SAND2013-4989C

Debugging a Running Qubit

or

Selling[1] Tomography[2] to the Inupiaq[3,(4,5)]

Robin Blume-Kohout

(Sandia National Labs)

[1] The (tortured) joke here is that I'm opening new "markets" [2] for tomography.

[2] Except, actually, I'm going to argue that this was always the natural "market" for tomography [5]

[3] This is the polite and correct word for "eskimo" [4].

[4] So this is supposed to be a reference to "Selling ice to the eskimos" [5]. Get it?

[5] But whereas the Inupiaq really don't need more ice, I'm claiming that tomography is a natural fit for the application I'll discuss [2]. This is supposed to be ironic, but I'm starting to think it's just confusing.

U.S. DEPARTMENT OF ENERGY

Sandia National Laboratories

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

How is Tomography used?

- To verify that a gate/state/device works.
  - For bragging rights (Nature, Science, etc.)
  - So we can actually *use* it. Many, many times.
- To verify that it *doesn't* work.
- To determine how *well* it works (and therefore how much error correction we need).
- To identify how it fails, so that we can sit around and cry.
- To identify how it fails, so that we can fix it.

When is Tomography used?

- Just once, after you've already designed really good gates, to prove that you're awesome.
- Never.
- A couple of times, iteratively while improving your device, until it works perfectly.
- Only in theory.
- Every now and then, because your device needs to be recalibrated.
- Constantly, because your device needs to be recalibrated.

My principles

- Characterization has no purpose but to achieve control.
  - Characterization needs to be *predictive*, *diagnostic*, and probably *targeted* (at important control parameters).
  - Tomography needs to be *robust* to bad calibration.
  - We don't need an *estimate*, just *instructions*.
- Characterization will be an *iterative* process.
- Because of drift, it will also be an *ongoing* process.

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Outline

- In this talk, I want to display two narrow thrusts into this vast problem area:

- 1. Can we characterize a running qubit in such a way as to counteract drift?

    - Yes! When a logical qubit is encoded an N-qubit code, we can detect and compensate for randomly drifting 1-qubit Hamiltonians on all N qubits (even in the presence of stochastic noise). This demonstrates <u>one</u> of the principles...

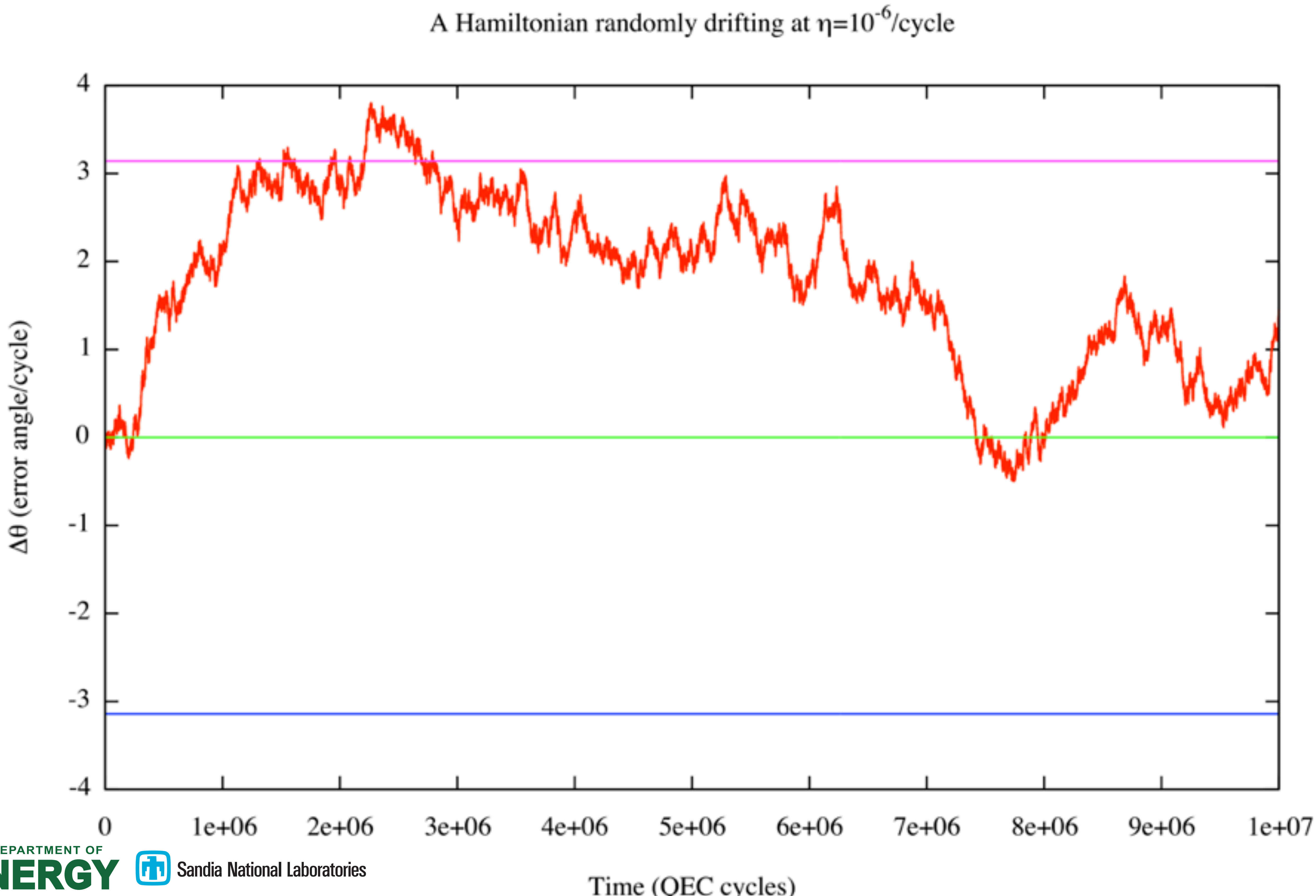
- 2. How do we *do* full characterization robustly & efficiently?

    - We can characterize all of our operations *relationally* in a highly robust (though not more efficient than process tomography) way. This is *gate-set tomography*, and it works -- but poses some interesting challenges. This demonstrates the other key principle behind the overall program.

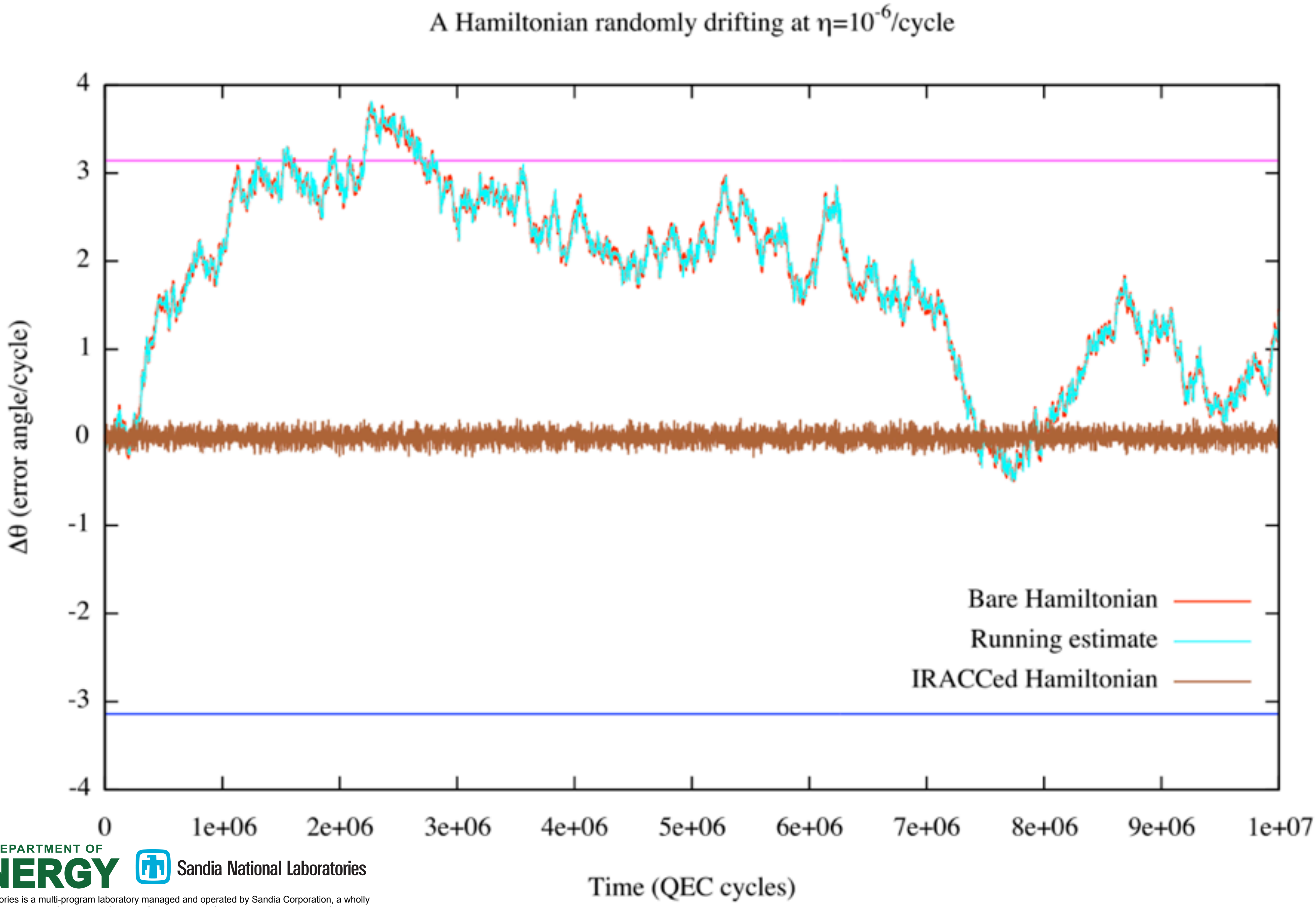
U.S. DEPARTMENT OF ENERGY
Sandia National Laboratories
Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Drift Control

Hamiltonians drift ☹



Fortunately, we can fix that ☺



How’d we do that?

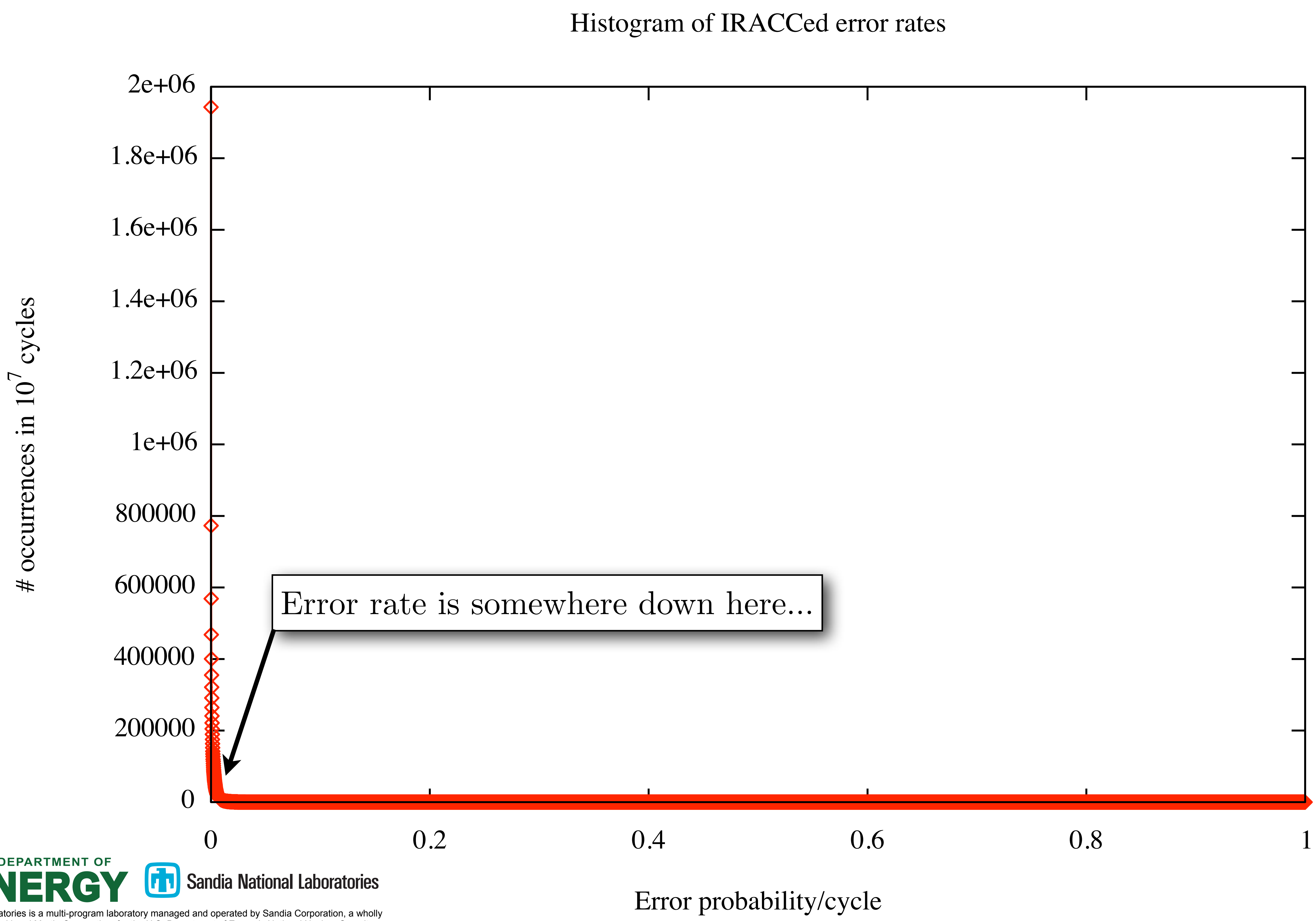
I applied a bunch of estimation theory, and tried a lot of algorithms.

The dumbest one still works best. 

