SANDIA REPORT

SAND2015-0224
Unlimited Release
Printed January, 2015

Turbocharging Quantum Tomography

R. Blume-Kohout, J. K. Gamble, E. Nielsen, P. Maunz, T. Scholten, K. Rudinger

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

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.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online ordering: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2015-0224
Unlimited Release
Printed January, 2015

Turbocharging Quantum Tomography

R. Blume-Kohout
Mail Stop 1322
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185

J. K. Gamble, E. Nielsen, P. Maunz, T. Scholten, K. Rudinger

Abstract

Quantum tomography is used to characterize quantum operations implemented in quantum information processing (QIP) hardware. Traditionally, *state* tomography has been used to characterize the quantum state prepared in an initialization procedure, while quantum *process* tomography is used to characterize dynamical operations on a QIP system. As such, tomography is critical to the development of QIP hardware (since it is necessary both for debugging and validating as-built devices, and its results are used to influence the next generation of devices). But tomography suffers from several critical drawbacks. In this report, we present new research that resolves several of these flaws. We describe a new form of tomography called *gate set tomography* (GST), which unifies state and process tomography, avoids prior methods critical reliance on precalibrated operations that are not generally available, and can achieve unprecedented accuracies. We report on theory and experimental development of *adaptive tomography* protocols that achieve far higher fidelity in state reconstruction than non-adaptive methods. Finally, we present a new theoretical and experimental analysis of process tomography on multispin systems, and demonstrate how to more effectively detect and characterize quantum noise using carefully tailored ensembles of input states.

Acknowledgment

The work described in this report was significantly aided by helpful discussions and collaborations with Kevin Young, Wayne Witzel, Toby Jacobson, Cort Johnson, Jonathan Mizrahi, Dohun Kim, Julian Kelly, Aephraim Steinberg, Dylan Mahler, Lee Rozema, Stephen Bartlett, Steve Flammia, Steven van Enk, Marcus Silva, and by project management support from Steve Rinaldi and John Aidun.

This work was funded under LDRD Project Number 165581, titled "Turbocharging Quantum Tomography".

Contents

1 Introduction **11**

2 Quantum measurement and tomography **13**

2.1 The nature of quantum measurements 13

2.2 Mathematical formalism of quantum measurement 15

2.2.1 Classical theory 15

2.2.2 Quantum measurement theory 16

2.2.3 Technical formalism of quantum theory 18

2.3 Quantum state tomography 21

2.3.1 How tomography can go wrong 21

2.3.2 Basic state tomography 22

2.3.3 Linear inversion tomography 23

2.3.4 Least-squares and maximum likelihood tomography 25

2.4 Quantum process tomography 28

2.4.1 The Choi-Jamiolkowski isomorphism 28

2.4.2 Mathematical formalism of process tomography 29

3 Gate-set tomography and the LGST algorithm **31**

3.1 Motivation 31

3.2 Foundations 32

3.2.1 Gate sets 32

3.2.2 The gauge 33

3.3 Linear inversion GST (LGST) 35

3.3.1 The LGST protocol 36
3.3.2 Why LGST works 37
3.3.3 Fiducial sequences: completeness and overcompleteness 38
3.4 Contraction to deal with CPTP-violations caused by finite-sample errors 40

4 Hyperaccurate GST using long circuits 43
4.1 Extended LGST 44
4.1.1 Computing the eLGST estimate 45
4.2 Base sequences and germ powers 47
4.2.1 Choosing germs 49

5 Adaptive tomography improves accuracy quadratically 51
5.1 Adaptive tomography 53
5.2 Simulation results 54
5.3 Experimental results 56
5.4 Discussion 59
5.5 Sources of systematic error 60

6 $SU(2)$-covariant probe states for quantum process characterization 67
6.1 Experimental Methods and Preliminary Validations 70
6.2 Detecting depolarization 73
6.3 Quantum Process Tomography 77
6.4 Conclusions 81

References 82

List of Figures

5.1 Two features of qubit tomography with Pauli measurements (shown for an equatorial cross-section of the Bloch sphere): **(a)** The distribution or "scatter" of any unbiased estimator $\widehat{\rho}$ (depicted by dull red ellipses) varies with the true state ρ (black stars at the center of ellipses); **(b)** The expected infidelity between $\widehat{\rho}$ and ρ as a function of ρ. Within the Bloch sphere, the expected infidelity is $O\left(1/N\right)$. But in a thin shell of nearly-pure states (of thickness $O\left(1/\sqrt{N}\right)$), it scales as $O\left(1/\sqrt{N}\right)$ – *except* when ρ is aligned with a measurement axis (Pauli X, Y, or Z). 55

5.2 Average infidelity $1 - F(\widehat{\rho}, \rho)$ vs. sample size N for Monte Carlo simulations of four different tomographic procotocols: standard tomography (black), the procedure proposed in [1] using $N_0 = N^{2/3}$ (red), our procedure using $N_0 = N/2$ (blue), and "known basis" tomography (green). Both adaptive procedures clearly outperform static tomography, but our procedure clearly outperforms the $N_0 = N^{2/3}$ approach, and matches the asymptotic scaling of known-basis tomography. The inset shows the dependence of the prefactor (β) on $\alpha = N_0/N$. 57

5.3 Spontaneous parametric downconversion is performed by pumping a nonlinear BBO crystal with linearly polarized light. One photon is sent directly to a detector as a trigger. A rotation using a quarter-half waveplate combination prepares the other photon in any desired polarization state. Finally, a projective measurement onto any axis of the Bloch sphere is performed by a quarter-half waveplate combination followed by a polarizing beamsplitter. The measurement waveplates are connected to a computer to enable adaptation. 58

5.4 Experimental data: a) The average infidelity $1 - F(\widehat{\rho}, \rho)$ for the three tomographic protocols shown in Fig. 5.2 vs. the number of samples N. Each average is over 150 different realizations of the experiment. b) Average infidelity $1 - F(\widehat{\rho}, \rho)$ for standard tomography (black) and *reduced* adaptive tomography (blue) is plotted versus N. Each average is over 200 different realizations of the experiment; error bars are standard deviation of the mean of these samples. Error bars are standard deviation of the mean of these samples. 59

5.5 Average infidelity vs. sample size (N) for simulations with systematic errors (of Model 1 type; see text) on the order of $E = 10^{-2}$. The infidelity decreases with increasing N up to a point, after which it flattens out after hitting a 'noise floor'. The noise floor occurs at a lower average infidelity for adaptive tomography than for static tomography. 61

5.6 Average infidelity vs. sample size (N) for simulations of adaptive tomography with systematic errors (of Model 1 type; see text) on the order of $E = 10^{-3}$. Also plotted is the region over which experimental data was taken (see main body of text) and a line of best fit for this region. 63

5.7 **Location of the noise floor ($1 - F$) for Models 1,2,3 (from top to bottom).** For each of the three Models, we plot the average infidelity as $N \to \infty$ (to ensure that systematic errors dominate) vs. E (the magnitude of systematic error). Data points are results of simulation, and lines are lines of best fit. 64

6.1 Schematic diagram of the experimental apparatus used to generate and measure different biphoton states. **a) State Preparation** – We can prepare any biphoton state in two steps. First, we prepare a state of the form $\sqrt{x}\,|2,0\rangle_{H,V} + e^{i\phi}\sqrt{1-x}\,|0,2\rangle_{H,V}$ by using a polarizing beam splitter (PBS) to combine a vertically polarized weak coherent state with the output of horizontally polarized type-I collinear down-conversion. The amplitude x is set by attenuating the laser intensity relative to the down-converted intensity. The angle ϕ is set by the relative phase between the two paths, which we control using a liquid-crystal wave plate (LCWP). Then, we apply any desired polarization rotation in $SU(2)$, using quarter- and half-waveplates, to produce any desired biphoton state. **b) Process** – 3 LCWPs, oriented as shown, are used to perform arbitrary polarization rotations. To implement depolarization, the retardances of the LCWPs are made to fluctuate during each measurement. **c) State Measurement** – The biphoton states pass through wave plates, are coupled into a polarization-maintaining fiber and sent to a polarizing beam splitter. The output of each port of the beam splitter is probabilistically split, using 50:50 fiber beam splitters, and sent to single-photon counting modules. **d) Input States** – A graphical representation of the set of states used for process tomography. A fiducial state is prepared and rotated to nine other states. These states (including the unrotated fiducial state) make up the set of ten input states used for process tomography. The fiducial state is represented by the red line, and it is rotated to each of the nine other points on the sphere. The rotations are chosen to be (approximately) uniformly distributed on the surface of the sphere. 69

6.2 Typical results of state tomography, and the effect of $SU(2)$-jitter depolarization. In plots **(a)-(c)**, we show experimentally reconstructed Wigner functions (plotted on the Poincare sphere) for three different states after they have been decohered by three different amounts of $SU(2)$ jitter. Row (a) shows spin-coherent states with x=0, row (b) shows "2-design" states with x=0.15, and row (c) shows nearly-N00N states with x=0.47. Each row shows the effect of applying depolarization with strength (see Eq. 6.4) $\gamma = 0, 0.5, 1.5$. In **(d)**, we plot the *purity* of the same reconstructed states shown in (a-c). The solid lines are the theoretical predictions given by simulations of the process (Eq. 6.4). 72

6.3 **Sensitivity of different states to depolarization:** This figure shows two different measures of the probe state's ability to detect depolarization, for three different probe states, and compares theoretical predictions to experimental data. Plot **(a)** shows the probability that depolarization is *not* detected, which is simply the probability of finding the system in the same state in which was prepared despite depolarization having happened. The theory (solid line) is simply the projection $\langle\psi_x|\,\rho\,|\psi_x\rangle$, where $|\psi\rangle_x$ is the probe state and ρ is the decohered probe state. Experimental data points are empirical probabilities of nondetection. Plot **(b)** shows the *sensitivity* (Eq. 6.12) scaled by $\sqrt{N}$ (where N is number of two-photon counts detected at each point) of each state to small changes in depolarization, the solid lines are calculated from the slopes of the fit to the detection data shown in (a), and the dashed lines are calculated for ideal input states. 74

6.4 **Inaccuracy of process reconstruction.** This figure shows the *average process infidelity* (API, see Eq. 6.13 – smaller is better) between the true process and its tomographic reconstruction, and its dependence on the fiducial state used to generate the set of probe states for QPT. Solid lines are simulations of the experiment using 10 randomly oriented *pure* input states, while dashed lines are simulations using the 10 experimental input states (as determined using state tomography, and in particular accounting for the decline in their purity as x increases). Squares are experimentally estimated process infidelities. Red and blue represent different depolarization strengths. **INSET: "Completeness" of input sets:** The inset plot shows the determinant of the probe states' Gram matrix (normalized to a maximum of 1), whose inverse appears in the tomographic reconstruction. It depends on x; larger determinants yield a more robust inversion, while zero determinant indicates a tomographically incomplete set. We observe that a N00N fiducial state ($x = \frac{1}{2}$) yields a probe set that is not tomographically complete, while for $x \approx 0.15$ the reconstruction should be as robust and accurate as possible. The main plot confirms this theoretical prediction. 78

Chapter 1

Introduction

The goal of this project, a 2-year Early Career LDRD called "Turbocharging Quantum Tomography", was to advance the state of the art in quantum tomography. Quantum tomography denotes the suite of techniques used for detailed characterization of quantum systems, and especially quantum logic devices such as qubits. While there are many variants of tomography, most approaches to date are either *state tomography* or *process tomography*.

State tomography is used to find the $d \times d$ density matrix that describes a repeatable procedure for system initialization. By repeating the procedure over and over, many (N) samples of a state ρ are generated; these samples are measured in various bases, and the resulting statistics are combined to obtain an estimate $\widehat{\rho}$ of ρ. It is this combining of many snapshots, in different bases, that inspired the name "tomography", by analogy to tomographic imaging (e.g. in medicine).

Process tomography is used to find the $d^2 \times d^2$ process matrix that describes a repeatable quantum operation (e.g. logic gate). This is done in close analogy to state tomography, except that a variety of different initial states ρ_i are prepared, and then the unknown process $\mathcal{E}$ is applied to them; state tomography is then used to estimate $\mathcal{E}[\rho_i]$ for each of the input states, and these results are combined to form an estimate $\hat{E}$ of the process $\mathcal{E}$.

The most important accomplishment of this project was to introduce a new form of tomography called *gate-set tomography* (GST), which subsumes both state and process tomography (along with the less well-known *measurement tomography*) into a single unified procedure for characterizing the full functionality of a quantum logic device. GST solves one of the main outstanding problems identified at the beginning of this project, the fragility of tomography to calibration errors. GST is, with no exaggeration at all, a revolutionary protocol, and its development occupies the first three technical chapters of this report.

In addition, this project led to two new developments in photonic tomography. While not as significant as GST, both are worthy accomplishments. They were performed in collaboration with Aephraim Steinberg's experimental group at the University of Toronto, and have produced high-quality papers (published in *Physical Review Letters* and *Physical Review X*).

Chapter 2

Quantum measurement and tomography

Quantum systems are interesting because they march to a different drummer. They obey a different set of rules from the classical systems we are used to. These rules are *technologically* interesting, because they permit new capabilities (fast algorithms) and forbid some others (perfect copying). But these new rules also create new engineering challenges. The curious nature of *measurements* (or observations) of quantum systems is responsible for many of these challenges.

2.1 The nature of quantum measurements

Classical systems may be observed (and thus measured) in various ways, but these different measurements are all *compatible*. They emphasize different properties of the classical system (e.g. mass, color, location) that exist simultaneously, and can in principle be measured simultaneously. That is, if we can measure a classical system's position x, and we can also measure its velocity v, then there is no obstacle to observing its location in phase space (x, v). A single observation can reveal every property, which is to say every aspect of its state.

In the previous paragraph, "state" means "all the information necessary to predict the outcome of future measurements on the system". This is sometimes called a *state of reality*. There is another notion of state, *states of knowledge*, which encapsulate how a partly-informed observer might describe the system. These are probability distributions over states of reality – e.g., rather than a single point (x, v), a distribution $p(x, v)$. Such states cannot be found out by a single observation; the probabilities can only be reconstructed (approximately) from observations on many identical samples. But it remains true that only a single measurement *procedure* is required – and there is no better strategy than repeating it many times and estimating the distribution from its statistics.

In contrast, there is no uniquely powerful, optimally informative measurement for a quantum system. Quantum systems have many properties that are mutually *incompatible*. Only one (or more generally a small subset) of these properties can have well-defined unique

values at any given time. Observing incompatible properties demands distinct and incompatible procedures. If x and v are incompatible properties, and a system currently has a well-defined (i.e., predictable) value of x, then the result of attempting to observe v is unpredictable. However, nontrivial probabilities can be assigned to its outcomes, so v is not necessarily "completely unpredictable" – just not perfectly predictable.

As a result of all this, identifying (or estimating) a quantum system's state is rather more challenging. If we happen to know the state $|\psi\rangle$ in advance, then we can choose and perform the measurement tailored to its well-defined properties. Its outcome will be predictable, and upon observing the (unique) predicted outcome, we gain confidence that indeed the system was in state $|\psi\rangle$ before we measured it. But quantum states are much like classical states of knowledge (probability distributions), because even after a measurement reveals the predicted value, we cannot be sure that the system's state was original $|\psi\rangle$. There are other states – similar, but not identical to $|\psi\rangle$ – that could have generated identical data. Only by measuring many identically prepared samples can we approach certainty that they were all prepared in $|\psi\rangle$, or in a nearby state.

