

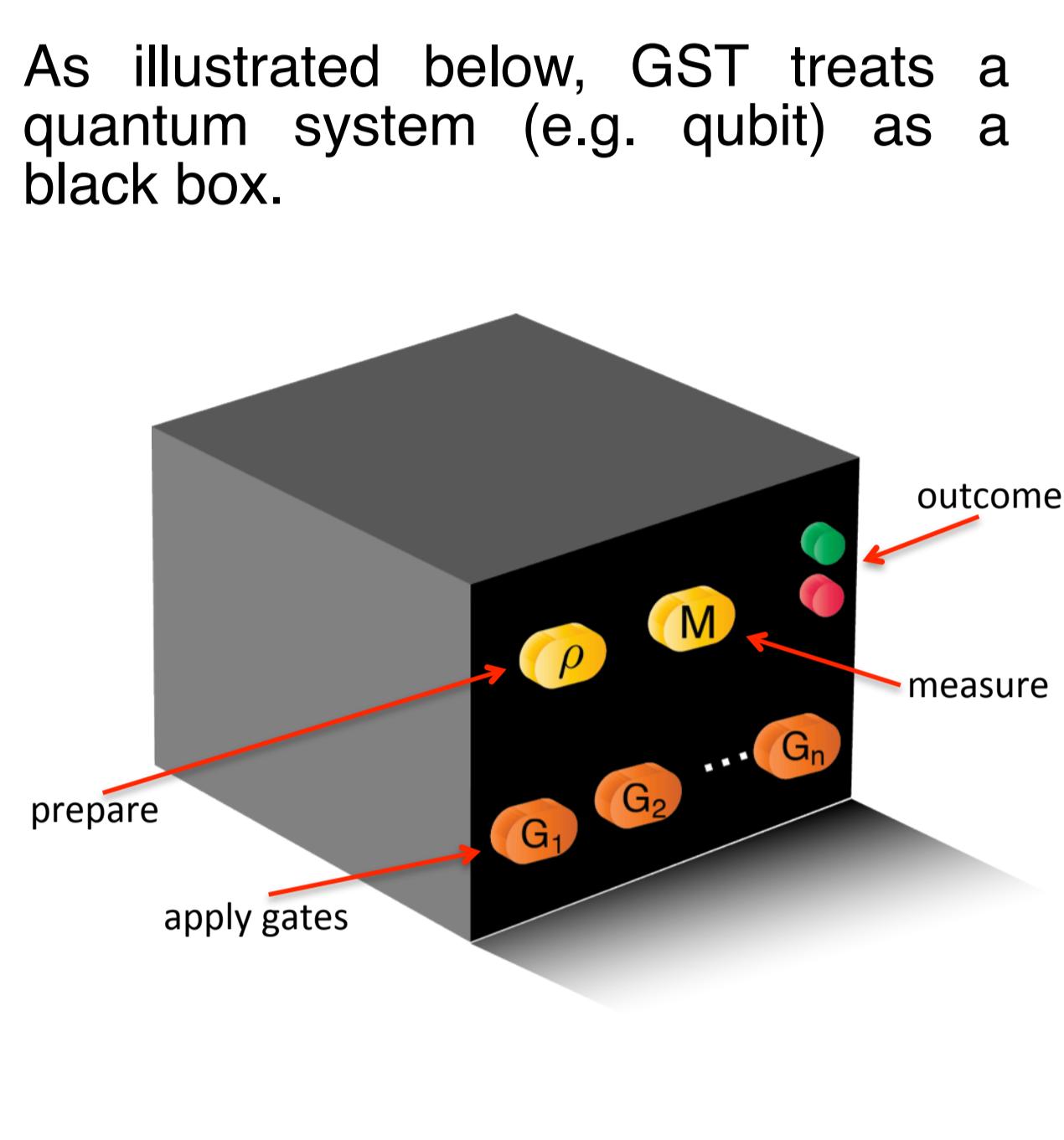


# Hyperaccuracy and Error Scaling in Gate Set Tomography

Kenneth Rudinger, Erik Nielsen, John K. Gamble, Robin Blume-Kohout

Gate set tomography is able to provide estimates of quantum states and processes with *unprecedented accuracy, consistency, and success*.