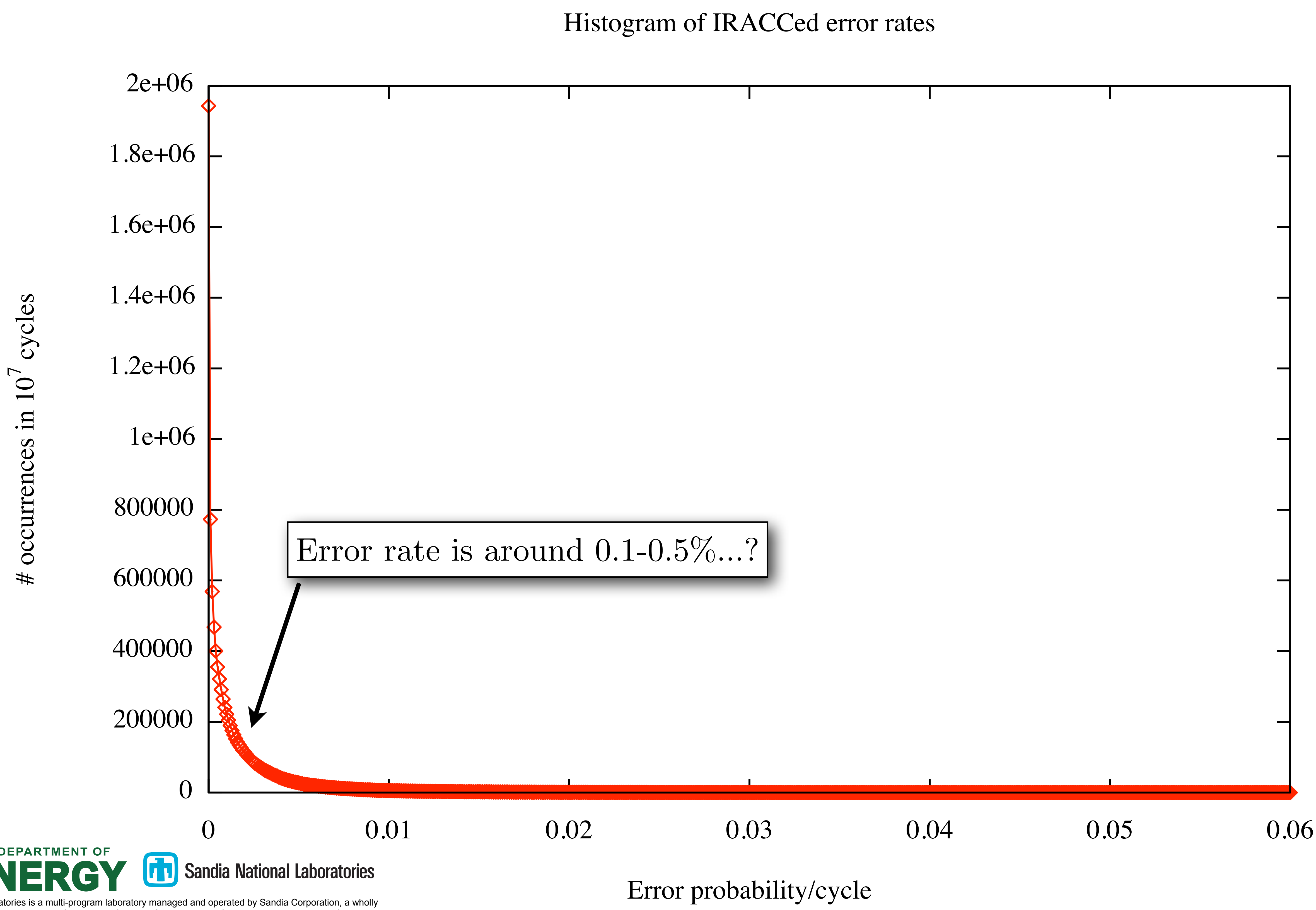
1. Count cycles (T) until $P \approx 1..5$ errors have occurred.
2. Guess $p_{\mathrm{err}} = \theta^2 = P/T \Longrightarrow \theta = \sqrt{P/T}$
3. Rotate by $\pm\theta$.
4. Next time, rotate the other way. GOTO 1.

So... how low can we keep the error rate?

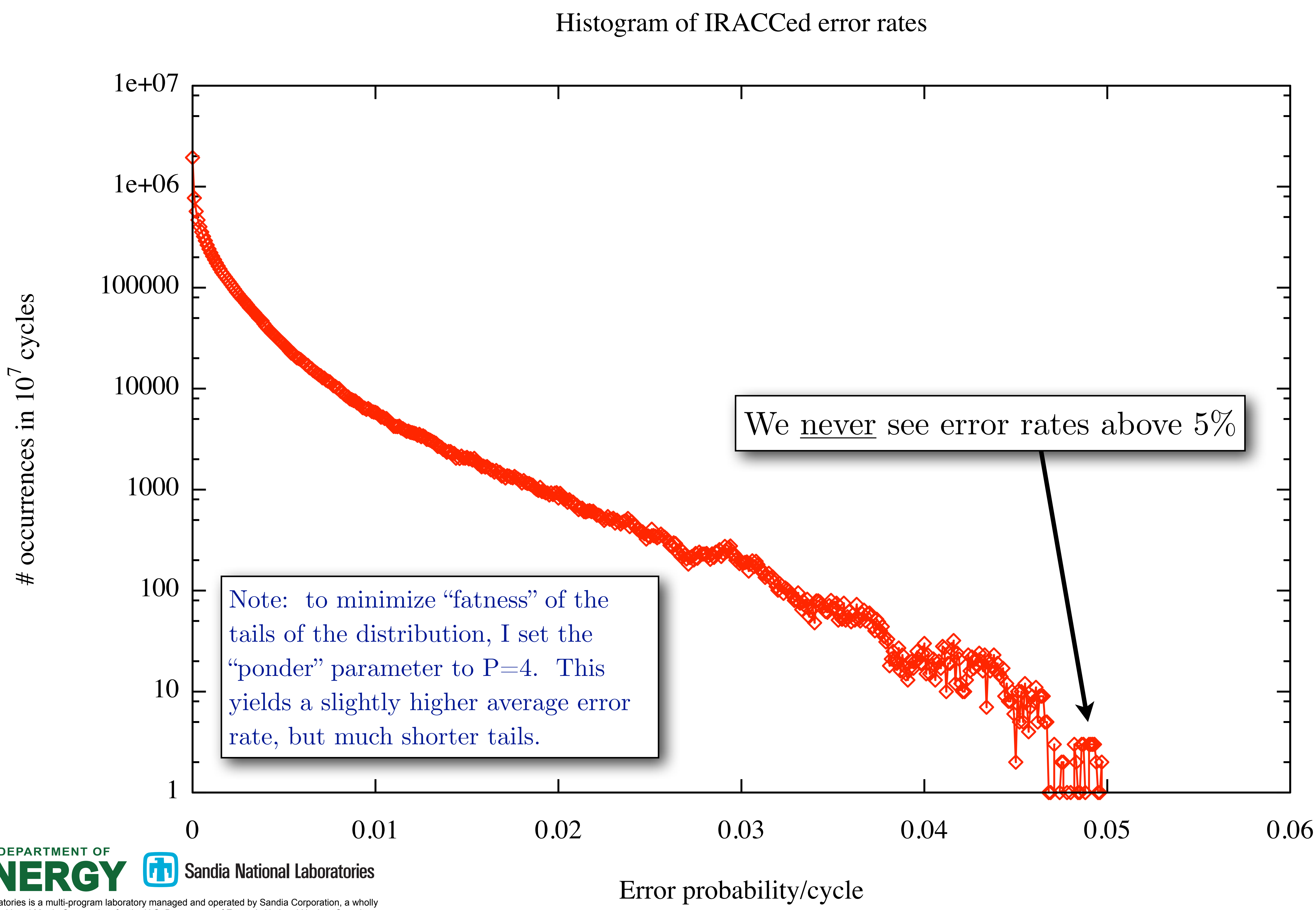
Hmm, that's hard to see



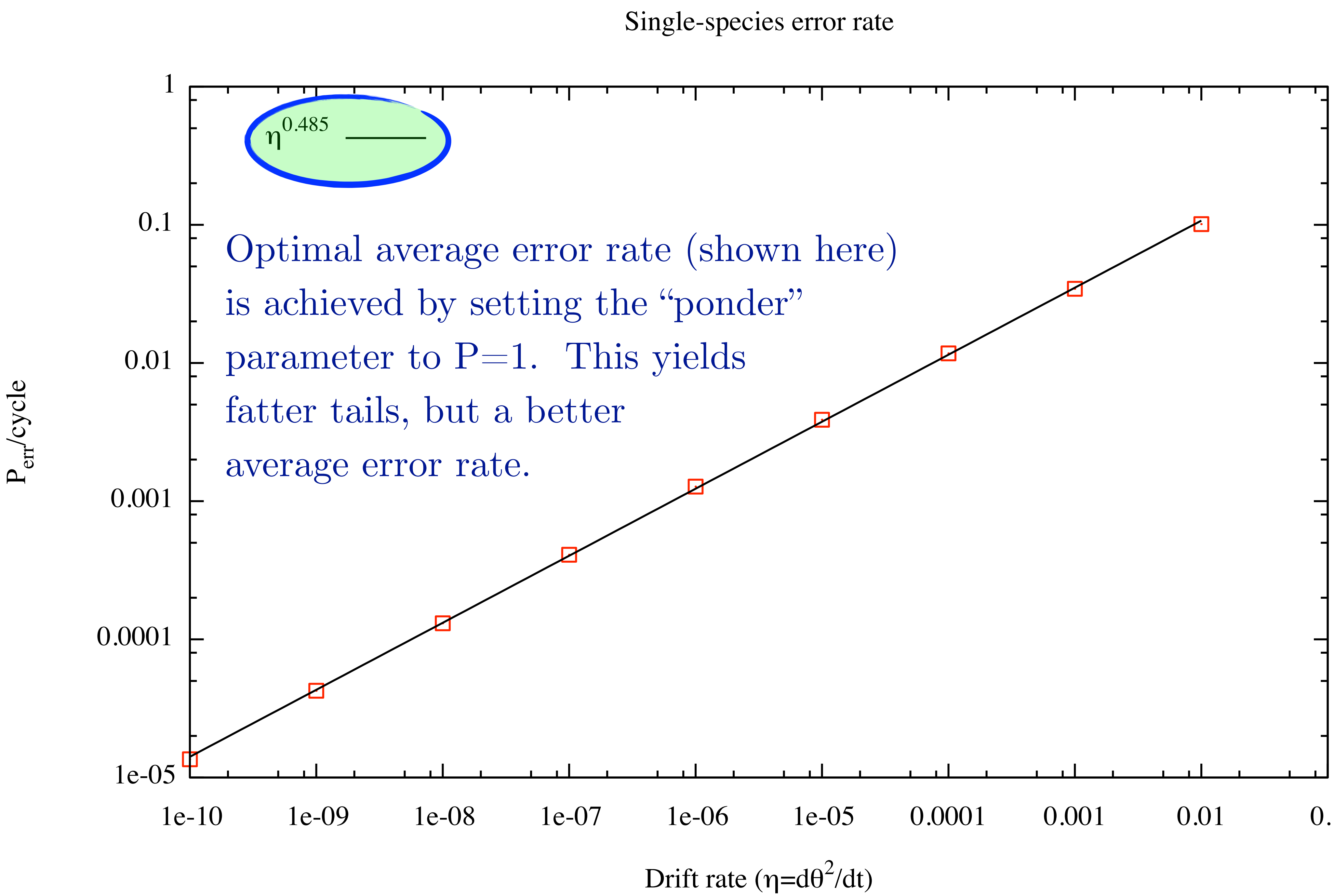
Still hard to see.



How about a log scale?



How does the stabilized error rate scale with η?



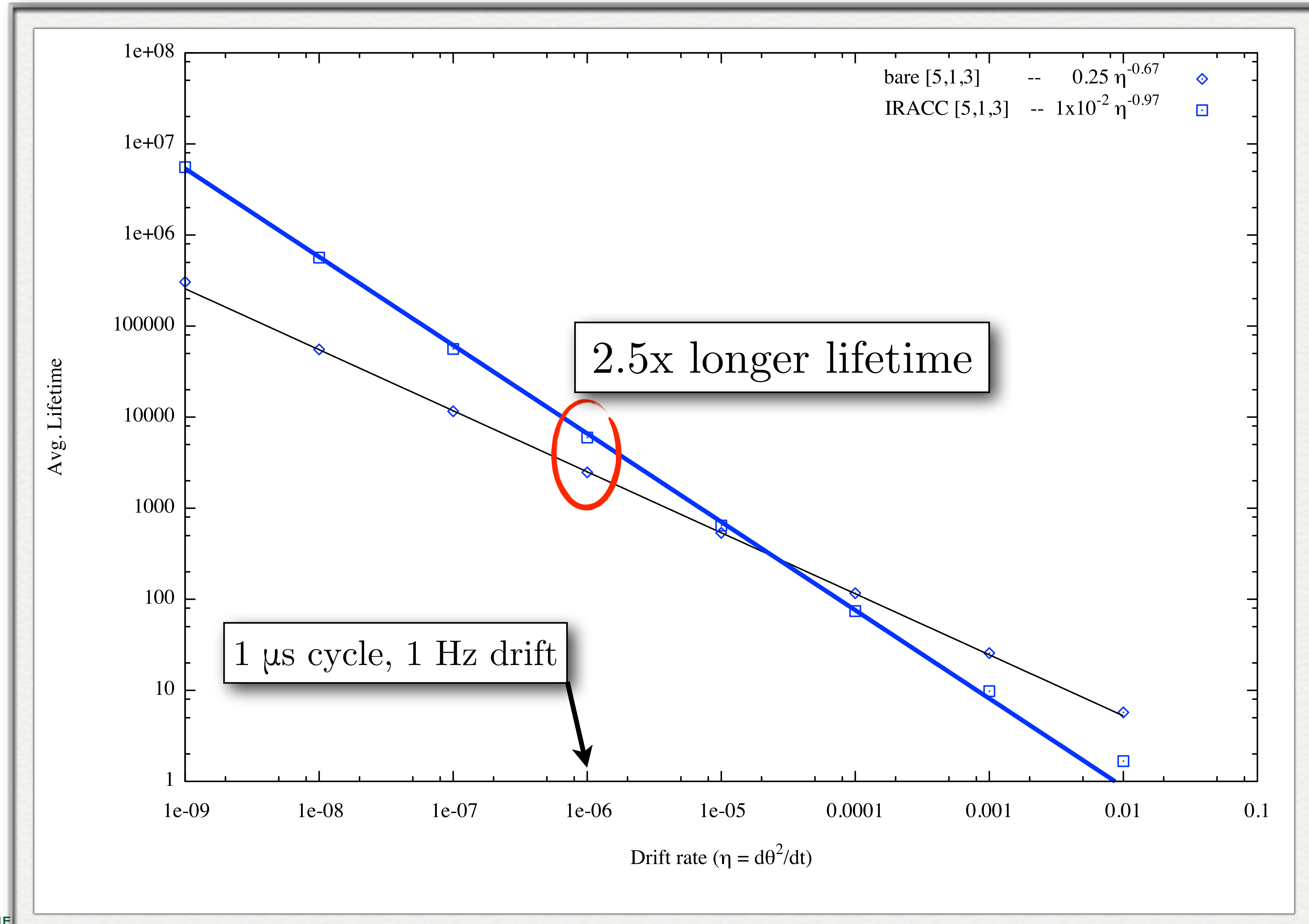
Okay. Monitoring syndrome measurements can stabilize drift. **But does it help QEC?**

Let's apply this to a real QECC. Like [5,1,3]. Ignore all implementation details -- syndrome measurements are magic, instant, & perfect.

