SAND2016-1803C

Episode IV:
Calibration-free characterization

Kenneth Rudinger, Erik Nielsen

Sandia National Laboratories

APS March Meeting 2016

This work was supported by the Laboratory Directed Research and Development program at Sandia National Laboratories. Sandia National Laboratories is a multiprogram 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

Introduction

What do we want? **Fault Tolerance!**

- (worst-case) error metrics
- more than process tomography…
- more than benchmarking…

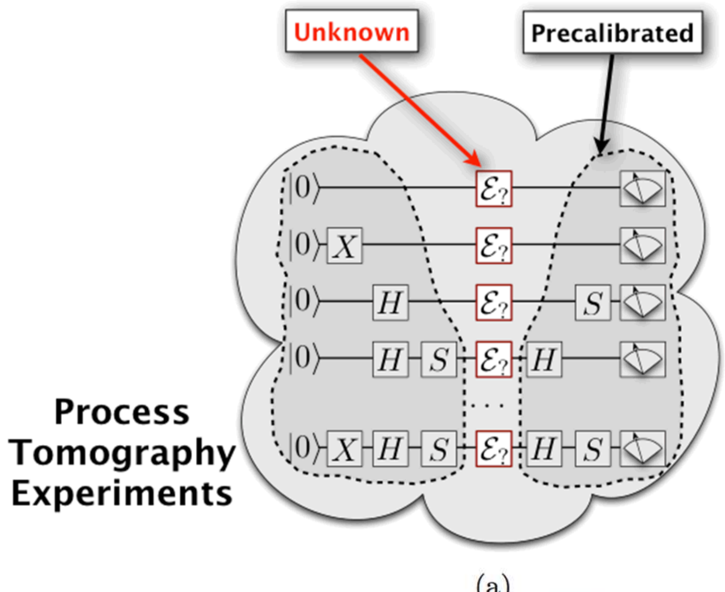
Then what do we need?

And how will we get it? **Calibration-free characterization!**

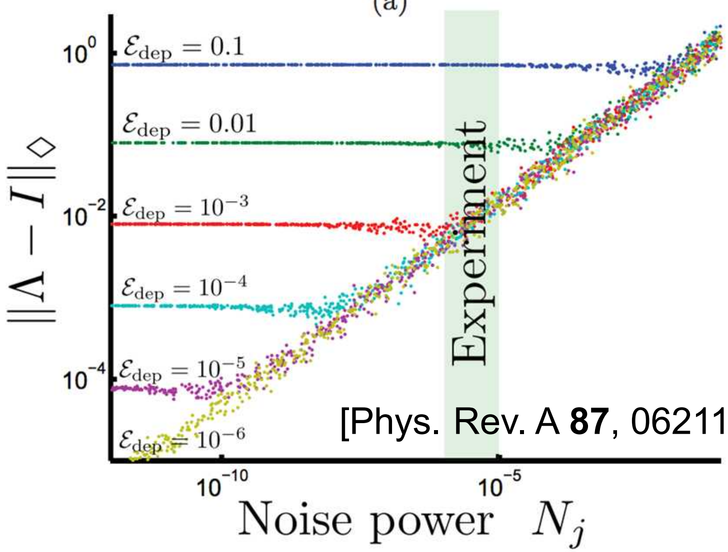
- Overview
  - Process tomography is insufficient (bad assumption that states are perfect, i.e. "calibrated")
  - We need "calibration-free" ways of extracting tomographic information.
  - In this episode, we'll review & contrast several modern methods:
    - Randomized Benchmarking Tomography
    - Robust Phase Estimation
    - **Gate Set Tomography**

Inadequacy of Process Tomography

- When the states and measurements used to interrogate the system are generated by gates that have systematic error (practically unavoidable), **process tomography is very inaccurate**.



- **(Right)** Error in process tomography estimate of identity gate when:
  - Depolarized SPAM (E_{dep})
  - Gaussian noise on measurement outcomes (N_j)



[Phys. Rev. A **87**, 062119]

Bibliography

Sandia National Laboratories

2009: Initial recognition of gap btwn tomography theory & exp.; importance of error bars

Relative tomography of an unknown quantum state, Mogilevtsev, Řeháček, and Hradil, Phys. Rev. A **79**, 020101(R) (Feb 2009)

Quantum tomographic reconstruction with error bars: a Kalman filter approach, Audenaert and Scheel, New J. Phys. 11 023028 (Feb 2009)

2012-2013: Initial "self-calibration" ideas, initial gate-set-tomography (GST) work

Self-calibration for self-consistent tomography, Mogilevtsev, Řeháček, and Hradil, New J. Phys. 14 095001 (Sept 2012)

Self-calibrating tomography for angular Schmidt modes in spontaneous parametric down-conversion, Straupe et al., Phys. Rev. A **87**, 042109 (April 2013)

Self-consistent quantum process tomography, Merkel et al., Phys. Rev. A **87**, 062119 (June 2013)

Self-calibrating tomography for multidimensional systems, Quesada, Brańczyk, and James, Phys. Rev. A **87**, 062118 (June 2013)

Tomography of a spin qubit in a double quantum dot, Takahashi, Bartlett, and Doherty, Phys. Rev. A **88**, 022120 (August 2013)

Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit, Blume-Kohout et al., arXiv:1310.4492 (2013)

2014: Several more new types of tomography; initial randomized-benchmarking-tomography (RBT) work

Robust Extraction of Tomographic Information via Randomized Benchmarking, Kimmel et al., Phys. Rev. X **4**, 011050 (March 2014)

Self-consistent tomography of the state-measurement Gram matrix, Stark, Phys. Rev. A **89**, 052109 (May 2014)

Quantum model averaging, Ferrie, New J. Phys. 16, 093035 (Sept 2014)

Randomized benchmarking with confidence, Wallman and Flammia, New J. Phys. 16 103032 (Oct 2014)

2015: More process tomography scrutiny; RBT in action, initial robust-phase-estimation (RPE) work

Systematic Errors in Current Quantum State Tomography Tools, Schwemmer et al., Phys. Rev. Lett. **114**, 080403 (Feb 2015)

Demonstration of robust quantum gate tomography via randomized benchmarking, Johnson et al., New J. Phys. 17 113019 (Nov 2015)

Robust calibration of a universal single-qubit gate set via robust phase estimation, Kimmel, Low, and Yoder, Phys. Rev. A **92**, 062315 (Dec 2015)

Desirable qualities

of a characterization method

- **Tomographic completeness**: can a method characterize the entire quantum process?
- **Efficiency**: how fast does the error decrease with each additional measurement (or other relevant unit, e.g. "hour in the lab")
  - Stochastic scaling- $1/\sqrt{N_{\text{resources}}}$ should be minimally expected.
  - Heisenberg scaling- $1/(\text{max. sequence length})$ is the best one can expect.
- **Error bars**: Does a method put error bars on its estimate?
- **Simplicity**: Is a method straightforward to implement?

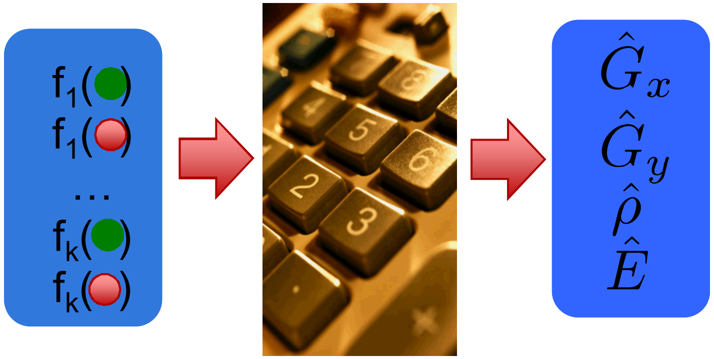
Comparison of mainstream methods

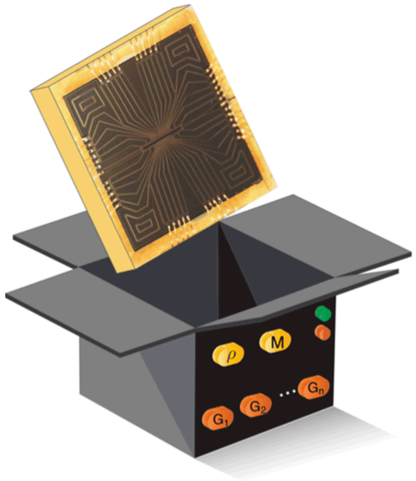
Method	Completeness	Efficiency	Error bars	Simplicity
Randomized Benchmarking Tomography (RBT)	**Only unital part of processes.**	**Non-Heisenberg**	**yes, via bootstrapping**	**Simple analysis** **(some work to construct sequences)**
Robust Phase Estimation (RPE)	**Only several specific parameters (e.g. the “phase”).**	**Heisenberg**	**yes, via bootstrapping**	**Simple analysis (trig.$^{-1}$ funcs)**
Gate Set Tomography (GST)	**Complete.**	**Heisenberg**	**computable**	**Fairly complex analysis.**

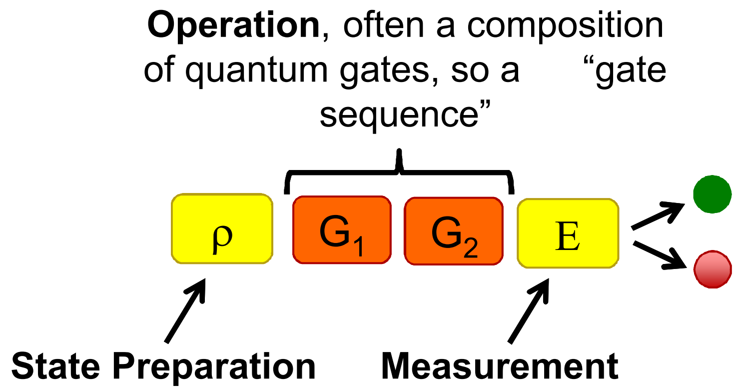
We focus on this method

Gate sequences, the common language

- Each characterization method utilizes experimental data.
- Each experiment (in this talk) consists of:
  - State preparation
  - (possibly) some operation on the state
  - Measurement
- Represented pictorially by:







Randomized Benchmarking Tomography

- **Main idea**: Use randomized benchmarking (RB) to estimate the *unital* part of **any** process matrix.
- **Reference**: Kimmel et al. Phys. Rev. X 4, 011050 (2014)
- Pros:
  - Estimates almost the entire process matrix (the unital part)
  - Error bars can be obtained by repeating by bootstrapping.
  - Analysis procedure is simple; very similar to doing RB.
- Cons:
  - Accuracy does not scale well with the number & length of gate sequences, and cannot even make use of long sequences b/c of fast decay.
  - Error bars are somewhat ad-hoc and have subtleties.