If we do not know that the system's state is $|\psi\rangle$ in advance, then things get harder. We have to measure a variety of different properties. None of them (typically) will be perfectly predictable, but if they are wisely chosen, then their outcomes' collective probabilities will uniquely identify the state. (This is not obvious, but it follows from the mathematics stated in the next section). Of course, as noted above in the context of classical states of knowledge, probabilities cannot be measured by observing a single sample. Nor can they be estimated exactly with any finite number of samples. So getting an approximate estimate of the quantum state requires repeated measurements, of various properties, on many identically prepare samples.

Generally, quantum systems are found not in pure quantum states $|\psi\rangle$, but in probabilistic mixtures of multiple different $|\psi_j\rangle$. These are *mixed* quantum states, and they are described by *density matrices* of the form

$$\rho = \sum_j p_j \, |\psi_j\rangle\langle\psi_j|,$$

which represent quantum "states of knowledge". That is, they represent imperfect knowledge about the system, and do not assign a unique value to *any* complete property of the system [1]. Noise and fluctuations are sufficiently ubiquitous for quantum systems that we expect to find systems in states that are at least slightly mixed, and therefore if we set out to estimate the state of an ensemble of identically prepared quantum systems, the estimate will be a density matrix $\hat{\rho}$, not a pure state. Fortunately, generic mixed states can be reconstructed

[1] A "complete" property is one whose value uniquely constrains the quantum state. In contrast, an incomplete property is a strict coarse-graining of a complete property, and therefore its value is consistent with multiple distinct states for the system. States are distinct if they predict different probability distributions (or even, as in this case, different well-defined and deterministic values) for some observation on the system. Thus, the point of the sentence from which this footnote originated is that a mixed quantum state may assign a unique value to an incomplete or coarse-grained property, but there will then exist a more finely resolved measurement, corresponding to the complete property that was coarse-grained to obtain the incomplete property, whose outcome is not predicted with certainty by that mixed state.

from a finite (if sometimes awkwardly large) set of probabilities for various properties, just like pure states as discussed in the previous paragraph.

One final aspect of quantum measurement is worth noting before we move on. Identifying an a-priori unknown quantum state requires, as noted above, measuring *multiple* properties and estimating the probabilities of their outcomes. However, it is always possible to combine these properties into a *single* property – but a property not of the system alone, but jointly of the system and an ancillary system. For example, states of quantum bits (qubits) are often estimated by independently measuring three binary-valued properties called X, Y, and Z. But these properties can be combined into a single joint property of the qubit *and* a 3-sided die. To measure this property, we first roll the 3-sided die, then (conditional on its result) measure either X, Y, or Z on the qubit. This experiment has 6 possible outomes, and their probabilities uniquely define the qubit's quantum state.

The feasibility of this process muddies the waters, calling into question the notion that quantum state estimation requires measurement of multiple properties – since it is sufficient to measure a *single* property (albeit a joint property of the system-of-interest and an ancillary random variable). This suggests the possibility that some such measurement is really fundamental, and can serve as the "one observation to rule them all" that would reveal all necessary information about any quantum state. One of the results reported here helps to resolve this confusion. We show in the discussion of adaptive tomography that while it is *possible* to estimate any arbitrary state using a single fixed measurement, is never *optimal*. For almost every state ρ, there is a different measurement – one that measures a property of the system-of-interest alone – that provides much sharper and more useful information about ρ than any "generic" measurement not tailored to ρ.

2.2 Mathematical formalism of quantum measurement

The mathematical structure of quantum measurements is often taken as a fiat construction. It is not. Instead, it is based upon (and best understood as a generalization of) the theory of *classical* systems and measurements.

2.2.1 Classical theory

A classical system is modeled by a *sample space*, which is a set containing one element for each outcome of the most fine-grained observation on that system. (Obviously, this is somewhat subjective in practice – it is always possible that we might discover, in the future, a further fine-graining, and this would enlarge the sample space). In the simplest and cleanest picture, an observation of that system is a process that yields a single outcome (element of the sample space). If repeated immediately, the observation will yield (by assumption and definition) the same outcome. We can thus predict the outcome of such a repeated measurement by recording the first outcome. This is a *state* for the system. Thus, in this

simple picture, both states and outcomes are labeled by elements of the sample space.

Both classical states and classical measurements can be defined more generally. In general, a state is *all the information required to predict, as well as possible, the outcome of the next observation on the system*. Clearly, an element of the sample space fits this bill. But some agents have imperfect knowledge, and cannot predict the outcome perfectly. In this case, a more general state is required, and this corresponds to a *measure* – or probability distribution – over outcomes. This sort of state is sometimes called an *epistemic state* (indicating that it describes a particular observer's knowledge) in contrast to *ontic states* (indicating that they describe a real property), which specify a unique outcome.

A general classical measurement is described by a set of *indicator functions*: $\{I_i(x)\}$. Here, x ranges over the sample space; each $I_i(x)$ is function taking values in the closed real interval $[0, 1]$; and the sum of all indicator functions is equal to $\sum_i I_i(x) = 1$ (for all values of x). This set of indicator functions has the following operational interpretation: Given that the system's state is a probability measure μ that we describe informally as $p(x)\mathrm{d}x$, the probability of outcome i is given by

$$Pr(i) = \int I_i(x)p(x)\mathrm{d}x.$$

This is a linear dot product between the indicator function I_i and the state $p(x)\mathrm{d}x$, and for finite sample spaces is more conveniently written as $Pr(i) = \vec{p} \cdot \vec{I_i} = \sum_k p(k)I_i(k)$. The simple theory stated earlier can be recovered (for a countable sample space) by choosing indicator functions that are Kronecker delta functions: $I_i(x) = \delta_{i,x}$.

In this framework, *pure states* are those that cannot be formed by convex combination of any other states, which means that they correspond to individual elements of the sample space. A *sharp* measurements, similarly, cannot be obtained by coarse-graining or convex combination, and corresponds (uniquely) to the maximally fine-grained measurement that reveals the state exactly. *Mixed* states are represented by probability distributions, and *fuzzy measurements* are represented by nontrivial indicator functions. A mixed state does not perfectly predict the outcome of a sharp measurement, while fuzzy measurements do not uniquely reveal the state (and therefore cannot be used to obtain information which perfectly predicts future sharp measurements).

2.2.2 Quantum measurement theory

There is a simple reason for presenting classical theory (and for the *way* it was presented) in the previous section. Quantum theory maintains the same structure. The mathematical *objects* are different, but the ways in which they combine to produce probabilities for observations are nearly identical. Thus, the presentation of classical theory in the previous sections can serve as a template for understanding the theory that underlies quantum tomography.

The single most important difference between quantum and classical theory is this: **there**

is no unique or preferred measurement on a quantum system. In classical theory, there is such a measurement, because classical theory is grounded on a *sample space* – which is nothing but the set of mutually exhaustive and exclusive outcomes of some measurement (a.k.a. observation). This (potentially implicit) measurement defines the system.

In a quantum system, the analogous role is played by a *Hilbert space* – a complex d-dimensional vector space $\mathcal{H} = \mathbb{C}^d$. (This is an oversimplification; much of the fun mathematics in quantum theory arises in the context of systems whose Hilbert space is *infinite*-dimensional, and therefore not isomorphic to $\mathbb{C}^d$. However, this report is largely concerned with the finite-dimensional spaces used in quantum information science. When we *do* consider infinite-dimensional quantum mechanics, none of their mathematical peculiarities will be relevant.)

Pure states and sharp measurement outcomes of the quantum system are represented by elements of $\mathcal{H}$ – i.e., complex vectors – exactly as their classical counterparts are represented by elements of the sample space. More precisely, pure states are given by elements of $\mathcal{H}$, denoted in Dirac's notation as

$$|\psi\rangle \in \mathcal{H},$$

while sharp measurement outcomes are given by elements of the dual space $\mathcal{H}^*$,

$$\langle\phi| \in \mathcal{H}^*.$$

However, because $\mathcal{H}$ is a Hilbert space, it is isomorphic to its dual space, and Dirac's bra/ket notation is only required to keep track of whether a given object is a row or a column vector.

The second most important difference between quantum and classical theory lies in how probabilities are calculated: **whereas in classical theory probabilities are bilinear in the elements of the sample space representing states and outcomes, quantum probabilities are biquadratic**:

$$Pr(\phi|\psi) = |\langle\psi|\phi\rangle|^2.$$

This is *Born's Rule*.

Depending on perspective, this seems like either a drastic or a trivial change. Instead, it lies somewhere in between. Most of the basic structure of classical probability survives the change from $Pr = \vec{p} \cdot \vec{I}$ to $Pr = |\langle\psi|\phi\rangle|^2$. However, there is at least one far-reaching difference. *Mixed* states and events, which in both theories are by measures over pure states and events, *cannot* be represented in quantum theory by convex combinations of elements of the Hilbert space. This is because probabilities are not bilinear in Hilbert space vectors. Instead, they are bilinear in the *projectors* onto those vectors:

$$Pr(\phi|\psi) = \mathrm{Tr}[|\psi\rangle\langle\psi| \;\; |\phi\rangle\langle\phi|].$$

Mixed states are therefore represented by convex combinations of projectors or *density matrices*:

$$\rho = \sum_k p_k \, |\psi_k\rangle\langle\psi_k|.$$

General events are represented by positive sums of projectors or *effects*:

$$E_i = \sum_k w_k \, |\phi_k\rangle\langle\phi_k|.$$

And any set of events that sum to unity represents a set of mutually exclusive and exhaustive events – i.e., a measurement or *POVM* (positive operator-valued measure):

$$\mathcal{M} = \{E_i\} : \quad E_i \geq 0, \; \sum_i E_i = \mathbb{1}.$$

The far-reaching consequence mentioned above is that, although we began with a d-dimensional Hilbert space, and therefore a *repeatable* measurement can have at most d outcomes, quantum states form a set of dimension $d^2 - 1$. Thus, there can be no single repeatable measurement whose statistics uniquely identify arbitrary quantum states. This is in direct contrast with the classical situation, and is the *raison d'ètre* for quantum tomography.

This rift between quantum behavior and classical intuition is usually ascribed to the superposition principle (which states that any two distinct quantum states can be linearly combined or "superposed" to produce additional states). This is not actually correct. Classical probability theory also obeys a superposition principle. Instead, the extra parameters in quantum states stem directly from the quadratic form of Born's Rule. Rather remarkably, Born actually got this wrong in his seminal paper, stating in the main text that $Pr = |\langle\psi|\phi\rangle|$. The correct formula is only stated in a footnote, added in the final proofs! (The necessity for the quadratic form of Born's Rule *can* be traced to the superposition principle, but only with the additional caveat that every superposition of pure states must *also* be a pure state – which is not true in classical probability theory. Another misconception is that it has something to do with complex numbers. This is not true; quantum theory with real amplitudes is well-defined, but actually slightly weirder than complex quantum theory. No truly compelling argument for the existence of complex amplitudes is known.)

2.2.3 Technical formalism of quantum theory

In the previous two subsections, we (1) motivated the structure of quantum theory using classical probability theory, and (2) motivated the details of quantum theory in direct contrast to classical theory. In this subsection, we abandon motivation and understanding, and simply state the *rules* of quantum probabilities, in a form that is particularly useful for presenting tomography results.

The structural properties of a quantum system are completely determined by associating a *Hilbert space* with it, $\mathcal{H} = \mathbb{C}^d$. The dimension d is equal to the maximum number of outcomes for any repeatable measurement on the system, and is therefore intrinsically subjective, and should be viewed as a hyperparameter of a *model* for the quantum system, rather than a real physical property of the system itself.

States are represented by positive semidefinite trace-1 *density matrices* $\rho : \mathcal{H} \to \mathcal{H}$. Measurement outcomes are represented by positive semidefinite *effect* matrices $E : \mathcal{H} \to \mathcal{H}$. Measurements or observations are represented by *positive operator-valued measures* (POVMs), which are sets of effects $\mathcal{M} = \{E_k\}$ that add to unity ($\sum_k E_k = \mathbb{1}$). All of these entities are given meaning exclusively by Born's Rule for probabilities, $Pr(E_k|\rho) = \mathrm{Tr}[E_k\rho]$.

It is very useful to represent these quantities using a mild generalization of Dirac's notation to describe *Hilbert-Schmidt space*. The system's Hilbert-Schmidt space is the d^2-dimensional vector space of linear operators on $\mathcal{H}$, denoted $\mathcal{B}(\mathcal{H})$ [2], equipped with the Hilbert-Schmidt inner product

$$\langle\langle A|B\rangle\rangle \equiv \mathrm{Tr}[A^\dagger B].$$

Both states ρ and effects E are elements of Hilbert-Schmidt space, and Born's rule is very conveniently written as an inner product:

$$Pr(E_k|\rho) = \langle\langle E_k|\rho\rangle\rangle.$$

Since both states and effects are Hermitian, we can almost always restrict focus to the subspace of Hermitian operators, which is a d^2-dimensional *real* vector space. In this case, the $\dagger$ in the inner product is irrelevant.

Tomography is much concerned with describing experimental (and thus imperfect) *quantum logic gates*. These are dynamical transformations of quantuym systems, usually called *quantum processes* or *quantum channels*. When quantum systems experience reversible dynamics, their states evolve according to unitary transformations on $\mathcal{H}$, which transform density matrices as

$$\rho \to U\rho U^\dagger.$$

This happens to be a linear map acting on ρ, as it must be in order to preserve the convex structure of probability. *Irreversible* dynamics can sometimes be modeled as a probabilistic mixture of unitary dynamics,

$$\rho \to G[\rho] = \sum_i p_i U_i \rho U_i^\dagger,$$

but most generally, *any* linear map on density matrices can be realized as a dynamical transformation,

$$\rho \to G[\rho], \text{ where } G[\rho_1 + \rho_2] = G[\rho_1] + G[\rho_2],$$

as long as it satisfies two conditions:

1. G must be *trace preserving* (TP),
2. G must be *completely positive* (CP).

[2]Technically, this denotes *bounded* operators on $\mathcal{H}$, and because boundedness is trivially true for all operators on $\mathbb{C}^d$, we ought to refer to it as $\mathcal{L}(\mathcal{H})$, the space of all linear operators on $\mathcal{H}$. But for infinite-dimensional spaces, $\mathcal{B}(\mathcal{H})$ and $\mathcal{L}(\mathcal{H})$ are distinct, and $\mathcal{B}(\mathcal{H})$ is the correct one. We therefore borrow, with apologies, the more heavy-duty notation.

Quantum dynamical transformations are thus described by *completely positive, trace-preserving linear maps*, or *CPTP maps*. The TP condition is simple:

$$\mathrm{Tr}\left(G[\rho]\right) = \mathrm{Tr}[\rho].$$

The CP condition is a bit more fraught; it is a generalization of *positivity* (not to be confused with positive semidefiniteness), which requires that

$$\mathcal{E}[\rho] \geq 0 \text{ whenever } \rho \geq 0.$$

But quantum dynamical process can act on *parts* of larger systems, while other parts are left alone, and this motivates the strong condition that

$$(\mathcal{E} \otimes \mathbb{1})[\rho] \geq 0 \text{ whenever } \rho \geq 0.$$

Somewhat surprisingly, there exist maps that are positive but not completely positive; the canonical example is the single-qubit linear map that maps $\sigma_z \to 0$ but leaves $\sigma_x, \sigma_y, \mathbb{1}$ unchanged. This "pancake map" is not physically possible.