Now we have to monitor 15 syndrome streams and adjust 15 terms in the Hamiltonian. Two errors in the same cycle => logical qubit dies.

The natural metric is <u>avg. logical qubit lifetime.</u>

Drift control does improve logical qubit lifetime -- moderately.



Maybe a better error-correcting code would make my results look more impressive?

The 5-qubit code is a distance 3 code.
It can only tolerate 1 error.

A distance-7 code can tolerate 2 errors.
A distance-9 code can tolerate 3 errors.

How many qubits do we need to tolerate 1,2,3 errors (d=3,5,7)?

Home

Linear Codes

QECC

Acknowledgments

How to cite

References

Links

Contact

Bounds on the minimum distance of additive quantum codes

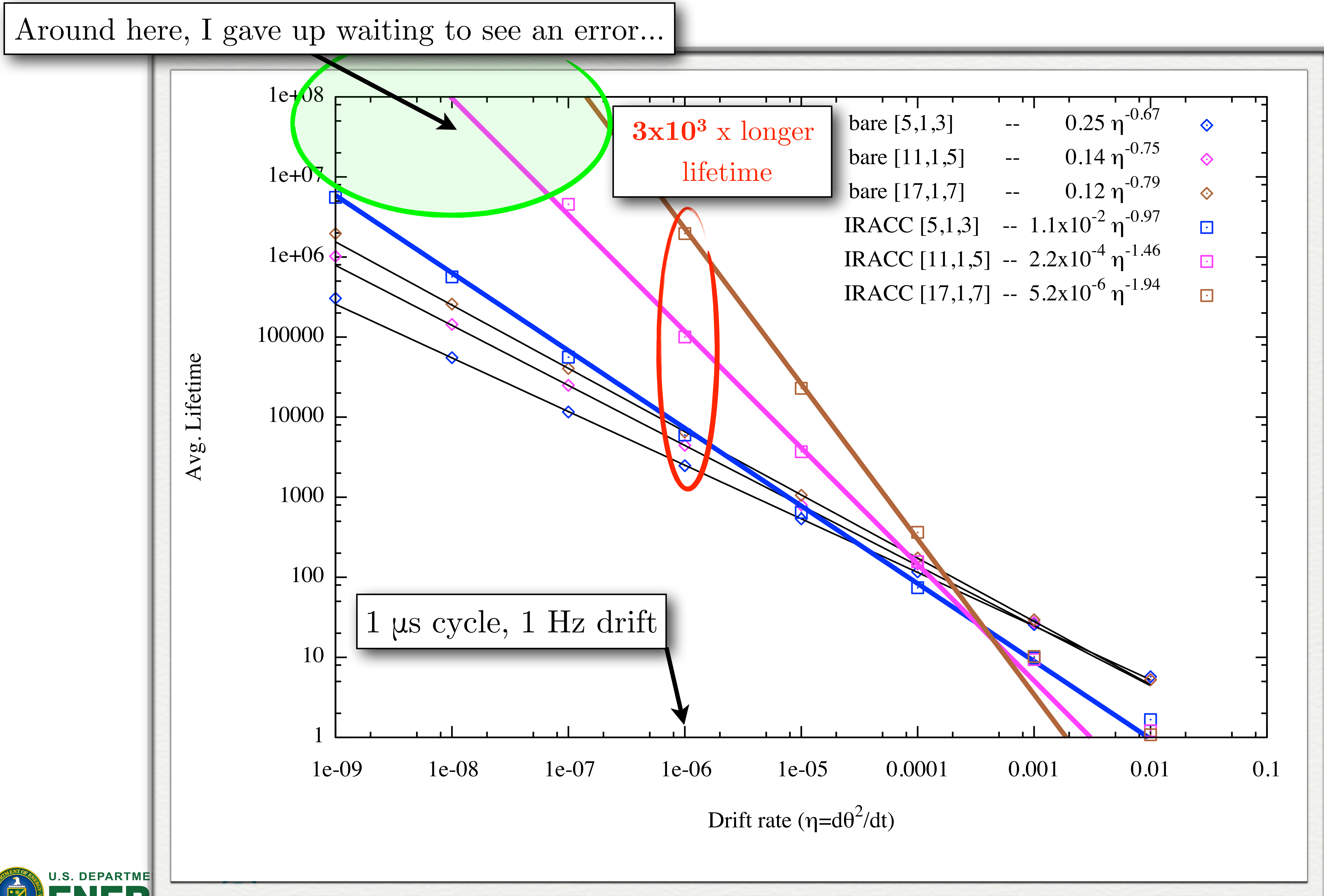
larger tables for n≤50, n≤128

n/k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	1	1																					
2	2	1	1																				
3	2	1	1	1																			
4	2	2	2	1	1																		
5	3	3	2	1	1	1																	
6	4	3	2	2	2	1	1																
7	3	3	2	2	2	1	1	1															
8	4	3	3	3	2	2	2	1	1														
9	4	3	3	3	2	2	2	1	1	1													
10	4	4	4	3	3	2	2	2	2	1	1												
11	5	5	4	3	3	3	2	2	2	1	1	1											
12	6	5	4	4	4	3	3	2	2	2	2	1	1										
13	5	5	4	4	4	3	3	3	2	2	2	1	1	1									
14	6	5	5	4-5	4	4	4	3	3	2	2	2	2	1	1								
15	6	5	5	5	4	4	4	3	3	3	2	2	2	1	1	1							
16	6	6	6	5	5	4-5	4	4	3	3	3	2	2	2	2	1	1						
17	7	7	6	5-6	5	4-5	4-5	4	4	4	3	3	2	2	2	1	1	1					
18	8	7	6	5-6	5-6	5	5	4	4	4	3	3	2	2	2	2	2	1	1				
19	7	7	6	5-6	5-6	5-6	5	4-5	4	4	3-4	3	3	2	2	2	2	1	1	1			
20	8	7	6-7	6-7	6	5-6	5-6	4-5	4-5	4	4	3-4	3	3	2	2	2	2	2	1	1		
21	8	7	6-7	6-7	6-7	6	5-6	5-6	4-5	4-5	4	4	3-4	3	3	3	2	2	2	1	1	1	
22	8	7-8	6-8	6-7	6-7	6-7	5-6	5-6	5-6	4-5	4-5	4	4	3-4	3	3	2	2	2	2	2	1	1

Borrowed from the internets (thanks to Markus Grassl)

Markus Grassl (grassl@ira.uka.de), IAKS, Fakultät für Informatik, Universität Karlsruhe (TH)

Yes, Virginia, there is a Santa Claus



Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Frankly, this is a slightly boring conclusion. The theory curves on the previous slide were, indeed, theory predictions based on counting logical errors and using $p_{\text{err}} = \eta^{0.485}$.

It works pretty much perfectly.

So, in summary, we can stabilize drift, we end up with an effective *incoherent* error rate of

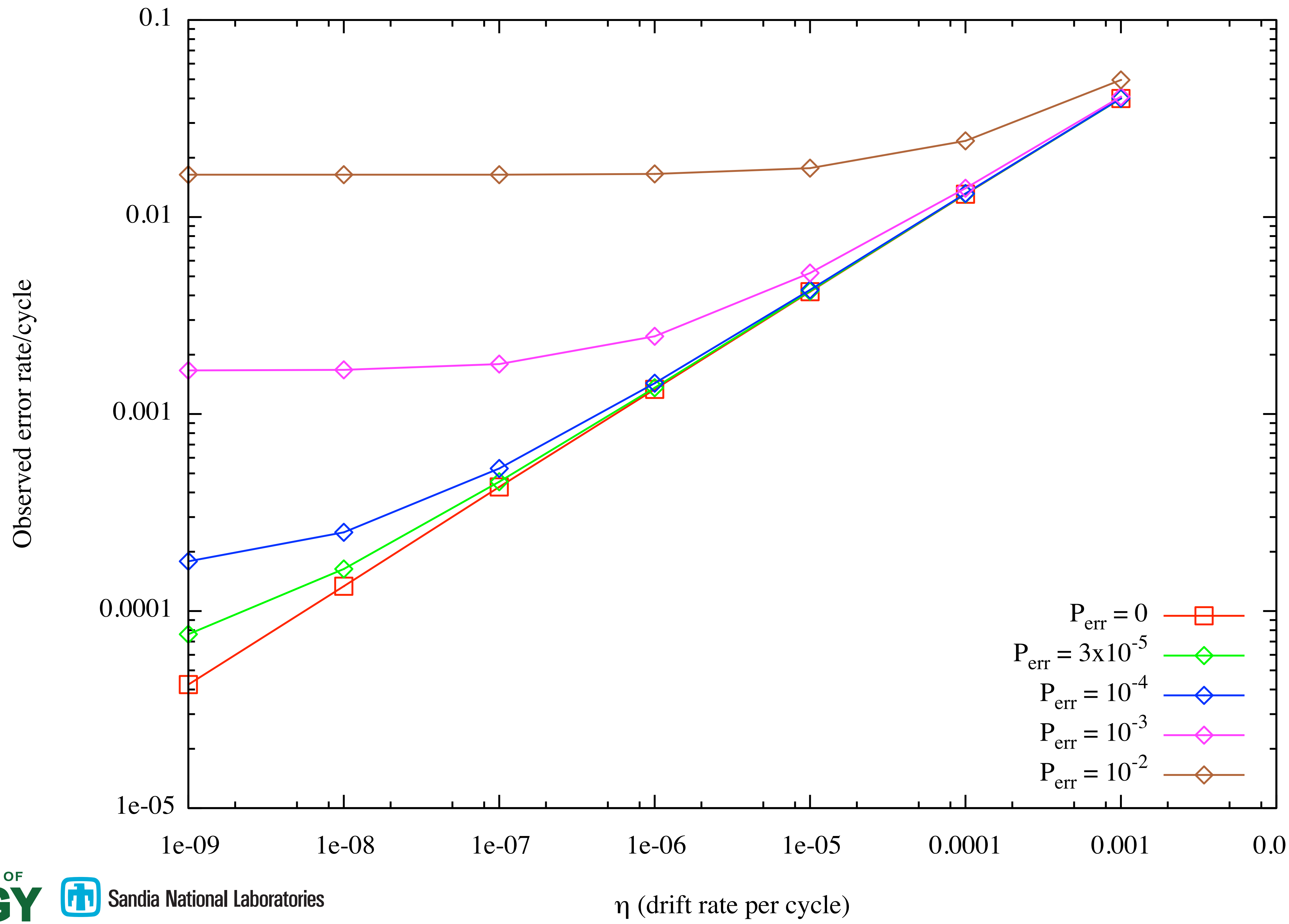
$$p_{\text{err}} = \eta^{0.485}$$ (why not 0.5? I have no idea, but that's the fit...)

and everything else (e.g. QEC) just follows from that.

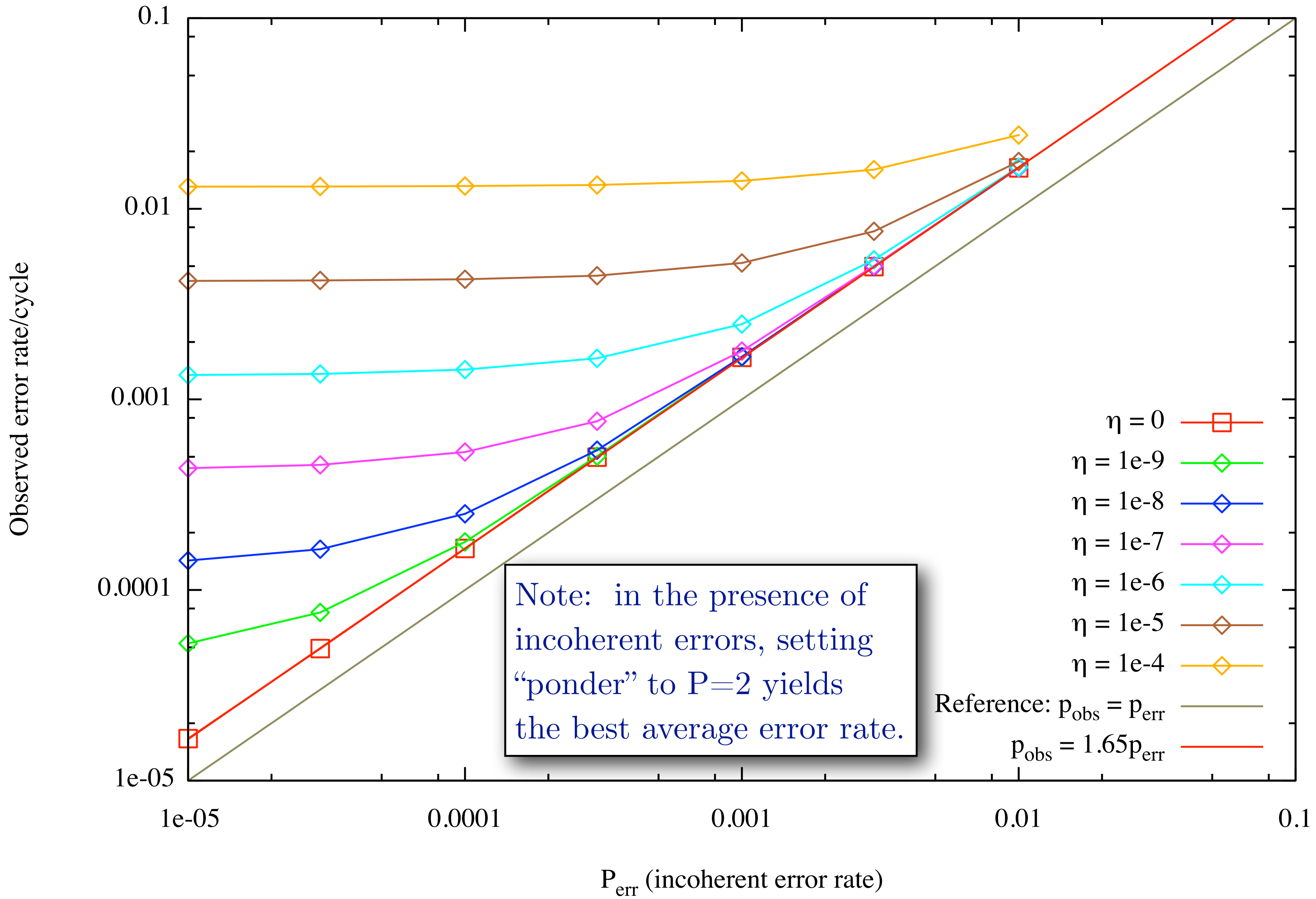
But what happens if we also have <u>incoherent</u> noise?