**Gate Set Tomography** (GST) is a new tomographic procedure for characterizing quantum states and processes. Unlike state and process tomography, GST does not rely on prior knowledge of either precalibrated states or operations. (E.g., process tomography requires prior knowledge of which different measurements are being performed.) Instead, GST seeks to characterize an experimental set of state preparations, processes, and measurements (the *gate set*) in a self-consistent manner.



GST protocols begin with experiments, each repeated some number of times to build up good statistics. The following steps constitute a single experiment:

1. State preparation (push the  $\rho$  button).
2. Execution of a fixed gate sequence (push buttons  $G_1 \dots G_k$ ).
3. Measurement (push the  $M$  button).

What we get from repeating each of these experiments is good estimates of many probabilities:

$$\text{Prob.} = \langle \langle M | G_k \dots G_1 | \rho \rangle \rangle$$

( $\langle \langle \dots \rangle \rangle$  denotes the vectorization operation.)



**GST can use both short and long gate sequences for gate set analysis.** Using a small set ( $\sim 100$ ) of short gate sequences ( $< 10$  gates per sequence), we can provide a rough estimate of the gate set, using only linear algebra. (For details on this procedure, cf. arXiv: 1310.4492; Robin Blume-Kohout's poster "Gate Set Tomography with Long Sequence of Gates".)

However, GST can also incorporate data from *long* gate sequences — and these can provide literally unprecedented tomography accuracy. In particular, sequences that repeat a short "germ" many times (in contrast to the *random* sequences used in randomized benchmarking) can amplify small deviations from ideal gates, proportional to the length ( $L$ ) of the sequence. We have developed algorithms to efficiently analyze the resulting data.



## Results (shown in following plots)

Using GST with long-sequence data, we demonstrate (for a single qubit) that:

1. In simulated data, estimate accuracy that scales linearly with gate sequence length (up to decoherence time).
2. In simulated data, gate set estimates succeed in approximating the true gates with high probability.
3. In simulated and experimental data, gate set estimates fit the data to a high degree; we can also use this fit data to diagnose non-Markovian behavior.

## We can achieve $1/L$ scaling in error, limited only by decoherence.

In (a), we compute gate set estimates from simulated data generated by a gate set with unitary noise present. The error in our estimates scales as  $1/L$ .