There are at least three useful ways to represent a quantum process G. The first and arguably most straightforward is to observe that density matrices are d^2-dimensional vectors in $\mathcal{B}(\mathcal{H})$, and any linear map on them is therefore a $d^2 \times d^2$ matrix acting on $\mathcal{B}(\mathcal{H})$:

$$|\rho\rangle\!\rangle \to G\,|\rho\rangle\!\rangle.$$

This is the *superoperator* or *Liouville* picture of a quantum process. It is very useful for tomography, because its matrix elements correspond directly to observable probabilities, and the use of Hilbert-Schmidt space notation is natural. For example, the trace of an operator is $\mathrm{Tr}(A) = \mathrm{Tr}(\mathbb{1}A) = \langle\!\langle \mathbb{1}|A\rangle\!\rangle$, and therefore

$$\mathrm{Tr}\left(G[\rho]\right) = \langle\!\langle \mathbb{1}|\, G\,|\rho\rangle\!\rangle,$$

and so the TP condition is simply

$$\langle\!\langle \mathbb{1}|\, G = \langle\!\langle Id| \Longrightarrow G^\dagger\,|\mathbb{1}\rangle\!\rangle = |\mathbb{1}\rangle\!\rangle.$$

The downside of this representation is that no simple statement of the CP condition is known.

Conversely, the *Kraus representation* of G is extremely well suited to enforcing complete positivity. A quantum process G is CP if and only if it can be written as

$$G[\rho] = \sum_i K_i \rho K_i^\dagger$$

for some set of *Kraus operators* $\{K_i\}$. The disadvantage of this representation is that while the map G itself is linear, it is not a linear function of the Kraus operator parameterization $\{K_i\}$. This makes it inconvenient for tomography.

The *Choi representation* provides a useful compromise. It is closely related to the Kraus representation, and represents G as

$$G[\rho] = \sum_{ij} \chi_{ij} B_i \rho B_j,$$

where the $\{B_i\}$ are any desired orthogonal basis of matrices, and the *Choi matrix* of χ_{ij} coefficients define G with respect to that basis. The Kraus representation is simply the diagonal form of the Choi representation, obtained by diagonalizing the χ matrix. The CP condition is simply that $\chi \geq 0$. Furthermore, G is linear in χ rather than quadratic, which makes it relatively well-suited to tomography.

2.3 Quantum state tomography

The purpose of quantum state tomography is to characterize a *preparation procedure* for a quantum system $\mathcal{S}$. The system has a d-dimensional Hilbert space, and it is assumed that the preparation procedure can be described by a $d \times d$ density matrix ρ. "Described by" implies that there is no *more* successful way to predict subsequent measurements on $\mathcal{S}$ than using Born's Rule and ρ. Given this assumption, the goal of quantum state tomography is to estimate ρ.

2.3.1 How tomography can go wrong

Before we go any further, it is worth noting that this ansatz is not absolutely reliable. For example, one preparation procedure that may violate the assumption is: "Initialize the system $\mathcal{S}$ by entangling it maximally with a distant reference system." The effect of this preparation procedure may be described by the $d \times d$ density matrix $\rho = \frac{1}{d}\mathbb{1}$ *if and only if* all subsequent experiments on $\mathcal{S}$ do not involve the reference system. If the reference system *is* involved, then this "reduced density matrix" is not sufficient to predict future experiments. This situation lies outside the scope of [standard] quantum state tomography.

For another example, consider the following qubit preparation procedure: "If the number of seconds elapsed since noon is even, prepare $|1\rangle\langle 1|$; if it is odd, prepare $|0\rangle\langle 0|$." Again, the system may be described by $\rho = \frac{1}{2}\mathbb{1}$ – *if and only if* future experiments are guaranteed to be performed at times that are uniformly distributed between "even" and "odd" time blocks. (This is quite a strong restriction; most experimental designs have some accidental periodicity built in). Otherwise, predicting measurements as well as possible requires additional information about time that is not incorporated in the single density matrix ρ.

So, when *is* the ansatz that underlies state tomography valid? It clearly applies when many (N) instantiations of the system, prepared either in series or in parallel according to the specified preparation procedure, *are* in some state $\rho^{\otimes N}$. If ρ is pure, then this is all the condition we need. If ρ is mixed, then we must address a foundational issue: mixed

states are (as previously noted) intrinsically *epistemic* (describing knowledge rather than reality) and are therefore subjective. Specifically, different observers may have different knowledge available, and may justifiably assign different mixed states. In this particular case, an observer who initially assigns $\rho^{\otimes N}$ might get access to an ancillary system that provides additional information, and by making use of that information assign a new state to the N copies that is not of the form $\rho^{\otimes N}$. This rather complicated-sounding situation merely generalizes the example given above, in which each sample was maximally entangled with an ancilla. Clearly, to an observer with access to these "purifications" of the samples, the samples would not be well-described by $\rho^{\otimes N}$! Both this situation and its generalized counterpart can be ruled out by a single condition: we require not only that a well-informed observer assign the state $\rho^{\otimes N}$, but also that *no other quantum systems correlated with the samples are, or will become, available for the future measurements we are trying to predict.*

These conditions are so restrictive that they never apply, strictly. The simplest and most pervasive cause is that preparation procedures *fluctuate* over time. If the fluctuations were strictly uncorrelated in time (i.e., white noise), this would be fine – the result of the fluctuations would be merely to add entropy to ρ. But this is never the case in real experiments; noise processes don't have perfectly flat spectral density, and this means that the time order of samples is relevant. This breaks the $\rho^{\otimes N}$ form.

Fortunately, this is exactly the sort of thing that cryptographers worry about, and a good theory for dealing with it has been developed in quantum cryptography under the name of *DeFinetti theorems*. The theory lies outside the scope of this report, but a simple summary of its implications for tomography is this: *almost all nontrivial time-dependence in the preparation procedure can be annihilated by randomly permuting the samples.* As long as the permutation is truly random, or at least unknown to the experimenters performing subsequent measurements, this operation (whether active or passive) "twirls" arbitrary states into statistical mixtures of $\rho^{\otimes N}$ (for various ρ), or DeFinetti states. Once this is done, tomography is justified. A more comprehensive discussion of this approach can be found in Ref. [2], which was partially supported by this LDRD project.

2.3.2 Basic state tomography

In this discussion, we assume that tomography is justified – i.e., we have N samples of an unknown ρ, so their state is $\rho^{\otimes N}$, but ρ is unknown. The goal of state tomography, then, is to find out ρ.

This is extremely simple in principle. Since ρ is a $d \times d$ Hermitian matrix, it is an element of $\mathcal{B}(\mathcal{H})$ (the d^2-dimensional Hilbert-Schmidt space), and can be written as $|\rho\rangle\!\rangle$. This can be uniquely specified by knowing d^2 distinct components

$$c_k = \langle\!\langle M_k|\rho\rangle\!\rangle, \tag{2.1}$$

where the $\{\langle\!\langle M_k|\}$ are linearly independent elements of $\mathcal{B}(\mathcal{H})$. Since we already know that

$$\langle\!\langle \mathbb{1}|\rho\rangle\!\rangle = \mathrm{Tr}\rho = 1, \tag{2.2}$$

only $d^2 - 1$ independent coefficients are required to specify ρ.

Perhaps the simplest example is single-qubit Pauli tomography. In this exercise, we observe that the normalized Pauli operators

$$\frac{1}{\sqrt{2}}\{X, Y, Z, \mathbb{1}\}$$

form a complete orthonormal basis for the single-qubit $\mathcal{B}(\mathcal{H})$, and therefore ρ can be identified by measuring or estimating the expectation values of the Paulis,

$$\begin{aligned} \rho &\Leftrightarrow \{\langle X\rangle, \langle Y\rangle, \langle Z\rangle\} && (2.3)\\ &= \{\langle\langle X|\rho\rangle\rangle, \langle\langle Y|\rho\rangle\rangle, \langle\langle Z|\rho\rangle\rangle\}. && (2.4) \end{aligned}$$

However, these expectation values cannot be measured directly. They are *averages* of a measurement that always yields ± 1. So instead the tomographer must estimate them, by:

1. Measuring each of the 3 Pauli operators on many distinct samples,
2. Collating the measurement results, and counting how many times each outcome is observed,
3. Estimating each of the three expectation values from these data.

The simplest and most obvious way to estimate (e.g.) $\langle Z\rangle$ is by the *empirical* expectation value:

$$\widehat{\langle Z\rangle} = \frac{n_+ - n_-}{n_+ + n_-}, \tag{2.5}$$

where $n_\pm$ is the number of times that $Z = \pm 1$ was observed (respectively).

2.3.3 Linear inversion tomography

The example given above provides a complete scheme for qubit Pauli state tomography, but it is rather specific to (1) single qubit states, and (2) Pauli measurements. For a more general picture, consider that ρ can be specified using *any* spanning set of expectation values corresponding to any convenient complete basis of operators. And, while the Paulis are elegant, they are not the operators whose expectation values are most directly inferrable. The most directly observed expectation values are the Born's Rule probabilities of measurement outcomes:

$$p_k = \langle\langle E_k|\rho\rangle\rangle, \tag{2.6}$$

where $E_k = |\phi_k\rangle\langle\phi_k|$, and the $\{\phi_k\}$ are the six eigenvectors of $\{X, Y, Z\}$ (e.g. $|+\rangle, |-\rangle, |0\rangle, |1\rangle$, etc.)

The set $\{E_k\}$ is indeed a basis (overcomplete) for $\mathcal{B}(\mathcal{H})$, and so we can represent

$$\rho \Leftrightarrow \{p_k\}$$

just as easily as we could represent ρ using the expectation values of Paulis. Furthermore, it is even easier to obtain an estimate of p_k from data; the obvious choice is

$$\widehat{p}_k = \frac{n_k}{N_k} \tag{2.7}$$

where n_k is the number of times E_k was observed, and N_k is the number of times it *could* have been observed – i.e., the number of times the measurement containing E_k was performed. Note that N_k is not generally the total number of samples. For example, in qubit Pauli tomography, three different measurements have to be made, so N samples will typically be divided into three subsamples of size $N_k = N/3$.

A convenient way to simplify this situation is to treat *all* the $\{E_k\}$ as outcomes of a single POVM (positive, operator-valued measure). This single POVM corresponds to the following measurement procedure: first, draw a classical random random number and use it to decide which observable to measure; then, perform that measurement and record the outcome. Mathematically, this means scaling each effect (E_k) down by a factor proportional to N_k/N. In the standard view, this is the fraction of the N samples on which the measurement containing E_k was measured. In the "single POVM" view, it is the probability of performing that measurement. These views are statistically equivalent, and the single-POVM view is simpler; it permits us to simply estimate

$$\widehat{p}_k = \frac{n_k}{N}, \tag{2.8}$$

where N is the total number of samples.

Finally, given that the $\{p_k\}$ specify ρ, and that the $\{\widehat{p}_k\}$ are a reasonable estimate of them, we need an algorithm to construct an estimate $\widehat{\rho}$ of the state. The most straightforward process is to solve the following set of linear equations:

$$\langle\langle E_k|\rho\rangle\rangle = \widehat{p}_k. \tag{2.9}$$

This can be done exactly if the effects $\{E_k\}$ are linearly independent, in which case it is said that the performed measurements are "complete", and form a "quorum". This is the motivating case for linear-inversion tomography.

If they are not linearly independent – i.e., there are more than d^2 of them – then the solution is overconstrained, and it is said that the performed measurements are "overcomplete". In this case, we are forced to acknowledge that the $\{\widehat{p}_k\}$ are only estimates – i.e., that in general $p_k \neq \widehat{p}_k$. We will return to this important fact in the next subsection. However, linear inversion tomography is simple precisely because it ignores this fact, and the simplest way to do so for overcomplete measurements is to (by fiat) choose the unweighted least-squares solution to these linear equations.

This solution can be found using simple linear algebra. To do so, we view the probability distribution of the observed effects as a vector in a vector space of probabilities, whose basis vectors we denote $|i\rangle$:

$$\{p_k\} \longrightarrow |p\rangle \equiv \sum_i p_i \, |i\rangle. \tag{2.10}$$

Now, Born's Rule is simply

$$p_k = \langle\langle E_k|\rho\rangle\rangle, \tag{2.11}$$

and by combining these two equations, we get

$$\begin{aligned} |p\rangle &= \sum_i |i\rangle\langle\langle E_i|\rho\rangle\rangle & (2.12)\\ &= M\,|\rho\rangle\rangle & (2.13) \end{aligned}$$

where $M \equiv \sum_i |i\rangle\langle\langle E_i|$. Now, it seems painfully obvious that to find $\widehat{\rho}$ from the $\{\widehat{p}_k\}$, we need only compute

$$|\widehat{\rho}\rangle\rangle = M^{-1}\,|\widehat{p}\rangle\,. \tag{2.14}$$

For overcomplete measurements, M is not square and therefore not invertible, so in order to find the least-squares solution we use the Moore-Penrose pseudo-inverse,

$$M^{-1} \equiv \left(M^\dagger M\right)^{-1} M^\dagger. \tag{2.15}$$

This defines the linear inversion estimate of ρ, the simplest of tomographic estimators.

2.3.4 Least-squares and maximum likelihood tomography

There are a number of problems with linear inversion tomography. All of them can be traced to a simple fact: linear inversion tomography is based on the assumption that $p_k = \widehat{p}_k$, which is patently not true. The $\{\widehat{p}_k\}$ are merely *estimates* of the true probabilities. For example, consider a two-outcome measurement – which we may conceptualize as a flip of a biased coin – and suppose it is performed just once. Assume, without loss of generality, that the "coin" comes up heads. The linear inversion estimate would be $\widehat{p}_{\text{heads}} = 1$. This is (a) probably not the correct value, and (b) rather an extreme conclusion to draw, as it implies that the probability of tails is zero! Thus, linear inversion is based upon a fallacy, and should be taken with a large grain of salt.

What conclusions *should* we draw from the data? This is a very deep question, encompassing most of statistical inference, and we do not even attempt to survey it here. Instead, we merely observe that if the event described by E_k is observed n_k times out of N trials, then *probably* p_k is *approximately* equal to n_k/N. Why? Because we can compute the probability of observing n_k events, and we find that if p_k were very different from n_k/N, then this probability is very low. Even this rather vague statement requires a bit of clarification, because we may (and often do) find that the probability of observing n_k events is *always*

rather small, no matter what p_k is! But if we look at the *relative* probability of observing what we actually observed, we find that it is much higher if $p_k \approx n_k/N$ than otherwise.

This is an intuitive statement of the *likelihood principle*, which (as generally interpreted) has two parts:

1. A theory (e.g., about p_k) is plausible in proportion to the probability it predicts for the data that were actually observed.

2. The *likelihood function*, $\mathcal{L}(p) \equiv Pr(\text{data}_{\text{actually observed}}|p)$, contains everything about the data that are relevant to inferring or estimating $\vec{p}$.