- This is a pretty big deal. There's going to be some depolarizing or random/incoherent noise too. We need a controller that's robust to it!
- Incoherent noise causes errors (nontrivial syndrome measurements) even when $\boldsymbol{H}=0$. There's no intrinsic way to tell these errors from ones caused by non-zero $\boldsymbol{H}$!
- It seems like we need a new control algorithm -- or at least a way of telling our old one about incoherent noise.
- To my great annoyance, the totally dumb algorithm achieves the best performance of anything I've tried.

This is what happens if we just run it with P_{err} incoherent noise.



Same error rates, vs. P_{err} (not η)

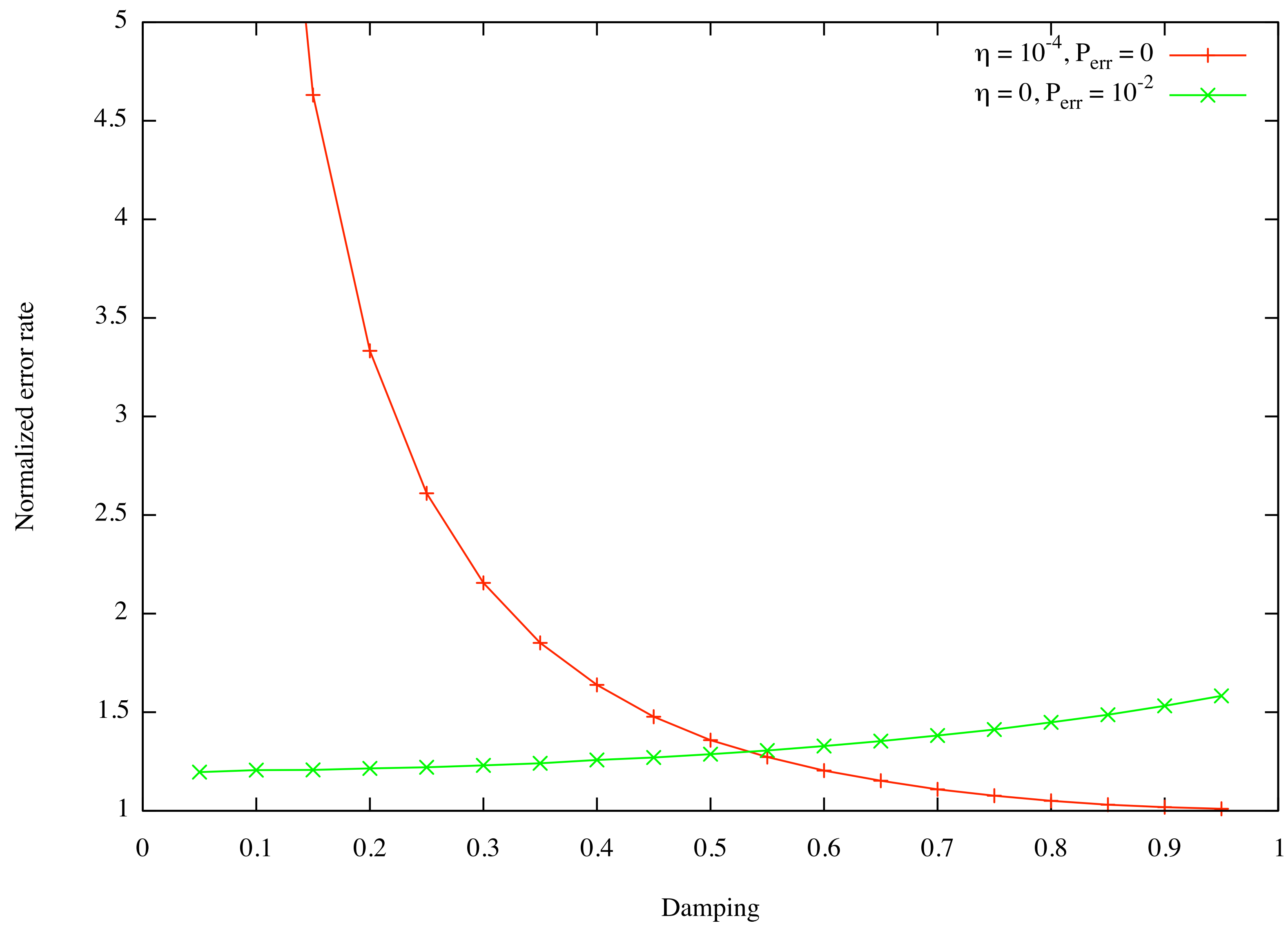


Note: in the presence of incoherent errors, setting "ponder" to P=2 yields the best average error rate.

Notice that even for η=0, our control introduces extra noise -- we see $p_{\text{obs}} = 1.65p_{\text{err}}$, not $p_{\text{obs}}=p_{\text{err}}$.

U.S. DEPARTMENT OF ENERGY
Sandia National Laboratories
Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

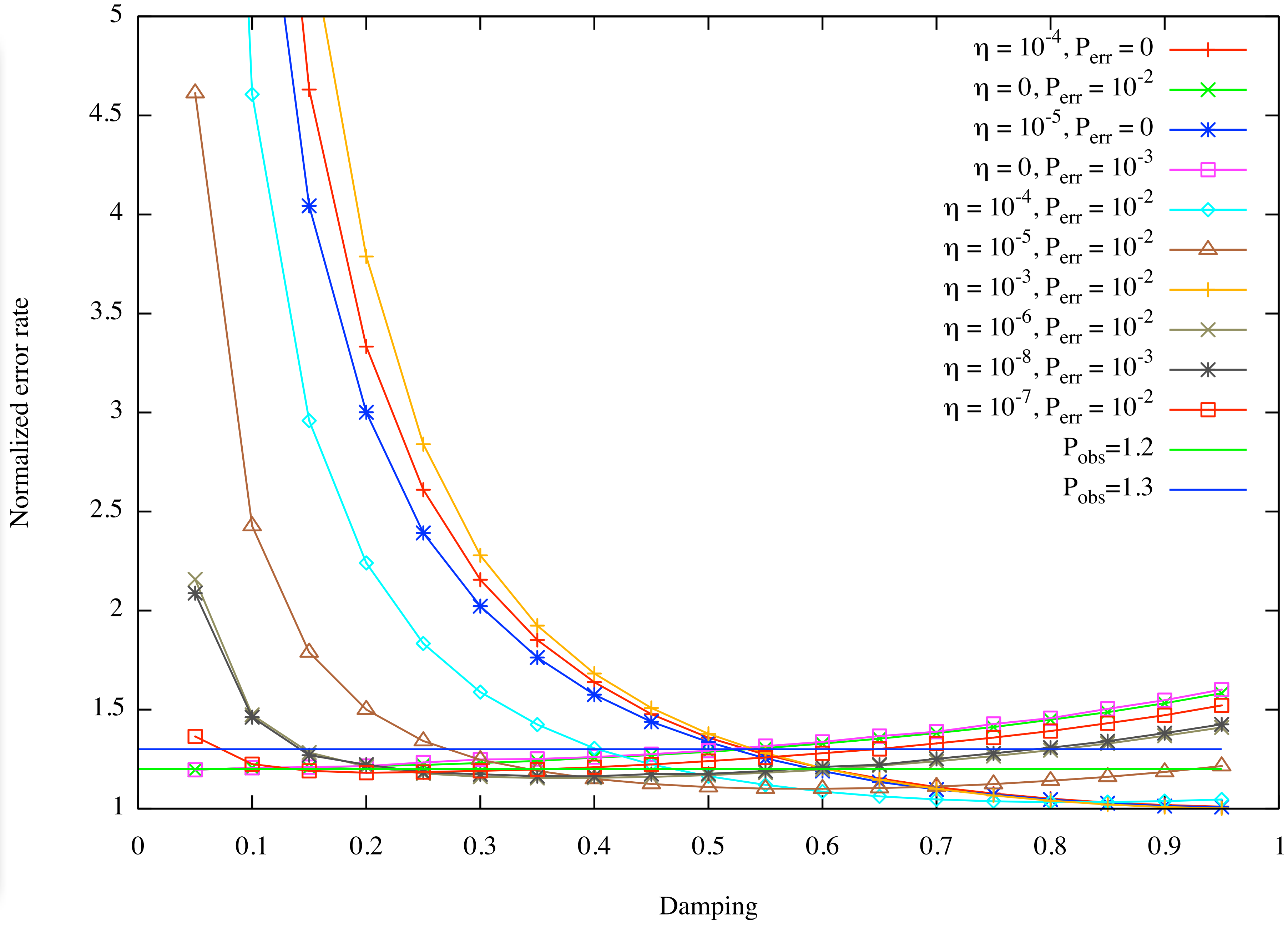
We can tweak the algorithm by "damping" each correction by a factor of d<1. This helps for incoherent errors.



Plotted error rates are *normalized* -- divided by the minimum achievable error rate over all damping values. Note that damping (d<1) is good for incoherent errors, but [very] bad for correcting drift.

A representative variety of normalized error rates vs. damping. We can't beat 1.2 for pure incoherent noise... but d=0.55 achieves ≤1.3 in all cases. ☺

Note: the "normalized" error rate shown here is obtained by adding together (1) the incoherent rate p_{err}, and (2) the lowest error rate obtainable with any strategy for η, which (as noted previously) appears to be $\eta^{0.485}$. It's actually somewhat remarkable that this sum (which should be a lower bound for the observed error rate) seems to be always achievable to within 20% for some damping value, and to within 30% with d=0.55.



Summary: Not only can we stabilize drift pretty well, but we can do it almost equally well even in the presence of incoherent noise.

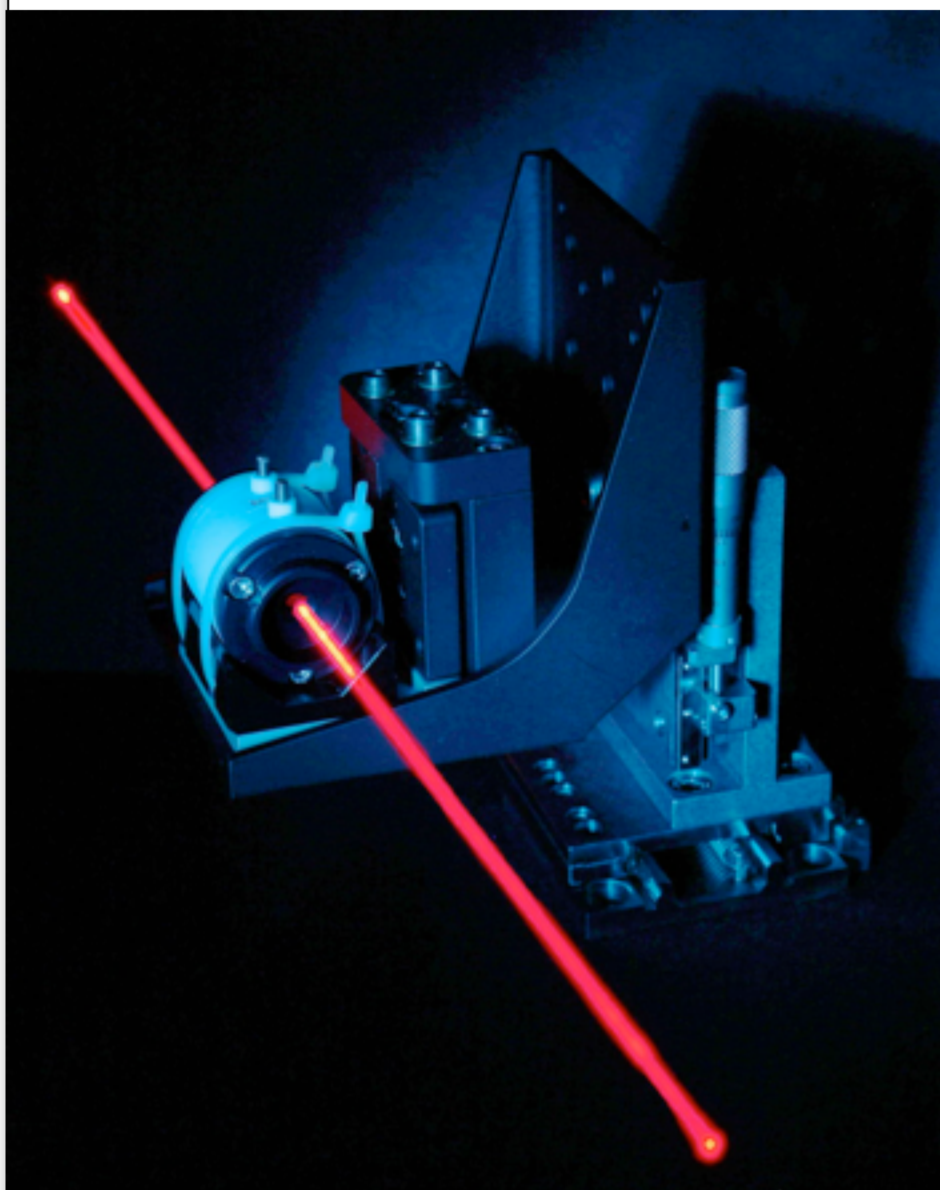
The best strategy I've found to do so is to run the dumb algorithm with "ponder" $P=2$ and damping $d=0.55$ or so.

This always gets within a factor of 2 (often less) of the theoretical limit for observed error rate.

