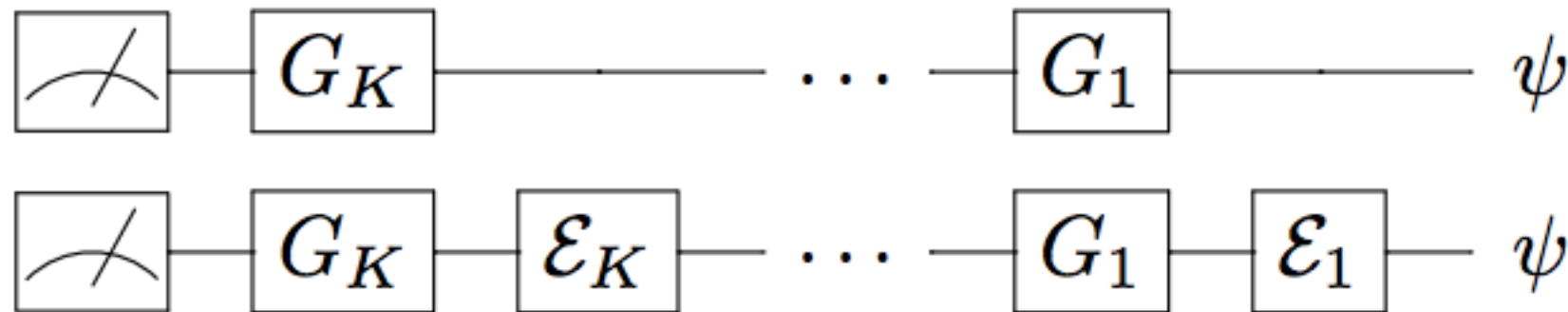
Amplitude damping is **not** the problem.



Using SDP methods, we get **exact** and analytic bounds

The main idea is to use semidefinite programs to find dual feasible points

How errors accumulate: coherence matters!



Given an *ideal* and a *noisy* implementation of the same circuit, what is the difference in the success probability?

Estimating the Coherence of Noise

For simplicity, let me restrict **trace-preserving** and **unital** noise, i.e. noise for which the maximally mixed state is a fixed point.

$$u(\mathcal{E}) = \frac{1}{d^2 - 1} [\text{Tr}(\mathcal{E}^\dagger \mathcal{E}) - 1]$$

The **unitarity** is a measure of noise that is **only** sensitive to incoherent noise. For any noise map \mathcal{E} and unitary \mathcal{U} ,

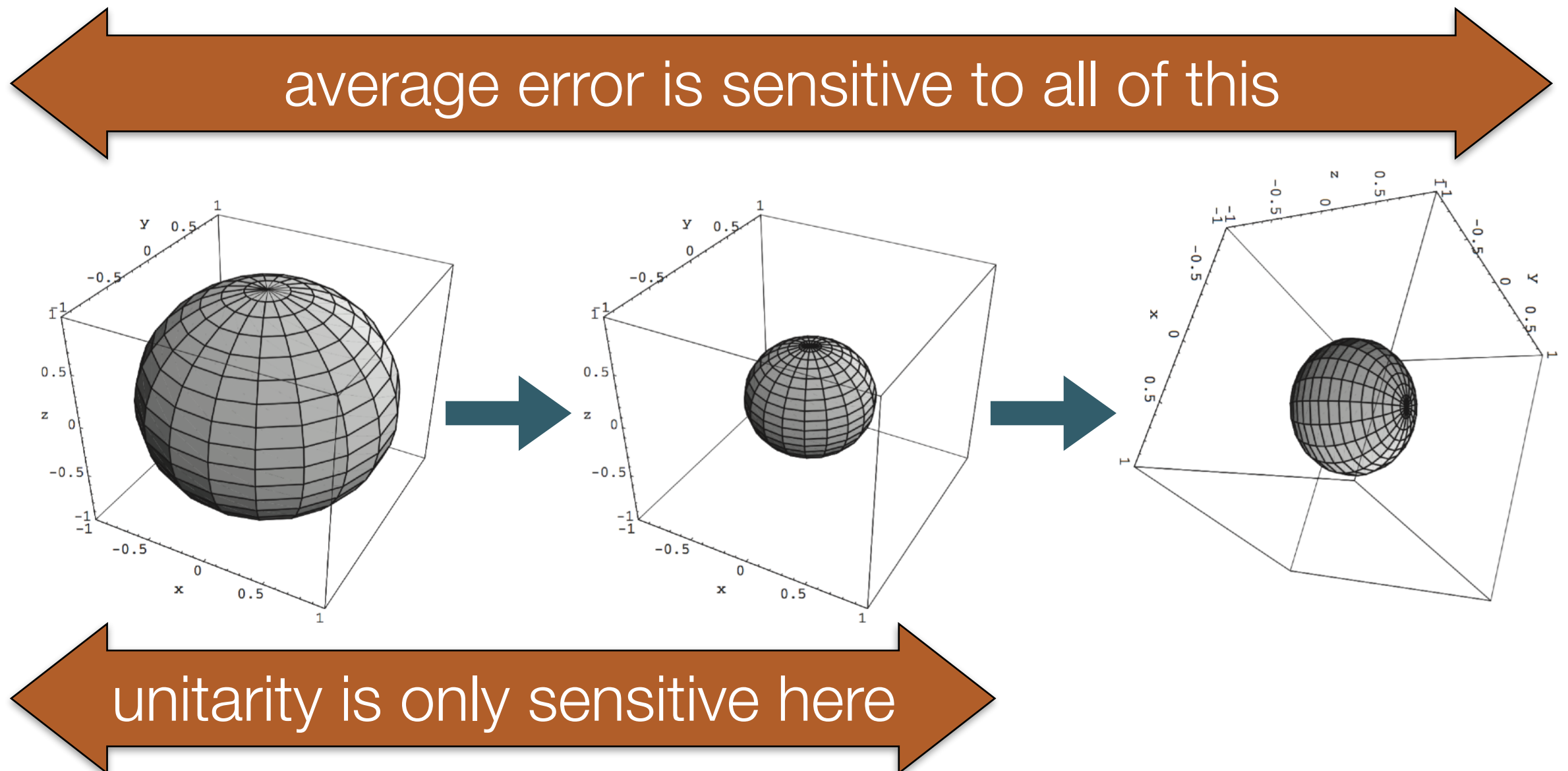
$$u(\mathcal{E}) = u(\mathcal{U}\mathcal{E}) = u(\mathcal{E}\mathcal{U})$$

Because this quantity is quadratic in the noise, this acts like second moment information about the noise, and helps us distinguish **average-** and **worst-case** behavior.

(The definition generalizes for arbitrary noise, but is marginally more complicated.)

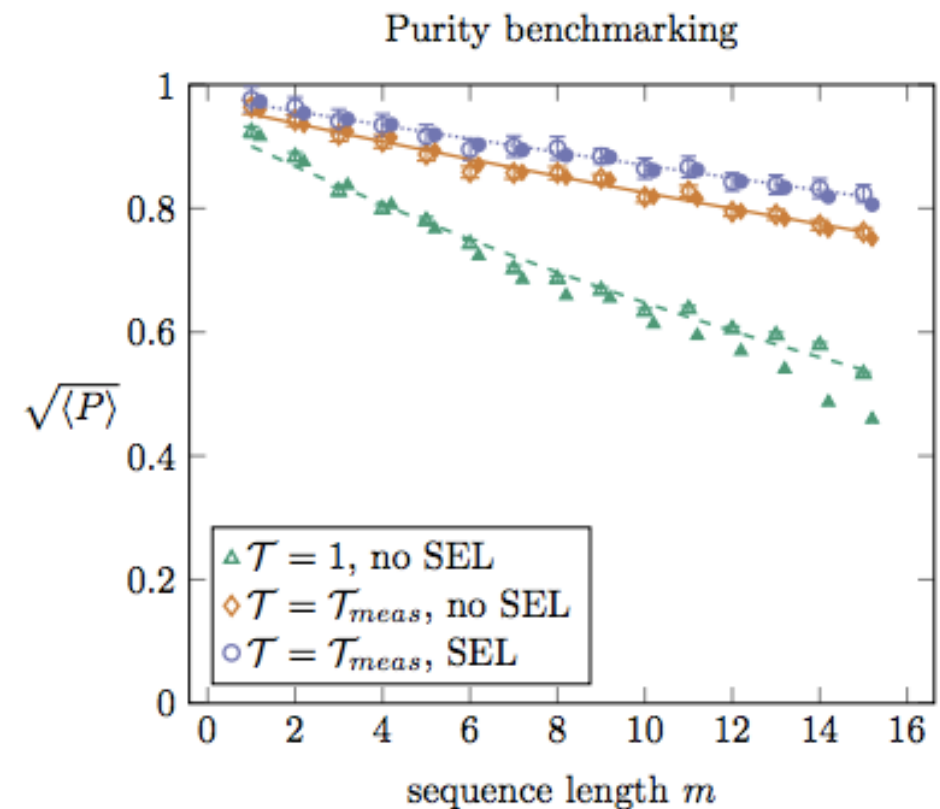
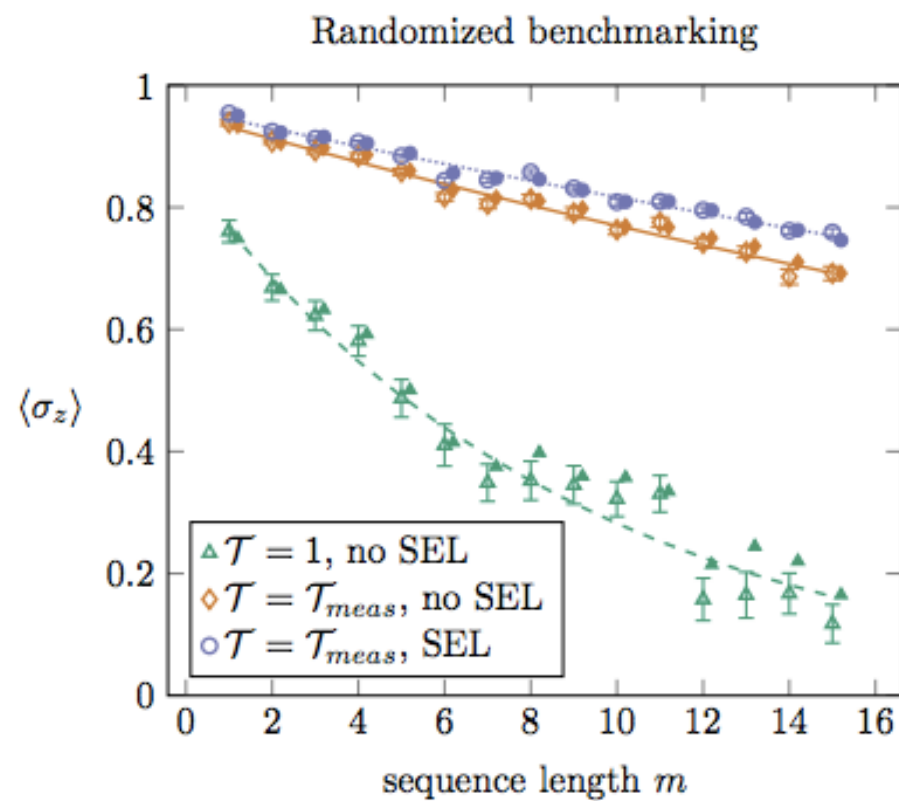
Estimating the Coherence of Noise

Ex: For a natural noise model with dephasing noise plus unitary rotation, the unitarity is sensitive to the **shrinkage** but not the **rotation**.



Purity Benchmarking

The unitarity can be estimated via **purity benchmarking**, an RB-like experiment that estimates a decay rate.



Purity Benchmarking

The unitarity can be estimated via **purity benchmarking**, an RB-like experiment that estimates a decay rate.

$$\mathbb{E}_s [\text{Tr}[\rho^2]] = A u(\Lambda)^{m-1} + B$$

Unitarity bounds the diamond distance up to constant factors:

$$c_d \sqrt{u + \frac{2dr}{d-1} - 1} \leq D \leq d^2 c_d \sqrt{u + \frac{2dr}{d-1} - 1},$$

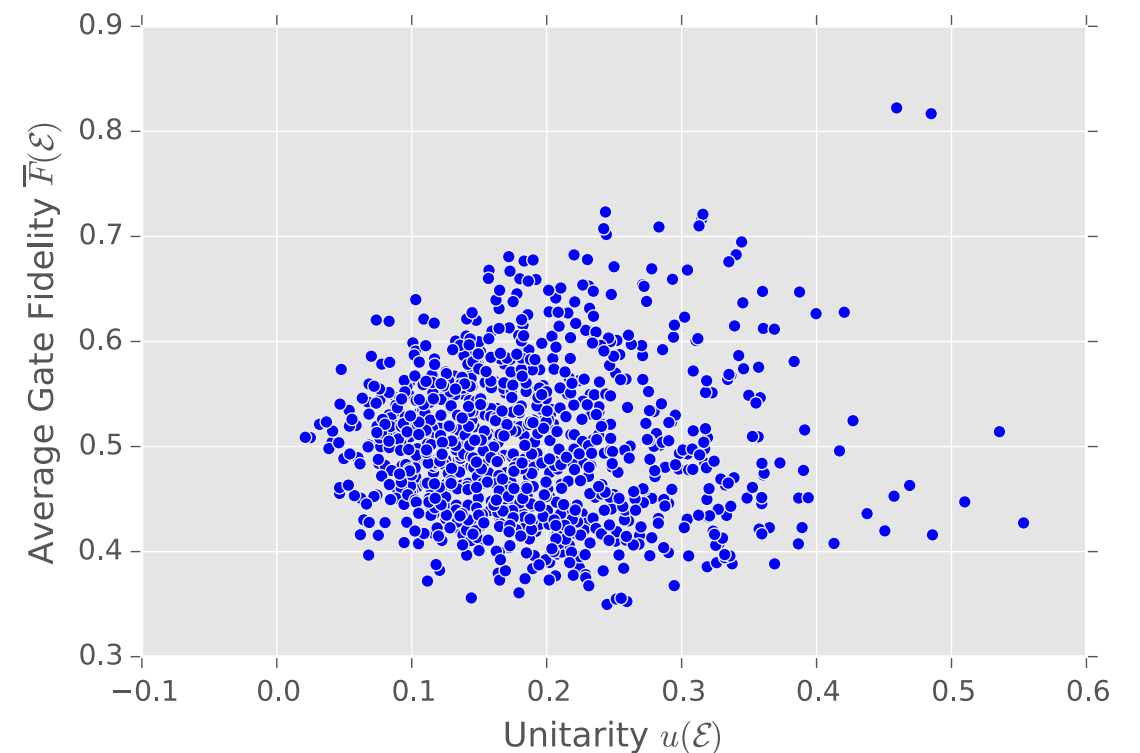
Purity Benchmarking

The unitarity can be estimated via **purity benchmarking**, an RB-like experiment that estimates a decay rate.