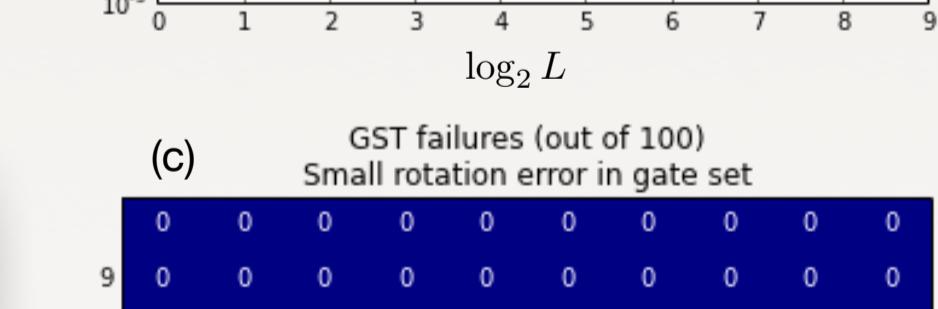
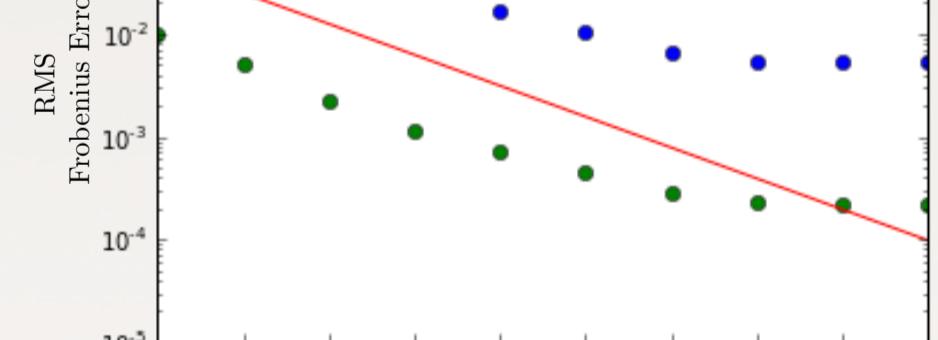
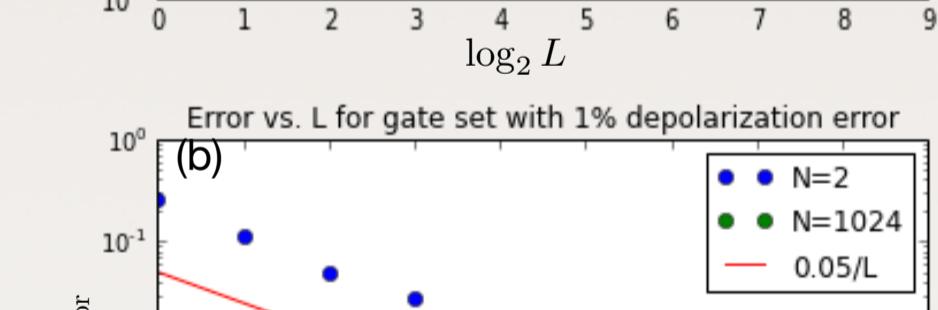
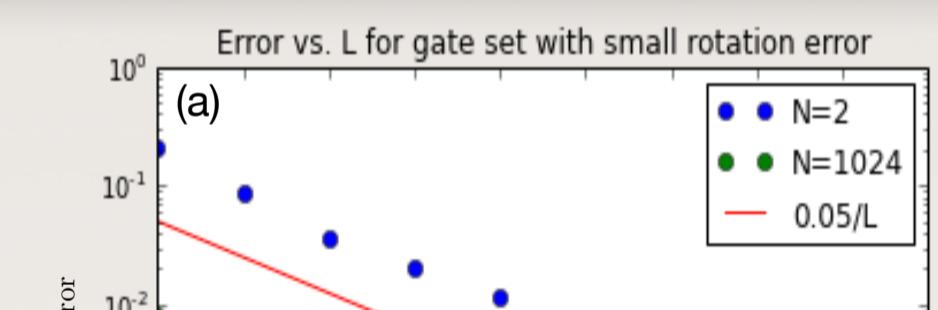
In (b), we perform the same computation, but using an underlying gate set with 1% depolarizing noise.

We see that  $1/L$  scaling is achieved until the sequences are long enough to significantly depolarize the qubit.

## GST will succeed in producing a high-accuracy estimate with high probability.

In (c) and (d), we compute what how many GST estimates succeed (by following  $1/\sqrt{N}$  error scaling) out of 100 trials. In (c), the underlying gate set is subject to small unitary error; the underlying gate set is subject to 1% depolarizing noise.

We see that, even with decoherence present, GST succeeds with high probability, failing at most 8% of the time. To guarantee success, only 32 experiments per gate sequence are required.



## Badness-of-fit.

We can measure *badness-of-fit* by computing  $\chi^2$  statistics. These diagnose whether and which data our estimate fails to fit.

Given a GST gate set estimate, one may calculate the expected probability of a measurement outcome following any gate sequence  $G_1 \dots G_k$ . If we denote the probability estimate of observing a +1 measurement following a particular gate sequence as  $p$  and the corresponding experimentally observed frequency as  $f$  then the  $\chi^2$  statistic for the gate set estimate with respect to that gate sequence is given by:

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

This statistic indicates how well the estimator fits the data, with a lower value indicating a better fit.