These statements, while not provable, are sensible in many circumstances and underly most forms of statistical inference (albeit with some major caveats and additions, especially in Bayesian inference). They arguably motivate *maximum likelihood estimation* (MLE), which (whatever else might be said about it) is simple, intuitive, and widely applicable. The MLE estimator is simply:

$$\widehat{p} = \text{argmax}\left(\mathcal{L}(p)\right). \tag{2.16}$$

In other words, we choose the value of p that maximizes the probability of the actually-observed data.

Interestingly, if MLE is applied to a coin-flip experiment (or to the data obtained from a quantum 2-outcome measurement, in isolation), then it leads to precisely the same conclusion as linear inversion:

$$\widehat{p}_{\text{MLE}} = \frac{n}{N}. \tag{2.17}$$

So, at the simplest level, it justifies this choice. But when multiple experiments are combined in a way that introduces constraints, MLE provides more flexibility. In quantum tomography, this provides an elegant way to deal with overcomplete measurements. Overcompleteness means that no $\widehat{\rho}$ can fit *all* the observed frequencies exactly. This is a constraint on the possible values of the $\{p_k\}$, and MLE handles it with aplomb. Unfortunately, except in special cases, there is no closed form solution to Eq. 2.16. It must be found by numerical optimization (fortunately, this is at least efficient; the likelihood function for quantum state tomography has convex level sets, so it can be maximized in poly(d) time.)

MLE solves – or at least addresses – another problem that bedevils linear inversion: *negativity*. Because the observed n_k are not generally given by Np_k, but are instead binomially distributed around it, the data exhibit *finite sample fluctuations* – and therefore so does the linear inversion estimate. If the true state lies close enough to the surface of the convex set of states in $\mathcal{B}(\mathcal{H})$, these fluctuations can (and in some cases almost certainly will) cause the linear inversion $\widehat{\rho}$ to fall outside the set of positive density matrices! When this occurs, $\widehat{\rho}$ has negative eigenvalues, which in turn correspond to negative probabilities for some measurement outcome. This is a rather serious conceptual problem. Moreover, it wreaks havoc with the calculation of entropic quantities (which tend to involve logarithms of the eigenvalues), and casts serious doubt on the reliability of the estimator in general.

MLE embodies a simple fix for negative estimates. If the domain of $\mathcal{L}(\rho)$ is taken to be all of $\mathcal{B}(\mathcal{H})$ (subject to the trace=1 constraint), then it yields very similar results to linear inversion (see "least squares" discussion below). But by simply restricting the domain of $\mathcal{L}(\rho)$ to only include *valid* density matrices – i.e., those with positive eigenvalues – we can ensure that argmax($\mathcal{L}(\rho)$) is always a valid state! This removes any possibility of even *approximating* $\widehat{\rho}_{\mathrm{MLE}}$ analytically, since the positivity constraint is complicated, but numerical methods still work.

MLE thus solves some of the problems with linear inversion; it is statistically well-motivated, it deals gracefully with overcomplete data, and it can be constrained to report a valid state. However, it retains a variety of pathologies, many of them stemming from the "zero eigenvalue problem". This was first presented in Ref. [3], which remains perhaps the best reference. Stated concisely, it is the fact that whenever linear inversion returns a $\widehat{\rho}$ with negative eigenvalues, MLE will produce a $\widehat{\rho}$ with at least one *zero* eigenvalue. This corresponds to a zero probability for some measurement, which is difficult to justify from finite data. It also presents problems for error bars, since it's difficult to argue that a probability is given by $p = 0 \pm \Delta p$. This issue is partly resolved in [4], where it is argued that the correct interpretation of the likelihood principle is *not* as a justification of MLE, but rather as motivating *region estimators* containing all high-likelihood states. However, this is also outside the scope of this report.

We conclude this section by pointing out that, although MLE generally requires numerical optimization, it can be analyzed rather nicely in the $N \to \infty$ limit of very much data. The central limit theorem takes hold in this limit, and the likelihood function converges to a Gaussian form. In this limit, if the positivity constraint is ignored (or if it is irrelevant because the maximum of the unconstrained $\mathcal{L}(\rho)$ happens to be non-negative), then MLE reduces to a *weighted* least-squares problem that can be solved using linear algebra. If the positivity constraint ($\rho \geq 0$) *is* relevant, then analytic techniques break down (the corresponding unweighted least-squares problem does have a semi-analytic solution [5], but weight factors break it), but the numerical maximization is much easier than in the general MLE case.

It is well worth noting that the weighted least-squares approximation to MLE is not an arbitrary one. In fact, it is precisely *minimum* χ^2 *estimation*, because as $N \to \infty$, the likelihood function approaches

$$\mathcal{L}(\rho) \to e^{-\chi^2(\rho)}. \tag{2.18}$$

The χ^2 function, in this context, is a function of (a) the probabilities p_k implied by ρ and the observed frequencies $f_k = n_k/N$, and it is given by

$$\chi^2(\rho) = \chi^2(\{p_k\}, \{f_k\}) = \sum_k \frac{(p_k - f_k)^2}{p_k}. \tag{2.19}$$

This equivalence is extremely useful for a variety of problems when $N \gg 1$.

2.4 Quantum process tomography

Quantum information processing relies far less heavily on precise preparations of quantum states than it does on precisely implemented quantum *gates*. In fact, state preparation can easily be described as just a particular kind of logic gate – one that, instead of transforming an existing d-dimensional state into another one, transforms a trivial 1-dimensional state into a particular d-dimensional one.

For this reason, quantum *process* tomography is far more important than state tomography. Fortunately – both for the relevance of state tomography, and for the theory of process tomography – the two are close cousins. Virtually every general feature of state tomography maps to an equivalent feature of process tomography. In fact, even the precise details are in close correspondence, because of the *Choi-Jamiolkowski isomorphism*.

2.4.1 The Choi-Jamiolkowski isomorphism

The CJ isomorphism states that the set of (CPTP) quantum processes on a d-dimensional system is in 1:1 correspondence with a subset of the quantum states on a d^2-dimensional system. The isomorphism is explicit: for each CPTP process G acting on $\mathcal{B}(\mathcal{H})$, there exists a corresponding *Jamiolkowski state* $\rho_J(G)$ on $\mathcal{H} \otimes \mathcal{H}$. (In fact, the CJ isomorphism can also be seen as a correspondence between $\mathcal{B}(\mathcal{H})$ and $\mathcal{H} \otimes \mathcal{H}$, which incidentally can be applied to densities over these spaces). It is given by:

$$\rho_J(G) = (G \otimes \mathbb{1})[|\Psi\rangle\langle\Psi|], \tag{2.20}$$

where $|\Psi\rangle$ is the symmetric maximally entangled state on two copies,

$$|\Psi\rangle = \frac{1}{\sqrt{d}} \sum_{j=0}^{d-1} |j\rangle \otimes |j\rangle. \tag{2.21}$$

To see that this is indeed an isomorphism, represent G in Choi form:

$$G[\rho] = \sum_{ij} \chi_{ij} B_i \rho B_j, \tag{2.22}$$

and choose the basis $\{B_i\}_{i=1\ldots d^2}$ to be the matrix units $\{|\alpha\rangle\langle\beta|\}_{\alpha,\beta=1\ldots d}$, so that

$$\begin{aligned} G[\rho] &= \sum_{\alpha,\beta,\gamma,\delta} \chi_{(\alpha,\beta),(\gamma,\delta)} \, |\alpha\rangle\langle\beta| \, \rho \, |\gamma\rangle\langle\delta| & (2.23) \\ &= \sum_{\alpha,\beta,\gamma,\delta} \chi_{(\alpha,\beta),(\gamma,\delta)} \rho_{\beta,\gamma} \, |\alpha\rangle\langle\delta|. & (2.24) \end{aligned}$$

Now, observe that

$$|\Psi\rangle\langle\Psi| = \frac{1}{d} \sum_{i,j=0}^{d-1} |i\rangle\langle j| \otimes |i\rangle\langle j|,$$

and so the Jamiolkowski state is

$$
\begin{aligned}
(G \otimes \mathbb{1})[|\Psi\rangle\langle\Psi|] &= \frac{1}{d} \sum_{i,j,\alpha,\beta,\gamma,\delta} \chi_{(\alpha,\beta),(\gamma,\delta)} \delta_{i,\beta}\delta_{j,\gamma} |\alpha\rangle\langle\delta| \otimes |i\rangle\langle j| && (2.25)\\
&= \frac{1}{d} \sum_{\alpha,\beta,\gamma,\delta} \chi_{(\alpha,\beta),(\gamma,\delta)} |\alpha\rangle\langle\delta| \otimes |\beta\rangle\langle\gamma| && (2.26)\\
&= \frac{1}{d} \sum_{\alpha,\beta,\gamma,\delta} \chi_{(\alpha,\beta),(\gamma,\delta)} |\alpha\beta\rangle\langle\delta\gamma| && (2.27)\\
&\simeq \chi. && (2.28)
\end{aligned}
$$

In other words, the elements of ρ_J and the χ matrix that defines G are identical up to a factor of $1/d$, and so each is a description of the other.

The CJ isomorphism is responsible for the general similarities between state and process tomography. For example, the CP constraint on G is isomorphic to the positivity constraint $\rho_J \geq 0$. Processes are slightly more constrained, however; the equivalent to the trace constraint $\mathrm{Tr}\rho = 1$ is the more powerful TP constraint $G^\dagger[\mathbb{1}] = \mathbb{1}$. This is a linear constraint on not one but d^2 matrix elements of G (or its corresponding χ), so whereas a general state on $\mathcal{H} \otimes \mathcal{H}$ has $d^4 - 1$ free parameters, a CPTP map has only $d^4 - d^2$ parameters.

But it is worth mentioning that the CJ isomorphism can be used to establish a *direct* correspondence between state and process tomography, at least in principle, in what is known as *ancilla-assisted process tomography*. It is very simple: to do tomography on G, we simply prepare a many samples of the maximally entangled state $|\Psi\rangle$ between the system and an ancilla, apply G to them (to prepare ρ_J) and then do tomography on the resulting bipartite states. In practice, this is extraordinarily impractical.

2.4.2 Mathematical formalism of process tomography

In practice, process tomography is performed not by applying G to a single entangled state, but by preparing a variety of states $\{\rho_j\}$, applying G to them, and then doing state tomography on the resulting output states $\rho'_j = G[\rho_j]$. In principle, this provides a set of linear equations

$$G\,|\rho_j\rangle\!\rangle = |\rho'_j\rangle\!\rangle, \tag{2.29}$$

and these can be solved if the $\{\rho_j\}$ form a complete (spanning) basis for $\mathcal{B}(\mathcal{H})$. However, while this is conceptually simple, it is more useful to break the state tomography down into individual experiments. Doing state tomography on ρ'_j means measuring many samples, and thus estimating various probabilities of the form

$$
\begin{aligned}
p_{k|j} = \langle\!\langle E_k|\rho'_j\rangle\!\rangle & && (2.30)\\
&= \langle\!\langle E_k|\,G\,|\rho_j\rangle\!\rangle && (2.31)\\
&= \mathrm{Tr}\left[|\rho_j\rangle\!\rangle\langle\!\langle E_k|\,G\right]. && (2.32)
\end{aligned}
$$

Stated this way, the problem looks very similar to state tomography – the final equation is simply an inner product in the vector space of *superoperators*, which is $O(d^4)$-dimensional. The only difference is that the "effects" in process tomography are not Hermitian (or symmetric, technically, as $\mathcal{B}(\mathcal{H})$ is a real space), being of the form $|\rho_j\rangle\!\rangle\langle\!\langle E_k|$ rather than $|\phi_k\rangle\langle\phi_k|$ as in state tomography.

In any case, this form permits every technique used for state tomography to be brought to bear on process tomography directly. Linear inversion, least squares, MLE, and other estimators can be applied. Theories that suggest what POVMs are most effective for extracting information about unknown ρ can be repurposed to construct experimental designs (i.e., $\{|\rho_j\rangle\!\rangle\langle\!\langle E_k|\}$) that are most effective at extracting information about unknown processes.

However, there is one **very** important difference between processes and states. For states, experiments that probe $\mathrm{Tr}[E_k\rho]$ are the only ones possible. But processes, unlike states, can be *composed*. If we can apply G, then we can apply G twice in a row – which means applying G^2. Or, if we desire, G^{99}. And this means we can estimate probabilities like

$$Pr = \langle\!\langle E_k| G^{99} |\rho_j\rangle\!\rangle,$$

which have no equivalent in state tomography (because ρ^2 or ρ^{99} is a meaningless object). This capability is extremely important for gate-set tomography.

Chapter 3

Gate-set tomography and the LGST algorithm

3.1 Motivation

We developed gate-set tomography (GST) to address one serious problem with standard state and process tomography: **characterizing an unknown state or process using tomography requires the use of known measurement, states, and/or processes.** Moreover, if the tomographer is *wrong* about what these reference objects are, then the tomographic estimate will be wrong too (and often undetectably so).

In the case of state tomography, the reference objects on which we rely are the effect operators $\{E_k\}$ describing measurement outcomes. It is assumed that we *know* what operator E_k corresponds to an observation labeled "k", and therefore how to invert the observed (or estimated, to be precise) probabilities $p_k = \langle\langle E_k|\rho\rangle\rangle$.

In the case of process tomography, we rely on prior knowledge both of the measurement effects $\{E_k\}$ and of the input states $\{\rho_j\}$. However, it's only really necessary to know what we are measuring; given this knowledge, state tomography can be used to characterize the $\{\rho_j\}$.

In practice, very few qubit technologies admit multiple distinct measurements natively. Instead, there is usually just one measurement that can be performed – often with only two outcomes – and other measurements are implemented by (1) applying some unitary gate U, and (2) performing the fixed "native" measurement. This means that when we rely on multiple precalibrated measurements, we are actually relying primarily on certain knowledge of the gates or operations used to implement those measurements.

This makes standard process tomography perilously circular. It is, after all, the tool used to characterize gates... but it relies on precalibrated gates, which must have been calibrated using process tomography. The entire framework of quantum gate characterization begins to look like an infinite sequence of turtles, all the way down.

Fortunately, it is not. This should come as no surprise to experimentalists, who have developed various methods to characterize gates *without* tomography. It is possible, both in

practice and theory, to characterize quantum gates self-consistently and *without* any prior calibration. This is what GST does.

3.2 Foundations

GST begins from the premise that *all* operations on the system of interest (generally a single qubit) are unknown and untrusted. We treat it as a black box, purportedly containing a qubit or other quantum system, with a very limited control interface. The following controls, aka "buttons", are available:

1. A button that initializes the system in an unknown state ρ.
2. A button that performs an unknown 2-outcome measurement $\mathcal{M} = \{E, \mathbb{1} - E\}$, and causes one of two labeled lights to flash depending on which outcome was observed.
3. Several buttons that apply unknown quantum operations $\{G_1, G_2 \ldots\}$.

Collectively, these operations (including state preparation and measurement, aka *SPAM*) form the *gate set* that is available to control the system – and which, in practice, define and describe its QIP capabilities.

The one major assumption made in GST is that each of these buttons can be associated to a quantum operation – a density matrix, POVM, or CPTP map, respectively. This seems rather uncontroversial (and it is certainly no stronger than the assumptions *always* made in standard tomography!), but experiments have shown that it is surprisingly unreliable. Experimental quantum hardware, at least at the current stage of maturity, exhibits detectable levels of non-Markovian noise (including but not limited to systematic drift) which violate this paradigm. For the purposes of this report, however, we will perservere with the assumption of Markovian quantum operations.