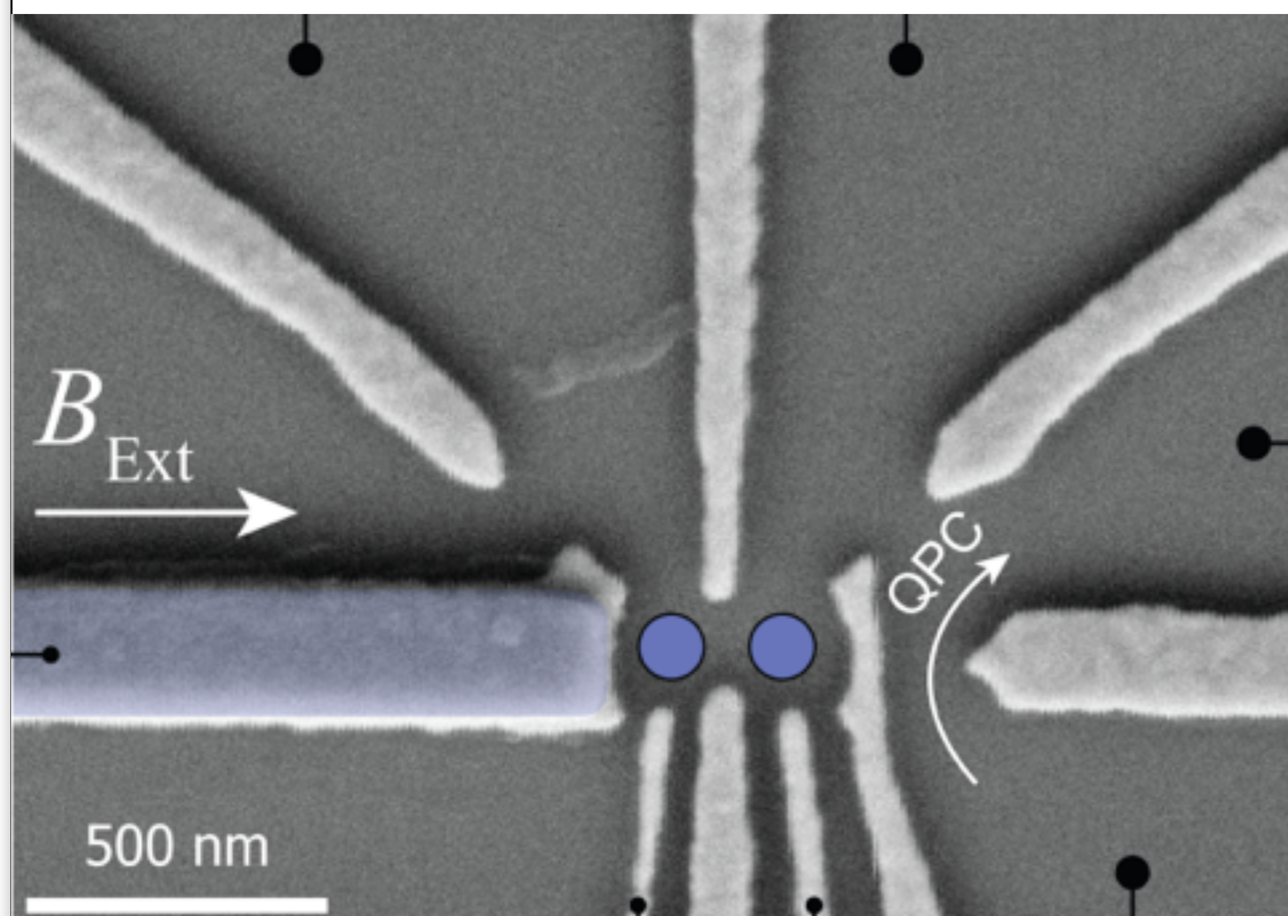
Gate-Set Tomography

The Problem With Tomography

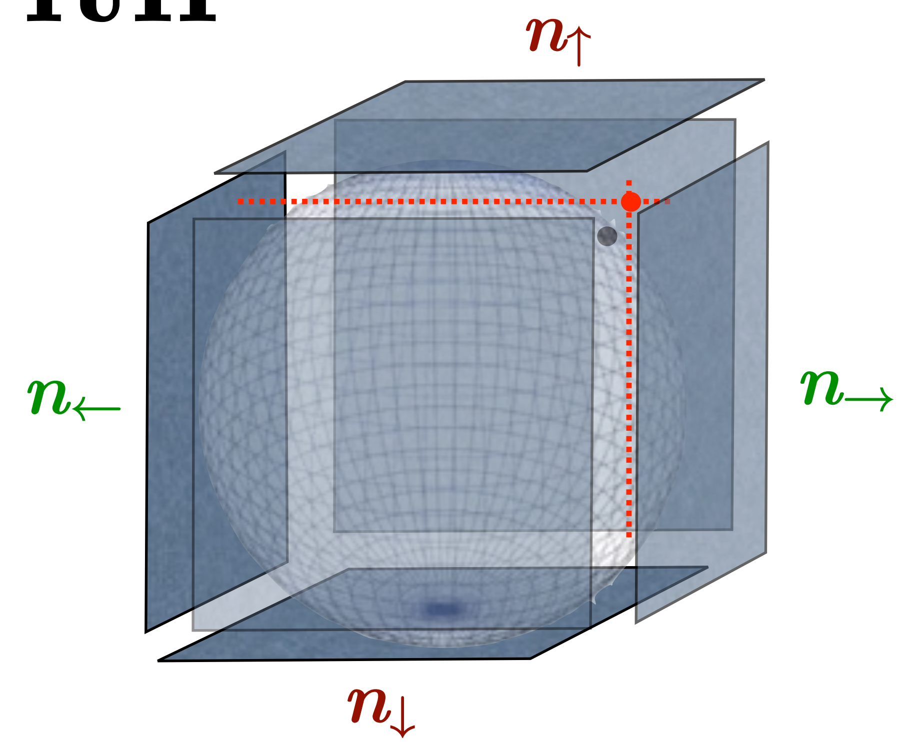
Invented here



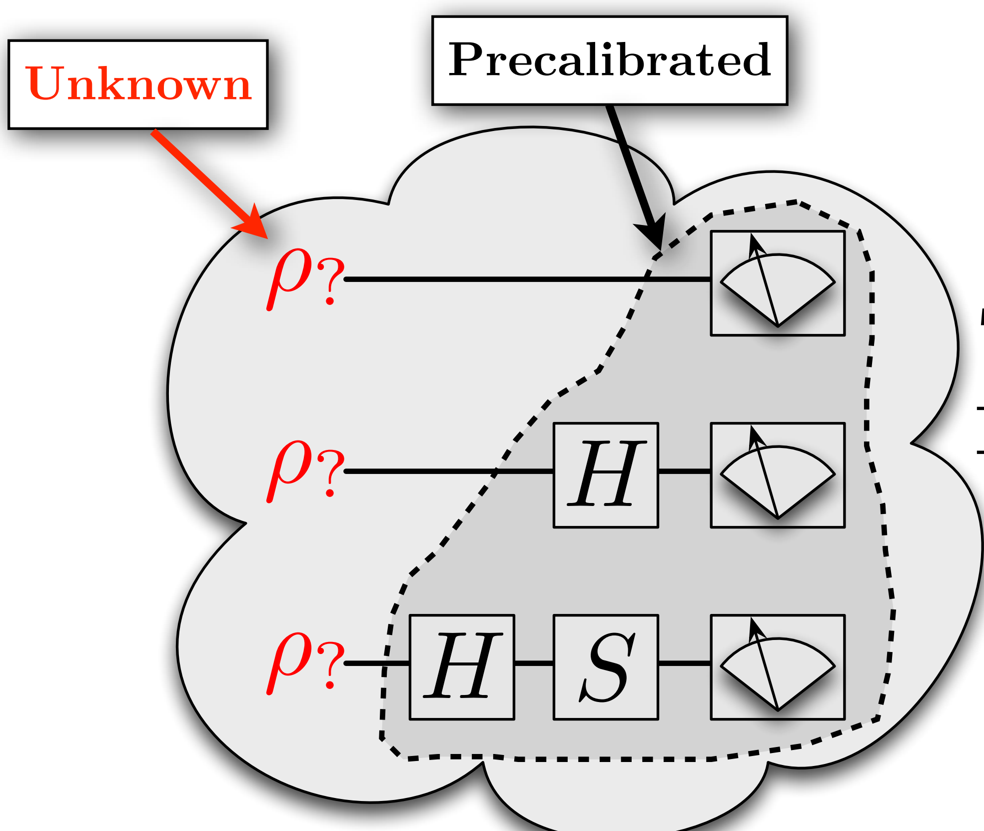
I need it here



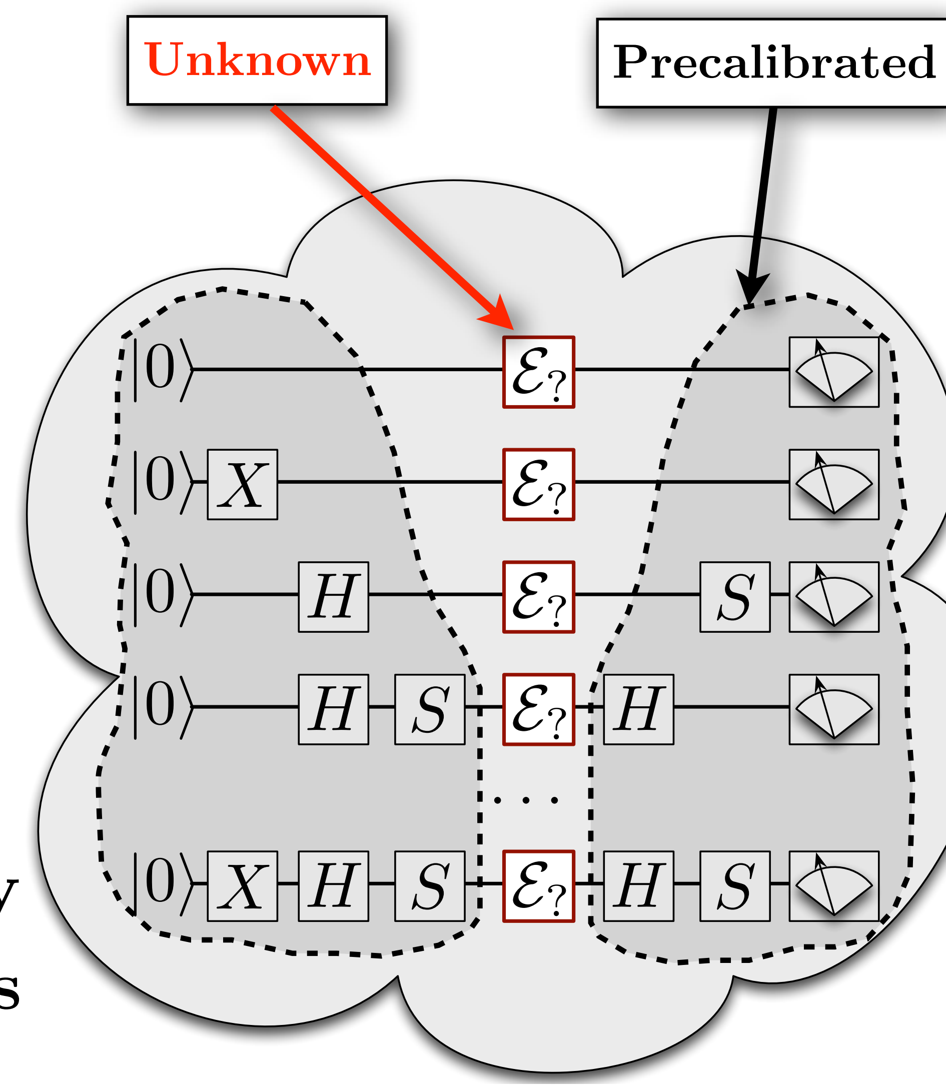
The Problem With Tomography



- Standard state / process tomography uses **precalibrated reference frames.**
- Most QIP technologies don't provide native X, Y, Z states/measurements.