It correlates with, but is **distinct** from average error rate:

Given a fixed average error rate r , the unitarity u cannot be too small:

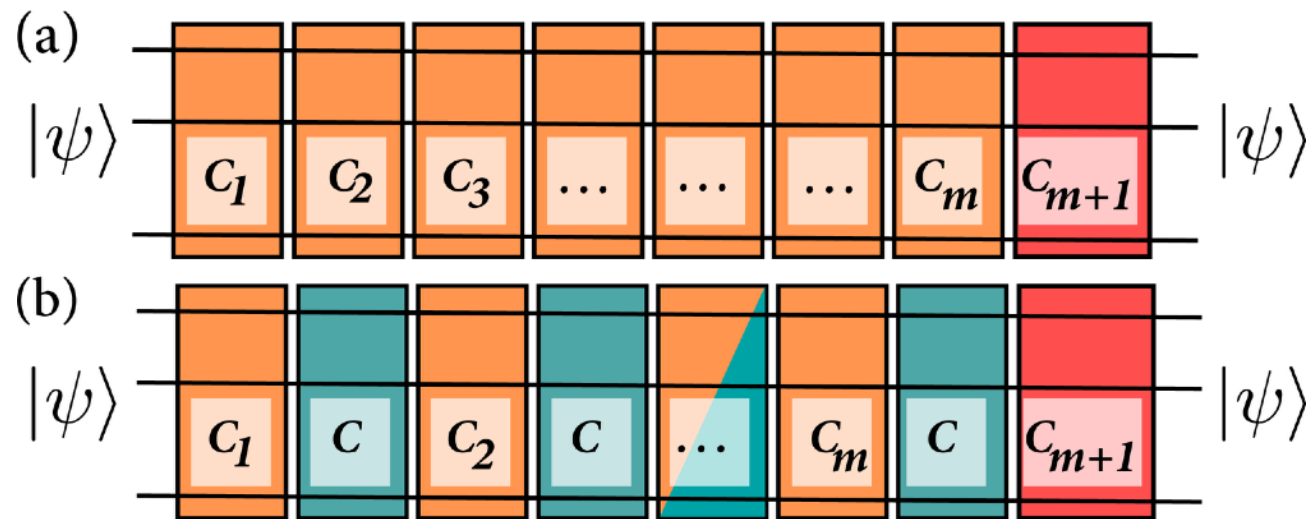
$$u(\mathcal{E}) \geq \left[1 - \frac{d}{d-1} r(\mathcal{E}) \right]^2$$



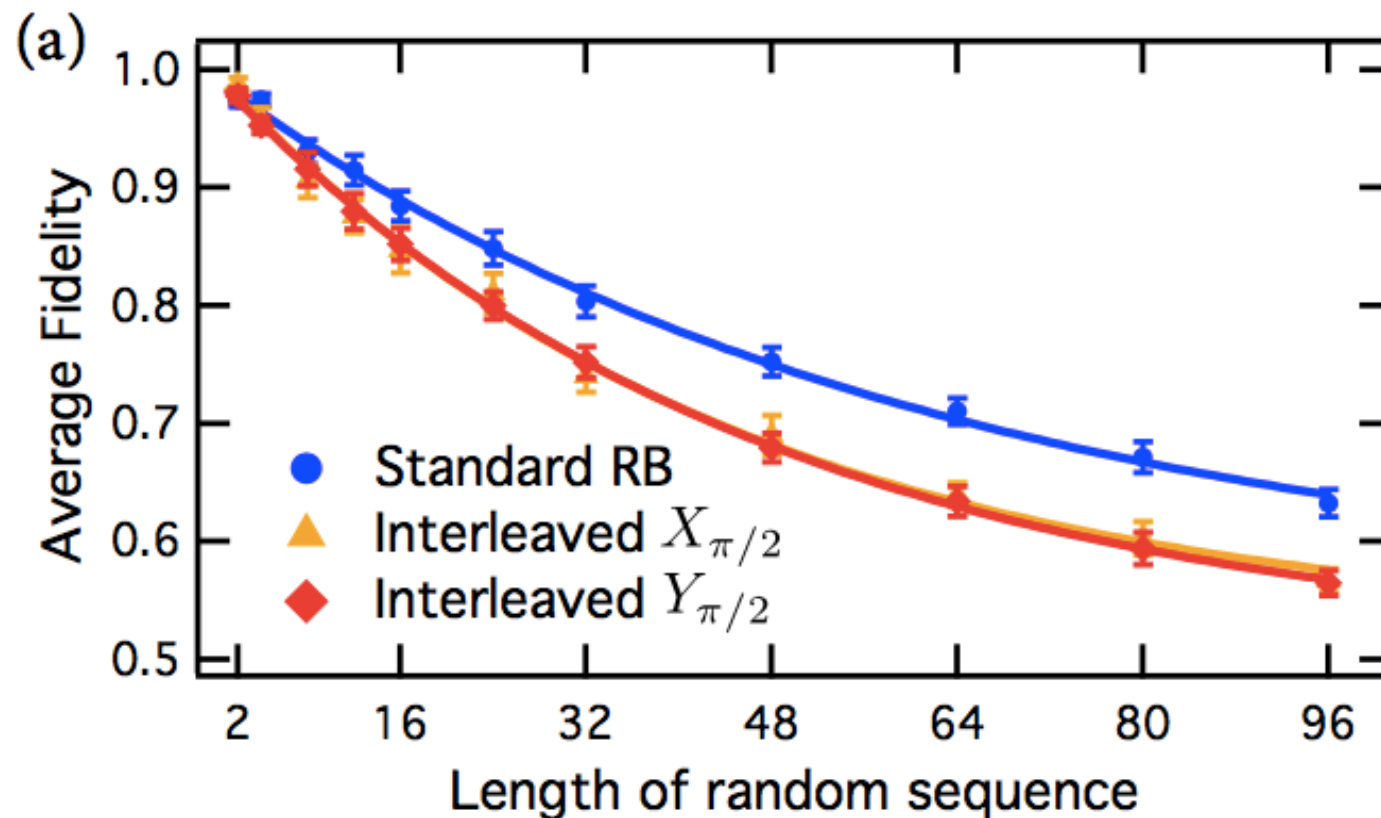
It also provides a bound on the best possible average error rate achievable via unitary control:

$$\frac{d-1}{d} \left[1 - \sqrt{u(\mathcal{E})} \right] \leq \min_{\mathcal{U}} r(\mathcal{U}\mathcal{E})$$

Interleaved RB



Do two experiments: a control experiment with random Cliffords, and a second “interleaved” experiment with a specific fixed gate after each random Clifford.

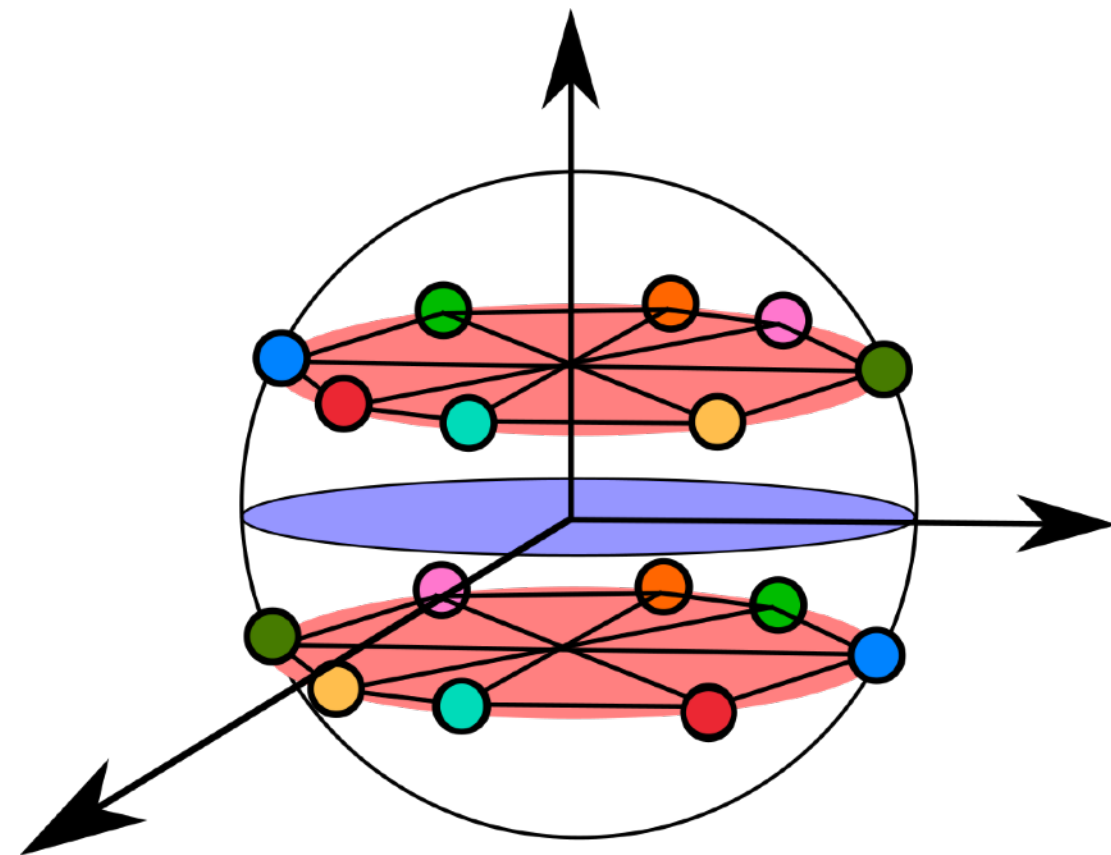


Under some assumptions, comparing the two decay curves allows us to back out an estimate of the fidelity of the **specific interleaved gate**.

Characterizing T gates

We also need to be able to characterize non-Clifford gates such as T .

- **Dihedral RB** (Dugas *et al.* 2015, Cross *et al.* 2016): Use dihedral or dihedral-CNOT groups instead of Clifford group to do benchmarking. The T gate is now directly an element of the group.

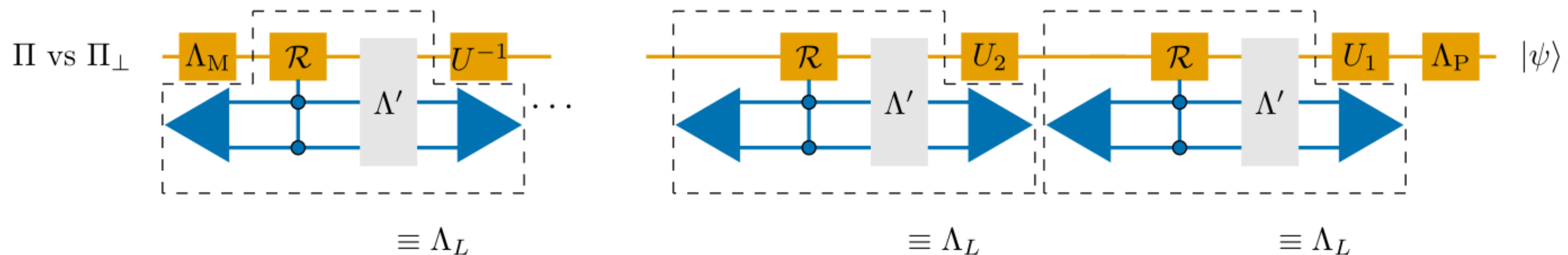


- **CTPT Benchmarking** (Harper & F. 2016): Carefully arrange Clifford and Pauli gates to take advantage of the fact that TPT is always a Clifford, giving traditional interleaved benchmarking.

Leakage and Logical RB

Two other extensions allow us to characterize leakage rates and logical error rates

- **Leakage RB** (Wallman *et al.* 2015): Change final measurement to identity, don't do final inverse. Decay then reflects leakage out of subspace represented by identity measurement.
- **Logical RB** (Lu *et al.* 2015 (exp); Combes *et al.* 2016 (th)): Perform randomized benchmarking on logical channels, rather than physical.



Direct Fidelity Estimation

We've seen that RB is a flexible and (in some cases) reliable method of estimating noise in quantum systems.

There is one other method that is well adapted to larger scale circuits, but has the drawback that it can be susceptible to SPAM errors

Main idea:

Monte Carlo estimation of fidelity by expressing the fidelity function in a simple operator basis.

“Direct” Fidelity Estimation: a trivial algorithm

First recall the definition of fidelity.