For convenience, we typically also make an assumption about the system's Hilbert space dimension d. This fundamental parameter determines the size of the vectors and matrices used to describe the elements of the gate set. However, it does *not* need to be assumed; it is possible to determine the system's effective Hilbert-space dimension (or, more precisely, the dimension of the Hilbert-Schmidt state space that it explores) dynamically from the data, using model selection techniques if necessary. We have not implemented this technique in experiments yet, so it remains outside the scope of this report. In this document, unless otherwise specified, we generally set $d = 2$.

3.2.1 Gate sets

A gate set is a tuple: $\{\rho, E, G_1 \ldots G_K\}$, where ρ and E are vectors in Hilbert-Schmidt space, and the G_k are operators on it. Gatesets are predictive *theories* about how a black-box

quantum device operates. Collectively, the set of all possible gatesets is a *model* for the black box.

We interact with and observe a black box by performing *experiments* on it. In GST, an experiment comprises a sequence of operations in which the first is always initialization, the last is always measurement, and between them a sequence of zero or more G_k is applied. Experiments are thus labeled by quantum circuits, or *gate sequences*, $S = \{s_1, s_2 \dots s_L\}$. Unless otherwise specified, L denotes the length of a gate sequence. For each experiment, E or $\mathbb{1} - E$ is observed with respective probability p_S or $1 - p_S$, where

$$p_S = \langle\langle E | \, G_{s_L} G_{s_{L-1}} \dots G_{s_2} G_{s_1} \, | \rho \rangle\rangle. \tag{3.1}$$

Each such probability can, of course, be estimated to any desired precision by repeating the corresponding experiment many times.

The basic principle behind GST is that each experiment whose probability is known provides some information about the gate set. More precisely, it constrains a single scalar parameter. A gate set has at most $2(d^2-1) + K(d^4 - d^2) = Kd^4 - (K-2)d^2 - 2$ parameters (just from counting matrix elements and constraints), so in principle we should be able to infer the entire gate set from roughly that many empirically estimated probabilities for distinct experiments. And, indeed, this is mostly correct. To perform GST, we choose a set of experiments $\{S_i\}$, repeat each one enough times to estimate the corresponding p_{S_i}, and reconstruct the entire gate set from these probabilities.

3.2.2 The gauge

The one main caveat is that *no* set of experimentally observable probabilities is sufficient to nail down the *entire* gate set. This is because gate sets have a *gauge*. That is, there are distinct gate sets – entire families, in fact – that are categorically indistinguishable using any experiment. Such families actually all represent the same physical device. The gauge freedom merely represents different ways of representing the same gate set. Some of the parameters in the gate set correspond directly to gauge freedoms, and since they have no physical consequence or meaning, they of course cannot be inferred from empirical observations. This *also* means, fortunately, that the tomographer is free to choose any convenient gauge.

It is not easy to divide gate set parameters into gauge and gauge-invariant sets (in fact, we flatly don't know how to do this). However, it *is* easy to state the full group of gauge transformations. Consider a gate set $\{|\rho\rangle\rangle, \langle\langle E| \, , \{G_k\}\}$ in which the G_k are $m \times m$ matrices (generally $m = d^2$, but other state spaces are possible). Let T be a real invertible $m \times m$ matrix, and consider the (generally different) gate set

$$\{|\rho'\rangle\rangle, \langle\langle E'| \, , \{G'_k\}\} = \{T \, |\rho\rangle\rangle, \langle\langle E| \, T^{-1}, \{T G_k T^{-1}\}\}. \tag{3.2}$$

This transformation leaves every observable probability unchanged:

$$\begin{aligned} p'_S &= \langle\langle E'| G'_{s_L} G'_{s_{L-1}} \dots G'_{s_2} G'_{s_1} |\rho\rangle\rangle & (3.3)\\ &= \langle\langle E| T^{-1} T G_{s_L} T^{-1} T G_{s_{L-1}} T^{-1} \dots T G_{s_2} T^{-1} T G_{s_1} T^{-1} T |\rho\rangle\rangle & (3.4)\\ &= \langle\langle E| G_{s_L} G_{s_{L-1}} \dots G_{s_2} G_{s_1} |\rho\rangle\rangle & (3.5)\\ &= p_S. & (3.6) \end{aligned}$$

This is an action of $GL(m)$, a Lie group with m^2 parameters, so there are (in general) m^2 gauge parameters in the gate set.

Since gauge transformations have no observable effect, there is no reason not to choose the most convenient gauge. The point of GST, of course, is to report a gate set that accurately describes the experimental device and predicts its future behavior, and any gauge would (by definition) serve equally well for this purpose. But in almost all cases, the experimentalist or tomographer has some pre-existing expectation of what the device is *supposed* to do, which is represented by a *target gate set*. Target gate sets are usually idealized (e.g., ρ and E are projectors, and the G_k are unitary), but one of the critical questions GST should answer is "How close is the device to the target?" followed closely by "What are the most significant discrepancies between the estimate and the target?" Gauge transformations on the estimate can have huge effects on the answers.

For example, consider the target gate set $\mathcal{T} = \{\rho = |0\rangle\langle 0|, E = |1\rangle\langle 1|, G_1 = G_X\}$, where G_X is the superoperator corresponding to $\rho \to \sigma_x \rho \sigma_x$. Any unitary change of basis is a valid gauge transformation, so the targets are gauge-equivalent to $\mathcal{T}' = \{\rho' = |+\rangle\langle +|, E = |-\rangle\langle -|, G_1 = G_Z\}$, where $G_Z[\rho] = \sigma_z \rho \sigma_z$. If an experimentalist *intended* to produce $\mathcal{T}$, but the tomographer reports the [equivalent] gateset $\mathcal{T}'$, then by all the obvious metrics the device is wildly out of spec. Of course, this is merely a misunderstanding resulting from different choices of gauge. To avoid such communication failures, it is necessary that all parties agree on a shared gauge.

Unfortunately, we do not know of any suitable way to specify "a gauge". Instead, in GST, we allow the target gateset to define the gauge. Given a GST estimate in an arbitrary gauge, we *optimize* the gauge to the targets by using gauge transformations to make the estimate as similar as possible to the target. More precisely, we attempt to find the gauge transformation that minimizes the total Euclidean distance between all the matrix/vector elements of (a) the gauge-transformed estimate, and (b) the target. Numerical experiments show that a local gradient optimization almost always finds the global optimum, provided that the gates are not completely different from the targets (so that the initial guess is reasonably good).

Gate sets should obey certain *positivity constraints*. In order to ensure that negative probabilities are not predicted, we generally demand that ρ and E are positive semidefinite, that $\mathrm{Tr}\rho = 1$ and $E \leq \mathbb{1}$, and that each of the G_k is completely positive and trace-preserving. A gateset satisfying these conditions is *explicitly CPTP*. But gauge transformations do not, in general, respect positivity. Their precise effect on positivity depends on what gateset

is being transformed. At one extreme, if each element of the gateset is strictly positive (i.e., does not lie on the boundary of its positivity constraint), then every sufficiently small gauge transformation (i.e., $T = \mathbb{1} + \epsilon Q$ for any matrix Q and ϵ sufficiently small) preserves positivity. But if the gateset is maximally extreme – i.e., ρ and E are rank-1, and each G_k is unitary – then the *only* gauge transformations that preserve positivity are basis changes, i.e., those corresponding to T such that $T[\rho] = U\rho U^\dagger$ for some unitary U.

In general, the interaction between gauge transformations and positivity is complex and messy. Given an explicitly CPTP gateset, the subgroup of unitary gauge transformations will always preserve its positivity, but other directions in the Lie algebra generating $GL(m)$ are truncated – small transformations may preserve positivity, but sufficiently large ones produce a gateset that is not explicitly CPTP. The harder problem is, given a gateset that is *not* explicitly CPTP, to determine whether it is gauge-equivalent to one that is. Such a gateset is *implicitly CPTP*, and at this time we know of no efficient and reliable algorithm for testing for this condition.

3.3 Linear inversion GST (LGST)

The GST framework was invented and fully fleshed out at Sandia National Labs, but a significant subset of ideas comprising GST can be traced back to work at IBM Research circa 2012 (see Ref. [6]). The IBM researchers realized the vulnerability of standard tomography to calibration errors, and observed that observable probabilities can always be written as expressions of the form $p_S = \langle\langle E| G \dots |\rho\rangle\rangle$. In order to fit data, they proposed a brute-force maximum likelihood approach. Given a gateset $\mathcal{G} = \{\rho, E, \{G_k\}\}$, and a dataset described by $\{n_s, N_s\}$ (where experiment S was performed N_S times, and "E" was observed n_S times), the likelihood is easy to compute:

$$\mathcal{L}(\mathcal{G}) = \prod_S p_S^{n_S}(1 - p_S)^{N_S - n_S}. \tag{3.7}$$

The IBM group proposed (and implemented) an exhaustive set of experiments corresponding to *all* gate sequences of 3 or fewer gates. To analyze this data, they maximized the likelihood function numerically using local optimization algorithms.

As they discovered, the likelihood function is not nice. It is strongly not quasi-convex (i.e., its level sets are not convex), and therefore cannot be reliably or efficiently maximized using standard methods. This stems largely from two causes: (1) the probabilities for the experimental observations are not linear functions of the parameters (in contrast to the situation in state or process tomography), and (2) the existence of the gauge group produces a very significant degeneracy in the maximum of the likelihood (instead of a unimodal "hill", a plot of the likelihood would resemble an assortment of winding "ridges", whose crest is perfectly level and traces out an orbit of the gauge group). As a result, the algorithm proposed by the IBM group relies rather critically on a good initial guess for the gates. Without a good initial guess, there is significant risk that it will end up in a local maximum.

We solved this problem by introducing *linear inversion GST*, or *LGST*. In the taxonomy of standard tomography, linear inversion is the oldest and most primitive method, largely replaced by MLE. For self-calibrating methods (including GST and IBM's "overkill" tomography), the relationship is partly reversed; LGST solved significant problems associated with brute-force maximum likelihood.

LGST is nontrivial precisely because experiments' probabilities are *not* linear in the gateset. The simplest experiment is the SPAM experiment:

$$p_{\text{SPAM}} = \langle\langle E|\rho\rangle\rangle,$$

and even this experiment is quadratic in the gate-set parameters. Fully characterizing the gateset requires far more information, which requires longer circuits, and ensures more nonlinearity.

To overcome this issue, LGST demands a specific set of experiments. The structure of this set guarantees a closed-form solution to the tomographic inversion problem – and, while the solution is not *linear*, it requires only linear algebra. LGST actually resembles standard process tomography rather closely, except that where process tomography relies on known probe states and measurements, LGST makes do with unknown ones.

3.3.1 The LGST protocol

We begin by choosing a set of $n = d^2$ *fiducial sequences* $\{F_i\}$, which are short gate circuits whose role is to transform the system's initial state ρ into a quorum of different input states *and* to transform the fixed measurement effect E into a quorum of distinct effects E_i:

$$|\rho_j\rangle\rangle = F_j\,|\rho\rangle\rangle, \tag{3.8}$$
$$\langle\langle E_i| = \langle\langle E|\,F_i. \tag{3.9}$$

Both sets should span $\mathcal{B}(\mathcal{H})$. Since we do not know ρ or the G_k in advance, we cannot guarantee that this will happen. Fortunately, it is possible to *test* for this condition after seeing the data, so if incomplete fiducials are chosen by accident, we can change them and redo the experiment.

Now, so equipped, we perform each of the $n^2(K+1)$ experiments

$$\langle\langle E|\,F_iG_1F_j\,|\rho\rangle\rangle \tag{3.10}$$
$$\dots \tag{3.11}$$
$$\langle\langle E|\,F_iG_KF_j\,|\rho\rangle\rangle \tag{3.12}$$
$$\langle\langle E|\,F_iF_j\,|\rho\rangle\rangle, \tag{3.13}$$

repeating each of them enough times to estimate the corresponding probability. We arrange these probabilities into matrices as

$$(\tilde{G}_k)_{ij} = \langle\langle E| F_i G_k F_j |\rho\rangle\rangle, \tag{3.14}$$

and also define

$$\tilde{\mathbb{1}}_{ij} = \langle\langle E| F_i F_j |\rho\rangle\rangle. \tag{3.15}$$

The LGST estimate of each gate is constructed as

$$\hat{G}_k = \tilde{\mathbb{1}}^{-1} \tilde{G}_k. \tag{3.16}$$

To obtain estimates of ρ and E, we perform additional experiments

$$\langle\langle E| F_i |\rho\rangle\rangle, \tag{3.17}$$

whose estimated probabilities are arranged into vectors

$$|\tilde{\rho}\rangle\rangle = \sum_i \langle\langle E| F_i |\rho\rangle\rangle |i\rangle\rangle, \tag{3.18}$$

$$\left\langle\langle \tilde{E}\right| = \sum_j \langle\langle E| F_j |\rho\rangle\rangle \langle\langle j|, \tag{3.19}$$

and then the LGST estimates are

$$|\hat{\rho}\rangle\rangle = \tilde{\mathbb{1}}^{-1} |\tilde{\rho}\rangle\rangle, \tag{3.20}$$

$$\left\langle\langle \hat{E}\right| = \left\langle\langle \tilde{E}\right|. \tag{3.21}$$

These LGST estimates are in an arbitrary gauge, so the final step before publishing an estimate of the gateset is to perform gauge optimization to maximize similarity to the target gates. (Note that the targets *only* appear at this stage – they only influence the gauge, not any physically meaningful aspect of the estimate).

3.3.2 Why LGST works

Now, let us show that this actually *works*. To do so, let us suppose for a moment that N (the number of repetitions of each experiment) is taken to infinity, and so the estimated probabilities are exactly correct. In that case, it is true that

$$\tilde{G}_k = \sum_{ij} \langle\langle E| F_i G_k F_j |\rho\rangle\rangle |i\rangle\rangle\langle\langle j| \tag{3.22}$$

$$= \sum_{ij} |i\rangle\rangle \langle\langle E| F_i G_k F_j |\rho\rangle\rangle \langle\langle j| \tag{3.23}$$

$$= A G_k B, \tag{3.24}$$

where the A and B matrices are defined as

$$A = \sum_i |i\rangle\!\rangle\langle\!\langle E_i| \tag{3.25}$$

$$B = \sum_j |\rho_j\rangle\!\rangle\langle\!\langle j| \tag{3.26}$$

using $|\rho_j\rangle\!\rangle \equiv F_j\,|\rho\rangle\!\rangle$ and $\langle\!\langle E_i| \equiv \langle\!\langle E|\,F_i$. Similarly, it is clear that

$$\tilde{\mathbb{1}} = AB \tag{3.27}$$

$$|\tilde{\rho}\rangle\!\rangle = A\,|\rho\rangle\!\rangle \tag{3.28}$$

$$\left\langle\!\langle \tilde{E}\right| = \langle\!\langle E|\,B, \tag{3.29}$$

and since $\tilde{Id}^{-1} = (AB)^{-1}$, we find that

$$\hat{G}_k = B^{-1}G_kB \tag{3.30}$$

$$|\hat{\rho}\rangle\!\rangle = B^{-1}\,|\rho\rangle\!\rangle \tag{3.31}$$

$$\left\langle\!\langle \tilde{E}\right| = \langle\!\langle E|\,B. \tag{3.32}$$

Thus, the estimate is gauge-equivalent to the true gateset (again, in the absence of finite-sample fluctuations). Moreover, we can even obtain a good guess as to the correct gauge by using the target gateset to define

$$B_{\text{target}} = \sum_j F_j^{(\text{target})} \left|\rho_j^{(\text{target})}\right\rangle\!\rangle\langle\!\langle j|, \tag{3.33}$$

and gauge-transforming the initial estimates by B_{target}.