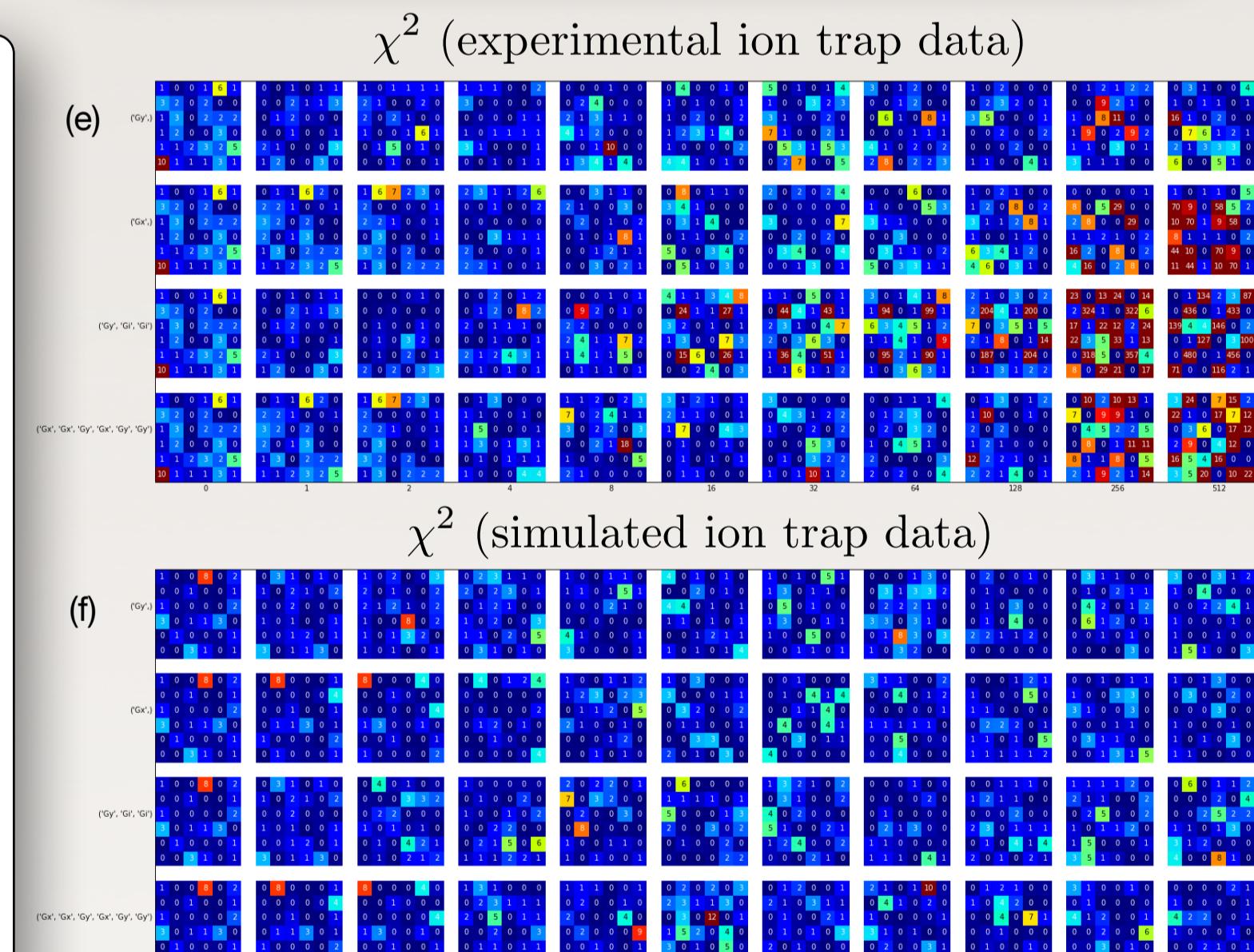
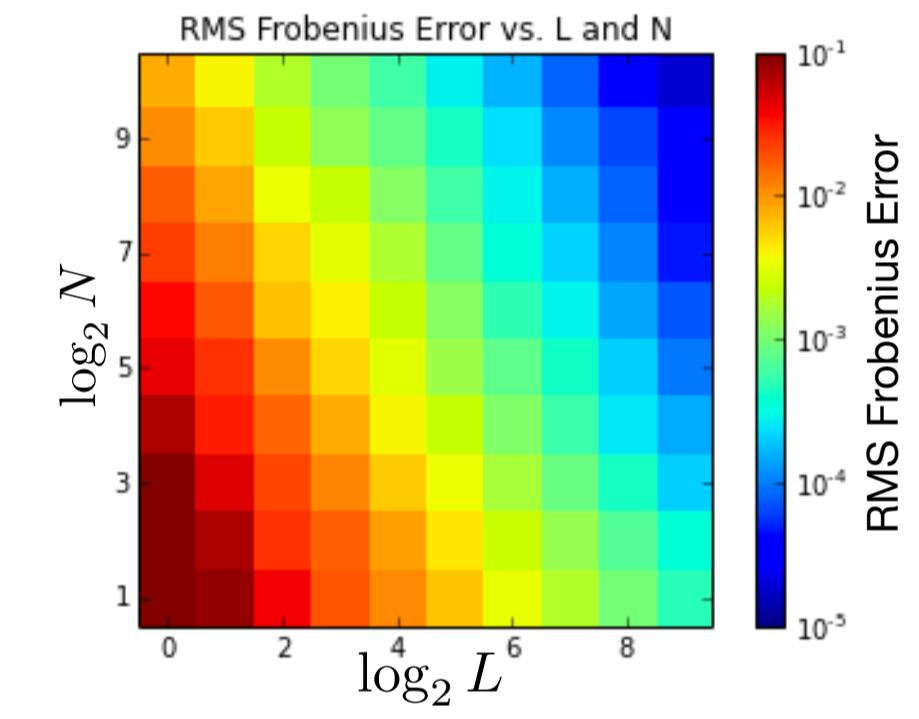
We show  $\chi^2$  values in figures (e) and (f). Each small box gives a  $\chi^2$  value for a single gate sequence. The sequences naturally organize into  $6 \times 6$  blocks (see Robin Blume-Kohout's poster for further detail).