The **fidelity** with respect to a **pure** state Ψ is given by:

$$F(\rho, \psi) = \text{Tr}(\rho\psi) = \langle \psi | \rho | \psi \rangle$$

If we allow arbitrary quantum operations, we can do the following:

- ✱ We are given many copies of the state ρ
- ✱ Make the two-outcome measurement $\{\psi, 1 - \psi\}$
- ✱ Repeat $O(1/\epsilon^2)$ times
- ✱ Average the measurements to get an estimate $\hat{F} = F \pm \epsilon$

This requires a quantum computer!

We want the quantum computational complexity to be $O(1)$!

Direct Fidelity Estimation

First **expand** ρ in the Pauli basis:

$$\rho = \sum_j \frac{\text{Tr}(\rho \hat{\sigma}_j)}{\sqrt{d}} \frac{\hat{\sigma}_j}{\sqrt{d}} = \sum_j \chi_\rho(j) \frac{\hat{\sigma}_j}{\sqrt{d}}$$

Direct Fidelity Estimation

First **expand** ρ in the Pauli basis:

$$\rho = \sum_j \frac{\text{Tr}(\rho \hat{\sigma}_j)}{\sqrt{d}} \frac{\hat{\sigma}_j}{\sqrt{d}} = \sum_j \chi_\rho(j) \frac{\hat{\sigma}_j}{\sqrt{d}}$$

The **fidelity** with respect to a pure state Ψ is given by:

$$F(\rho, \psi) = \text{Tr}(\rho\psi) = \langle \psi | \rho | \psi \rangle = \sum_j \chi_\rho(j) \chi_\psi(j)$$

Direct Fidelity Estimation

First **expand** ρ in the Pauli basis:

$$\rho = \sum_j \frac{\text{Tr}(\rho \hat{\sigma}_j)}{\sqrt{d}} \frac{\hat{\sigma}_j}{\sqrt{d}} = \sum_j \chi_\rho(j) \frac{\hat{\sigma}_j}{\sqrt{d}}$$

The **fidelity** with respect to a pure state Ψ is given by:

$$F(\rho, \psi) = \text{Tr}(\rho \psi) = \langle \psi | \rho | \psi \rangle = \sum_j \chi_\rho(j) \chi_\psi(j)$$

For a *pure* state Ψ , $\text{Tr}(\Psi^2)=1$, so we have a natural **probability distribution**: $\text{Pr}(j) = [\chi_\psi(j)]^2$

Direct Fidelity Estimation

First **expand** ρ in the Pauli basis:

$$\rho = \sum_j \frac{\text{Tr}(\rho \hat{\sigma}_j)}{\sqrt{d}} \frac{\hat{\sigma}_j}{\sqrt{d}} = \sum_j \chi_\rho(j) \frac{\hat{\sigma}_j}{\sqrt{d}}$$

The **fidelity** with respect to a pure state Ψ is given by:

$$F(\rho, \psi) = \text{Tr}(\rho\psi) = \langle \psi | \rho | \psi \rangle = \sum_j \chi_\rho(j) \chi_\psi(j)$$

For a *pure* state Ψ , $\text{Tr}(\Psi^2)=1$, so we have a natural **probability distribution**: $\text{Pr}(j) = [\chi_\psi(j)]^2$

Rewrite the fidelity: $F(\rho, \psi) = \mathbb{E}_j [X]$ with $X = \frac{\chi_\rho(j)}{\chi_\psi(j)}$

Direct Fidelity Estimation

Thus, the fidelity can be computed by *sampling* the random variable X and averaging over many trials!

Moreover, the variance is small:

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \text{Tr}(\rho^2) - F^2 \leq 1$$

By Chebyshev's inequality, to achieve $\Pr[|\hat{F} - F| \geq \epsilon] \leq \delta$

we only need $O(1/\epsilon^2\delta)$ independent samples, which is *constant*.

Thus, the *measurement complexity* depends only on the precision and the confidence and not on the system size.

Direct Fidelity Estimation: the caveats

Caveat 1:

The (classical) computational complexity depends on d .

Sampling from the relevance distribution is in general a **hard computation task** and will take exponential time on a classical computer for a generic random state.

Direct Fidelity Estimation: the caveats

Caveat 1:

The (classical) computational complexity depends on d .

The silver lining:

Direct Fidelity Estimation: the caveats

Caveat 1:

The (classical) computational complexity depends on d .

The silver lining:

- ✱ Sampling is all done as preprocessing

Direct Fidelity Estimation: the caveats

Caveat 1:

The (classical) computational complexity depends on d .

The silver lining:

- ✱ Sampling is all done as preprocessing
- ✱ No need for complicated data analysis

Direct Fidelity Estimation: the caveats

Caveat 1:

The (classical) computational complexity depends on d .

The silver lining:

- ✱ Sampling is all done as preprocessing
- ✱ No need for complicated data analysis
- ✱ Sampling can be done in parallel

Caveat 2:

The sample complexity depends on d .

We can only learn X up to some finite precision by repeatedly measuring Pauli operators... in general we will need to **measure many times** to resolve the bias if the intended state was chosen completely at random.

Direct Fidelity Estimation: complexity

Direct Fidelity Estimation: complexity

- ✿ If the noise is **dephasing** (or depolarizing) then we get a very favorable scaling:

$$O\left(\frac{d}{\epsilon^2} \log \frac{1}{\delta}\right)$$

Direct Fidelity Estimation: complexity

- ✿ If the noise is **dephasing** (or depolarizing) then we get a very favorable scaling:

$$O\left(\frac{d}{\epsilon^2} \log \frac{1}{\delta}\right)$$

Direct Fidelity Estimation: complexity

- ✱ If the noise is **dephasing** (or depolarizing) then we get a very favorable scaling:

$$O\left(\frac{d}{\epsilon^2} \log \frac{1}{\delta}\right)$$

- ✱ By truncating negligibly small probabilities from the relevance distribution, we can **improve the worst-case** scaling at the cost of adding a tiny (also negligible) bias to our estimate.

Direct Fidelity Estimation: complexity

- ✱ If the noise is **dephasing** (or depolarizing) then we get a very favorable scaling:

$$O\left(\frac{d}{\epsilon^2} \log \frac{1}{\delta}\right)$$

- ✱ By truncating negligibly small probabilities from the relevance distribution, we can **improve the worst-case** scaling at the cost of adding a tiny (also negligible) bias to our estimate.
- ✱ Works with more **general operator bases**, like coherent states.

Direct Fidelity Estimation: complexity

- ✱ If the noise is **dephasing** (or depolarizing) then we get a very favorable scaling:

$$O\left(\frac{d}{\epsilon^2} \log \frac{1}{\delta}\right)$$

- ✱ By truncating negligibly small probabilities from the relevance distribution, we can **improve the worst-case** scaling at the cost of adding a tiny (also negligible) bias to our estimate.
- ✱ Works with more **general operator bases**, like coherent states.
- ✱ If the nonzero Pauli expectations are only inverse polynomially small, then the number of copies is **polynomial**

Direct Fidelity Estimation: complexity

- ✱ If the noise is **dephasing** (or depolarizing) then we get a very favorable scaling:

$$O\left(\frac{d}{\epsilon^2} \log \frac{1}{\delta}\right)$$

- ✱ By truncating negligibly small probabilities from the relevance distribution, we can **improve the worst-case** scaling at the cost of adding a tiny (also negligible) bias to our estimate.
- ✱ Works with more **general operator bases**, like coherent states.
- ✱ If the nonzero Pauli expectations are only inverse polynomially small, then the number of copies is **polynomial**
 - ✱ Lots of interesting states satisfy this: e.g. stabilizer states, W states, Dicke states (fixed k), and Van den Nest's "computationally tractable" states.

Direct Fidelity Estimation: complexity

- ✱ If the noise is **dephasing** (or depolarizing) then we get a very favorable scaling:

$$O\left(\frac{d}{\epsilon^2} \log \frac{1}{\delta}\right)$$

- ✱ By truncating negligibly small probabilities from the relevance distribution, we can **improve the worst-case** scaling at the cost of adding a tiny (also negligible) bias to our estimate.
- ✱ Works with more **general operator bases**, like coherent states.
- ✱ If the nonzero Pauli expectations are only inverse polynomially small, then the number of copies is **polynomial**
 - ✱ Lots of interesting states satisfy this: e.g. stabilizer states, W states, Dicke states (fixed k), and Van den Nest's "computationally tractable" states.
- ✱ For Clifford circuits, the entire estimate is achievable in **poly(n)** time

Hamiltonian and Phase Estimation

Learns generators instead of gates, allowing for RB-like amplification to aid estimation.

- **Robust Phase Est.:** (Kimmel *et al.* 2015). Proves that Heisenberg-limited accuracy is *achievable* with non-adaptive measurements, using modification to Higgins *et al.* 2009 binary search algorithm. Robust to additive errors.
- **Filtering:** Particle (Granade *et al.* 2012), rejection (Wiebe and Granade 2016), guaranteed-cost (Roy *et al.* 2016). All provide time-dep estimation, with varying tradeoffs in implementation, generality and robustness.

Phase est. as resource for characterization / control design: Kimmel *et al.* 2015 also applies to calibration problem: represent over-rotation as phase accumulation, estimate.

Learning States and Gates

Several of these examples highlight the special role of stabilizer states and Clifford gates.

A more general question is can we **efficiently learn** states that have **concise descriptions**? There are many candidates for which we know how to do this

- Low rank states and gates (Gross *et al.* 09)
- Clifford gates (Low 09)
- Permutationally invariant states (Toth *et al.* 10)
- Matrix product states, or MPS (Cramer *et al.* 10)
- Sparse process matrices (Shabani *et al.* 11)
- MERA tensor networks (Landon-Cardinal *et al.* 12)
- Stabilizer states (Montanaro *et al.* 13)
- ...

Learning States and Gates

Several of these examples highlight the special role of stabilizer states and Clifford gates.

A more general question is can we **efficiently learn** states that have **concise descriptions**? There are many candidates for which we know how to do this

- **Low rank states and gates (Gross *et al.* 09)**
- Clifford gates (Low 09)
- Permutationally invariant states (Toth *et al.* 10)
- **Matrix product states, or MPS (Cramer *et al.* 10)**
- Sparse process matrices (Shabani *et al.* 11)
- MERA tensor networks (Landon-Cardinal *et al.* 12)
- Stabilizer states (Montanaro *et al.* 13)
- ...

Compressed tomography

Gross *et al.* 09; Gross 10; Liu 11; F. *et al.* 12

Suppose a state is well approximated by a rank r density matrix.
Can we improve the complexity of tomography?

Compressed tomography is a nearly ideal solution!

*For process tomography, replace d with d^2

Compressed tomography

Gross *et al.* 09; Gross 10; Liu 11; F. *et al.* 12

Suppose a state is well approximated by a rank r density matrix.
Can we improve the complexity of tomography?

Compressed tomography is a nearly ideal solution!

- ✱ Reduces the measurement complexity for states which are approximately rank r to $\sim O(rd)$ (instead of d^2)

*For process tomography, replace d with d^2

Compressed tomography

Gross *et al.* 09; Gross 10; Liu 11; F. *et al.* 12

Suppose a state is well approximated by a rank r density matrix.
Can we improve the complexity of tomography?

Compressed tomography is a nearly ideal solution!

- ✱ Reduces the measurement complexity for states which are approximately rank r to $\sim O(rd)$ (instead of d^2)
- ✱ Reduces sample complexity to $\sim O(r^2d^2)$ (instead of d^4)

*For process tomography, replace d with d^2

Compressed tomography

Gross *et al.* 09; Gross 10; Liu 11; F. *et al.* 12

Suppose a state is well approximated by a rank r density matrix.
Can we improve the complexity of tomography?

Compressed tomography is a nearly ideal solution!

- ✱ Reduces the measurement complexity for states which are approximately rank r to $\sim O(rd)$ (instead of d^2)
- ✱ Reduces sample complexity to $\sim O(r^2d^2)$ (instead of d^4)
- ✱ Requires only simple Pauli measurements

*For process tomography, replace d with d^2

Compressed tomography

Gross *et al.* 09; Gross 10; Liu 11; F. *et al.* 12

Suppose a state is well approximated by a rank r density matrix.
Can we improve the complexity of tomography?

Compressed tomography is a nearly ideal solution!

- ✱ Reduces the measurement complexity for states which are approximately rank r to $\sim O(rd)$ (instead of d^2)
- ✱ Reduces sample complexity to $\sim O(r^2d^2)$ (instead of d^4)
- ✱ Requires only simple Pauli measurements
- ✱ Convex cost function (local optima are global optima)

*For process tomography, replace d with d^2

Compressed tomography

Gross *et al.* 09; Gross 10; Liu 11; F. *et al.* 12

Suppose a state is well approximated by a rank r density matrix.
Can we improve the complexity of tomography?

Compressed tomography is a nearly ideal solution!

- ✱ Reduces the measurement complexity for states which are approximately rank r to $\sim O(rd)$ (instead of d^2)
- ✱ Reduces sample complexity to $\sim O(r^2d^2)$ (instead of d^4)
- ✱ Requires only simple Pauli measurements
- ✱ Convex cost function (local optima are global optima)
- ✱ Runtime is about $O(d^{1+\epsilon})$ (impossible for dense methods)

*For process tomography, replace d with d^2

Compressed tomography

Gross *et al.* 09; Gross 10; Liu 11; F. *et al.* 12

Suppose a state is well approximated by a rank r density matrix.
Can we improve the complexity of tomography?

Compressed tomography is a nearly ideal solution!