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \tau_x & \lambda_{xx} & \lambda_{xy} & \lambda_{xz} \\ \tau_y & \lambda_{yx} & \lambda_{yy} & \lambda_{yz} \\ \tau_z & \lambda_{zx} & \lambda_{zy} & \lambda_{zz} \end{pmatrix}$$

unital part: G'

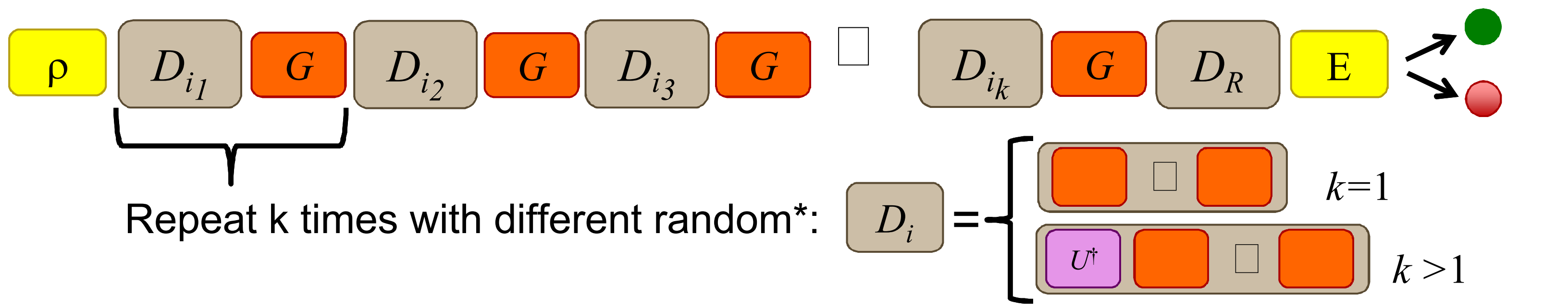
- Prep for what's next:
  - recall RB yields <F(U,G)> for unknown G and unitary U:

$$\overline{F}(G,U) = \int d\mu(\psi) \langle \psi | (U^+ \circ G(|\psi\rangle\langle\psi|)) | \psi \rangle$$

Randomized Benchmarking Tomography

How to do RBT:

- Find $G \circ N$ by fitting data from the gate sequences (for different k):

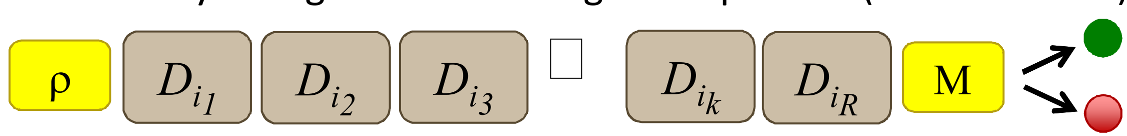


*assume:
1) $D_i = N \circ D_i^{perfect}$
2) D_R s.t. $\rho_i = \rho_f$ when $G=U$

Fit data(k) to $A_0 p^k + B_0$ → $p = \dfrac{d\,\bar{F}(G \circ N, U) - 1}{d-1}$ → $\boldsymbol{G \circ N}$

{U} spans cliffords

- Find N by fitting data from the gate sequences (for different k):



Fit data(k) to $A_0 p^k + B_0$ → $p = \dfrac{d\,\bar{F}(N, U) - 1}{d-1}$ → $\boldsymbol{N}$ → $\boldsymbol{G'}$

Use the result:

$G' = (G \circ N)' \circ (N')^{-1}$

(X' =unital part of X)

Randomized Benchmarking Tomography

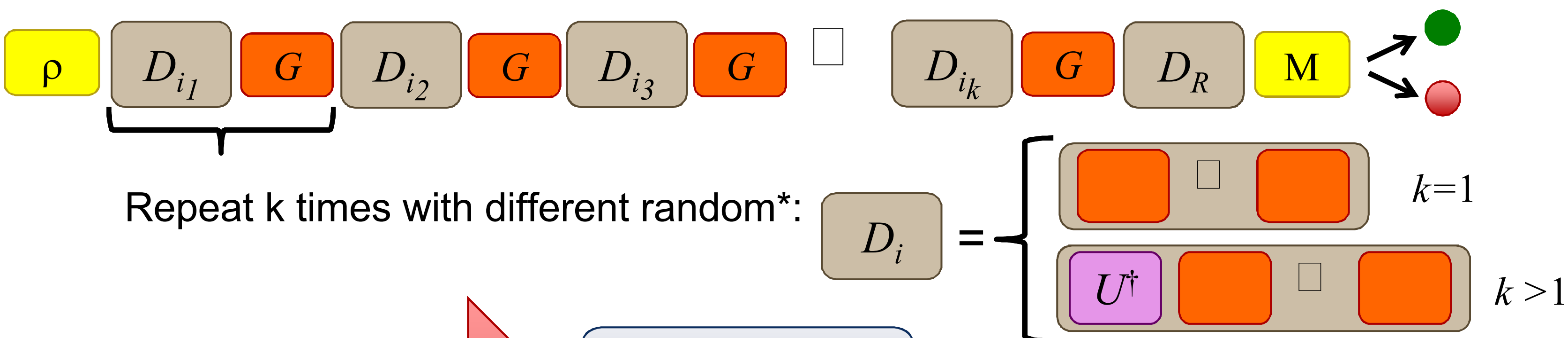
- **Main Idea:** use RB to estimate the unital part of **any** process matrix.
- RB gives access to average Fidelity between unknown process G and unitary U:

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \tau_x & \lambda_{xx} & \lambda_{xy} & \lambda_{xz} \\ \tau_y & \lambda_{yx} & \lambda_{yy} & \lambda_{yz} \\ \tau_z & \lambda_{zx} & \lambda_{zy} & \lambda_{zz} \end{pmatrix}$$

unital part

$$\bar{F}(G,U) = \int d\mu(\psi) \langle \psi | (U^{+} \circ G(|\psi\rangle\langle\psi|)) | \psi \rangle$$

  - Find $F\,(G,U)$ by fitting data from the gate sequences (for different k):



Fit data(k) to $A_0 p^k + B_0$

$$p = \frac{d\,\bar{F}(G \circ N, U) - 1}{d - 1}$$

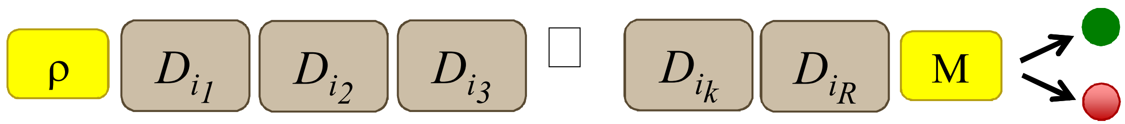
*assume:

- $D_i = N \circ D_i^{perfect}$
- D_R s.t. $\rho_{i=}\rho_f$ when $G=U$

- Knowing $F(G \circ N, U)$ between G and a spanning set of Clifford gates lets **you determine $G \circ N$**

Randomized Benchmarking Tomography

- **(continued…)** Performing the same sequences without G:



gives an estimate for just the noise process $\boldsymbol{N}$, and using the result:

$$G' = (G \circ N)' \circ (N')^{-1}$$

(where X' means the unital part of X)

One obtains an **estimate for G', the *unital* part of G.**

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \tau_x & \lambda_{xx} & \lambda_{xy} & \lambda_{xz} \\ \tau_y & \lambda_{yx} & \lambda_{yy} & \lambda_{yz} \\ \tau_z & \lambda_{zx} & \lambda_{zy} & \lambda_{zz} \end{pmatrix}$$

G’

- **Error bars** can be obtained by repeating by bootstrapping.