3.3.3 Fiducial sequences: completeness and overcompleteness

LGST's key element is the use of fiducial sequences ($\{F_i\}$) to implement diverse input states and measurements. The resulting sets $\{\rho_j\}$ and $\{E_i\}$ are *supposed* to be informationally complete (IC), which means that they span $\mathcal{B}(\mathcal{H})$. Moreover, they should should be close to *uniformly informationally complete* (UIC), which means that there is no element of $\mathcal{B}(\mathcal{H})$ that is nearly orthogonal to all of them. Both the state- and effect-sets are acting as *frames* (non-orthogonal generalizations of orthogonal bases); $\tilde{\mathbb{1}}$ is the corresponding *frame operator*, and the multiplication by $\tilde{\mathbb{1}}^{-1}$ undoes the distortion caused by non-orthogonality in the frame.

Of course, the whole point of GST is that we don't know the gate set in advance, and so there's no way to ensure that any given set of fiducials produces state- and effect-sets that really are close to UIC (or even IC). Fortunately, it's easy to check *ex post facto*. We simply diagonalize $\tilde{\mathbb{1}}$, and examine its eigenvalues. It is full-rank iff the sets are IC. UIC implies that none of its eigenvalues are close to zero. It should be intuitively obvious that small eigenvalues in $\tilde{\mathbb{1}}$ are a problem, since they yield *large* eigenvalues in $\tilde{\mathbb{1}}^{-1}$, which in turn amplify any finite-sample error that is present. In a perfect world, $\tilde{\mathbb{1}}$ would be proportional to the identity operator; in practice, this is not achievable because both the $\{|\rho_j\rangle\rangle\}$ and the $\{\langle\langle E_i|\}$ are heavily biased toward the $|\mathbb{1}\rangle\rangle$ vector in $\mathcal{B}(\mathcal{H})$. The best *achievable* behavior is achieved when the state- and effect-sets are rank-1 projectors that form *2-designs* for the system's Hilbert space (see Ref. [7] for a reasonably painless introduction to frame theory and 2-designs for quantum tomography).

While there exist single-qubit 2-designs comprising just $d^2 = 4$ states, the best-known (and easiest to implement) comprises 6 states. They are the eigenstates of the three Pauli operators $\{X, Y, Z\}$, and they form a full set of *mutually unbiased bases*. Since the native gates in most qubit experiments are precisely the Clifford operations that generate these 6 states, it's very tempting to use them as state- and effect-sets. For the [extremely popular] target gateset defined by

$$\rho = |0\rangle\langle 0|\,,\ E = |1\rangle\langle 1|\,,\ G_x = e^{-i\frac{\pi}{4}\sigma_x},\ G_y = e^{-i\frac{\pi}{4}\sigma_y},$$

this is achieved by the fiducial sequences

$$\begin{aligned}
F_1 &= \{\}, && (3.34)\\
F_2 &= G_x, && (3.35)\\
F_3 &= G_y, && (3.36)\\
F_4 &= G_xG_x, && (3.37)\\
F_5 &= G_xG_xG_x, && (3.38)\\
F_6 &= G_yG_yG_y. && (3.39)
\end{aligned}$$

The only problem with this whole scheme is that now there are *six* fiducials. The LGST protocol given above assumed $n = d^2$, in which case the $\{\rho_j\}$ and $\{E_i\}$ are linearly independent vectors in $\mathcal{B}(\mathcal{H})$. With $n > d^2$ fiducials, these sets are *overcomplete* – they should still span $\mathcal{B}(\mathcal{H})$, but are not linearly independent. The A and B matrices are 6×4 and 4×6, respectively, and all the $\tilde{G}_k$ are 6×6. $\tilde{\mathbb{1}}$ is also $6x6$, but since it is the Gram matrix for sets of vectors lying in a 4-dimensional space $[\mathcal{B}(\mathcal{H})]$, its rank should be only 4.

Overcomplete fiducials add some complications to LGST, but they are also very useful. They allow us to *test* our assumption that the system is described by Markovian operations on a d-dimensional Hilbert space. If the rank of $\tilde{\mathbb{1}}$ is greater than d^2, then it is a clear sign that the LGST experiments are exploring a larger state space than expected. This is not entirely unexpected; it can be caused by leakage levels, coupling to ancillary qubits, or non-Markovian noise.

Generically, finite-sample errors will cause $\tilde{\mathbb{1}}$ to be full-rank (even if all model assumptions are perfectly satisfied). Heuristically, we can simply look at its spectrum to see whether there are d^2 "large" eigenvalues and $n - d^2$ "small" ones that are explainable by finite-sample fluctuations. A more rigorous procedure is to use model selection techniques to test whether $\tilde{\mathbb{1}}$'s rank is *significantly* greater than d^2. If it is, then the qubit is out of spec, and manual reconsideration is required. Typically, we either proceed with LGST to obtain the best possible d-dimensional model of the data (current practice), or we use model selection techniques to choose an appropriate dimension d and re-analyze (under development).

The simplest way to proceed (whether or not the system passes the "rank-d^2" test) is to truncate all of the empirical vectors and matrices ($\tilde{\mathbb{1}}$, $\tilde{G}_k$, $|\tilde{\rho}\rangle\rangle$, $\left\langle\langle \tilde{E} \right|$) to a suitable d^2-dimensional subspace, and then proceed as described above. Currently, we do this by performing an SVD on $\tilde{\mathbb{1}}$ and selecting the singular vectors corresponding to the d^2 largest singular values. More sophisticated methods, involving the Moore-Penrose pseudo-inverse of A and B, are under development. Simulations and experiments show that our current SVD-based approach works very well, although theory indicates that it is moderately suboptimal.

3.4 Contraction to deal with CPTP-violations caused by finite-sample errors

As shown above, LGST characterizes the device perfectly – *if* the probabilities for the specified experiments can be measured exactly. Unfortunately, they can't. When a 2-outcome experiment is performed N times, and one outcome is observed n times, the natural estimator of its probability is $\hat{p} = n/N$. This is the best possible unbiased estimator, but n is not generally equal to pN. Instead, it is binomially distributed around it:

$$n = pN \pm \sqrt{p(1-p)}\sqrt{N}.$$

Therefore, $\hat{p}$ will fluctuate by $\pm\sqrt{p(1-p)}/\sqrt{N}$ around the true value of p. This unavoidable estimation error is known as *finite sample fluctuation*.

The obvious consequence of finite-sample fluctuations is that the estimated gateset will not be identical to the true one. Each matrix element will typically be off by $O(1/\sqrt{N})$. This is not in itself a terribly big deal, but matters are complicated by the fact that if the true gateset is close enough to a boundary – i.e., one or more of its elements has an eigenvalue smaller than the fluctuations – then there is a high probability that the estimated gateset will violate positivity.

When this happens, the tomographer has several choices. The easiest is to do nothing, and report the nonpositive gate set. In some cases, this is perfectly fine, but if and when the estimated gateset is used to predict future experiments, it may predict negative probabilities. This is a serious problem whenever entropies, fidelities, or failure probabilities are calculated.

A second choice is to modify the gauge-optimization step. Instead of varying the gauge

to make the estimated gates as close as possible to the targets, we can instead minimize their negativity (the sum of the negative eigenvalues of the various gates). We call this *optimizing toward CPTP*, since it attempts to minimize the distance to the set of explicitly CPTP gatesets. It is much more difficult than optimizing to the targets, because the set of explicitly CPTP gatesets is not convex, and is not in any way congruent to the orbits of gauge transformations. Global optimization is required. When the true gates (or the estimate) are not actually close to the targets, this strategy can be very valuable, and can result in estimates that are far less negative than those obtained by optimizing to the (wrong) target. However, it can also introduce severe pathologies (although we have not observed this to happen yet). It may occur that the estimate is quite close to the target *but* non-positive due to fluctuations, and also happens to be gauge-equivalent to an explicitly CPTP gateset that is far from the target. In this case, successful gauge optimization to CPTP will find the far-away but explicitly CPTP gateset, and lead to the erroneous conclusion that the true gates are not close to the targets.

A third option is to *contract* the gateset, forcing it to be CPTP. This is always done after a gauge optimization either toward CPTP or toward a specific CPTP target, in order to ensure that the pre-contraction gateset is not very negative (i.e., not far from the CPTP set). Then, the estimate is projected to the nearest CPTP gateset. This is done in three steps:

1. ρ and E are contracted to the nearest valid state and effect as measured by Hilbert-Schmidt (Euclidean) distance,

2. Each of the G_k is written in Choi form (where violations of CP correspond to negative eigenvalues of the χ matrix), and then contracted to the nearest valid CP map, as measured by Euclidean distance, and transformed back to the superoperator representation.

3. Finally, each of the resulting G_k is projected linearly onto the affine hyperplane of TP maps (since TP is a linear constraint on the superoperator).

A simple algorithm for finding the nearest positive matrix in Hilbert-Schmidt distance is given in [5].

We make use of the 1st solution for intermediate steps (see next chapter), and the 3rd for final estimates that *must* be physically valid gatesets to avoid negative probabilities, imaginary entropies, and the like. The second solution (optimizing to CPTP) is useful only when the estimate is very bad in the first place (generally because the gates are far from the targets). However, none of these solutions are fully satisfactory. What is really needed is an explicitly gauge-invariant parameterization of gate sets, but we have not yet found one.

Chapter 4

Hyperaccurate GST using long circuits

GST was invented to solve a specific problem, the reliance of tomography on precalibrated operations. LGST solved this problem. To make it work, we needed to develop analysis methods that could deal with "nonlinear data" – that is, results of experiments whose probabilities are highly nonlinear functions of the parameters. By resolving this problem, we opened the road to a completely different – and genuinely revolutionary – improvement in tomography.

Standard process tomography characterizes a gate G using experiments in which G appears just once. This makes analysis easy, but it also limits the precision that can be achieved. Each matrix element of G corresponds more or less 1:1 to an observed probability, and the estimation error in each such probability (due to finite sample fluctuations) is approximately

$$\hat{p} = p \pm \frac{O(1)}{\sqrt{N}}.$$

As we show in Chapter 5, it is possible under *some* circumstances to improve this scaling to $O(1)/N$, but while this can offer substantial improvements, it is still limited by SPAM noise, as it relies on choosing experiments for which $p \approx 0$. Fortunately, GST offers a completely different way to break the $1/\sqrt{N}$ boundary, and achieve what we refer to as *hyperaccuracy*.

The key idea here is that in GST, we can quite easily perform and incorporate data from *long* circuits, in which a gate appears many times. The sensitivity of the corresponding probability to the elements of G_k increases proportional to the number of times G_k appears in the circuit. So, for example, if we estimate the probability for

$$Pr = \langle\langle E| \, G_x G_x G_x G_x \, |\rho\rangle\rangle,$$

then it will provide 4 times as much precision in estimating G_x as would

$$Pr = \langle\langle E| \, G_x \, |\rho\rangle\rangle.$$

However, in order to turn this idea into a practical, reliable algorithm, we need to address several critical questions:

1. What experiments (circuits, or gate sequences) should be performed?

2. How can the resulting data be efficiently analyzed?

3. How reliable is the resulting algorithm?

The answers are entangled – e.g., the choice of experiments is influenced significantly by the requirement that the resulting data be easily and reliably analyzable. We experimented with a wide variety of approaches, eventually developing the *extended LGST* (eLGST) approach presented here. In the interest of clarity and brevity, we provide only a cursory overview of the various things that *didn't* work.

Our original approach was entirely unstructured. We chose a variety of random gate sequences (much like those used in randomized benchmarking, but without any underlying compiled-Clifford-gate structure), and then analyzed the resulting data using MLE. The GST likelihood function is very messy, but still easy to evaluate, and its derivative can be computed analytically. We found two problems with this approach:

1. Random gate sequences did not appear to give as much precision in the result as we hoped for. Accuracy increased with L (the length of the longest or typical sequence), but not proportional to L. Furthermore, with no underlying theory of how this "MLEGST" should perform, we could not determine how close it came to optimality.

2. The lack of structure in the likelihood function made MLE very difficult. Local gradient optimization worked only unreliably, and was highly dependent on starting location. We achieved reasonable success by beginning with LGST (which is closed-form and totally reliable, but not very accurate), and then refining this estimate by adding successively longer circuits to the likelihood function. However, when we applied this technique in experiments (where the data are less clean, and the underlying model less reliable), we found it to be still unreliable – basically, as we ran the optimizer longer, and incorporated more global-optimizing techniques, it continued to find higher local maxima of the likelihood, rather than converging to a global maximum.

We concluded that (a) as with LGST, we needed to choose our gate sequences carefully to make the analysis easier, and (b) brute-force MLE was not an appealing solution.

4.1 Extended LGST

We have settled on a protocol that reliably, consistently provides both extremely high accuracy *and* reliable analysis. We call it *extended LGST* (eLGST), and it is based around three critical modifications to the original "unstructured MLE" approach.

First, instead of implementing random sequences of gates, we choose a relatively small set of *base sequences*, denoted $\{S_l : l = 1 \dots n\}$. A base sequence S represents a sequence of

gates from the gate set, as $S = [s_1, s_2, \ldots s_L]$, where L is the length of S. Applying the base sequence implements an operation given by the *gate sequence product* $\sigma(S)$:

$$\sigma(S) = G_{s_L} G_{s_{L-1}} \ldots G_{s_2} G_{s_1}. \tag{4.1}$$

For each base sequence S_l, we perform LGST on its gate sequence product, by performing the n^2 experiments

$$\langle\langle F_i | \, \sigma(S_l) \, | F_j \rangle\rangle. \tag{4.2}$$

(That is, we prepare ρ, then apply the short fiducial sequence F_j, then all the gates in S_l, then F_i, and then measure.) Given these data, we can use basic LGST methods to *directly* estimate $\sigma(S_l)$, up to a gauge that is the same for all base sequences.

Second, the base sequences themselves are chosen carefully, not randomly. We discuss this in detail below, but each base sequence consists of a short "germ" sequence (e.g., $G_x G_y G_y$) repeated several (perhaps many) times in a row. This structure ensures that small deviations from ideal behavior in each gate are amplified, and thus have an effect proportional to L in the observed probabilities. This allows eLGST to achieve consistent, reliable, predictable accuracy that scales as $1/L$.

Third, we do not use MLE directly. Instead, we use least-squares methods. As noted in the introductory material, least-squares optimization is a good proxy for MLE when the number of samples (N) is large. It is generally a bit suboptimal for small N, but is both much faster *and* more amenable to analytic understanding (although in general it still requires numerical implementation). eLGST relies partly on direct LGST (which is itself a least-squares method, albeit one implemented analytically using only linear algebra), and partly on numerical least-squares to collate the results of LGST on different base sequences.