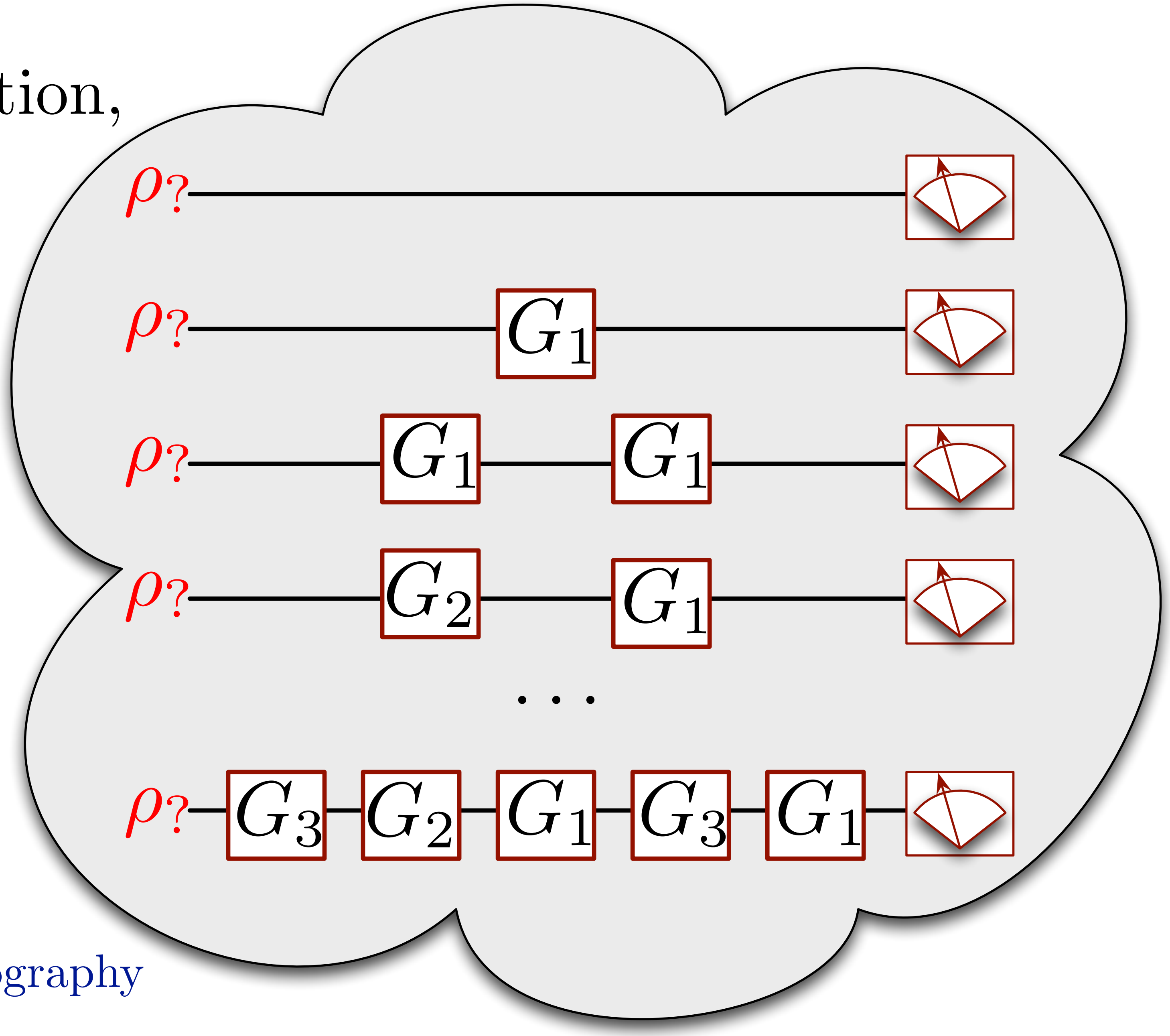
State Tomography Experiments



Process Tomography Experiments

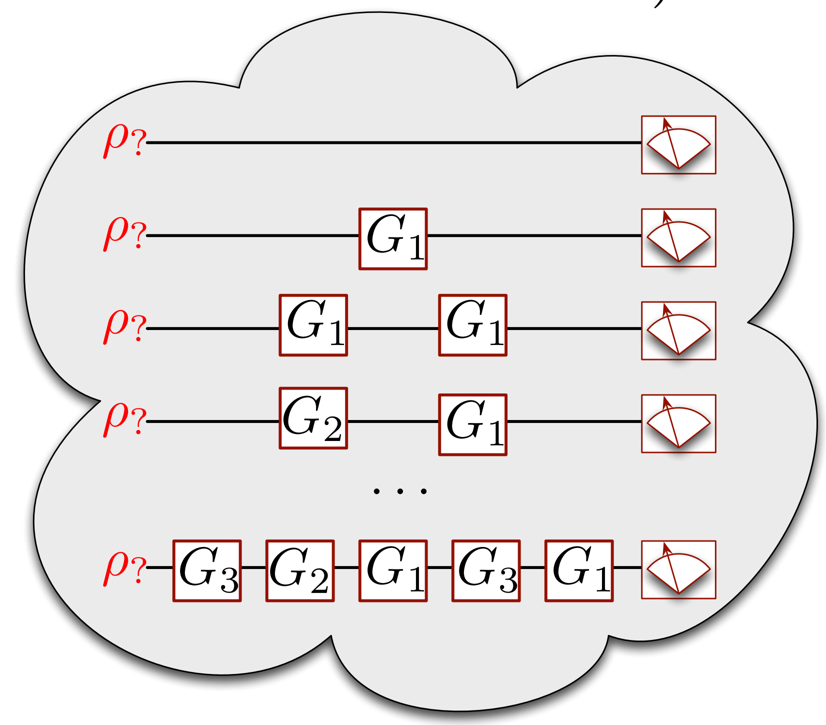
Gate-Set Tomography

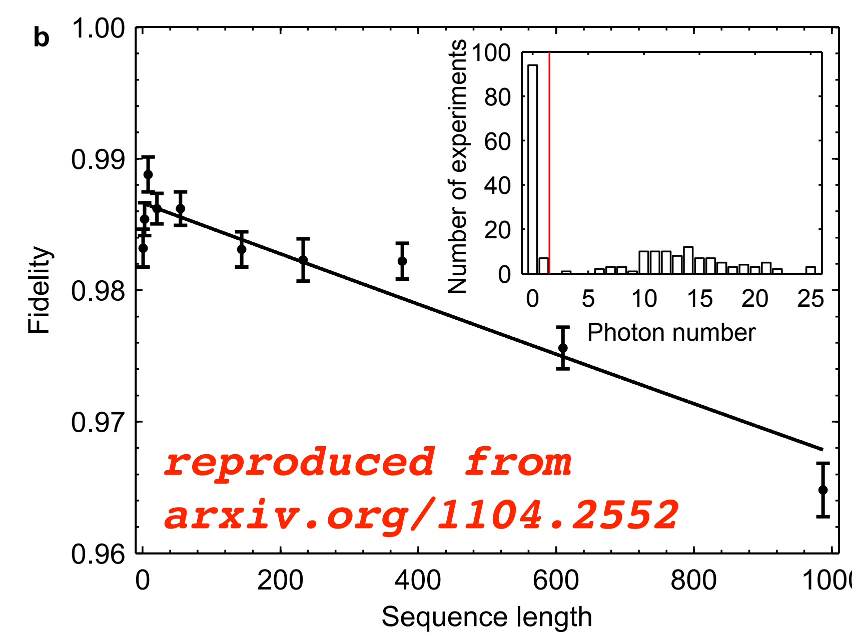
- Assume *nothing* about preps, operations, or measurements.
- Everything (prep, operation, measurement) is a *gate*.
- Do lots of different *sequences* of gates.
- Estimate the entire *gate set* at once.
- "Self-consistent" (Toronto), "Overkill" (IBM), other groups...
- This subsumes state/process tomography



Randomized Benchmarking

- Assume *nothing* about prep/measurements (robust to "SPAM error").
- Assume that dynamical *gates* are pretty good Clifford operations.
- Do many <u>random</u> gate *sequences*, all of which <u>would</u> perform the identity operation <u>**if**</u> the Clifford gates were perfect.
- Measure the rate of decay of success probability.
- RB efficiently yields "per-gate error rate"... but provides absolutely <u>***no***</u> diagnostic/debugging information.





RB vs. GST

- **RB and GST share a common foundation:**
  - robust to SPAM (preparation/measurement) error,
  - rely on [long] *sequences* of gates (unlike traditional tomography)
  - designed to test & verify real quantum hardware.

- RB is simpler than GST. Much easier to crunch the data.

- But **RB provides no diagnostic info** (just a per-gate error rate).
  GST tells us <u>exactly</u> what operations are being performed.

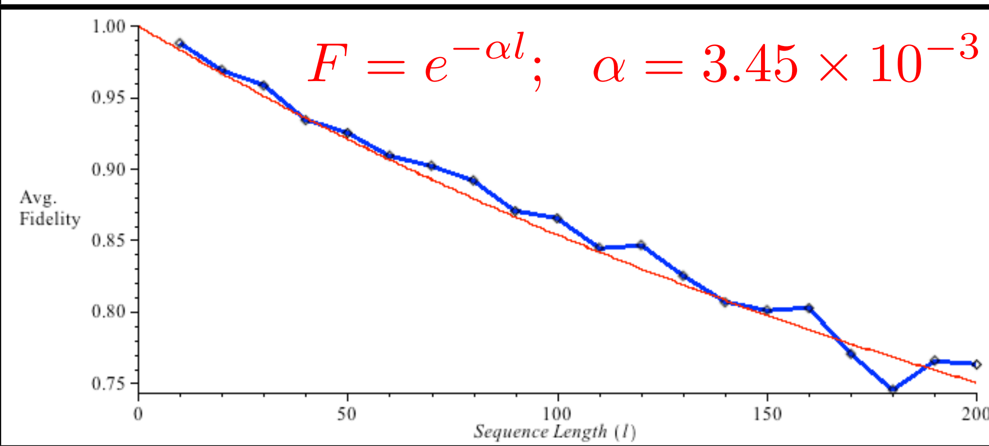
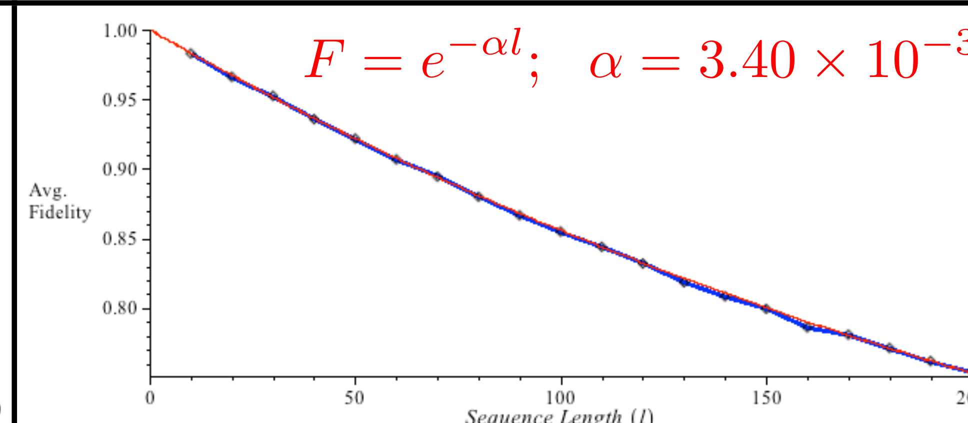
- RB *requires* that sequences of gates be chosen uniformly at random:
  "error rate" is an average over all strings.

- GST only requires a fixed # of sequences (~20 for 1 qubit) to identify the gates exactly -- and they can be chosen to optimize performance.

RB vs. GST (simulation)

We applied RB and GST to two different error models on a typical gate set:

$$\{H = e^{i\frac{\pi}{4}Y},\ S = e^{i\frac{\pi}{4}Z}\}$$

	Overrotation (2% for H, 4% for S) + 10^{-3} depolarizing noise	**Unequal depolarizing noise ($5\text{x}10^{-3}$ for H, $1\text{x}10^{-3}$ for S)**
Randomized Benchmarking averaged over K=130 strings at each length $\boldsymbol{l}$		
Gate-Set Tomography K=63 strings total; M=1000 reps each	**<u>True</u> superoperators (acting on ⟨X,Y,Z⟩)** $\Lambda_H = \begin{pmatrix} -0.0314 & 0 & -0.9985 \\ 0 & 0.9990 & 0 \\ 0.9985 & 0 & -0.0314 \end{pmatrix}$ $\Lambda_S = \begin{pmatrix} -0.0627 & 0.9970 & 0 \\ -0.9970 & -0.0627 & 0 \\ 0 & 0 & 0.9990 \end{pmatrix}$ $\widehat{\Lambda_H} = \begin{pmatrix} -0.0306 & 0.0100 & -0.9988 \\ 0.0027 & 0.9982 & -0.0002 \\ 0.9984 & 0.0129 & -0.03928 \end{pmatrix}$ $\widehat{\Lambda_S} = \begin{pmatrix} -0.0637 & 0.9972 & 0.0052 \\ -0.9975 & -0.0610 & -0.0066 \\ -0.0169 & -0.0067 & 0.9987 \end{pmatrix}$ **GST estimated superoperators**	**<u>True</u> superoperators** $\Lambda_H = \begin{pmatrix} 0 & 0 & -0.9950 \\ 0 & 0.9950 & 0 \\ 0.9950 & 0 & 0 \end{pmatrix}$ $\Lambda_S = \begin{pmatrix} 0 & 0.9990 & 0 \\ -0.9990 & 0 & 0 \\ 0 & 0 & 0.9990 \end{pmatrix}$ $\widehat{\Lambda_H} = \begin{pmatrix} -0.0020 & 0.0090 & -0.9951 \\ -0.0009 & 0.9905 & -0.0044 \\ 0.9949 & 0.0126 & -0.0126 \end{pmatrix}$ $\widehat{\Lambda_S} = \begin{pmatrix} -0.0013 & 0.9994 & 0.0152 \\ -0.9995 & 0.0051 & -0.0121 \\ -0.0211 & 0.0001 & 0.9990 \end{pmatrix}$ **GST estimated superoperators**

Conclusion: RB does not distinguish overrotation from depolarization. GST *does* distinguish them, and accurately identifies the error so it can be fixed.

GST: Necessities

1. Preparation in a consistent (unknown) state ρ.
2. At least 2 different (unknown) noncommuting operations $\{G_k\}$.
3. Some measurement $\{E_m\}$ -- perhaps just 2 outcomes $\{E, \mathbf{1}\text{-}E\}$.

4. A quorum of different experiments, corresponding to distinct *gate strings*: $S_i = G_{i_1} \circ G_{i_2} \circ \ldots \circ G_{i_L}$.
5. Enough repetitions to estimate

$$\begin{aligned} Pr(E_m|\rho, S_i) &= \mathrm{Tr}\left[E_m S_i[\rho]\right] \\ &= \langle\langle E_m|\, G_{k_1} \circ G_{k_2} \circ \ldots \circ G_{k_L}\, |\rho\rangle\rangle \end{aligned}$$

Gauge/slack parameters

- A natural description of the unknowns we are estimating is:

  - $\rho = d\times d$ density matrix (d^2-1 params)
  - $G_k =$ CP-map (d^4-d^2 params)
  - $E_m =$ POVM effect (d^2 params)

- But some of these parameters are *invisible!*

  -- observables are invariant under transformations:

$$\begin{aligned} |\rho\rangle\rangle &\rightarrow \mathcal{U}\,|\rho\rangle\rangle \\ \langle\langle E_m| &\rightarrow \langle\langle E_m|\,\mathcal{U}^{-1} \\ G_k &\rightarrow \mathcal{U} G_k \mathcal{U}^{-1} \end{aligned}$$

  - $\mathcal{U}$ unitary ⇒ *gauge group/parameter*
  - $\mathcal{U}$ invertible ⇒ similarity transform ⇒ *slack parameter*
    - Slack parameters are gauges constrained by positivity.
    - SPAM noise is an example of a slack parameter.

Gates are relational

- Inability to estimate gauge/slack variables is ~~a real problem~~.
  utterly irrelevant!

- Parameters with no observable consequences are "ghosts".
  Any model (description of gates) that describes a sufficiently rich set of experiments will predict future experiments (e.g. quantum circuits) equally well.

- Absence of a reference frame is its own solution -- any set of gates that *acts* like $\{H,\ T,\ CNOT\}$ in all circumstances *is* $\{H,\ T,\ CNOT\}$!