- ✱ Reduces the measurement complexity for states which are approximately rank r to $\sim O(rd)$ (instead of d^2)
- ✱ Reduces sample complexity to $\sim O(r^2d^2)$ (instead of d^4)
- ✱ Requires only simple Pauli measurements
- ✱ Convex cost function (local optima are global optima)
- ✱ Runtime is about $O(d^{1+\epsilon})$ (impossible for dense methods)
- ✱ Comes with a certificate, so you know if it worked

*For process tomography, replace d with d^2

Compressed tomography

Gross *et al.* 09; Gross 10; Liu 11; F. *et al.* 12

Suppose a state is well approximated by a rank r density matrix.
Can we improve the complexity of tomography?

Compressed tomography is a nearly ideal solution!

- ✱ Reduces the measurement complexity for states which are approximately rank r to $\sim O(rd)$ (instead of d^2)
- ✱ Reduces sample complexity to $\sim O(r^2d^2)$ (instead of d^4)
- ✱ Requires only simple Pauli measurements
- ✱ Convex cost function (local optima are global optima)
- ✱ Runtime is about $O(d^{1+\epsilon})$ (impossible for dense methods)
- ✱ Comes with a certificate, so you know if it worked
- ✱ Comes with rigorous confidence intervals

*For process tomography, replace d with d^2

Main results

Gross, Liu, F., Becker, & Eisert, PRL 2009; Gross IEEE TIT 2009.

For an arbitrary matrix ρ of dimension $d \times d$, rank r , sample a set Ω of $m = O(rd \log^2 d)$ iid random Paulis P .

Let $\mathcal{R}(x) = \text{Tr}(Px) \quad \forall P \in \Omega$ be the *sampling operator*.

Estimate $b(P) \approx \text{Tr}(P\rho) \pm \frac{\epsilon}{\sqrt{m}}$

Compute: $\hat{\rho} = \arg \min \|x\|_{\text{tr}} \text{ s.t. } \|\mathcal{R}(x) - b\|_2 \leq \epsilon$

Theorem: with high probability over Ω and the data

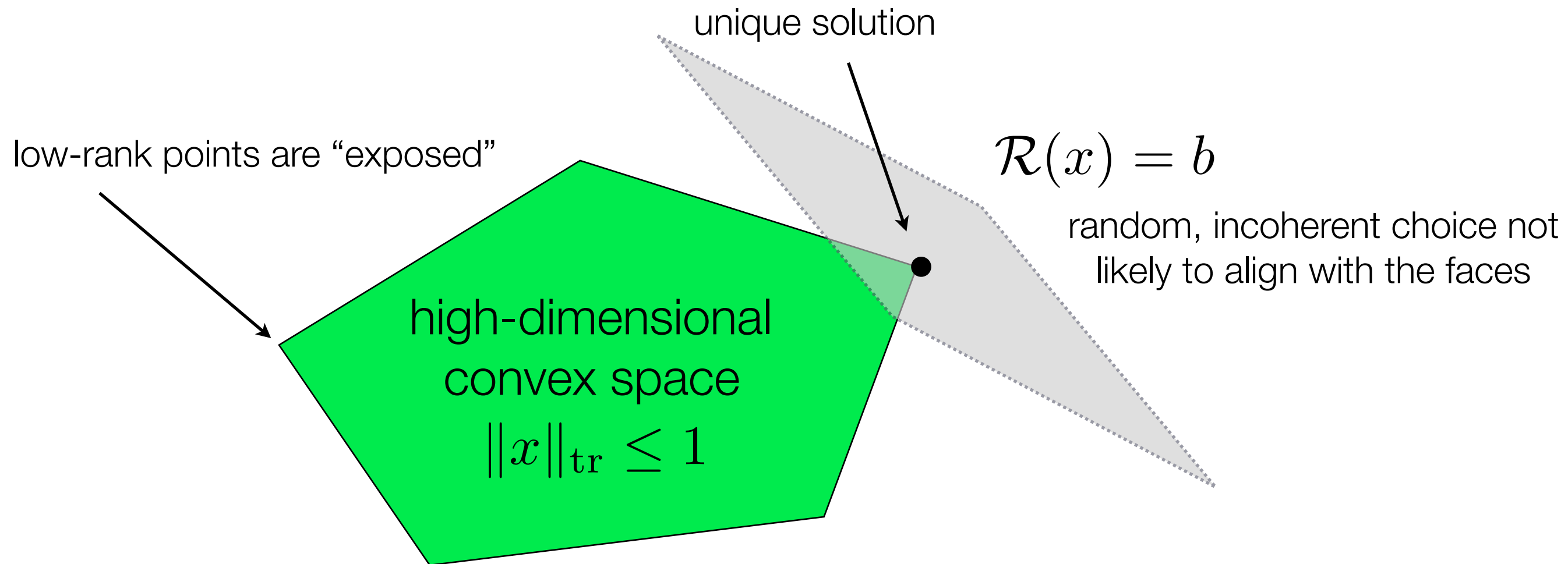
$$\|\hat{\rho} - \rho\|_{\text{tr}} \leq O(\epsilon \sqrt{rd})$$

*uses $O(d^4)$ samples!

“local” results

Gross, Liu, F., Becker, & Eisert, PRL 2009; Gross IEEE TIT 2009.

Argument is based on **local** properties (dual certificate).



Any perturbation around the true state either increases $\|x\|_{\text{tr}}$ or changes the value of $\mathcal{R}(x)$

Main results

Gross *et al.* 09; Gross 10; Liu 11; F. *et al.* 12

Sample a set Ω of $m = O(rd \log^6 d)$ iid random Paulis P .
For every matrix ρ of dimension $d \times d$, rank r ,

$$\text{Estimate: } b(P) \approx \text{Tr}(P\rho) \pm \epsilon \frac{\log^{2.5} d}{\sqrt{rd}}$$

$$\text{Compute: } \hat{\rho} = \arg \min \|x\|_{\text{tr}} \text{ s.t. } \|\mathcal{R}^*(\mathcal{R}(x) - b)\| \leq \epsilon$$

Theorem: with high probability over Ω and the data:

$$\|\hat{\rho} - \rho\|_{\text{tr}} \leq O(\epsilon).$$

This uses $t = O\left(\frac{r^2 d^2 \log(d)}{\epsilon^2}\right)$ copies... far few than previously!

*For process tomography, replace d with d^2

Main results

Gross et al. 09; Gross 10; Liu 11; F. et al. 12

“Universal” set Ω that holds for **all** ρ

Sample a set Ω of $m = O(rd \log^6 d)$ iid random Paulis P .
For every matrix ρ of dimension $d \times d$, rank r ,

$$\text{Estimate: } b(P) \approx \text{Tr}(P\rho) \pm \epsilon \frac{\log^{2.5} d}{\sqrt{rd}}$$

$$\text{Compute: } \hat{\rho} = \arg \min \|x\|_{\text{tr}} \text{ s.t. } \|\mathcal{R}^*(\mathcal{R}(x) - b)\| \leq \epsilon$$

Theorem: with high probability over Ω and the data:

$$\|\hat{\rho} - \rho\|_{\text{tr}} \leq O(\epsilon).$$

This uses $t = O\left(\frac{r^2 d^2 \log(d)}{\epsilon^2}\right)$ copies... far few than previously!

*For process tomography, replace d with d^2

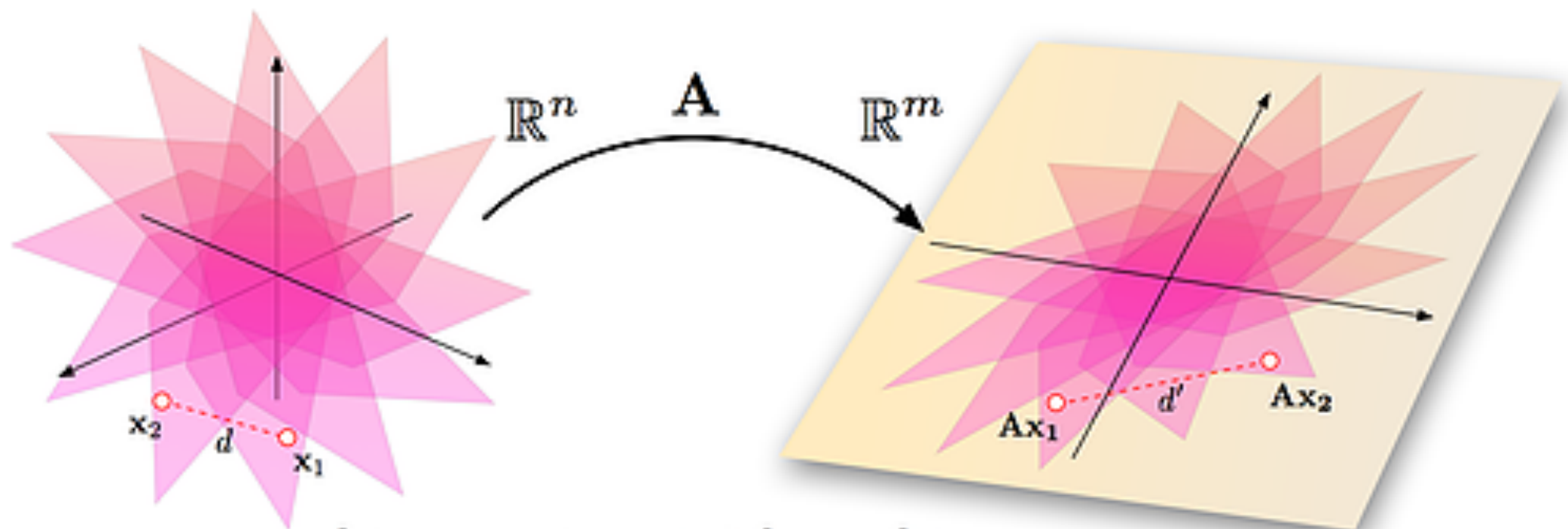
Restricted Isometry Property

Key result to achieve this is “RIP for Paulis”

A sampling operator obeys the (r, δ) RIP if
for all x with rank r ,

$$|\|\mathcal{R}(x)\|_2 - \|x\|_2| \leq \delta \|x\|_2$$

“Projection onto a subspace approximately preserves length”



And if the state is full rank?

F., Gross, Liu, & Eisert 12

As before, suppose we measure t copies, with $t = O\left(\frac{r^2 d^2 \log(d)}{\epsilon^2}\right)$

Now decompose ρ into the best rank r approximation plus the residual tail: $\rho = \rho_r + \rho_t$

Theorem: same method as before gives a reconstruction error

$$\|\hat{\rho} - \rho\|_{\text{tr}} \leq \epsilon + C \|\rho_t\|_{\text{tr}}$$

Thus, we get the **optimal error bound** modulo the constant in front of the truncation error!

Sample complexity

How good is this result? Can the sample complexity be improved? Let's define the *minimax risk*:

$$M(\alpha) = \inf_{\langle \hat{\rho}, P_i \rangle} \sup_{\rho} \Pr \left[\|\hat{\rho} - \rho\|_1 > \alpha \right]$$

We now get another **Theorem**, a lower bound on the sample complexity for some fixed risk tolerance.

$$\text{If } M = O(1) \text{ then } t = \Omega \left(\frac{r^2 d^2}{\log d} \right)$$

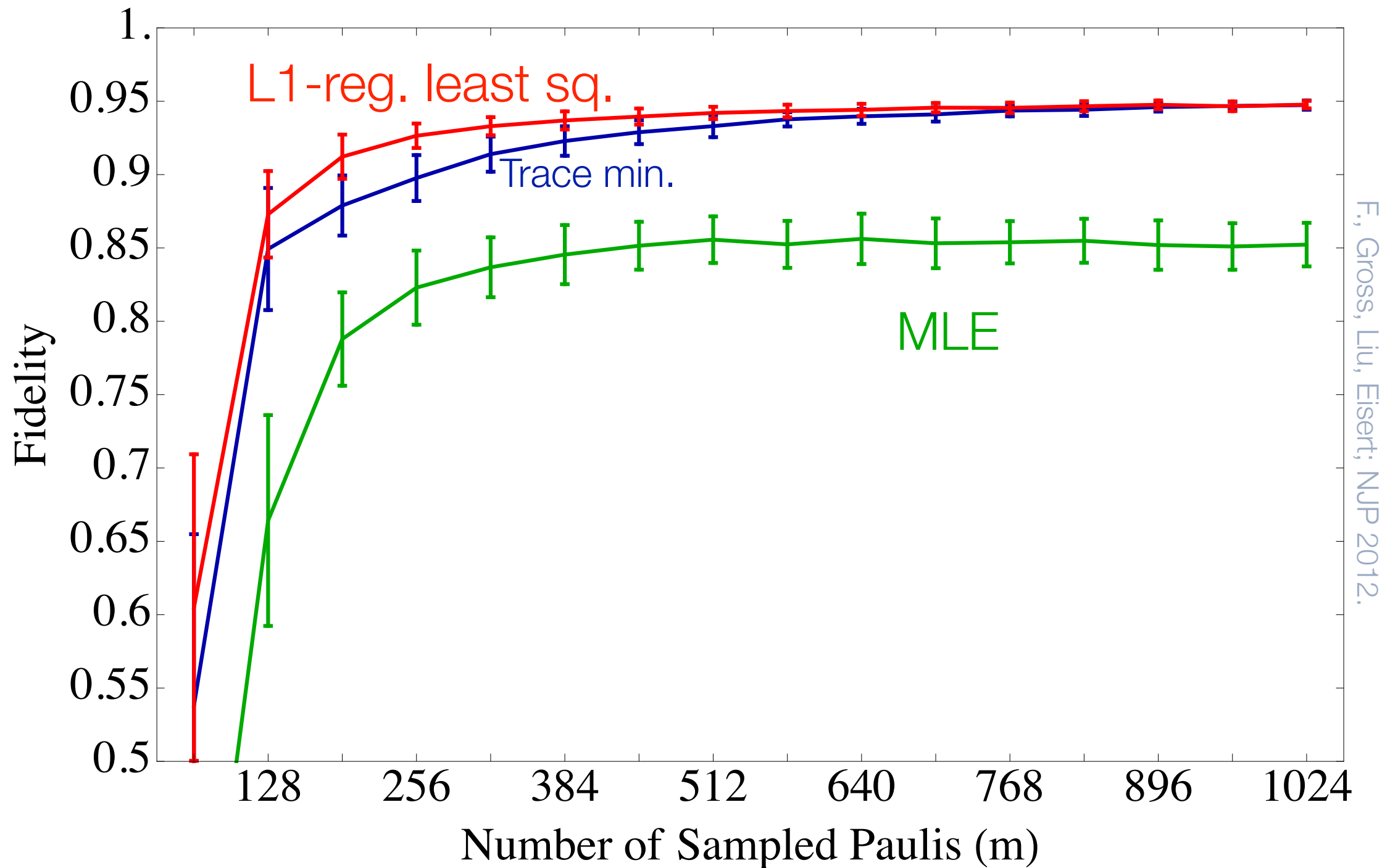
Our sample complexity is **optimal** up to log factors!