- Reference: [PHYS. REV. X 4, 011050 (2014)]

Robust Phase Estimation

- **Main idea**: a characterization protocol for estimating a few (specific) parameters of a set of quantum gates.
- **Reference:** Kimmel, Low, Yoder Phys. Rev. A 92, 062315 (2015);
- Aim is **to estimate a few parameters as efficiently as possible** (contrast with "tomography", which seeks to estimate the *entire* process matrix)
- Pros:
  - Parameters are learned with **optimal efficiency (Heisenberg scaling)**
  - **Non-adaptive** (simple!)
  - Accommodates a additional errors (depol. Noise)
- Cons:
  - Need **bounds** on state preparation
  - Need **bounds** on gate control

Robust Phase Estimation

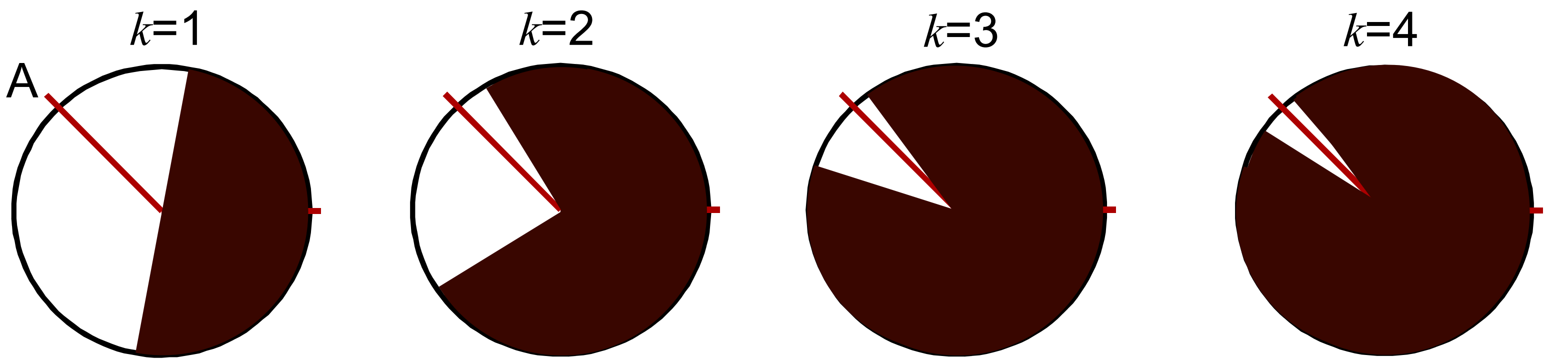
- Main Idea: one can estimate rotation of angle A in [-pi,pi] efficiently when you have access to “coins” with heads-probabilities

$$p_0(A,k) = \frac{1+\cos(kA)}{2} + \delta_0(k),$$
$$p_+(A,k) = \frac{1+\sin(kA)}{2} + \delta_+(k).$$

(k an integer)

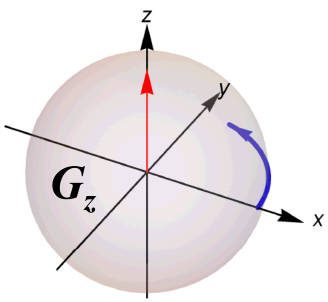
as long as the d errors aren’t too large ($\sup_k\{|\delta_0(k)|,|\delta_+(k)|\} < 1/\sqrt{8}.$)

- Start with k=1 and increment. At each k, can rule our half of remaining angular space.



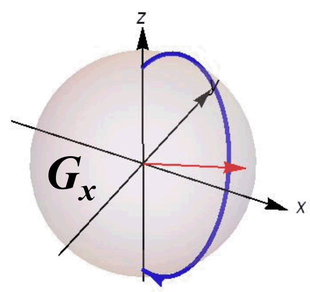
Robust phase estimation

- **Example**; 2 gates (but works for almost any Z,X-like):
  - G_z: rotation about *z-axis by* $\pi/2(1+\alpha)$
  - G_x: rotation about $(cos(\phi)x+sin(\phi)z)$*-axis by* $\pi/2(1+\varepsilon)$



Amplitude of Rotation:

Axis of Rotation:



Amplitude of Rotation:

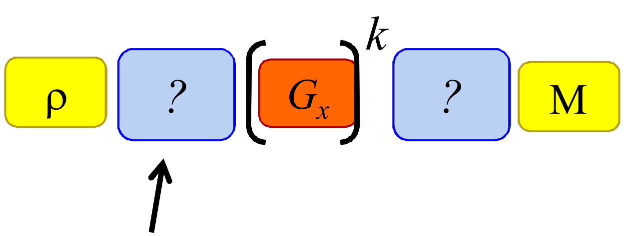
Axis of Rotation:

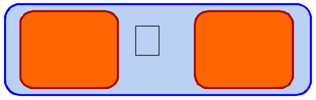
  - G_z case:

$$|\langle +|Z_{\pi/2}(\alpha)^k|+\rangle|^2 = \frac{1+\cos[-k(\pi/2)(1+\alpha)]}{2},$$

$$|\langle +|Z_{\pi/2}(\alpha)^k|\rightarrow\rangle|^2 = \frac{1+\sin[-k(\pi/2)(1+\alpha)]}{2}.$$

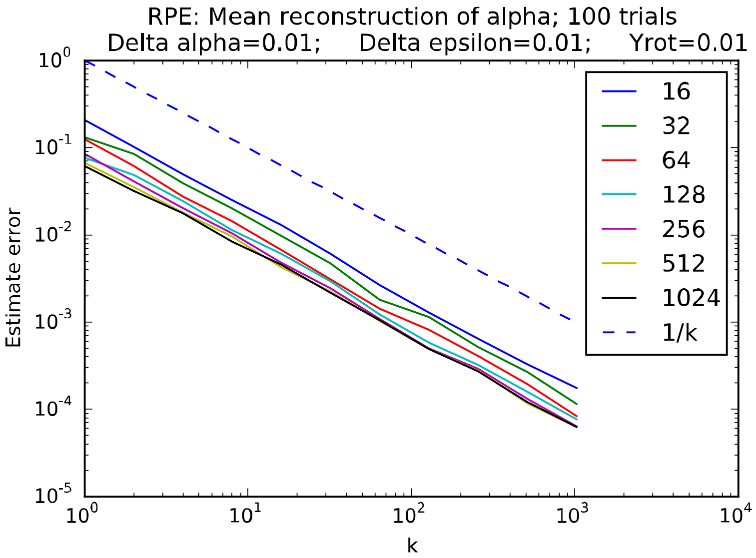
  - Other gates similar…

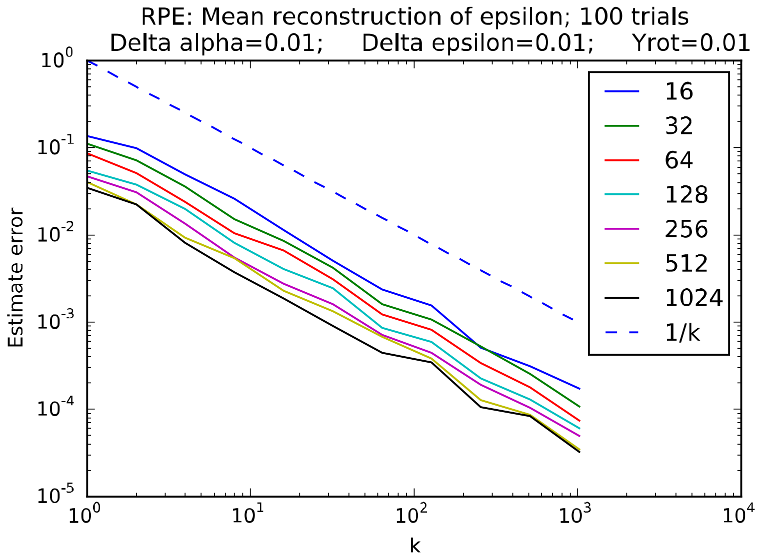


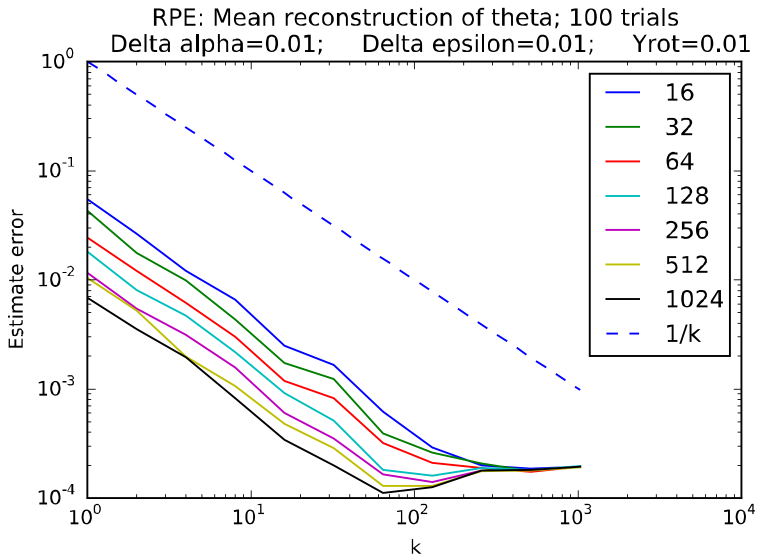


Extra gates may be required in practice for prep/measure.

Robust phase estimation







Numerical simulations demonstrate Heisenberg scaling in accuracy/

Accuracy of $2*10^{-4}$ with only 1056 experiments!

For more details:
E44.06, E44.07
(Tuesday 9:24 AM, 9:26 AM)

Gate Set Tomography

- **Main idea**: Estimates an entire set of gates, along with state preparation and measurement, *very very* accurately – but it takes some work to do it.

- **Originated from ideas in**: Phys. Rev. A **87**, 062119
- **“Modern” GST Reference**: Blume-Kohout et al., arXiv:1310.4492

- Pros:
  - Provides full tomographic estimates
  - Uses long-sequence data efficiently, allowing very high accuracy in estimates.
  - Provides rigorous confidence-interval error bars on “raw” gate estimates as well as derived quantities (e.g. fidelity, diamond-norm,…)
  - Clearly detects model violation (when data is “bad”, i.e., it doesn’t fit *any* qubit model)

- Cons:
  - Analysis of data is complicated; results in lots of estimate data to interpret.