4.1.1 Computing the eLGST estimate

Let us postpone the discussion of exactly how the base sequences are chosen. Suppose, for now, that we simply perform direct LGST on base sequences of the form G_k^L – that is, L repetitions of each individual gate in the gateset. Let us also assume that we repeat each experiment N times, and the result of the LGST analysis is that we obtain a reasonably accurate estimate of the gate sequence product (σ) for each base sequence. We denote this estimate by $\widehat{G_k^L}$, and assume that

$$\widehat{G_k^L} = G_k^L \pm \frac{O(1)}{\sqrt{N}}.$$

Now, if each G_k were a real scalar, then getting a high-precision estimate would be very simple. We would take a low-precision estimate $\widehat{G_k^L}$ for some very large L and compute its Lth root,

$$\widehat{G_k} = \left(\widehat{G_k^L}\right)^{1/L} = G_k \pm \frac{O(1)}{L\sqrt{N}}.$$

This is the scaling we do eventually achieve, but it's not that easy, because the G_k are matrices.

The first problem is demonstrated even by the simple case where G_k is a complex scalar,

$$G_k = e^{i\theta}.$$

Now $G_k^L = e^{i(L\theta \bmod 2\pi)}$, and so the Lth root is multivalued (i.e., $\theta \to \theta + n\frac{2\pi}{L}$ leaves all observable probabilities invariant). We have to choose the right *branch*. This is impossible without further information.

We solve this problem by bootstrapping up from $L = 1$ to $L = L_{max}$, choosing base sequences for $L = 1, 2, 4, 8, 16 \ldots, L_{max}$. The additional cost is only logarithmic in L_{max}, and as long as N is large enough, it completely solves the problem of choosing which branch. We begin by using the $L = 1$ data to get a decent estimate of $G_k \pm O(1)/\sqrt{N}$. Then, we use the $L = 2$ data to deduce that

$$G_k \approx \pm\sqrt{\widehat{G_k^2}},$$

and use the $L = 1$ estimate to identify unambiguously which of the two branches indicated by the $\pm$ symbol is correct. We repeat this process recursively for each successively larger L.

The second problem is peculiar to matrices. The procedure given above works very well for scalar $G_k = e^{i\theta}$, but not for matrix-valued G_k. To see this, consider the example of

$$G_k = \sigma_Z.$$

Clearly, $G_k^2 = \mathbb{1}$. If we perform LGST on G_k^2, we will generally obtain $\widehat{G_k^2} = \mathbb{1} \pm O(1)/\sqrt{N}$, where the error term is a small, random matrix. Suppose (without loss of generality) that we get $\widehat{G_k^2} = \mathbb{1} + \epsilon\sigma_x$. There are multiple square roots of $\widehat{G_k^2}$, but this isn't the main problem; the main problem is that none of them are even *close* to the true value of $G_k = \sigma_z$! Instead, every square root of $\mathbb{1} + \epsilon\sigma_x$ is diagonal in the σ_x basis.

The root of the problem is that the matrix square-root function is highly non-smooth, and can rip apart the topology of matrices. To fix it, we observe that we should be looking for a $\widehat{G_k}$ such that $\widehat{G_k}^2 \approx \widehat{G_k^2}$ – *not* such that $\widehat{G_k} \approx \sqrt{\widehat{G_k^2}}$. These two equations are not even remotely equivalent (thanks to the topological violence that can be hiding in the matrix root).

We implement these tricks using the following iterative algorithm:

1. Use LGST on base sequences of the form G_k^L for $L = 1, 2, 4, 8, 16 \ldots L_{max}$ to obtain approximate estimates $\widehat{G_k^L}$.

2. For each G_k, set the *initial* $\widehat{G_k}$ equal to $\widehat{G_k^1}$.

3. Use least-squares optimization to vary $\widehat{G_k}$ so as to minimize

$$\delta_2 = |\widehat{G_k} - \widehat{G_k^1}|_2^2 + |\widehat{G_k}^2 - \widehat{G_k^2}|_2^2. \tag{4.3}$$

4. Repeat Step 3 with a new cost function $\delta_4 = \delta_2 + |\widehat{G_k}^4 - \widehat{G_k^4}|_2^2$.

5. Continue repeating Step 3 with cost functions $\delta_8, \delta_{16}, \ldots \delta_{L_{max}}$ until all data has been incorporated.

This algorithm is easy to implement, fast (in many cases it runs in $O(1)$ second), reliable, and accurate. Its accuracy can be improved (see below) by being more statistically careful and using *weighted* least squares, but this generally provides at most a factor of 2. This is utterly dwarfed by the hyperaccuracy that stems from the $1/L$ factor.

However, one significant problem remains. It stems not from the analysis algorithm, but from the base sequences we have (naively, so far) chosen to perform LGST on.

4.2 Base sequences and germ powers

The reason for long-circuit GST (and hence the eLGST protocol) is to achieve very high accuracy, by amplifying the gate parameters so that they can be measured very accurately using relatively imprecise estimates of probabilities. Intuitively, we want to take each parameter θ in the gateset and map it to a probability that depends on it as $p \approx L\theta$. The protocol outlined above does this – but only for *some* parameters of the gateset.

To see this, consider the simple (but realistic) example of a gateset containing just one gate that implements an x-rotation by angle θ, $G[\rho] = U\rho U^\dagger$ where $U = e^{i(\theta/2)\sigma_x}$. Repeated L times, this gate performs a rotation by $L\theta$, and so indeed we can estimate the rotation angle θ to great precision $(1/L)$. But what if the *axis* of the rotation changes? Suppose, for example, that $\theta = \pi$, so that two rotations equate to the identity operation. Suppose that we perturb U by changing the axis around which it rotates:

$$U \to U' = e^{i(\theta/2)\vec{\sigma}\cdot\vec{n}},$$

where $\vec{n} = \cos\epsilon\sigma_x + \sin\epsilon\sigma_z$. This is an $O(\epsilon)$ change, so we expect that it will become significant for $L \approx 1/\epsilon$. But it does nothing of the sort, because U' is also a π rotation (around a different axis) and so $(U')^2 = 1\!\!1$ (up to an unobservable global phase). And therefore $(U')^L = U^L$ if L is even, or $(U')^L = U'U^{L-1}$ if L is odd. In both cases, no matter how large L is, the deviation is never amplified *at all* – it is always of order ϵ, with no L-dependence at all.

The problem here is that the dominant X rotation is actually echoing away the ϵZ perturbation on it. The circuit G^L amplifies X errors (which commute with G), but performs dynamical decoupling on Z errors (which anticommute with G). In fact, it is relatively

straightforward to show that for nearly-unitary gates, the *only* deviations that can be amplified by repeating G many times are those that commute with G itself. Since G is a $d^2 \times d^2$ matrix, it has d^4 parameters that could vary, but only d^2 of them (in general) can commute with G and therefore be amplified by G^L.

As a result, we find that when we use only base sequences of the form G_k^L, we obtain extremely precise information about the eigenvalues of each G_k, but no advantage proportional to L at all for the eigenvectors or off-diagonal elements of G_k.

In order to achieve hyperaccuracy for *all* parameters, we add additional base sequences. They consist of short sequences called *germs*, repeated many times. Each germ, when repeated, amplifies a particular linear combination of the parameters in the gateset (*not* typically parameters of any single gate, since multiple gates appear in most germs). We denote the set of germs by $\{g_m\}$, and for each germ g we define base sequences of [approximate] length $L = 1, 2, 4, 8, \ldots L_{max}$ as

$$S_{g,L} = g^{\lfloor L/l(g) \rfloor}, \tag{4.4}$$

where $l(g)$ is the length of germ g. Here is a specific real-world example. The most common gateset we encounter includes three gates, denoted $\{G_x, G_y, G_i\}$ which are intended to be $\pi/2$ rotations around the X and Y axes and an idle/identity gate (respectively). A good set of germs for this gateset (see next section) is

$$\begin{aligned}
g_1 &= G_x \\
g_2 &= G_y \\
g_3 &= G_i \\
g_4 &= G_xG_y \\
g_5 &= G_xG_yG_i \\
g_6 &= G_xG_iG_y \\
g_7 &= G_xG_iG_i \\
g_8 &= G_yG_iG_i \\
g_9 &= G_xG_xG_iG_y \\
g_{10} &= G_xG_yG_yG_i \\
g_{11} &= G_xG_xG_yG_xG_yG_y.
\end{aligned} \tag{4.5}$$

The germ-powers for $L = 1$ are

$$\begin{aligned}
S_{1,1} &= G_x \\
S_{2,1} &= G_y \\
S_{3,1} &= G_i.
\end{aligned}$$

For $L = 16$, they are

$$\begin{aligned}
S_{1,16} &= (G_x)^{16} \\
S_{2,16} &= (G_y)^{16} \\
S_{3,16} &= (G_i)^{16} \\
S_{4,16} &= (G_x G_y)^8 \\
S_{5,16} &= (G_x G_y G_i)^8 \\
S_{6,16} &= (G_x G_i G_y)^5 \\
S_{7,16} &= (G_x G_i G_i)^5 \\
S_{8,16} &= (G_y G_i G_i)^5 \\
S_{9,16} &= (G_x G_x G_i G_y)^4 \\
S_{10,16} &= (G_x G_y G_y G_i)^4 \\
S_{11,16} &= (G_x G_x G_y G_x G_y G_y))^2.
\end{aligned}$$

This more elaborate set of base sequences requires a slightly more sophisticated eLGST algorithm. We maintain the iterative structure of the algorithm, but in the nth step, we include all sequences of length $L \leq 2^n$. The objective function at the nth step is

$$\delta_n = \sum_{\text{all } S_{m,L} \text{ of length } \leq 2^n} \delta(\{\widehat{G_k}\}, \widehat{S_{m,L}}), \tag{4.6}$$

where $\widehat{S_{m,L}}$ is a direct LGST estimate of the gate sequence product for the base sequence $S_{m,L}$, and $\delta()$ is the squared 2-norm distance between the gate sequence product *predicted* by the estimate $\{\widehat{G_k}\}$ and the one directly *estimated* by LGST.

4.2.1 Choosing germs

The point of doing experiments that correspond to germ-power sequences is to ensure that *every* parameter of the gateset is amplified. We choose germs with this in mind. As of this writing, we do not know of a purely theoretical way to choose germs. In principle, we could simply simulate GST using various sets of germs, and see which ones work well. Fortunately, this isn't necessary; we can use theory to simplify the task a great deal (although ultimately some numerics *are* required). We simplify the problem by considering two idealizations:

1. We assume that the gates are nearly unitary,
2. We focus on the effect of very long ($L \gg 1$) germ powers.

For each germ g, let us define the gate sequence product $\sigma(g)$. The gate sequence product for germ powers of g with $L \gg 1$ will then be simply $\sigma(g)^l$ for some (relatively large) l. These

matrices, $\sigma(g)^l$, are the ones on which we do LGST, and for which we get low-precision estimates $\widehat{\sigma(g)^l}$.

Each such estimate will constrain certain parameters of the gateset. To determine *which* parameters – and how precisely it constrains them – we ask how $\sigma(g)^l$ changes if the gateset is varied just slightly. Let $\mathcal{G}$ be a vector representing the entire gateset (ρ and E are not relevant for this calculation, so $\mathcal{G} = \{G_k\}$), with all Kd^4 matrix elements of the $\{G_k\}$ arranged in a column. The *sensitivity* of the [directly observable] $\sigma(g)^l$ to $\mathcal{G}$ is simply the $Kd^4 \times d^4$ matrix Jacobian

$$J_g = \frac{\partial \sigma(g)^l}{\partial \mathcal{G}}. \tag{4.7}$$

We could compute this Jacobian numerically for each g and L, but it is simpler to observe that

$$\frac{\partial \sigma(g)^l}{\partial \mathcal{G}} = l\sigma(g)^{l-1}\mathcal{T}_{\sigma(g)}[\frac{\partial \sigma(g)}{\partial \mathcal{G}}], \tag{4.8}$$

where, for $l \gg 1$, $\mathcal{T}_{\sigma(g)}$ is the superoperator corresponding to *twirling* by $\sigma(g)$. (This expression is somewhat nontrivial to derive, but only because $\sigma(g)$ and $\frac{\partial \sigma(g)}{\partial \mathcal{G}}$ do not commute – which is critical to deriving the twirling portion). The effect of the twirling is to annihilate all elements of $\frac{\partial \sigma(g)}{\partial \mathcal{G}}$ that do not commute with $\sigma(g)$, or (equivalently) to project it onto the commutant of $\sigma(g)$.

So for each germ, we can compute a Jacobian which identifies exactly what parameters of the gateset $\mathcal{G}$ will influence $\sigma(g)^l$. In the limit of large l, every parameter is either amplified proportional to l or echoed away. Given two or more germs, their joint sensitivity is described by simply stacking the corresponding Jacobians next to each other:

$$J_{g_1,g_2\ldots} = \left(J_{g_1}\ J_{g_2}\ \cdots\right). \tag{4.9}$$

And a set of germs is *complete* – i.e., amplifies every observable parameter of $\mathcal{G}$ – if and only if the rank of its Jacobian (which can be obtained by SVD) is equal to the number of observable parameters in $\mathcal{G}$. There is actually a hidden caveat here; some of the parameters of $\mathcal{G}$ are gauge parameters, so the number of observable parameters is actually less than Kd^4, and generally equals $(K-1)d^4+1$ (because there are usually d^4-1 gauge parameters). So, we can simply add germs until the Jacobian's numerically-computed rank reaches this threshold.

In practice, we want the germs to be not just complete but *uniformly complete* – i.e., they should amplify each parameter as much as possible. Amplification is proportional to the singular values of the Jacobian, so we want to ensure that all $(K-1)d^4+1$ singular values of the Jacobian are not just nonzero, but as large as possible. Although we have no closed-form algorithm for achieving this, we can do so rather effectively by adding and removing germs from the set until the spectrum of the resulting Jacobian stabilizes. It is this heuristic algorithm that produced the germ set given above in Eqs. 4.5.

Chapter 5

Adaptive tomography improves accuracy quadratically

Quantum information processing requires reliable, repeatable preparation and transformation of quantum states. *Quantum state tomography* is used to identify the density matrix ρ that was prepared by such a process. No finite ensemble of N samples is sufficient to uniquely identify ρ, so we *estimate* it, reporting either a single state $\widehat{\rho}$ that is "close" to ρ with high probability [8, 9, 3, 10, 11], or a confidence region of nonzero radius that contains ρ with high probability [12, 4]. Both approaches must accept some inaccuracy (the discrepancy between $\widehat{\rho}$ and ρ) or imprecision (the diameter of the confidence region). The universal goal of state tomography is to minimize this discrepancy, which has been quantified with various metrics (e.g., trace norm, fidelity, relative entropy, etc.). In this analysis, we focus on the particularly well-motivated *quantum infidelity*,

$$1 - F(\widehat{\rho}, \rho) = 1 - \mathrm{Tr}\left(\sqrt{\sqrt{\rho}\widehat{\rho}\sqrt{\rho}}\right)^2, \tag{5.1}$$

and show that as $N \to \infty$, *adaptive* tomography reduces expected infidelity from $O(1/\sqrt{N})$ to $O(1/N)$.