The implicit “constant” **depends** on ϵ !

Simulated performance

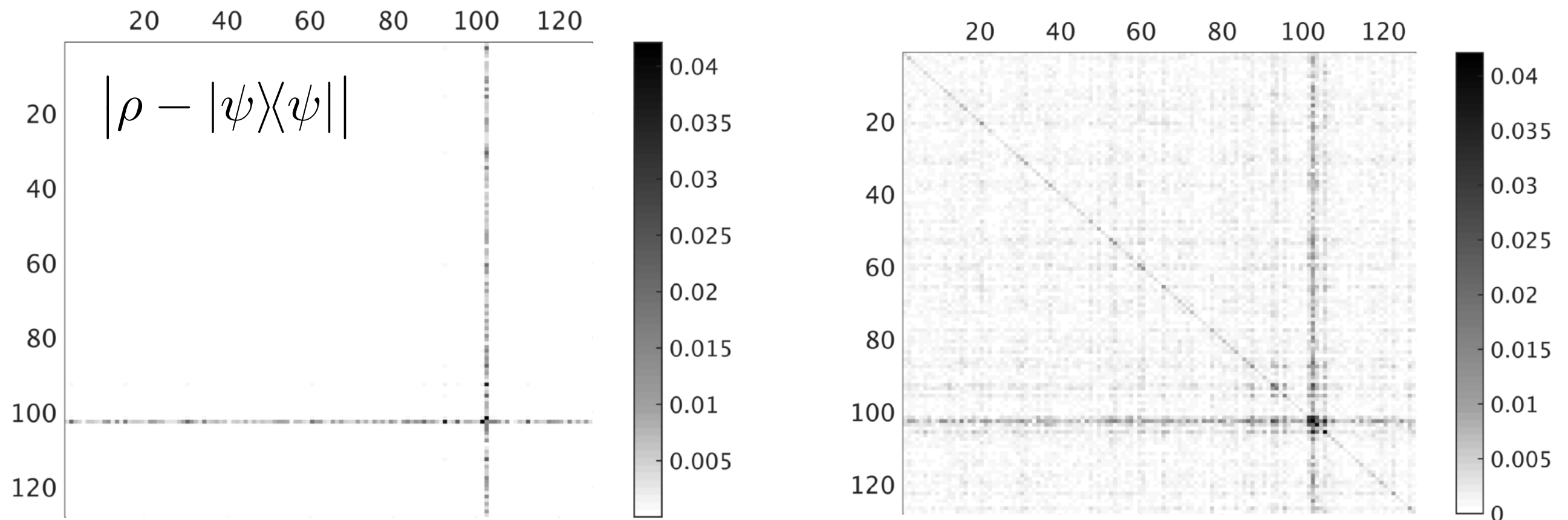
Fidelity and Trace Distance $T=80000$, $c=20$



Experimental performance

Baldwin *et al.* 14; Riofrio *et al.* 16

7-qubit ion-trap experiment (Innsbruck) using
~130 random Pauli measurements (Riofrio *et al.* 16).



The compressed sensing method yields a **massive reduction** in spurious signal compared to traditional estimators and accurately estimates the **coherent noise**.

What does compressed sensing “mean”?

We now see that compressed sensing involves **two separate ideas**:

(1) using an *incomplete* set of observables

(2) using *regularized* estimators to get low-rank solutions

We normally do both at the same time.

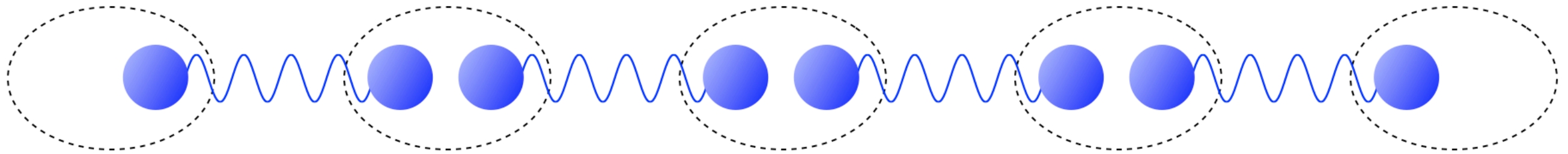
Our results show that (2) can be used *irrespective* of (1)...

...however, at the same time, there is **no penalty** for choosing to do (1).

And there is a **big practical incentive** to do (1), since sparse data can be processed faster.

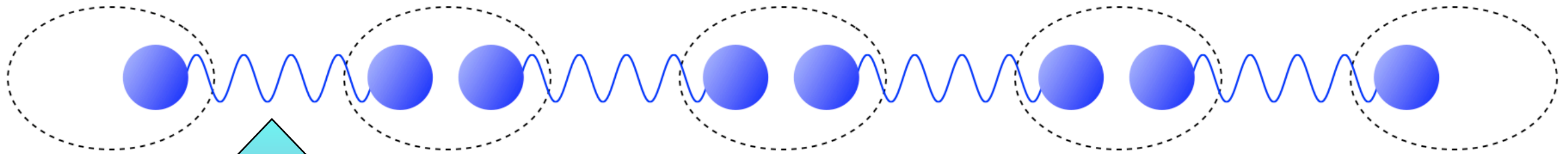
Matrix Product States (MPS)

Cramer *et al.* 10



Matrix Product States (MPS)

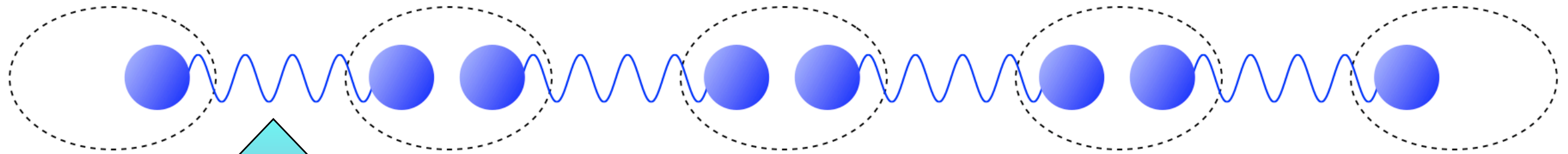
Cramer *et al.* 10



$$|\Phi\rangle = \sum_{\alpha=1}^D |\alpha\rangle |\alpha\rangle$$

Matrix Product States (MPS)

Cramer *et al.* 10

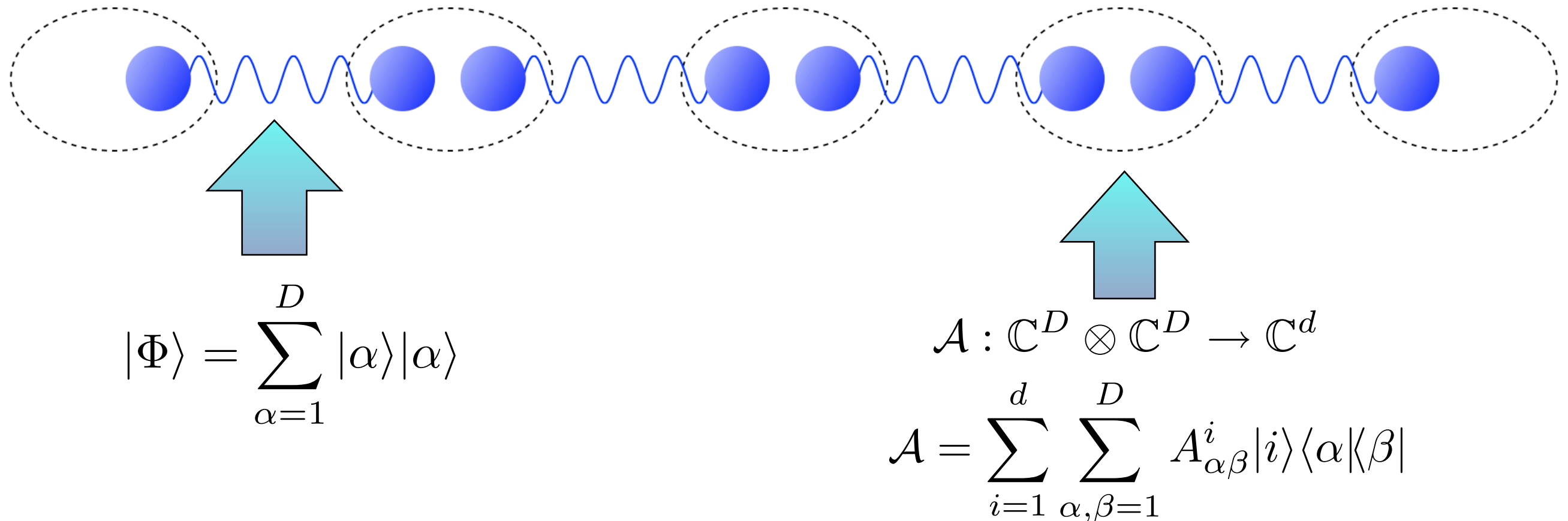


$$|\Phi\rangle = \sum_{\alpha=1}^D |\alpha\rangle |\alpha\rangle$$

$$\mathcal{A} : \mathbb{C}^D \otimes \mathbb{C}^D \rightarrow \mathbb{C}^d$$
$$\mathcal{A} = \sum_{i=1}^d \sum_{\alpha, \beta=1}^D A_{\alpha\beta}^i |i\rangle \langle \alpha| \langle \beta|$$

Matrix Product States (MPS)

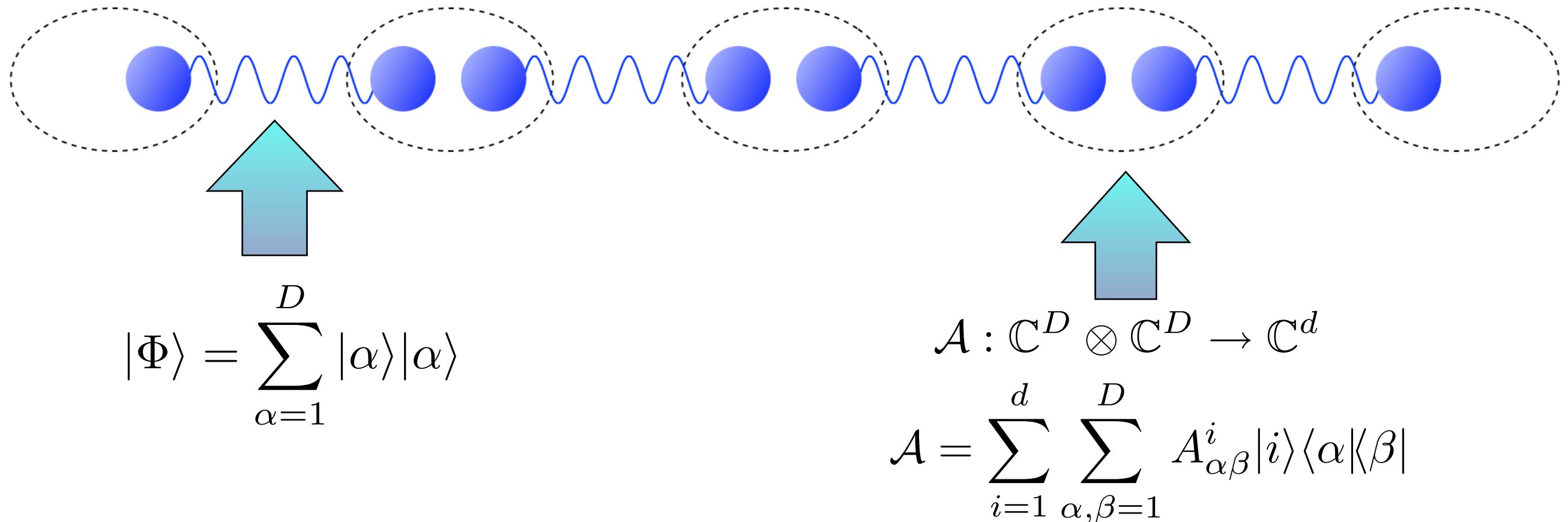
Cramer *et al.* 10



MPS form a variational ansatz for ground states of 1-dimensional quantum systems that systematically goes beyond the mean-field approximation.