Review of Lioville representation

- We can vectorize density matrices, and can do so in the basis of Pauli matrices:

$$\rho = \begin{pmatrix} a & b \\ \bar{b} & a \end{pmatrix} = \rho_I \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \rho_x \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \rho_y \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} + \rho_z \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \Rightarrow \begin{pmatrix} \rho_I \\ \rho_x \\ \rho_y \\ \rho_z \end{pmatrix}$$

- Operators M on density matrices ("super-operators"; "maps") are matrices, which act on density matrix vectors by matrix multiplication:

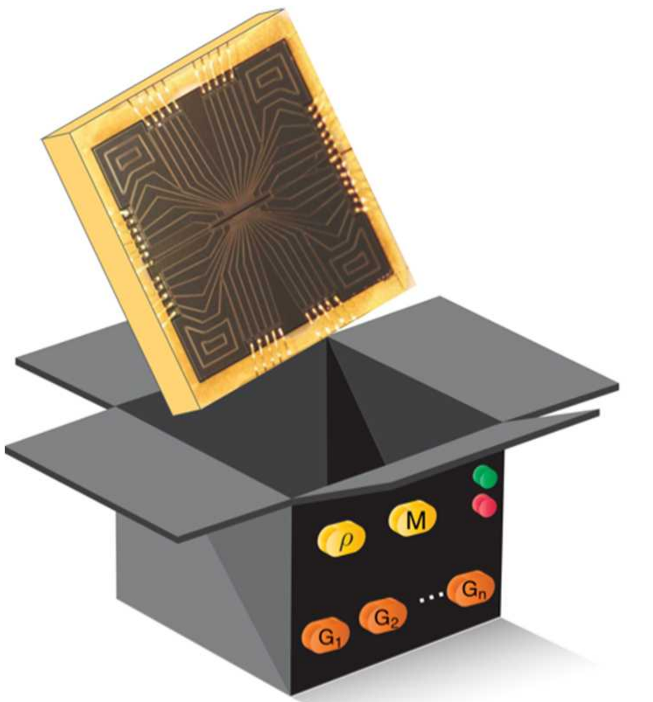
$$M : \rho \mapsto \rho' \qquad M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \tau_x & \lambda_{xx} & \lambda_{xy} & \lambda_{xz} \\ \tau_y & \lambda_{yx} & \lambda_{yy} & \lambda_{yz} \\ \tau_z & \lambda_{zx} & \lambda_{zy} & \lambda_{zz} \end{pmatrix}$$

- Composition of maps is just matrix multiplication:

Gate Set Tomography

Sandia National Laboratories

Experiment	Model ("gate set")



"gate sequence"

Press: (x 100) ρ $\underbrace{G_1 \; G_2}$ M → ●, ●

$= n$ ●, $N\text{-}n$ ●

$\boxed{p = n/N \text{ estimates } p_0}$

ρ (prep) $= \begin{pmatrix} \rho_I \\ \rho_x \\ \rho_y \\ \rho_z \end{pmatrix}$

E● (outcome) $= \begin{pmatrix} E_I & E_x & E_y & E_z \end{pmatrix}$

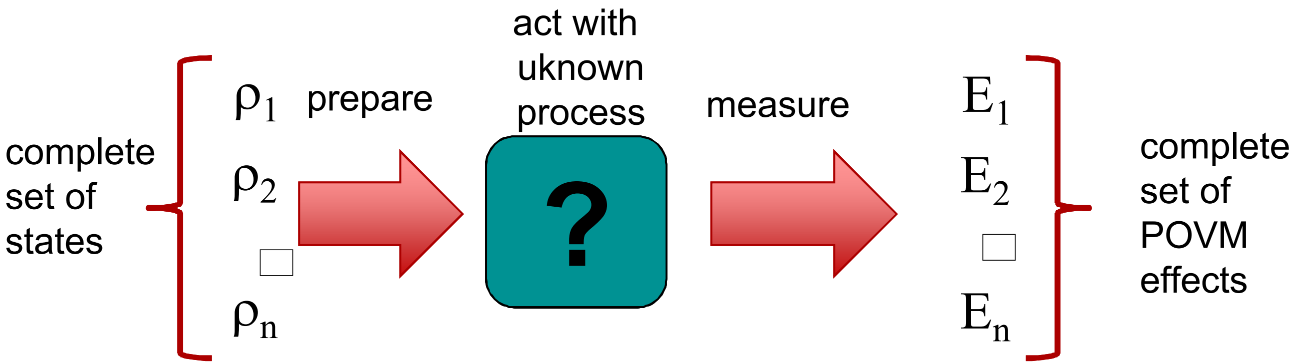
G_i (gate) $= \begin{pmatrix} 1 & 0 & 0 & 0 \\ \tau_x & \lambda_{xx} & \lambda_{xy} & \lambda_{xz} \\ \tau_y & \lambda_{yx} & \lambda_{yy} & \lambda_{yz} \\ \tau_z & \lambda_{zx} & \lambda_{zy} & \lambda_{zz} \end{pmatrix}$

Matrix mult. (note order reversal):

$\langle\langle E_{\bullet} | G_2 G_1 | \rho \rangle\rangle = p_0$

Linear GST: robust process tomography

Process Tomography:



Linear Gate Set Tomography:

Choose "fiducial" gate sequences $\boldsymbol{F_i}$:

$$|\rho_i\rangle\rangle = F_i|\rho\rangle\rangle \qquad \rho_i = \rho\,\overbrace{G_{i_1} \cdots G_{i_n}}^{F_i}$$

$$\langle\langle E_i| = \langle\langle E|F_i \qquad E_i = \underbrace{G_{i_1} \cdots G_{i_n}}_{F_i}\,E$$

And perform experiments:

$$\rho \quad F_i \quad G_k \quad F_j \quad E$$

Estimates: $\langle\langle E|F_j G_k F_i|\rho\rangle\rangle$

$$\rho \quad F_i \quad F_j \quad E \qquad \& \qquad \rho \quad F_i \quad E$$

$$\langle\langle E|F_j F_i|\rho\rangle\rangle \qquad \& \qquad \langle\langle E|F_j F_i|\rho\rangle\rangle$$

Linear GST: (cont.)

Sandia National Laboratories

Define "theory" matrices:

$$A = \begin{bmatrix} \langle\langle E_1 | \\ \langle\langle E_2 | \\ \vdots \\ \langle\langle E_n | \end{bmatrix}$$

$$B = \begin{bmatrix} |\rho_1\rangle\rangle & |\rho_2\rangle\rangle & \cdots & |\rho_n\rangle\rangle \end{bmatrix}$$

Define "data" matrices (w/tilde)

$$\tilde{G}_k = \left[f_{ji}^{k} \right]_{ji}$$

$$\tilde{I} = \left[I_{ji} \right]_{ji}$$

$$\tilde{S} = \left[S_i \right]_i$$

Experiments to estimate G_k :

ρ | F_i | G_k | F_j | E

Estimate of $\langle\langle E | F_j G_k F_i | \rho \rangle\rangle = f_{ji}^{k}$

ρ | F_i | F_j | E

Estimate of $\langle\langle E | F_j F_i | \rho \rangle\rangle = I_{ji}$

ρ | F_i | E

Estimate of $\langle\langle E | F_i | \rho \rangle\rangle = S_i$

Then:

is estimated by…

$AB \longrightarrow \tilde{I}$

$B^{-1}A^{-1} \longrightarrow \tilde{I}^{-1}$

$AG_kB \longrightarrow \tilde{G}_k$

$A|\rho\rangle\rangle \longrightarrow \tilde{S}$

$B|\rho\rangle\rangle \longrightarrow \tilde{I}^{-1}\,\tilde{S}$

$\langle\langle E|B \longrightarrow \tilde{S}^{T}$

$B^{-1}G_kB \longrightarrow \tilde{I}^{-1}\tilde{G}_k$

Estimates for $\langle\langle E|B$, $B^{-1}G_kB$, and $B|\rho\rangle\rangle$ constitute an **estimate of the entire gate set!**

Gauge Freedom

Estimates for $\langle\langle E|B$, $B^{-1}G_kB$, and $B|\rho\rangle\rangle$ constitute an **estimate of the entire gate set!**

What??

(we don't know B !)

- Altering a gate set by mapping

$$\langle\langle E| \longrightarrow \langle\langle E|B$$

$$|\rho\rangle\rangle \longrightarrow B|\rho\rangle\rangle$$

$$G_k \longrightarrow B^{-1}G_kB$$

"Gauge Transformation"

Does not, ***for any invertible matrix B***, affect any of the physical probabilities predicted by the gate set. (Because all probabilities are computed by:

$$\langle\langle E|G_s \cdots G_k G_j G_i|\rho\rangle\rangle = p_0$$

- We call this degree of freedom a "**gauge**" freedom. It defines equivalence classes of gate sets (where all the gate set representations in a class correspond to the same physical gate set) .
- Gates are **relational**. You tomograph a gate set, not just a gate…

GST Fiducial Selection

How do we select the F_i fiducial sequences (for use with LGST)?

- Require that $\{\, |\rho_i\rangle\rangle = F_i|\rho\rangle\rangle \}$ and $\{ \langle\langle E_i| = \langle\langle E|F_i \}$ each span the Hilbert-Schmidt space of d x d density matrices.