Fig. (e): We compute a gate set estimate using experimental data (from an SNL ion trap). The plot in (e) shows the estimate's  $\chi^2$  badness-of-fit for individual gate sequences, using sequences of lengths  $L=1, \dots, 512$  with a subset of the underlying periodic sequences.

Fig. (f): Using the gate set estimate from (e) as the "true gate set", we simulate and analyze GST data. The plot in (f) shows the simulated estimate's  $\chi^2$  badness-of-fit for individual gate sequences, using sequences of lengths  $L=1, \dots, 512$  with a subset of the underlying periodic sequences.

Note that the gate set estimate of experimental data (e) performs much worse at long sequences than the gate set estimate of the simulated data. This indicates a degree of non-Markovianity in the gates. (As the simulated data is by fiat Markovian, we can see what how a gate set estimate performs on Markovian data, and we see that its performance is insensitive to sequence length, unlike the experimental data.)

Through simulation, we can explore how maximum sequence length ( $L$ ) and number of experiments ( $N$ ) affect accuracy.



## Conclusions

- GST is ready for use, robust, and capable of unprecedented accuracy.
- Long-sequence experiments reveal non-Markovian noise in multiple qubit technologies.
- Long-sequence experiments allow the accuracy of GST to scale with sequence length.
- GST succeeds with high probability, even with short sequences and few experiments.