- Gate-set tomography is akin to randomized benchmarking -- to benchmark circuit elements, *use* them -- but:
  (1) more powerful because we keep track of which string was applied,
  (2) more robust -- no need for precalibrated Clifford gates.

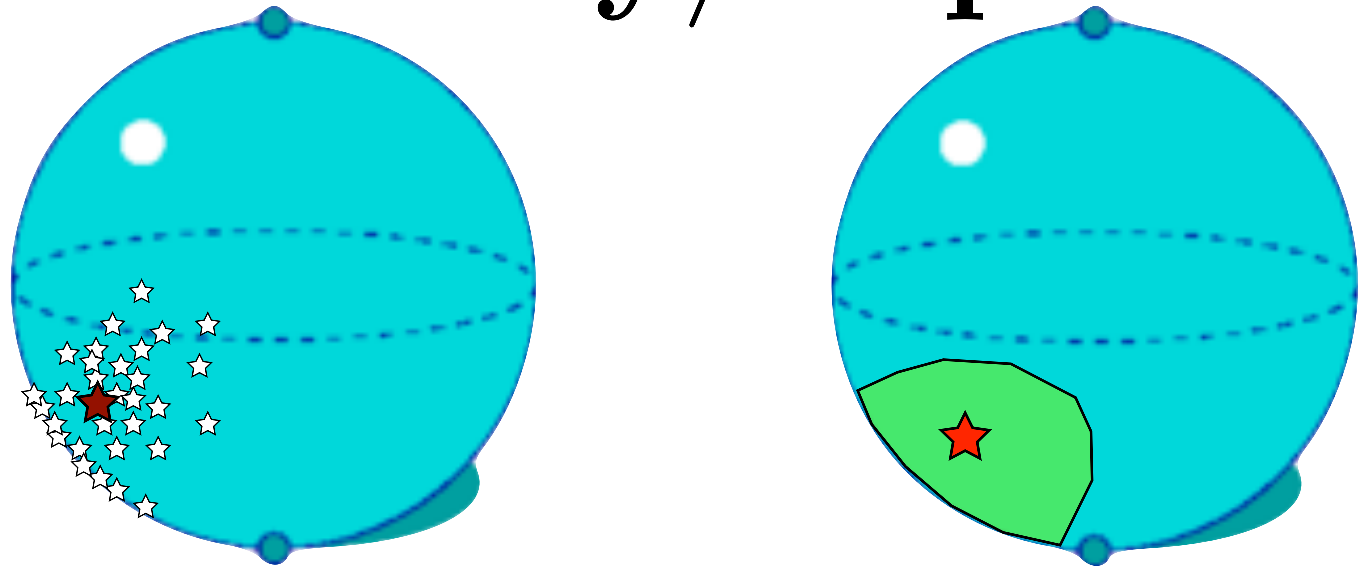
How to describe gate sets?

Natural description of the unknowns we are estimating:

- $\rho = d \times d$ density matrix (d^2-1 params)
- G_k = CP-map (d^4-d^2 params)
- E_m = POVM effect (d^2 params)

- This description is redundant and non-gauge-invariant.
- Open problem: find a gauge-invariant description of gate sets. **Why?**
  - More elegant -- might make estimation [much] simpler/faster.
  - We badly need an operationally meaningful measure of "fidelity" between: (1) real & ideal gate sets; (2) real & estimated gate sets.
    - N.B. Jamiolkowski-based quantities fail even for single processes because gates can be used *in sequence*, not just in parallel.

Inaccuracy/imprecision



- Two words for *almost* the same issue:
  point estimates are inaccurate; region estimates imprecise.
  Note that it's distinct from *error* (real vs. ideal gate).

- Minimizing imprecision (inaccuracy) should be our constant goal.

- Quantifying it is really important -- what metric should we use to measure the radius of a confidence region?
  -- *but first we need a framework to describe gate sets!!!*

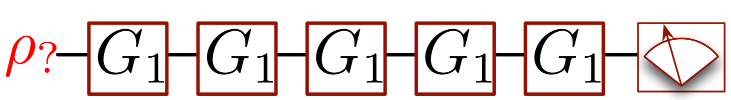
- General rule: "smaller" likelihood-ratio regions ⇔ better data ☺.

<u>Incomplete</u> gate sets: Rabi oscillation experiments

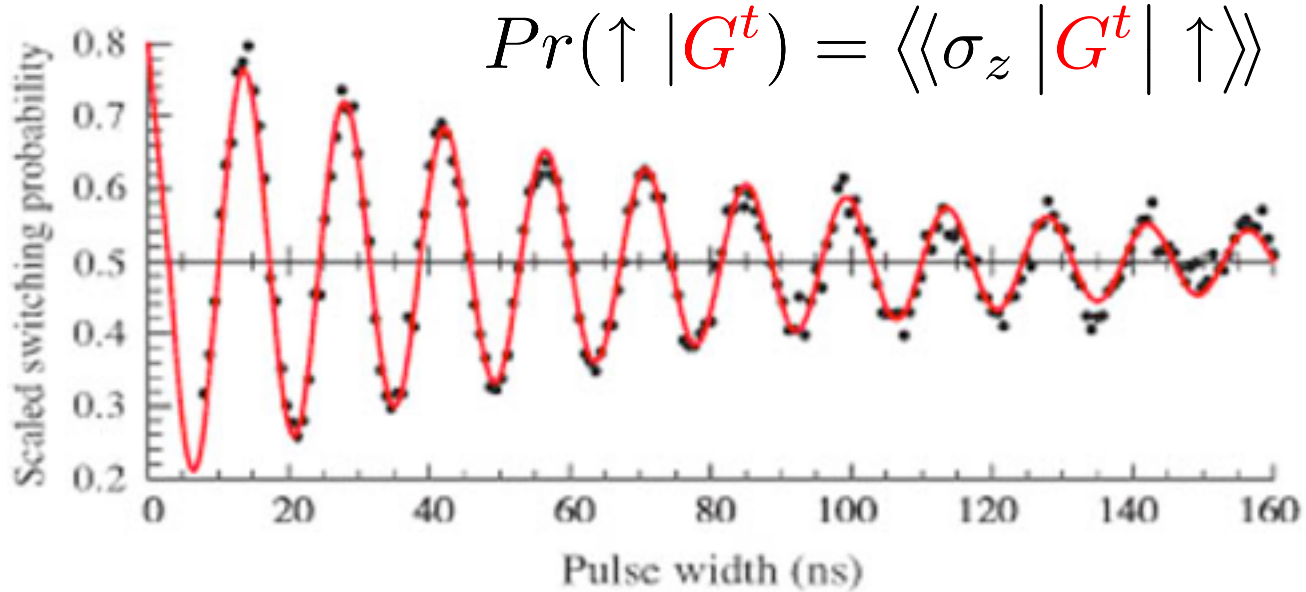
- What's new in gate-set tomography: **nonlinearity** in the $Pr(\text{data}|\{G_k\}) \Leftrightarrow \{G_k\}$ relationship.

$$\begin{aligned} Pr(E_m|\rho, S) &= \mathrm{Tr}\left[E_m S[\rho]\right] \\ &= \langle\langle E_m|\, G_{k_1} \circ G_{k_2} \circ \ldots \circ G_{k_L}\, |\rho\rangle\rangle \end{aligned}$$

- But for a *single* gate ($\{G_1\}$), this is just Rabi oscillations:



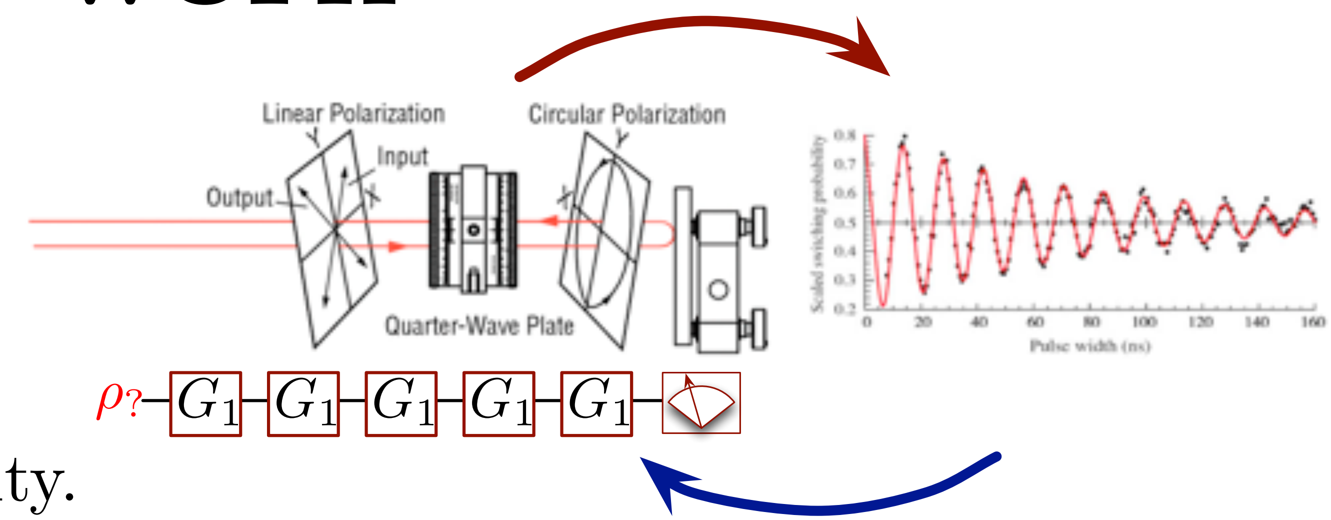
$$Pr(\uparrow|G^t) = \langle\langle \sigma_z | G^t | \uparrow\rangle\rangle$$



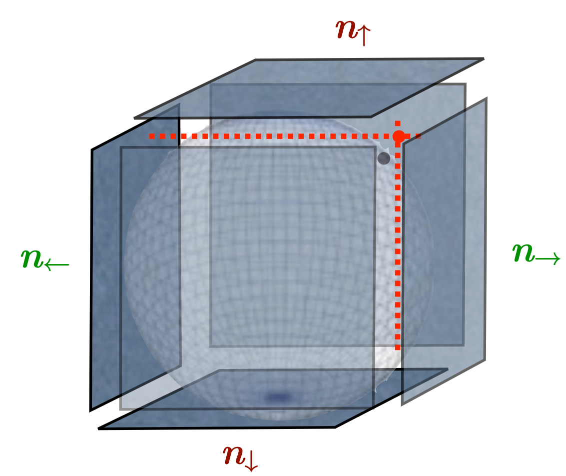
Borrowed from `http://physics.syr.edu/~bplourde/bltp-qcohere.htm`

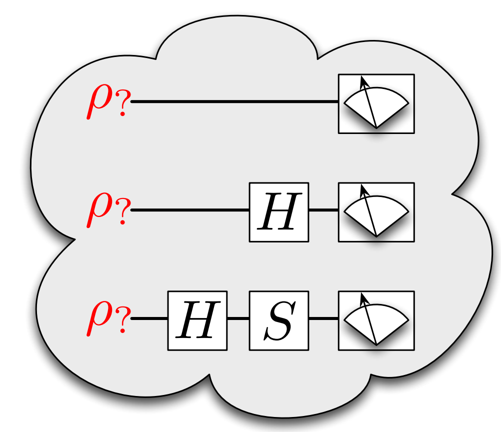
Why gate-set tomography <u>must</u> work

- Traditional tomography:
  (1) Calibrate essential gates using "other" methods that incorporate nonlinearity.



  (2) Use those calibrated resources to do tomography (linear).





- So all the information:
  (1) must be present *in the entire experimental chain*;
  (2) can be extracted using (even suboptimal) statistical analysis.

- Gate-set tomography is just the obvious step of *integrating* these parts to get (potentially huge) improvements in efficiency and precision.

Adaptive Measurements

Adaptive State Tomography

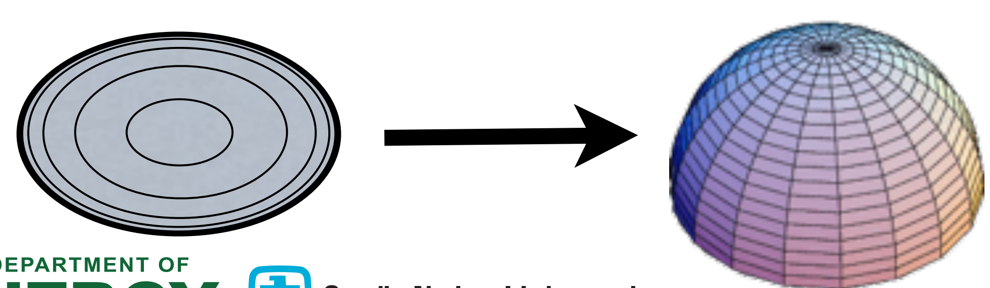
- **Punchline**: *adaptive* state tomography can achieve dramatic improvements in infidelity -- $\color{red}{1/\sqrt{N}} \rightarrow \color{blue}{1/N}$.