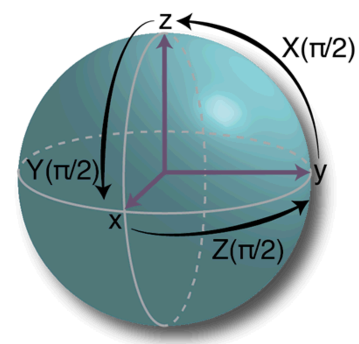
- In general, almost any set of d^2 matrices will be linearly independent (and so span the space) – but we want them to be *as linearly dependent as possible*. This is quantified by the Gram matrix, defined by:

$$\mathrm{Gram}_{ij} = \langle\langle E_i | \rho_j \rangle\rangle$$

(If either set contains linear dependences, the Gram matrix will be rank deficient.)

- For uniform informational completeness, and smallest singular value of the Gram matrix should be as large as possible.
    - Holds when sets for 2-designs
    - One 2-design of 4-elements for single-qubits (SIC POVM)
    - More convenient **1-qubit 2-design with 6-elements**:

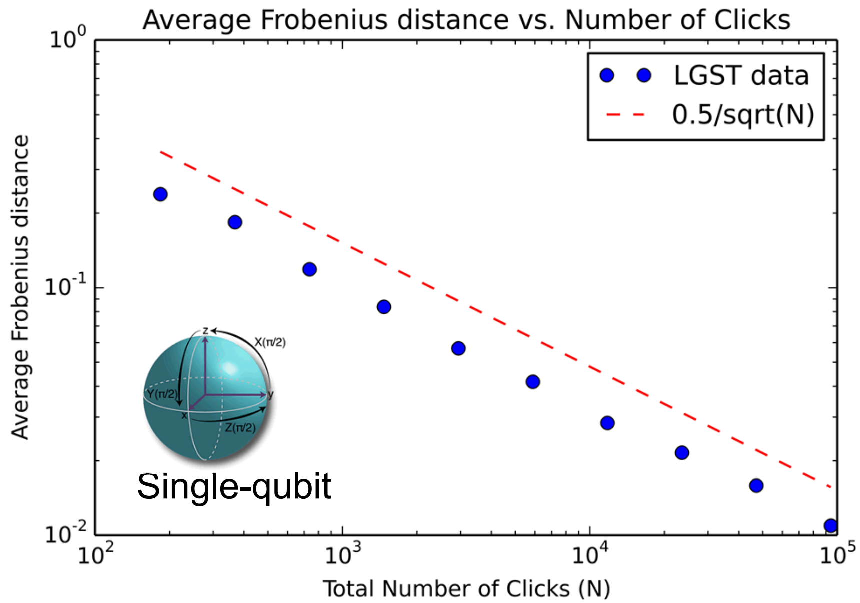
$$\emptyset,\ G_x,\ G_y,\ G_xG_x,\ G_xG_xG_x,\ G_yG_yG_y$$

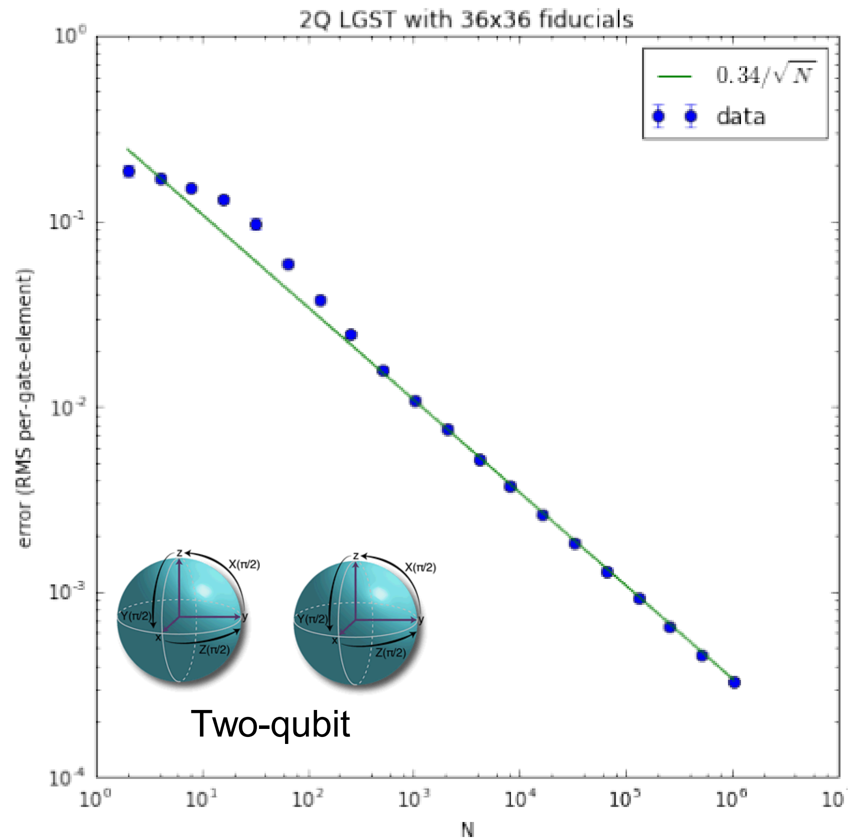


LGST Simulations: $1/\sqrt{N}$ Error scaling

- **Plots** show error between data-generating gate set and estimated gate set, as a function of N (the number of samples).
- Expect $1/\sqrt{N}$ behavior, as uncertainty in the mean of a distribution decreases as $1/\sqrt{N}$.

GST performance on **simulated data** (data generated from a *hidden* gateset)



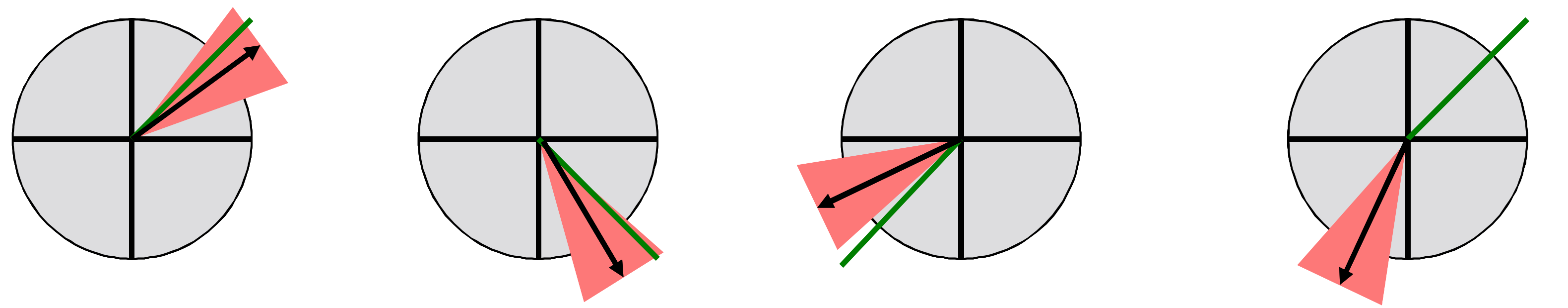


Coherent Error Amplification

- Error amplification – amplify **coherent** errors in gates by repeating them.
- Example: A $\pi/8$-rotation gate which over-rotates by θ:

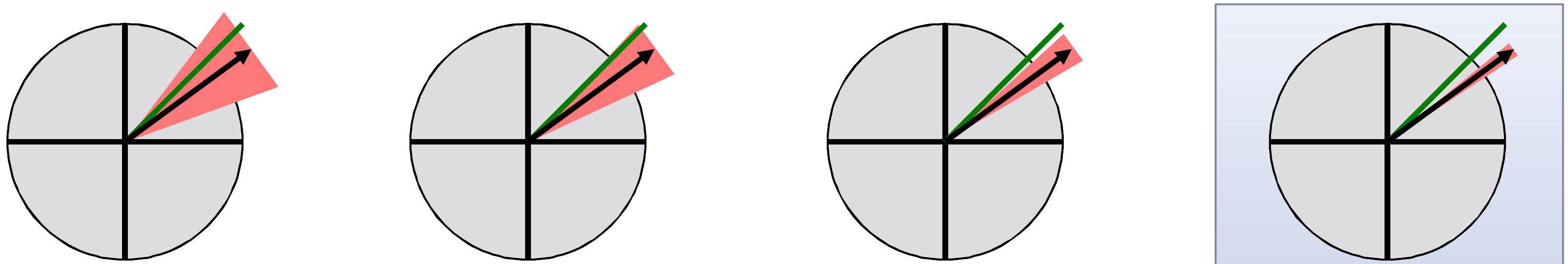
Repeats:	1	2	3	R

Repeated gate:



Estimated Angle:	$(\pi/8+\theta)+2\pi l$ $\pm\alpha/\sqrt{N}$	$(\pi/4+2\theta)+2\pi l$ $\pm\alpha/\sqrt{N}$	$(3\pi/8+3\theta)+2\pi l$ $\pm\alpha/\sqrt{N}$	$(R\pi/8+\mathrm{R}\theta)+2\pi l$ $\pm\alpha/\sqrt{N}$

Single gate:

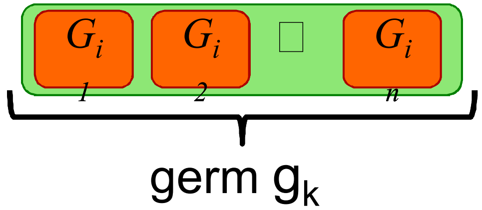


Estimated Angle:	$(\pi/8+\theta)+2\pi l$ $\pm\alpha/\sqrt{N}$	$(\pi/8+\theta)+\pi l$ $\pm\alpha/\left(2\sqrt{N}\right)$	$(\pi/8+\theta)+2\pi l/3$ $\pm\alpha/\left(3\sqrt{N}\right)$	$(\pi/8+\theta)+2\pi l/R$ $\pm\alpha/\left(\mathrm{R}\sqrt{N}\right)$

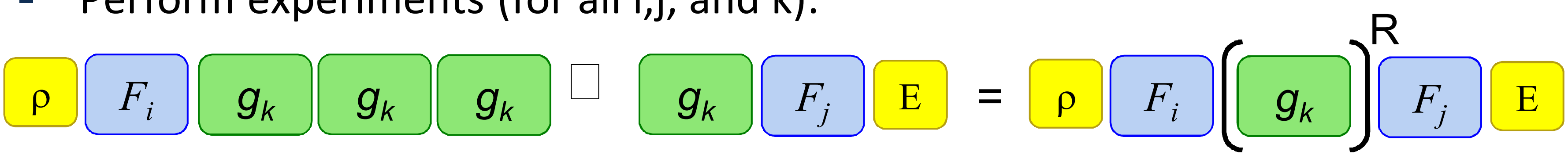
Uncertainty ~ 1/R !