Matrix Product States (MPS)

Cramer *et al.* 10

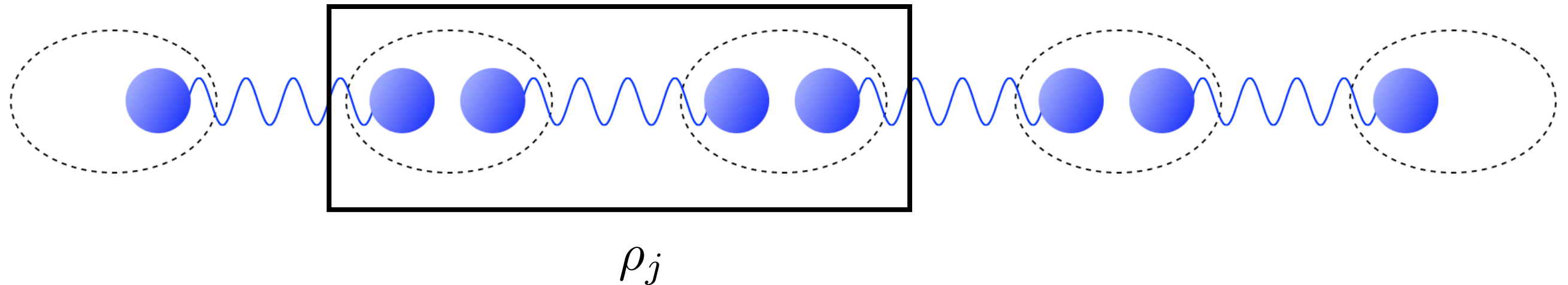


MPS form a variational ansatz for ground states of 1-dimensional quantum systems that systematically goes beyond the mean-field approximation.

$$|\psi\rangle = \sum_{j_1, \dots, j_n=1}^d \text{Tr}(A^{j_1} A^{j_2} \dots A^{j_n}) |j_1, j_2, \dots, j_n\rangle \quad O(ndD^2) \text{ parameters!}$$

State tomography for non-degenerate MPS

Cramer *et al.* 10

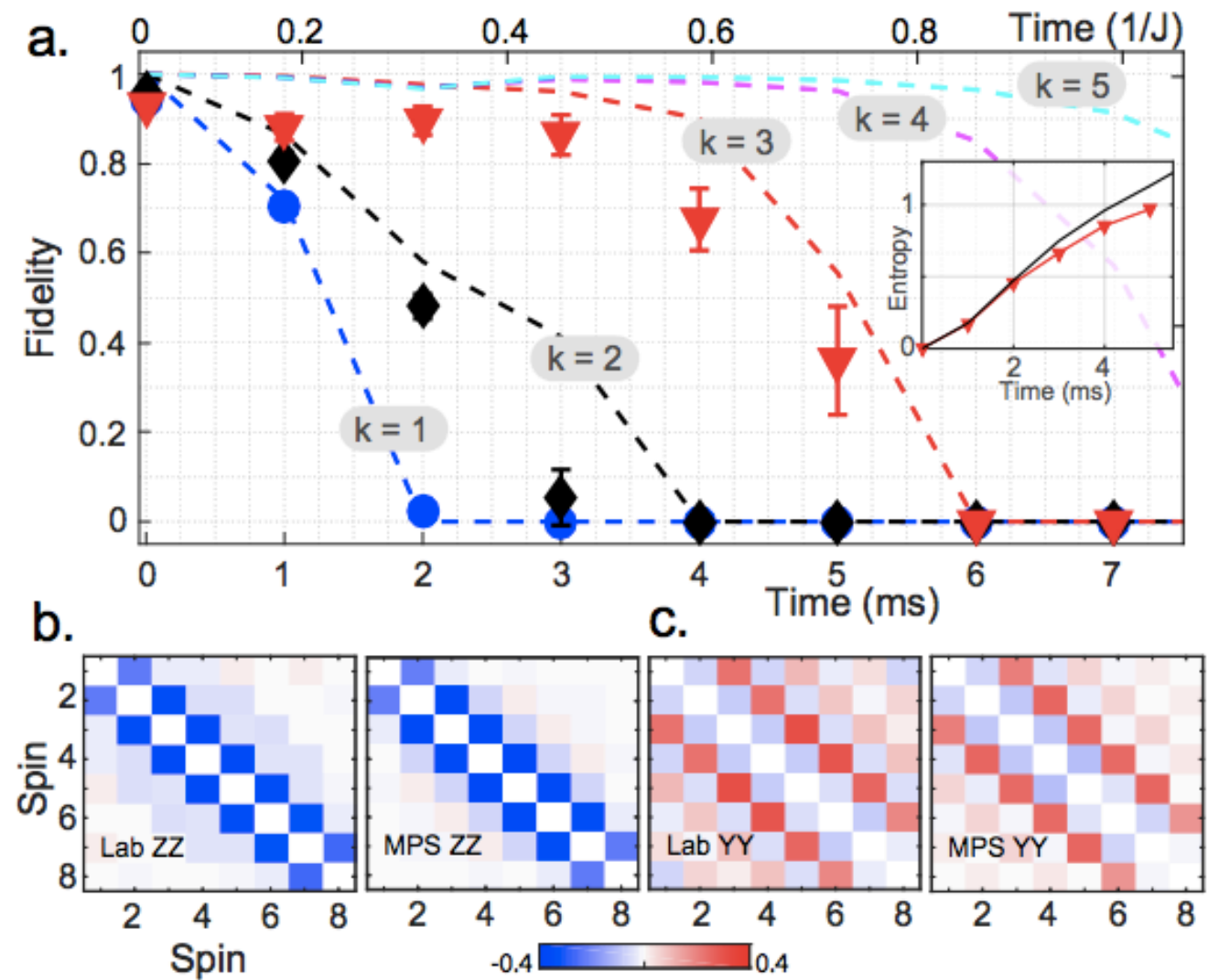
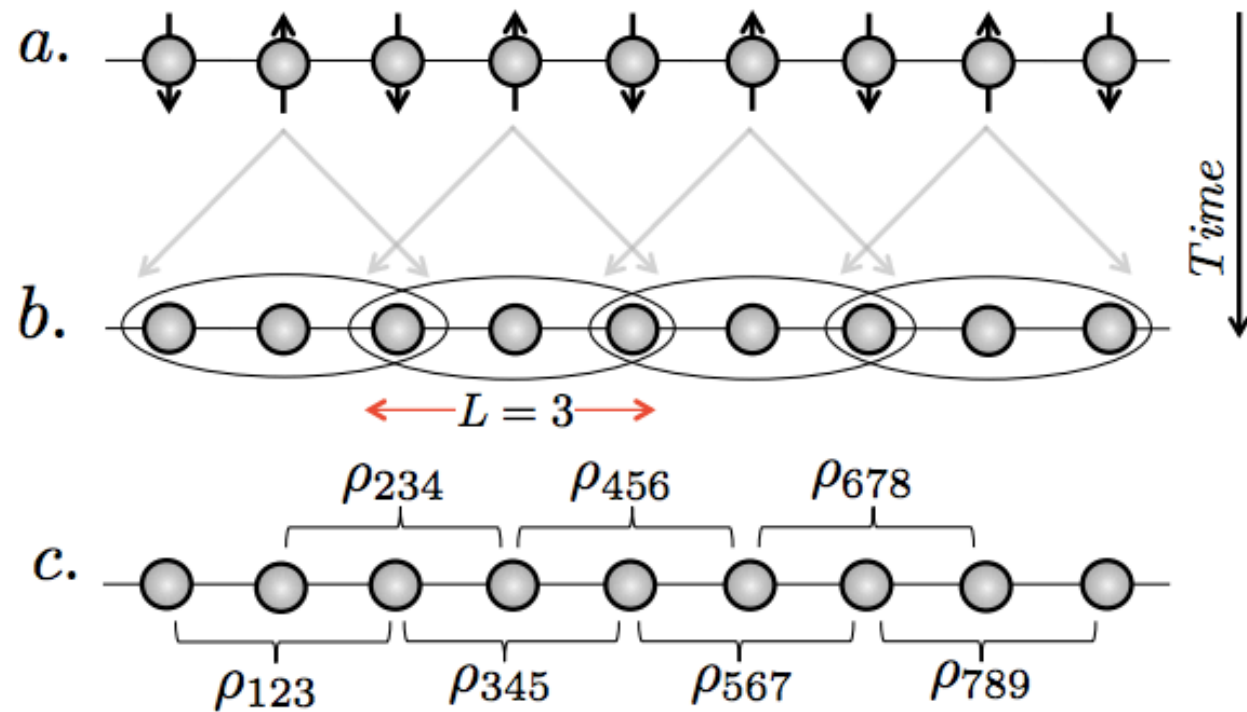


Measure local reduced density operators and use tools from many-body physics (DMRG) to solve the local consistency problem

This is QMA hard in general, but if the “true” state is indeed described by an injective MPS, then the method will work in time polynomial in n , the length of the chain.

Experimental MPS tomography

Lanyon *et al.* 16

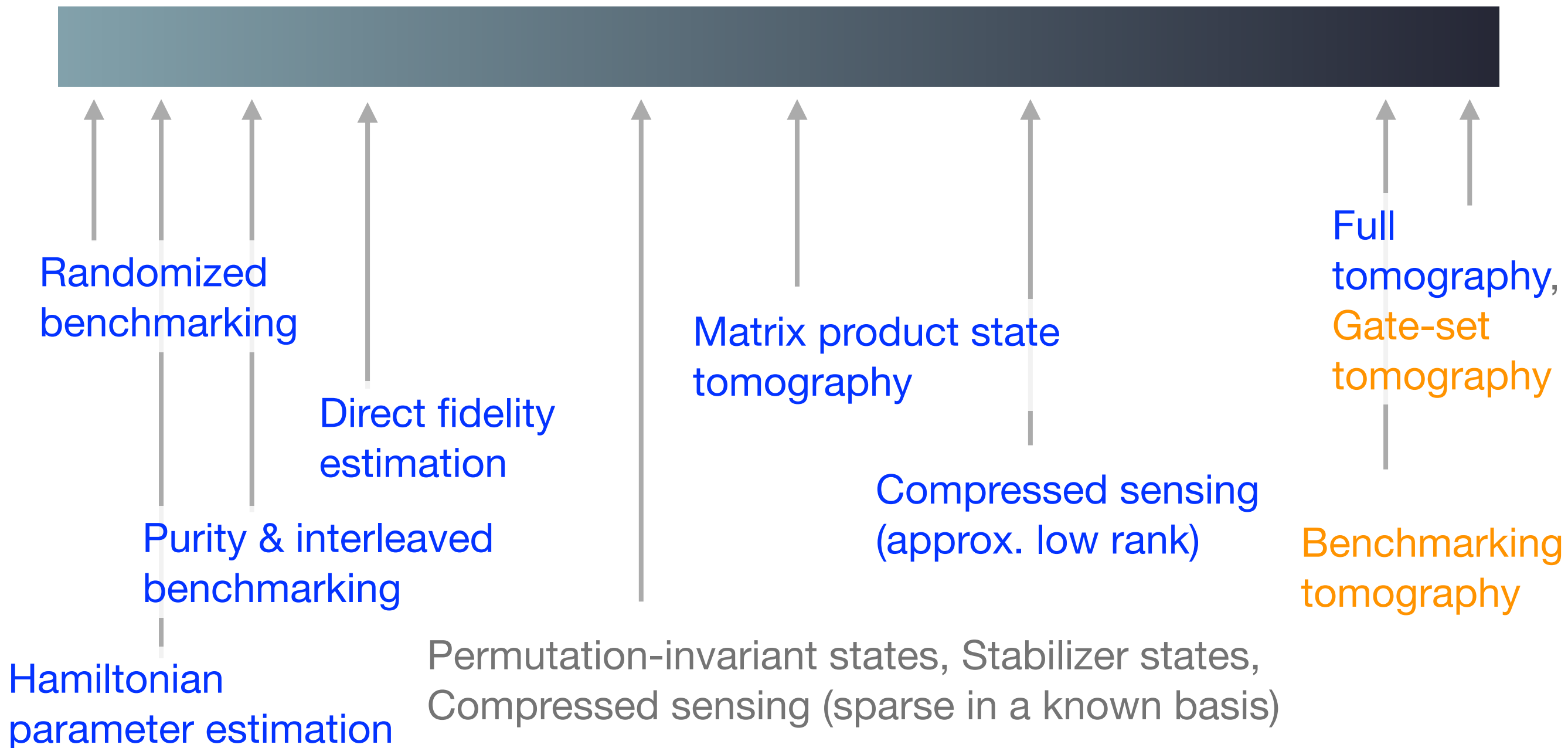


Measure properties of a quantum quench using MPS tomography

Strategies for Characterizing Noise

Low complexity,
Less information

High complexity,
More information

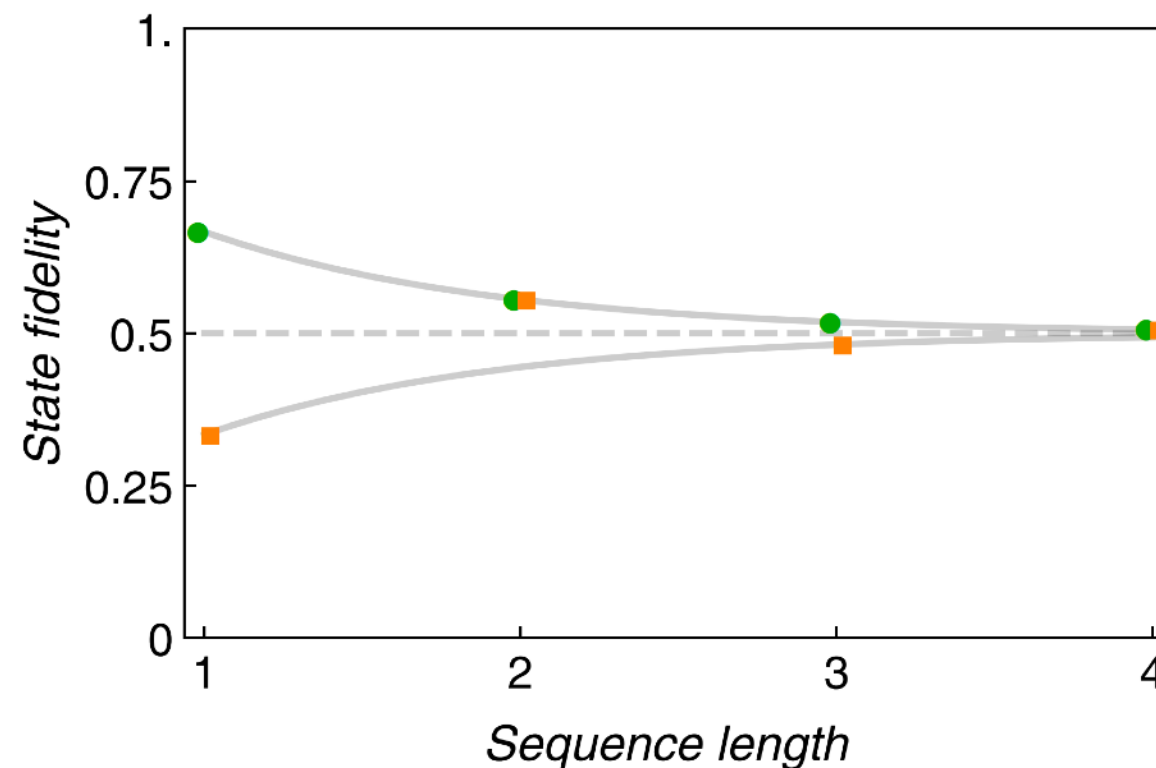


Randomized Benchmarking Tomography

Kimme et al. 14

Combine the SPAM-free advantages of RB with the debugging power of quantum tomography.