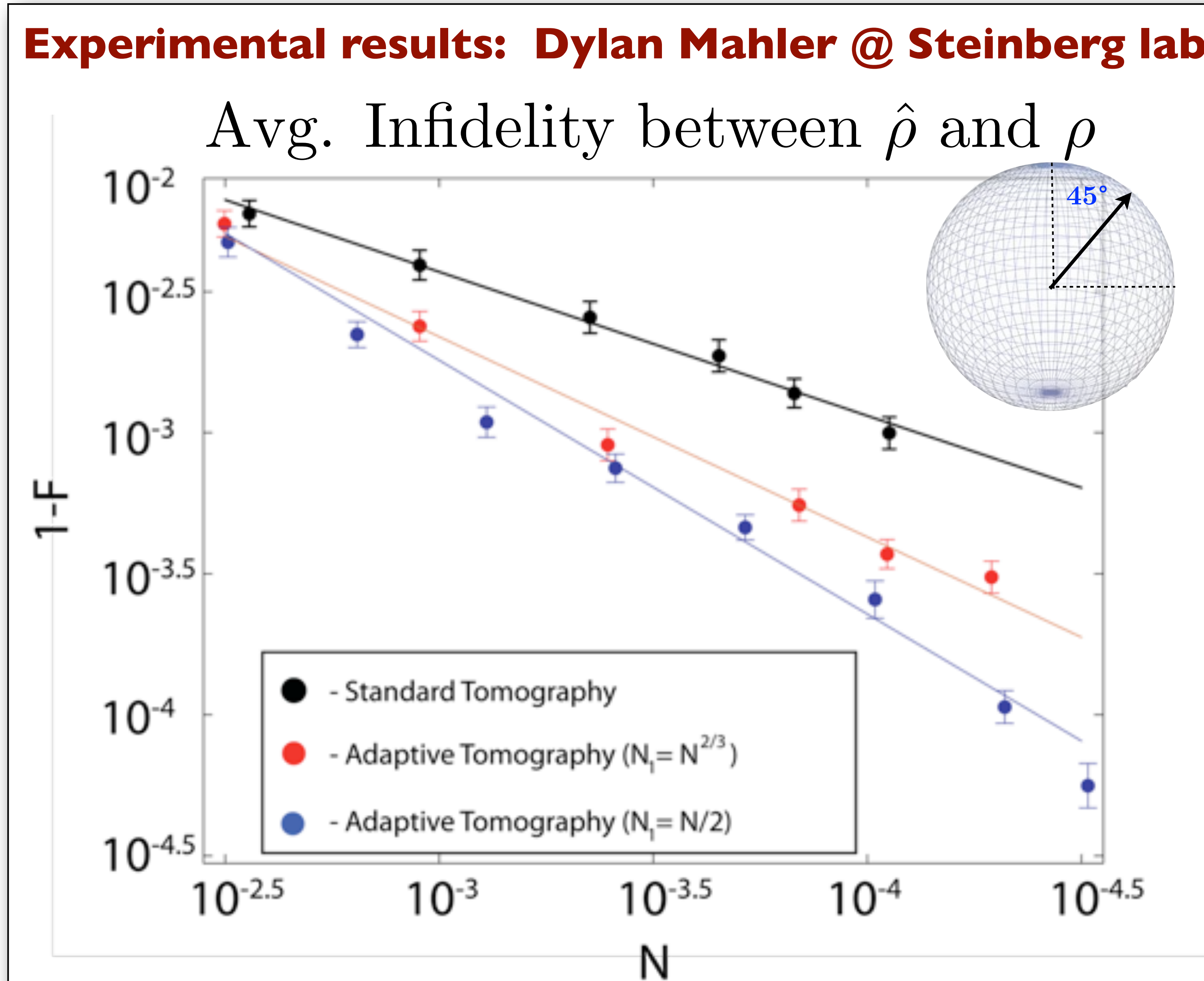
- First observed by Bagan et al, *PRL 97, 130501 (2006)*.

- Further theory + experiment; *Dylan Mahler, RBK et al* `arxiv:0303.1346`

- **How?** Measure eigenbasis of ρ.

- **Why?**
  (1) Zero probabilities can be estimated very accurately (1/N).
  (2) **Fidelity is not trace norm!**

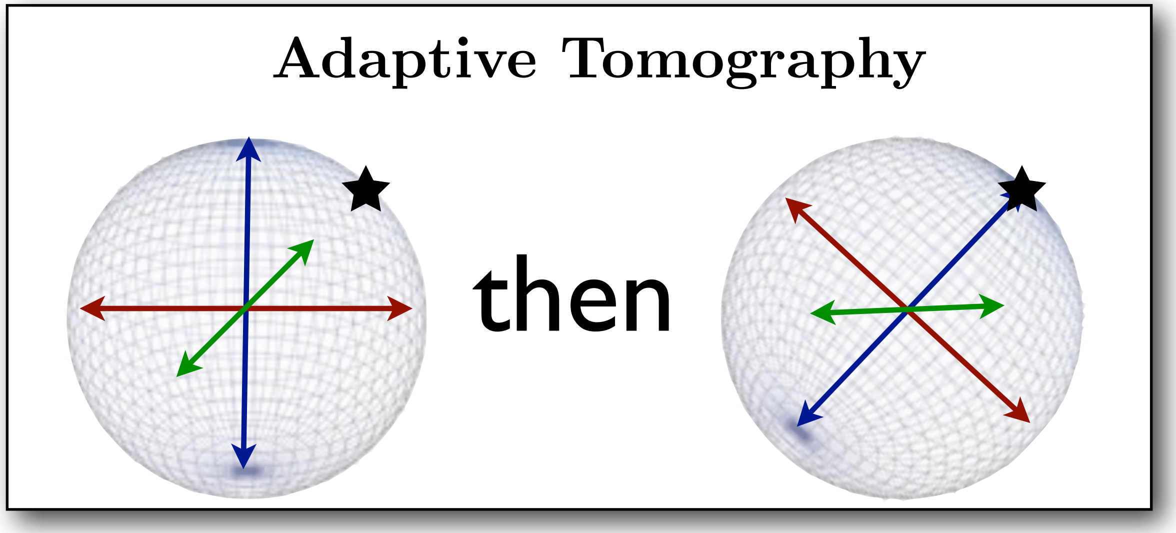




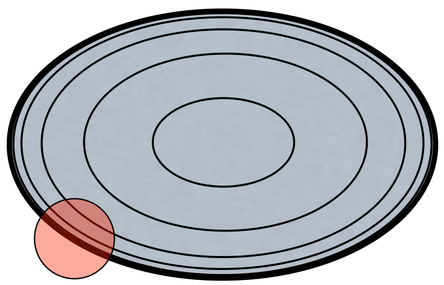
Adaptive State Tomography

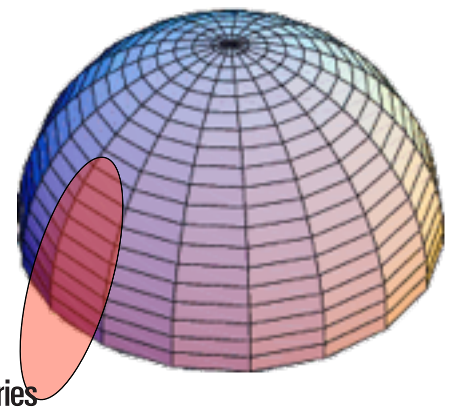
- **Punchline**: *adaptive* state tomography can achieve dramatic improvements in <u>infidelity</u> -- $1/\sqrt{N} \to 1/N$.

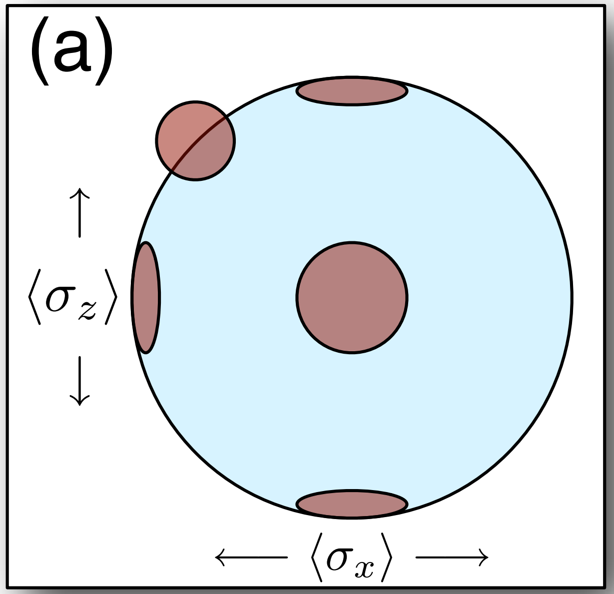
- **How?**
  (1) standard tomography on N/2 copies to get a preliminary estimate $\hat{\rho}_0$.
  (2) Measure eigenbasis of $\hat{\rho}_0$ on remaining N/2 copies.

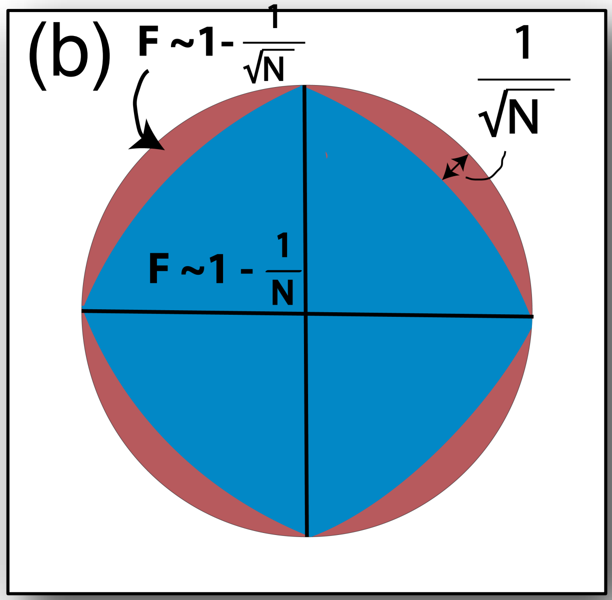


- **Why?**
  (1) Zero probabilities can be estimated very accurately (1/N).
  (2) **Fidelity is not trace norm!**



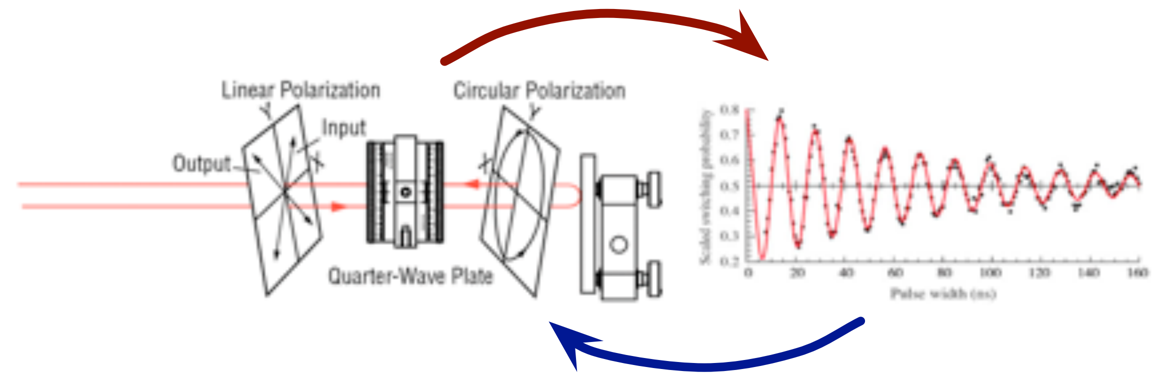






Adaptive Gate-Set Tomography

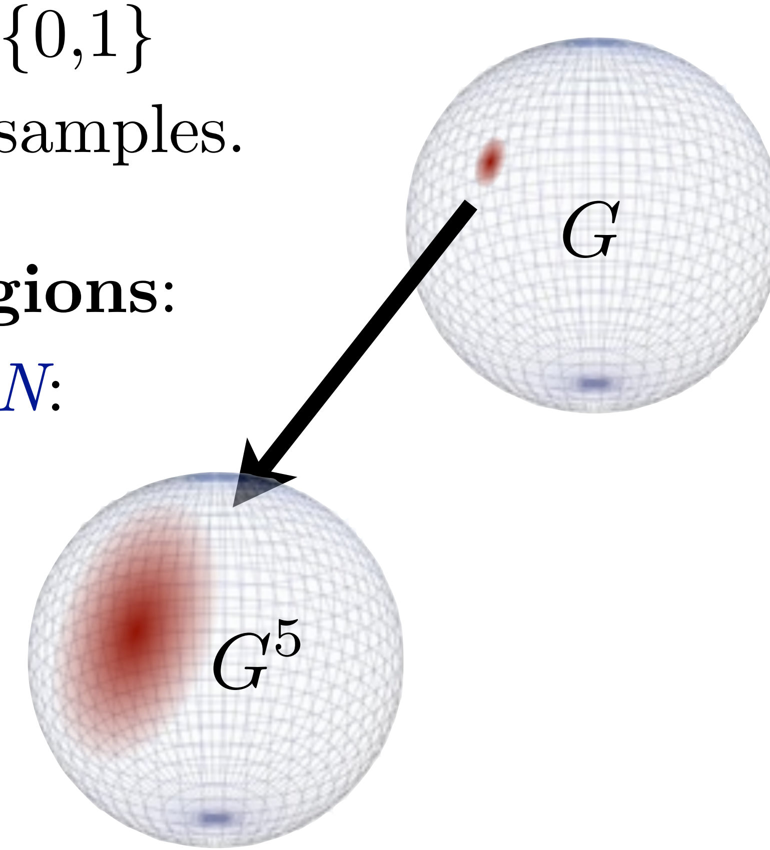
Three reasons for adaptivity:



1. Gate calibration traditionally involves feedback (adjust, measure, rinse, lather, repeat...)

2. **Same reason as for state tomography**: error metrics are fidelity-like, and **zero probabilities are easier to estimate**.
   => prescription: do experiments that yield $Pr \approx \{0,1\}$
   => improvement: quadratic, N samples => $N^{1/2}$ samples.

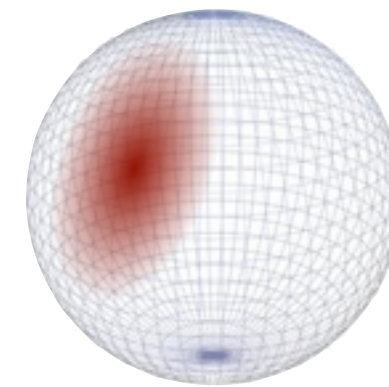
3. **Studying G^n for large n "blows up" small regions**:
   => improvement: **exponential**, N expts. => $\log N$:
   => requires adaptivity to choose the experiments (sequences) that will efficiently resolve the current uncertainty region at each stage.



U.S. DEPARTMENT OF ENERGY Sandia National Laboratories

How to choose gate strings

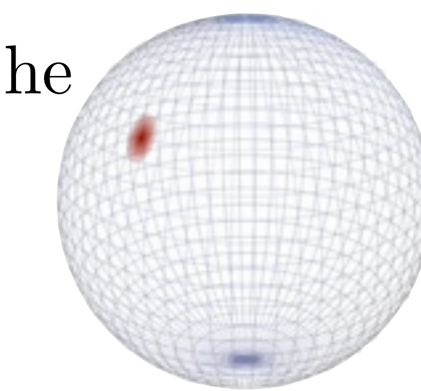
1. Perform a *quorum* of short-sequence experiments (e.g. all sequences of up to L=3..6 gates).



2. Get preliminary confidence region from likelihood function.

3. Generate candidate sequences of ~L^2 gates (how? something clever -- genetic algorithm?).

4. Select sequences that are *expected* to yield data that *sharpen* the likelihood function *for all* plausible states (in CR).



5. Iterate.

6. ???

7. Profit.