Long-sequence GST

- Use error amplification to amplify **all** possible gate errors by repeating not just the gates themselves but a set of short gate sequences $\{g_k\}$ called “germs”.



- Perform experiments (for all i,j, and k):



- Results in estimates (“frequencies”) for $\langle\langle E|\mathrm{F}_j(g_k)^R\,\mathrm{F}_i|\rho\rangle\rangle$ which we compare with the probabilities predicted by the model using the log-likelihood or χ^2 statistic:

$$\log L = \sum_i N f_i \log(p_i) \qquad \chi^2 = \sum_i N \frac{(p_i - f_i)^2}{p_i}$$

  - N = #samples, f = frequency, p = probability, and i ranges over gate sequences *and* outcomes.
  - Maximizing the likelihood or minimizing χ^2 gives an estimate for the gate set.

Interlude: Germ Selection

Sandia National Laboratories

How does one select a “complete” set of germs?

- **Jacobian** of the germs to power L w.r.t. gate set parameters, should collectively have n singular values which grow linearly with L (indicating they amplify the error given by the corresponding right singular vector), where n is the number of gauge-invariant gateset parameters.
- Define:

$$\nabla^{(L)}_{g_k} \equiv \frac{1}{L}\frac{\partial\left[g_k^L\right]}{\partial \vec{G}}\bigg|_{\vec{G}=\vec{G}_{\text{target}}}$$

we care about infinite-L limit of:
(does it have right singular rank = n?)

$$J = \begin{pmatrix} \nabla^{(L)}_{g_1} \\ \vdots \\ \nabla^{(L)}_{g_n} \end{pmatrix}$$

Gateset parameters (process mx elements)

Elements of g_k product

$$\nabla^{(L)}_{g_1}$$

“**herald vector**”: Did any germ amplify the trial error?

$$= \begin{bmatrix} \nabla^{(L)}_{g_1} \\ \nabla^{(L)}_{g_2} \\ \nabla^{(L)}_{g_3} \end{bmatrix}$$

Trial gateset error

- Insight: infinite-L limit = twirl = projection onto commutant :

$$\nabla^{(L)}_g = \frac{1}{L}\sum_{n=0}^{L-1}\sigma(g)^n\frac{\partial\sigma(g)}{\partial\vec{G}}\sigma(g)^{L-1-n}$$
$$= \left[\frac{1}{L}\sum_{n=0}^{L-1}\sigma(g)^n\nabla^{(1)}_g(\sigma(g)^\dagger)^n\right]\sigma(g)^{-(L-1)}$$

$$\lim_{L\to\infty}\nabla^{(L)}_g = \underbrace{\Pi_{\sigma(g)}\left[\nabla^{(1)}_g\right]}_{\text{projection onto commutant}}$$

Interlude: Likelihood and χ^2 statistics

- The likelihood function: $L = \Pr(\text{data}|\text{model}) = \Pr(\text{data}|\text{gateset})$
  - Often sharply peaked so use logarithm, the "log-likelihood"
  - The log-likelihood approximated to 2nd order using χ^2

- **Example**: N coin flips, f percent come up heads. If our model is a coin with probability p of heads, then

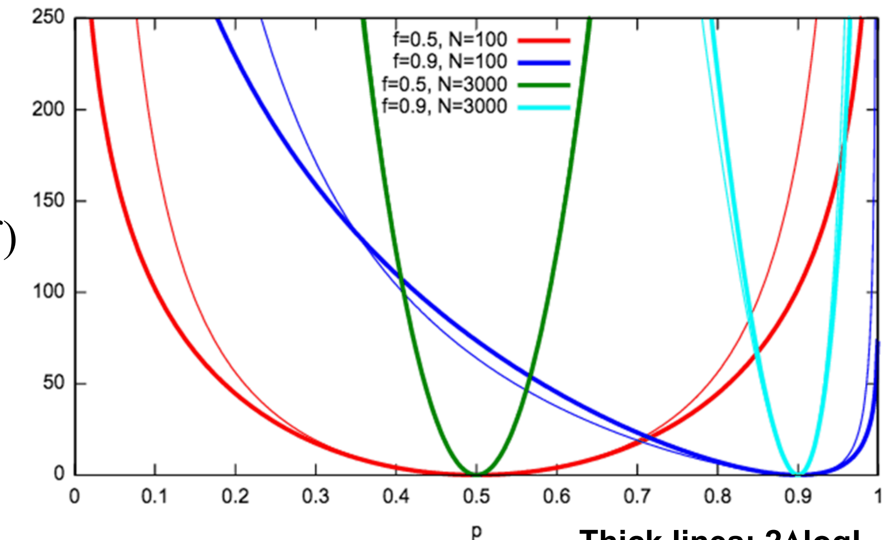
$$L = p^{Nf}(1-p)^{N(1-f)}$$

$$\log L = Nf\log(p) + N(1-f)\log(1-p)$$

$$\max_p(\log L) = Nf\log(f) + N(1-f)\log(1-f)$$

$$\chi^2 = N\frac{(p-f)^2}{p(1-p)}$$

Can check: $2(\max_p(\log L) - \log L) \cong \chi^2$



Thick lines: 2ΔlogL
Thin lines: χ^2

- In GST, we just have lots of coin flips (or dice rolls):

$$\log L = \sum_i Nf_i\log(p_i) \qquad \chi^2 = \sum_i N\frac{(p_i - f_i)^2}{p_i}$$

Long-sequence GST: iterations

- To perform “long sequence GST” we do the following:
  - **Run LGST** to get an initial gate set estimate, $\mathcal{G}_0$
  - **Iteratively** maximize the log-likelihood to obtain sequentially better estimates:

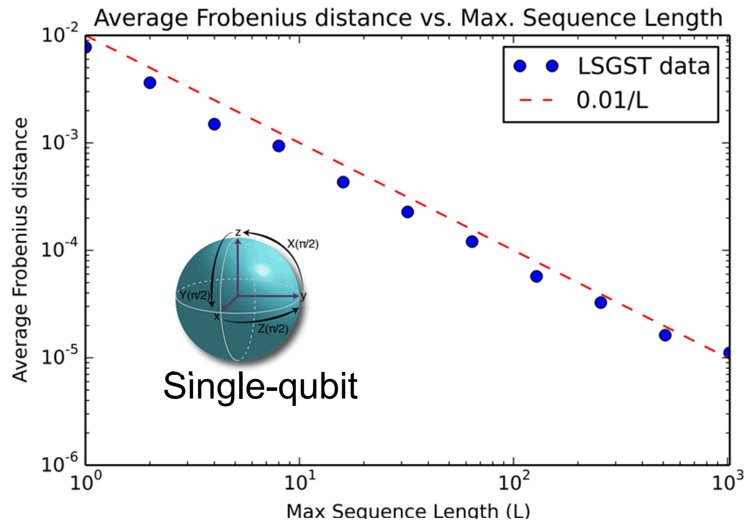
Iteration	Gate Sequences used in log L	Starting gate set	Max. Likelihood Estimate
0	$\rho \; F_i \; g_k \; F_j \; E$	$\mathcal{G}_0$	$\mathcal{G}_1$
1	$\rho \; F_i \; g_k \; g_k \; F_j \; E$	$\mathcal{G}_1$	$\mathcal{G}_2$
2	$\rho \; F_i \; g_k \; g_k \; g_k \; g_k \; F_j \; E$	$\mathcal{G}_2$	$\mathcal{G}_3$
r	$\rho \; F_i \; \left(g_k \right)^{2^r} F_j \; E$	$\mathcal{G}_{(r-1)}$	$\mathcal{G}_r$