- Interleave target gate with RB, estimate $C \circ \Lambda$ for each of several different Clifford gates C . Need 10 Clifford gates to span qubit unitary subspace. Note that this can yield *negative* decay probabilities. Ex:



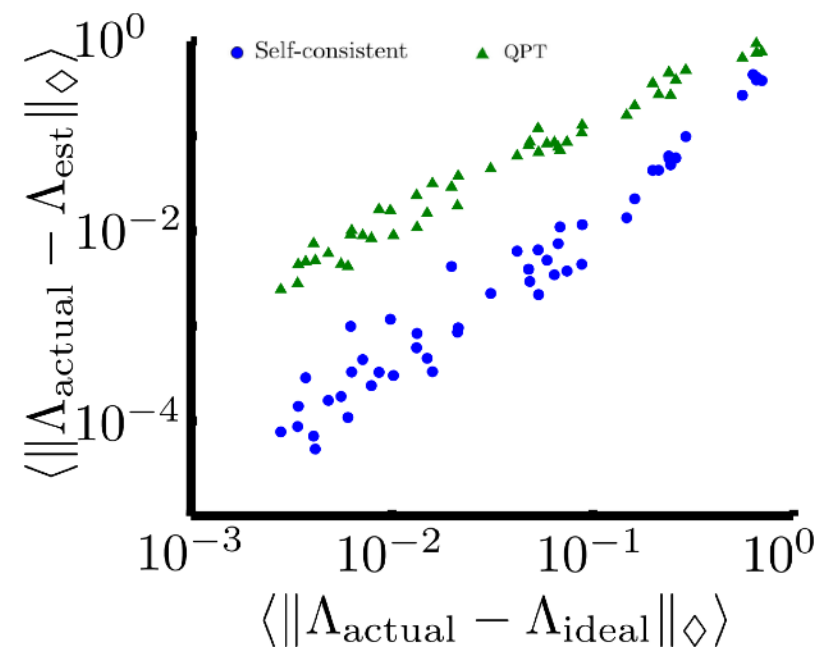
- Reconstruct Λ from the estimated overlaps.

Gate Set Tomography

Merkel *et al.* 13; Blume-Kohout *et al.* 13; Blume-Kohout *et al.* 16

Idea: “calibration-free” tomography that simultaneously estimates all gates in an ensemble, as well as SPAM errors.

- Merkel *et al.* 13; Expands to higher dimensions to include gate-dependent and correlated errors. Reconstructs entire gate set at once to ensure self-consistency, linearizing near target. Better \diamond predictions than QPT!

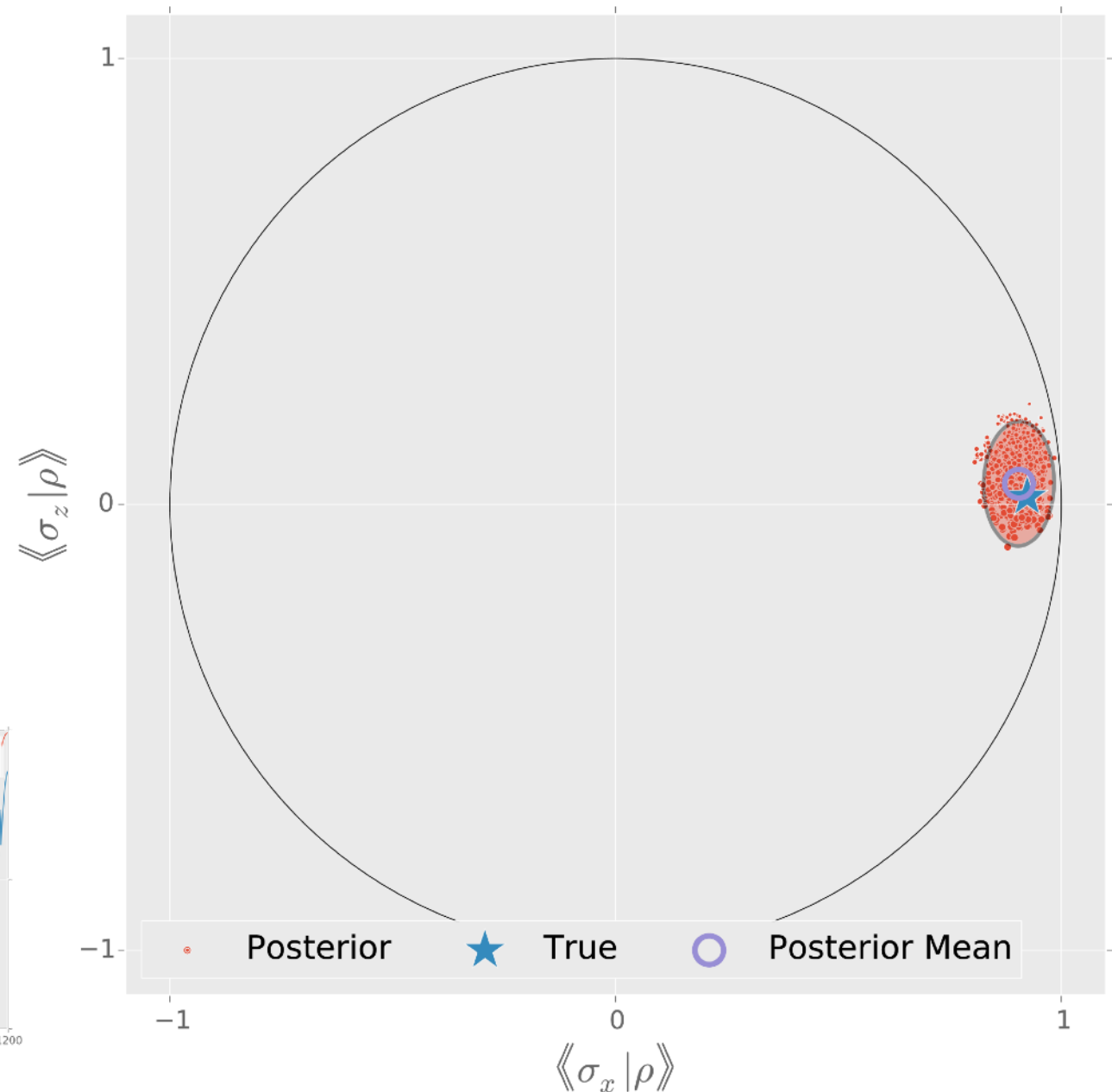
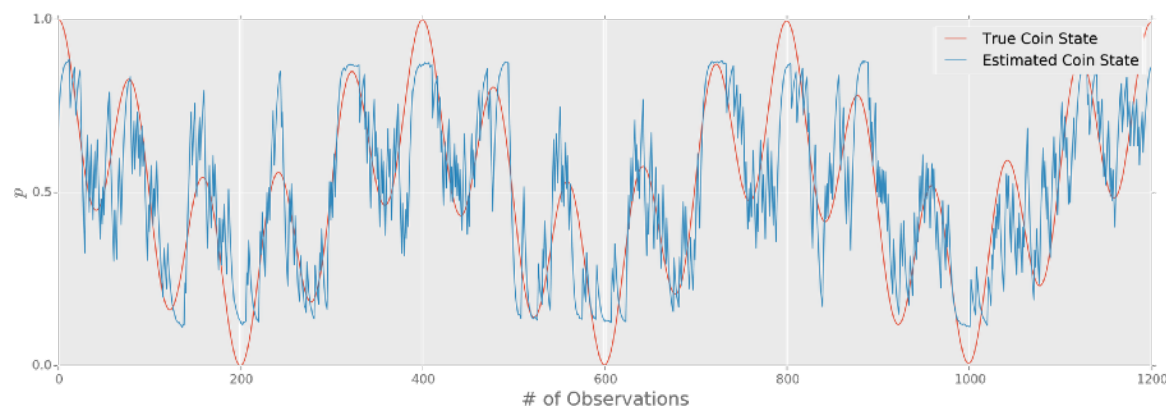


- Blume-Kohout *et al.* 13, 16. Treat linear est as starting point for MLE reconstruction. Application to trapped ions, ESR qubits. Software implementation (pyGSTi • pygsti.info).