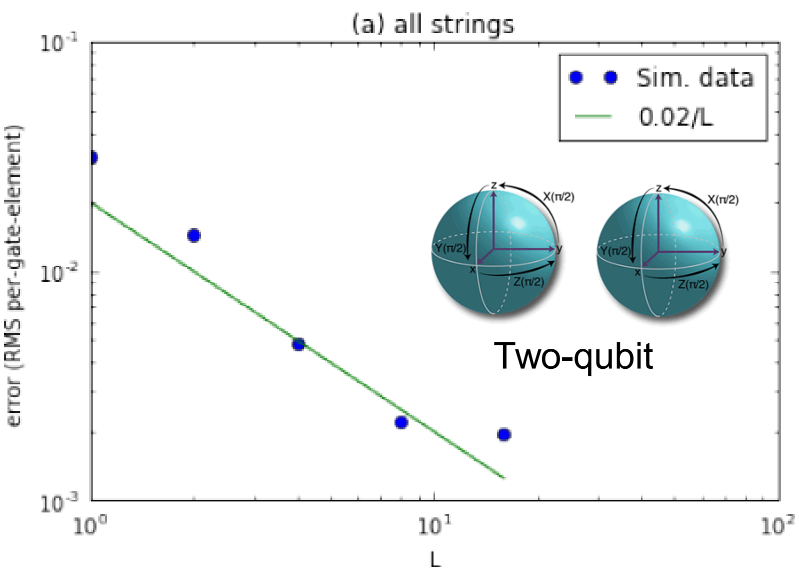
Final GST Estimate

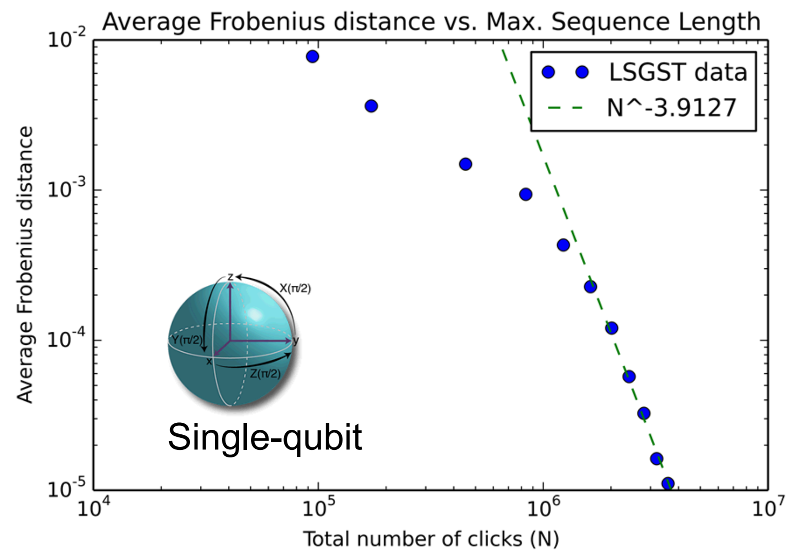
- Why iteratively? To avoid wrong “branch” of solution
- *Technical Point: actually use the χ^2 instead of log L for all but the final iter.

LS-GST Simulations: $1/L$ Error Scaling

- Long-sequence GST on simulated data shows desired $1/L$ scaling, where L=length (or exponent) of germ power.
- Even better scaling ($N^{-3.9}$?) with the total number of experiments.

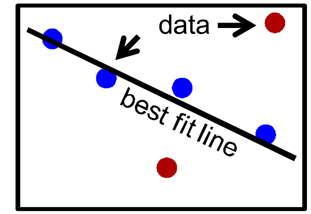




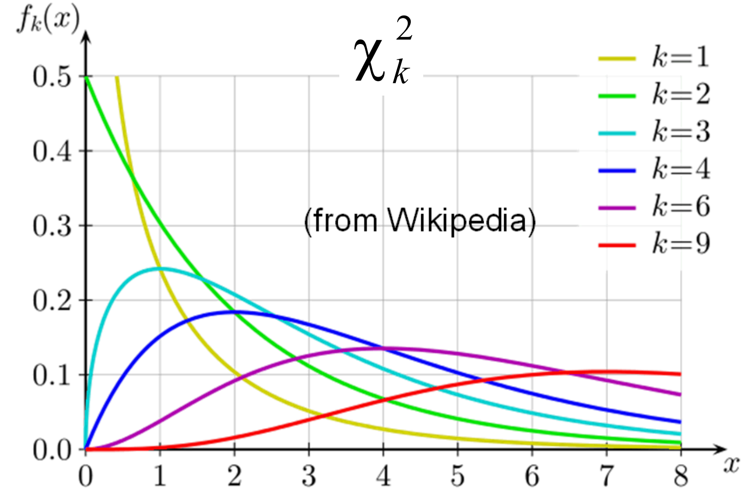


Likelihood ratio test and Wilks Theorem

- When we maximize the log-likelihood, it would be nice to know what values indicate the model is fitting "well".



- **Wilk's theorem** states: $2(\log L_1 - \log L_2)$ will be asymptotically (as $N \to \infty$) χ^2_k - distributed where k equals the difference in dimensionality of the models used to compute L_1 and L_2 when the models are valid.



- In our case we use a "**maximal model**" in which each data point is fit exactly to compute L_1. This model has the ma nd has $n_{\max} \equiv n_{\text{gatestrings}}(n_{\text{outcomes}} - 1)$ parameters

  - Thus, we expect the quantity (readily computable by GST):

$$2\left(\max_p(\log L) - \log L\right)$$

will be χ^2 – distributed with degrees of freedom equal to the difference between $n_{\max}$ and the number of (gauge invariant) gate set parameters.

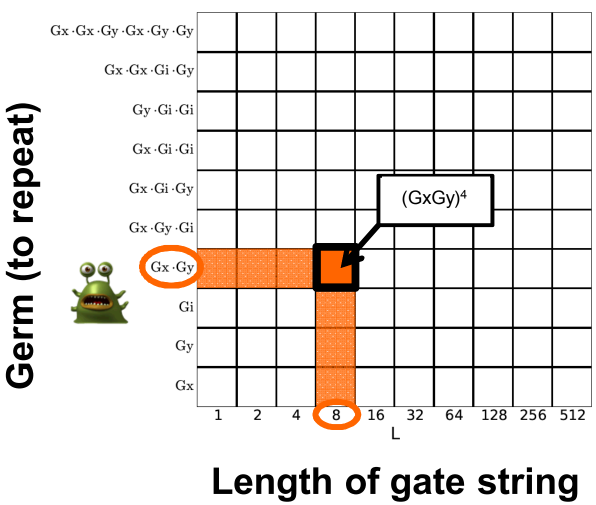
  - We expect the log L contribution of **single sequences** to be χ^2_1-distributed.
  - The extent to which this is not true indicates **model violation**, which, in this way, is easily detected.

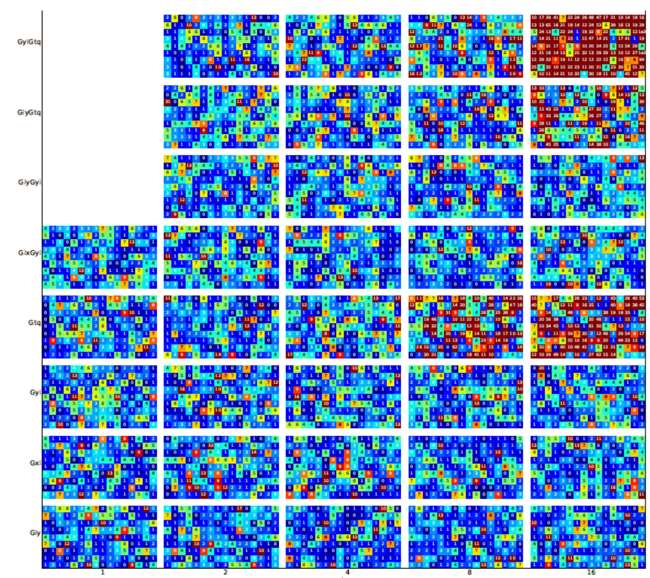
Detecting model violation in GST

- Example summary goodness-of-fit table from a GST report:

L	$2\Delta\log(\mathcal{L})$	k	$2\Delta\log(\mathcal{L})-k$	$\sqrt{2k}$	P	N_s	N_p	Rating
1	1523.267	1485	38.2672	54.49771	0.24	2508	1023	★★★★★
2	4791.623	4752	39.623	97.48846	0.34	5775	1023	★★★★★
4	19047.72	18684	363.7215	193.308	0.03	19707	1023	★★★★
8	52886.24	49356	3530.241	314.1847	0	50379	1023	★★★★
16	97747.73	80028	17719.73	400.07	0	81051	1023	★★★★

- Example "color box plots", showing contributions to $2\left(\max_p(\log L) - \log L\right)$ from **individual** gate sequences:

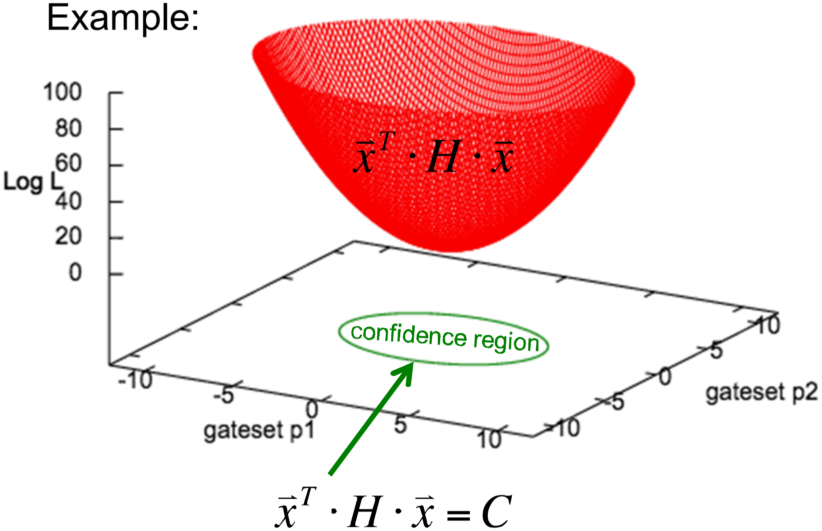




Likelihood-ratio error bars

- Everybody wants error bars…
- GST computes error bars by:
  - approximating the log-likelihood function as being quadratic (ok since we don't constrain it to hard boundaries).
  - Evaluating the Hessian at the maximum-log L point = covariance tensor.
  - Scale the Hessian H appropriately to find a valid 1-α confidence region.