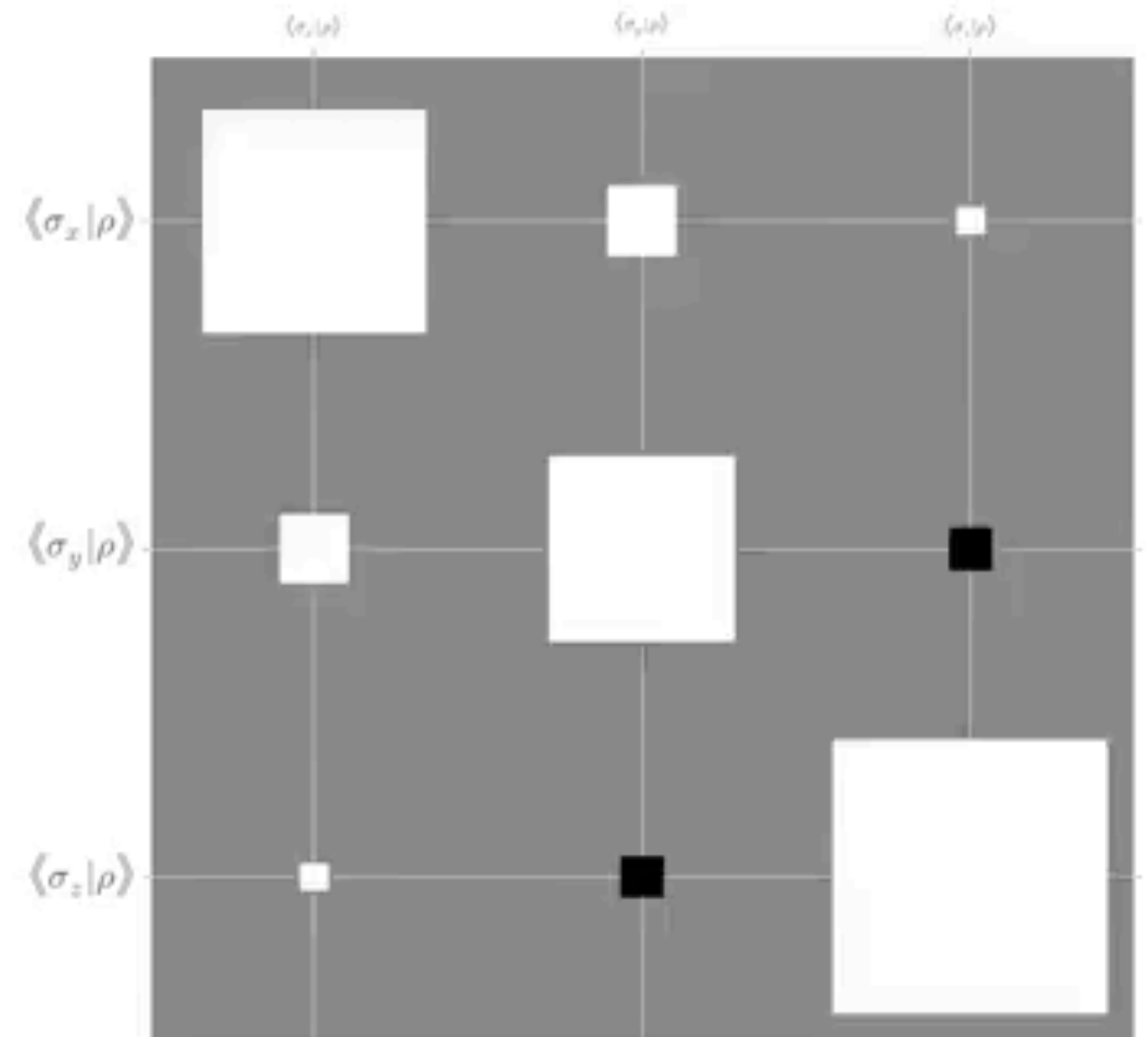
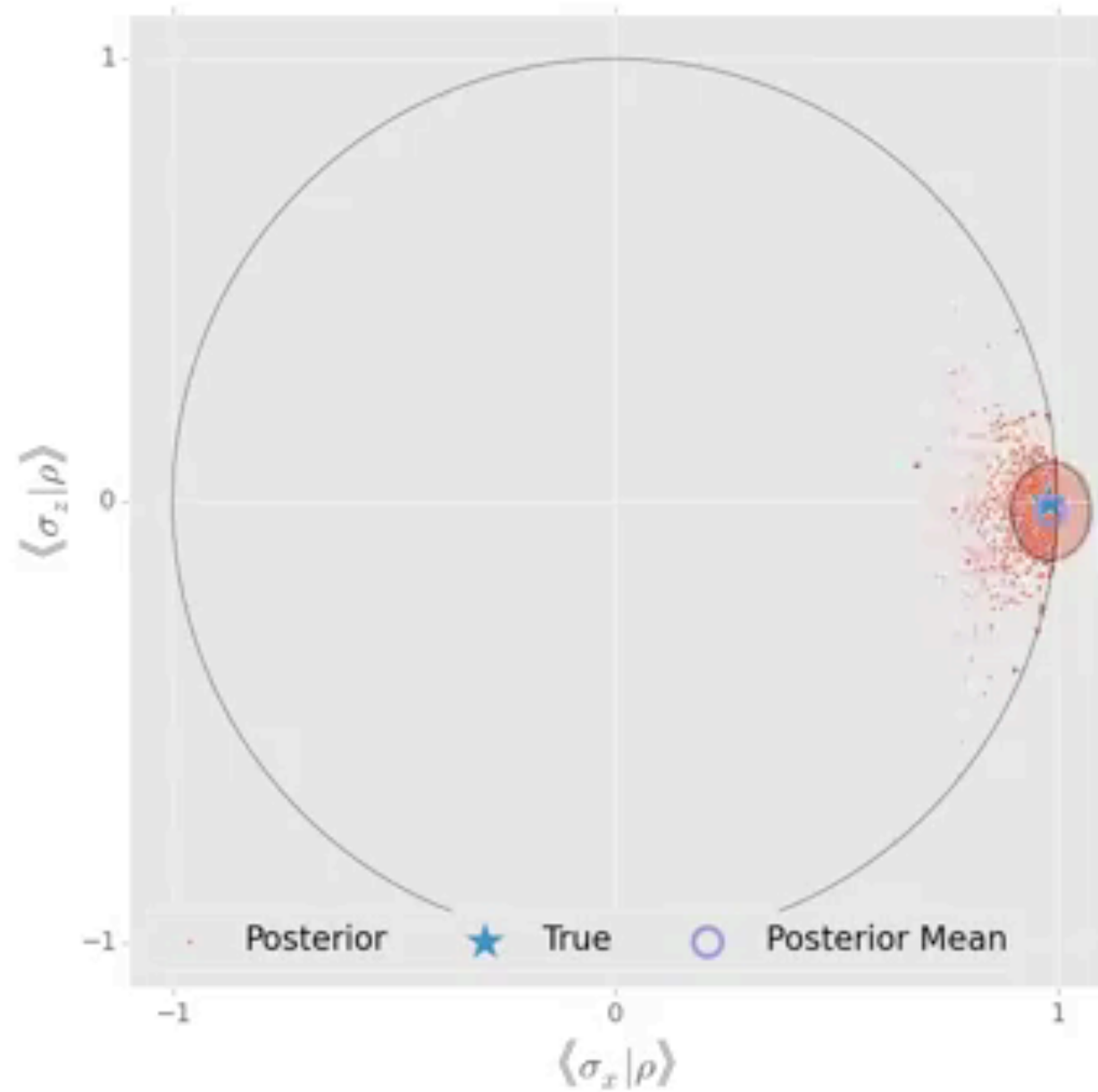
Drift, time dependence, and adaptivity

Tracking drift in a tomography experiment using sequential Monte Carlo and particle filters.

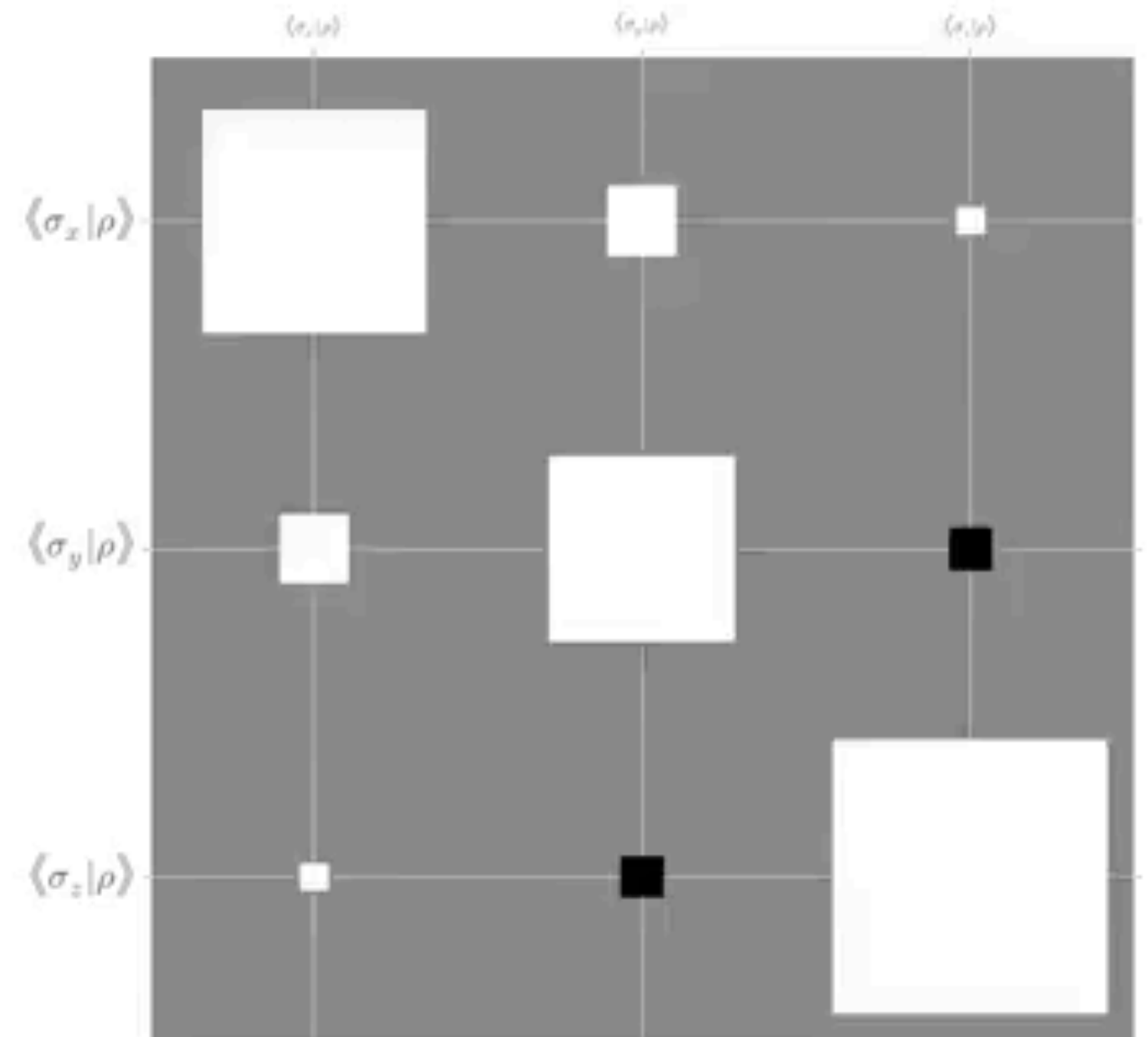
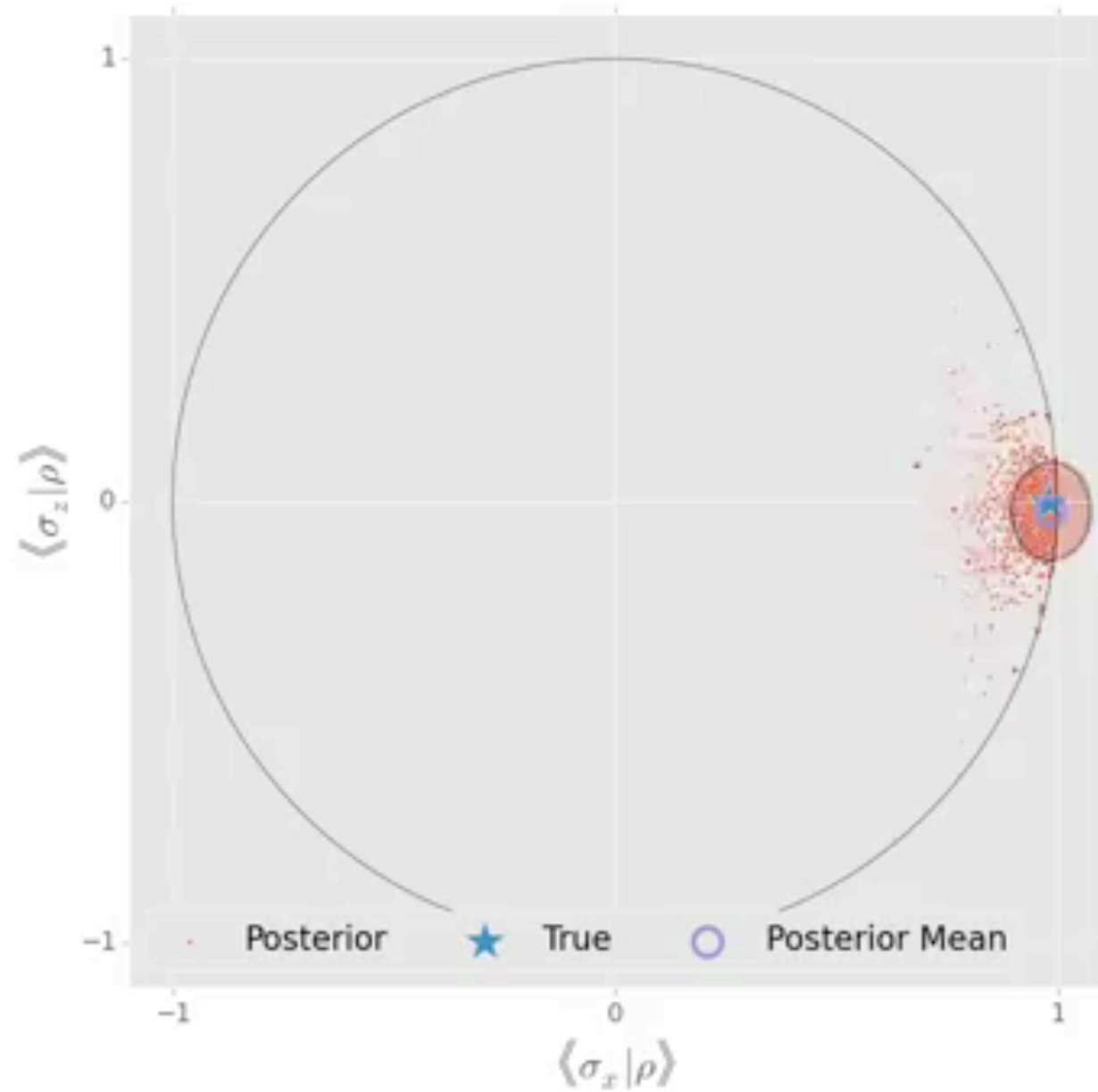
Time-dependence included as *diffusion*, similar to previous applications in classical computer vision.



Drift, time dependence, and adaptivity



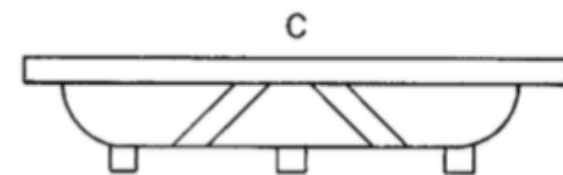
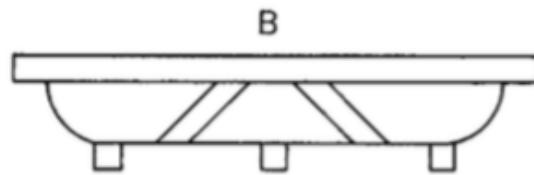
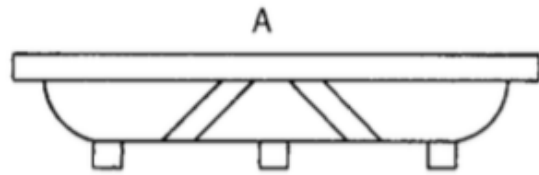
Drift, time dependence, and adaptivity



Conclusion

- There are a variety of powerful protocols for characterizing errors in quantum devices, depending on question of interest.
 - Tomography / RB at opposite extremes of model complexity
 - Many compelling intermediate models: compressed sensing, RB tomography, learning ansätze, etc.
 - Readily available estimation tools to support QCVV in practice.
- **Characterization isn't a solved problem.**
 - Gaps between theoretical assumptions and experimental reality, motivating relaxed / generalized approaches.
 - Need new theoretical and statistical tools for better comparisons with quantities of interest (e.g. \diamond -norm).
 - Close the engineering cycle by applying QCVV diagnostics to *new* experiments.

Fixed points and symmetry



- Start with three approximate planes
- Rub A against B and C until A matches both
- Compare B to C: relative defects are now exposed
- Permute and repeat