Example:



gate set parameters *relative* to MLE point

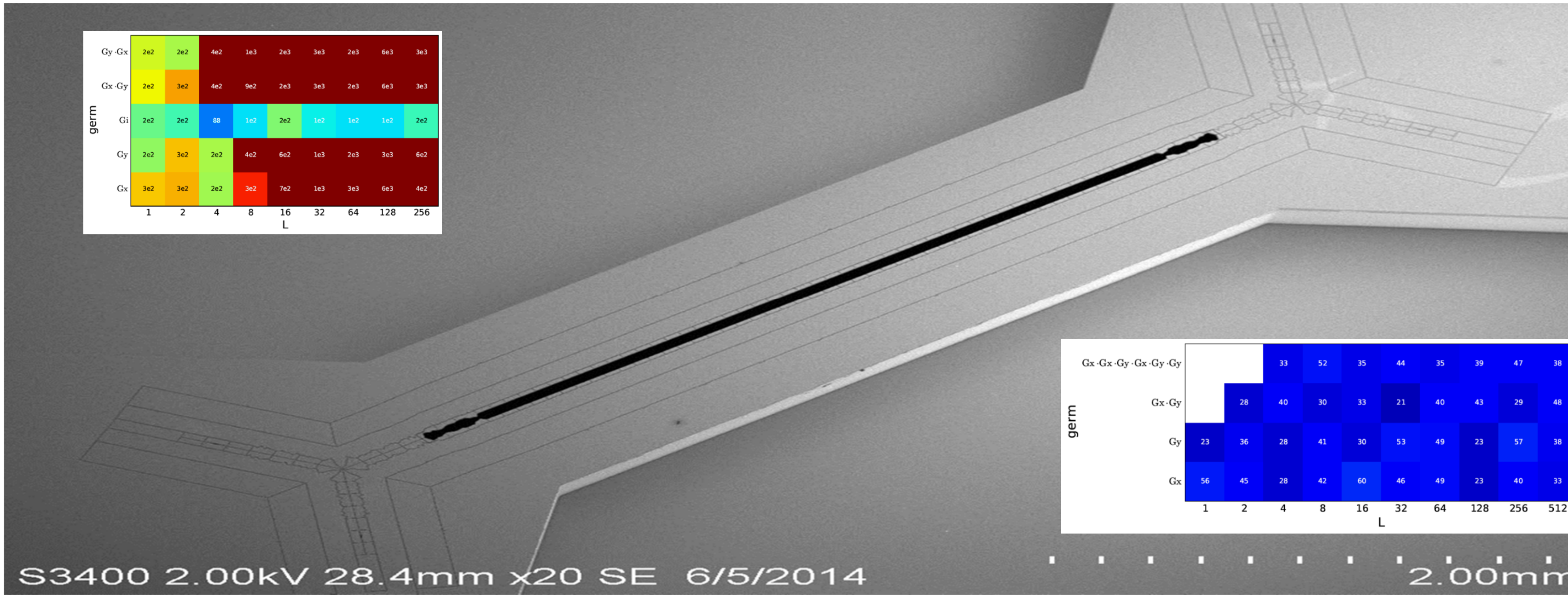
$$\vec{x}^T \cdot H \cdot \vec{x} = C \quad \text{(where } \vec{x} = d\vec{G} = \vec{G} - \vec{G}_0 \text{)}$$

Describes the boundary of a 1-α confidence region when $C = \mathrm{CDF}^{-1}[\chi_k^2](\alpha)$, where k is the dimensionality of the region being constructed.

arxiv:1202.5270

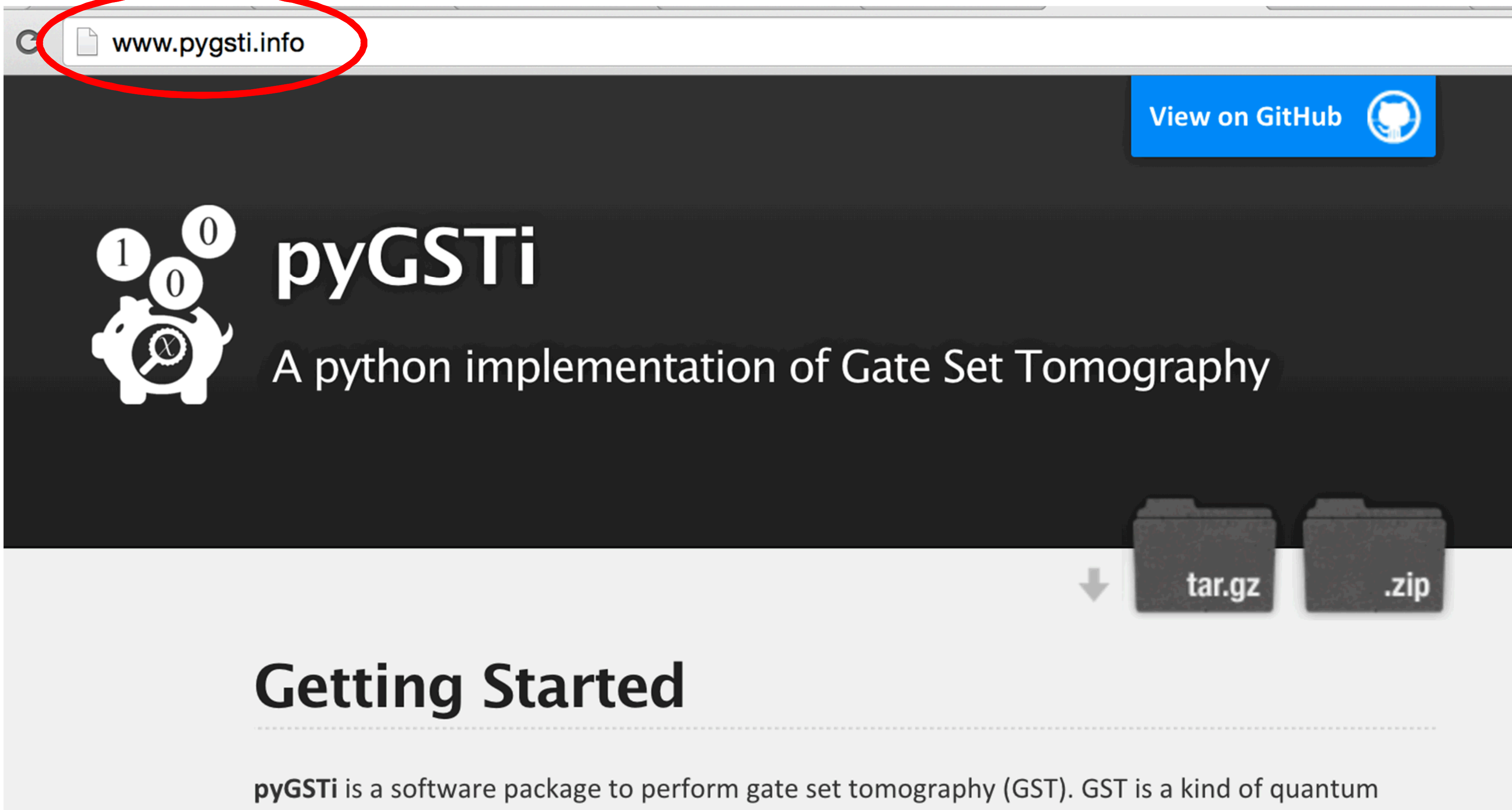
Experimental Demonstrations

- SNL Ion Trap (1Q and 2Q)



- SNL donor qubit
- UNSW Silicon donor single qubit
- Raytheon BBN transmon qubits (1Q and 2Q)
- Wisconsin-Madison double quantum dot
- More…

Open source GST software:



`import pygsti` (today!)

Recap: comparison of mainstream methods

Method	Completeness	Efficiency	Error bars	Simplicity	Sequences
Randomized Benchmarking Tomography (RBT)	Only unital part of processes.	Non-Heisenberg	yes, via bootstrapping	Simple analysis	$\rho \; \left[D_{i_1} \; G \right]^{k} \; D_R \; E$
Robust Phase Estimation (RPE)	Only several specific parameters (e.g. the "phase").	Heisenberg	yes, via bootstrapping	Simple analysis	$\rho \; ? \; \left[G_x \right]^{k} \; ? \; M$
Gate Set Tomography (GST)	Complete.	Heisenberg	computable	Fairly complex analysis.	$\rho \; F_i \; \left(g_k \right)^{2^r} \; F_j \; E$

Summary

- Calibration-free characterization methods are essential to creating fault-tolerant qubits.
- Common currency: gate sequences
- Key idea: exploit **long**-sequences for accuracy

- Many methods exist for calibration-free characterization; we covered a few of the more mainstream ones:
  - **Randomized Benchmarking Tomography** (RBT): Simple to implement & data-drive, but lacks some important advantages of GST.
  - **Robust Phase Estimation** (RPE): lightweight and efficient; intended for rapid characterization of a few gate-set parameters.
  - **Gate Set Tomography** (GST): efficient and tomographically complete; currently the best method for full-gate-set characterization. Complex implementation, but downloading free software is easy.

GST **Myths – a leftover from another of Erik's talks – probably don't need this….**

- **Myth**: GST is only for *good* qubits
  - GST is intended to operate in "harsh environments", and extract as much information as possible from noisy data.
- **Myth**: GST is slow
  - GST takes minutes to run on typical datasets
- **Myth**: GST results are hard to interpret
  - GST reports include detailed explanation of what results mean.