Unlike alternative metrics, $1 - F(\widehat{\rho}, \rho)$ quantifies an important operational quantity: *how many copies are required to reliably distinguish $\widehat{\rho}$ from ρ?*. Without doing justice to the rich body of research behind this simple statement (e.g., [13, 14, 15, 16, 17, 18]...), we summarize as follows. The discrepancy between $\widehat{\rho}$ and ρ given a *single sample* is well described by the trace distance, $|\widehat{\rho} - \rho|_1$. But tomography (i) requires $N \gg 1$ samples; (ii) is used to predict experiments on $N \gg 1$ samples; and (iii) yields errors that cannot be detected without $N \gg 1$ samples. So the operationally relevant quantity is $\left|\widehat{\rho}^{\otimes N} - \rho^{\otimes N}\right|_1$, which for $N \gg 1$ behaves as $1 - e^{-D(\widehat{\rho},\rho)N}$. The exponent D is the *quantum Chernoff bound* [18], and $N \approx D \log(1/\epsilon)$ samples are necessary and sufficient to distinguish ρ from $\widehat{\rho}$ with confidence $1 - \epsilon$. D is tightly bounded by the logarithm of the fidelity (see [17], Eq. 28); when $1 - F(\widehat{\rho}, \rho) \ll 1$ (which should always be true in tomography!), $-\log(F) \approx 1 - F$ and

$$\frac{1 - F}{2} \leq D \leq 1 - F. \tag{5.2}$$

Thus, $1 - F$ really does (almost uniquely) quantify tomographic inaccuracy; $N \approx [1 - F(\widehat{\rho}, \rho)]^{-1}$ samples are (up to a factor of 2) necessary and sufficient [1] to falsify $\widehat{\rho}$. In contrast, Hilbert-Schmidt- and trace-distance have no such N-sample meaning, and give wildly misleading metrics of tomographic error.

We show that standard tomography with static measurements can't beat $1 - F = O(1/\sqrt{N})$ as $N \to \infty$ for a large and important class of states, then introduce and explain a simple adaptive protocol that achieves $1 - F = O(1/N)$ for every state. Finally, we demonstrate this effect in a linear optical experiment, achieving a 10-fold improvement in infidelity (from 0.1% to 0.01% with $N = 3 \times 10^4$ measurements) over standard tomography. We believe this protocol will have wide application, particularly in situations where the rate

[1] Remarkably, for large N, local measurements can discriminate almost as well as joint measurements on all N samples. If D_Q and D_C are the optimal error exponents for joint and local measurements (respectively), then $(1 - F)/2 \leq D_C \leq D_Q \leq 1 - F$ [17].

of data collection is small, such as post-selected optical systems (e.g. [19], where data were collected at approximately 9 measurements per hour).

Adaptivity has been proposed in various contexts. Single-step adaptive tomography was first analyzed by [20], then refined in [21, 1, 22]. A scheme similar to ours (and its efficacy for *pure* states) was analyzed in [23]. Ref. [24] recently treated state estimation as parameter estimation, obtaining results complementary, but largely orthogonal, to those reported here. Here, we present both an experimental demonstration and simple, self-contained derivation of: (1) why quantum fidelity is significant; (2) why adaptive tomography achieves far better infidelity; and (3) how the adaptation should be done. We optimize *worst-case* infidelity over all states, not just pure states [23] or specific ensembles of mixed states (e.g. Ref. [1] achieved high average fidelity, but low fidelity on nearly-pure states).

5.1 Adaptive tomography

Static tomography uses data from a fixed set of measurements. Different measurements yield subtly different tomographic accuracy [25], but to leading order, "good" protocols for single-qubit tomography provide equal information [26] about every component of the unknown density matrix ρ,

$$\rho = \frac{1}{2}\left(\mathbb{1} + \langle\sigma_x\rangle\,\sigma_x + \langle\sigma_y\rangle\,\sigma_y + \langle\sigma_z\rangle\,\sigma_z\right). \tag{5.3}$$

The canonical example involves measuring the three Pauli operators (σ_x, σ_y, σ_z). This minimizes the variance of the estimator $\widehat{\rho}$ – but not the expected infidelity, for two reasons.

First, *the variance of the estimate $\widehat{\rho}$ depends also on ρ itself.* Consider the linear inversion estimator $\widehat{\rho}_{\rm lin}$, defined by estimating $\langle\sigma_z\rangle = \frac{n_\uparrow - n_\downarrow}{n_\uparrow + n_\downarrow}$ (and similarly for $\langle\sigma_x\rangle$ and $\langle\sigma_y\rangle$), and substituting into Eq. 5.3. Each measurement behaves like $N/3$ flips of a coin with bias $p_k = \frac{1}{2}(1 + \langle\sigma_k\rangle)$, and yields

$$\hat{p_k} = p_k \pm \sqrt{\frac{3}{N}}\sqrt{p_k(1-p_k)} \tag{5.4}$$

$$\Rightarrow \langle\sigma_k\rangle_{\rm estimated} = \langle\sigma_k\rangle_{\rm true} \pm \sqrt{\frac{3}{2N}}\sqrt{1 - \langle\sigma_k\rangle^2}. \tag{5.5}$$

When $\langle\sigma_k\rangle \approx 0$, its estimate has a large variance – but when $\langle\sigma_k\rangle \approx \pm 1$, the variance is very small. As a result, the variance of $\widehat{\rho}$ around ρ is anisotropic and ρ-dependent (see Fig. 5.1a).

Second, *the dependence of infidelity on the error, $\Delta = \widehat{\rho} - \rho$, also varies with ρ.* Infidelity is hypersensitive to misestimation of small eigenvalues. A Taylor expansion of $1 - F(\widehat{\rho}, \rho)$ yields (in terms of ρ's eigenbasis $\{|i\rangle\}$),

$$1 - F(\rho, \rho + \epsilon\Delta) = \frac{1}{4}\sum_{i,j}\frac{\langle i|\,\Delta\,|j\rangle^2}{\langle i|\,\rho\,|i\rangle + \langle j|\,\rho\,|j\rangle} + O(\Delta^3). \tag{5.6}$$

Infidelity is quadratic in Δ – except that as an eigenvalue $\langle i|\,\rho\,|i\rangle$ approaches 0, its sensitivity to $\langle i|\,\Delta\,|i\rangle$ diverges; $1-F$ becomes *linear* [2] in Δ:

$$1 - F(\rho, \rho + \epsilon\Delta) = \epsilon \sum_{i:\ \langle i|\rho|i\rangle = 0} \langle i|\,\Delta\,|i\rangle + O(\Delta^2). \tag{5.7}$$

To minimize infidelity, we must accurately estimate the small eigenvalues of ρ, particularly those that are (or appear to be) zero. For states deep within the Bloch sphere, static tomography achieves infidelity of $O(1/N)$ [20, 27]. Typical errors scale as $|\Delta| = O(1/\sqrt{N})$ (Eq. 5.5), and infidelity scales as $1 - F = O(|\Delta|^2)$. But for states with eigenvalues less than $O(1/\sqrt{N})$, infidelity scales as $O(1/\sqrt{N})$. Quantum information processing relies on nearly-pure states, so this poor scaling is significant.

To achieve better performance, we observe that if ρ is diagonal in one of the measured bases (e.g., σ_z), then infidelity *always* scales as $O(1/N)$. The increased sensitivity of $1-F$ to error in small eigenvalues (Eq. 5.6) is precisely canceled by the reduced inaccuracy that accompanies a highly biased measurement-outcome distribution (Eq. 5.5). This suggests an obvious (if naïve) solution: we should simply ensure that we measure the diagonal basis of ρ!

This is unreasonable – knowing ρ would render tomography pointless. But we *can* perform standard tomography on $N_0 < N$ samples, get a preliminary estimate $\widehat{\rho}_0$, and measure the remaining $N - N_0$ samples so that one basis diagonalizes $\widehat{\rho}_0$. This measurement will not diagonalize ρ exactly, but if $N_0 \gg 1$ it will be fairly close. The angle θ between the eigenbases of ρ and $\widehat{\rho}_0$ is $O(|\Delta|) = O(1/\sqrt{N_0})$. This implies that if ρ has an eigenvector $|\psi_k\rangle$ with eigenvalue $\lambda_k = 0$, then corresponding measurement outcome $|\phi_k\rangle\langle\phi_k|$ will have probability at most $p_k = \sin^2\theta \approx \theta^2 = O(1/N_0)$. Since we make this measurement on $O(N - N_0)$ copies [3], the final error in the estimated $\hat{p_k}$ (and therefore in the eigenvalue λ_k) is $O(1/\sqrt{N_0(N-N_0)})$. So using a constant fraction $N_0 = \alpha N$ of the available samples for the preliminary estimation should yield $O(1/N)$ infidelity for *all* states.

A similar protocol was suggested in Ref. [1], but that analysis concluded that $N_0 \propto N^p$ for $p \geq \frac{2}{3}$ would be sufficient. This works for *average* infidelity over a particular ensemble, but yields $1 - F = O(N^{-5/6})$ for almost all nearly-pure states.

5.2 Simulation results

We performed numerical simulations of single-qubit tomography using four different protocols: (1) standard fixed-measurement tomography; (2) adaptive tomography with $N_0 = N^{2/3}$, as proposed in [1]; (3) adaptive tomography with $N_0 = \alpha N$ (for a range of α); and (4) "known

[2] Because ρ lies on the state-set's boundary, the gradient of F need not vanish in order for $\widehat{\rho} = \rho$ to be a local maximum.

[3] The "O" notation is necessary here because some of the remaining $N - N_0$ copies may be measured in other bases that make up a complete measurement frame.

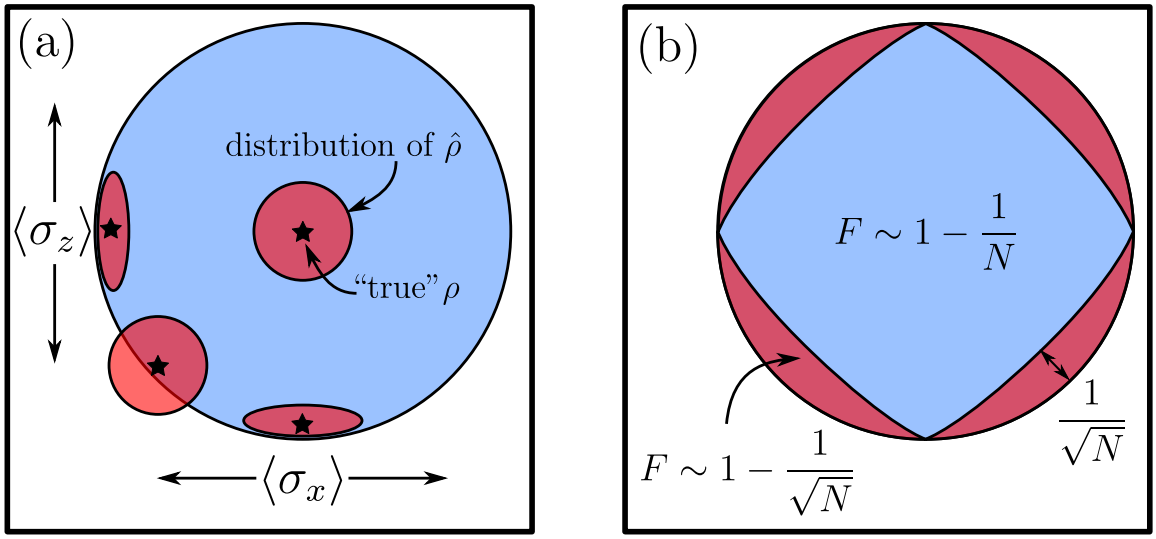

Figure 5.1. Two features of qubit tomography with Pauli measurements (shown for an equatorial cross-section of the Bloch sphere): **(a)** The distribution or "scatter" of any unbiased estimator $\widehat{\rho}$ (depicted by dull red ellipses) varies with the true state ρ (black stars at the center of ellipses); **(b)** The expected infidelity between $\widehat{\rho}$ and ρ as a function of ρ. Within the Bloch sphere, the expected infidelity is $O\left(1/N\right)$. But in a thin shell of nearly-pure states (of thickness $O\left(1/\sqrt{N}\right)$), it scales as $O\left(1/\sqrt{N}\right)$ – *except* when ρ is aligned with a measurement axis (Pauli X, Y, or Z).

basis" tomography, wherein we cheat by aligning our measurement frame with ρ's eigenbasis (for all N samples). We simulated many true states ρ, but present a representative case: a pure state with $(\langle\sigma_x\rangle, \langle\sigma_y\rangle, \langle\sigma_z\rangle) = (0.5, 1/\sqrt{2}, 0.5)$

$$|\nearrow\rangle = \frac{1}{2}\begin{pmatrix} \sqrt{3} \\ \frac{1}{\sqrt{3}} - \frac{2i}{\sqrt{6}} \end{pmatrix} \tag{5.8}$$

Our results are not particularly sensitive to the exact estimator used; we used maximum-likelihood estimation (MLE) with a quadratic approximation to the negative loglikelihood function:

$$\updownarrow(\rho) = -\log \mathcal{L}(\rho) \approx \sum_{k=1}^{3} \frac{N_k(\mathrm{Tr}[\rho E_k] - f_k)^2}{f_k(1-f_k)}, \tag{5.9}$$

where $f_k = n_k/N_k$ are the observed frequencies of the +1 eigenvectors of the three Pauli operators σ_k, E_k is the corresponding projector, and N_k is the number of samples on which σ_k was measured. Convex optimization (in MATLAB [28]) was used to find $\widehat{\rho}_{\rm MLE}$. Results were averaged over many (typically 150) randomly generated measurement records.

Figure 5.2 shows average infidelity versus N. We fit these simulated data to power laws of the form $1 - F = \beta N^p$, and found $p = -0.513 \pm 0.006$ (for static tomography), $p = -0.868 \pm 0.008$ (for adaptive tomography with $N_0 = N^{2/3}$), $p = -0.980 \pm 0.006$ (for adaptive tomography with $N_0 = 0.5N$), and $p = -0.993 \pm 0.09$ (for known-basis tomography). These results are not significantly different [4] from predictions of the simple theory ($p = -\frac{1}{2}, -\frac{5}{6}, 1$, and 1, respectively). The borderline-significant discrepancy is, we believe, due to boundary effects ($\widehat{\rho}_{\rm MLE}$ is constrained to be positive). We also varied $\alpha = N_0/N$ (Fig. 5.2, inset) and found that $\alpha = \frac{1}{2}$ optimizes the prefactor (β).

5.3 Experimental results

We implemented our protocol experimentally in linear optics (Fig. 5.3). Using type-1 spontaneous parametric down conversion in a nonlinear crystal, photon pairs were created. One of these photons was sent immediately to a single photon counting module (SPCM) to act as a trigger. The second photon was sent through a Glan-Thomson polarizer to prepare it in a state of very pure linear polarization. Computer-controlled waveplates were first used to prepare the polarization state of the photon, and subsequently used in tandem with a polarization beamsplitter to project onto any state on the Bloch sphere.

We compared static and adaptive tomography protocols on a measured state given (in the H/V basis) by

$$\rho = \begin{pmatrix} 0.7711 & 0.2010 + 0.3624i \\ 0.2010 - 0.3624i & 0.2289 \end{pmatrix}, \tag{5.10}$$

which has purity $\mathrm{Tr}(\rho^2) = 0.991$ and fidelity $F = 0.992$ with $|\nearrow\rangle$ (see Eq. 5.8). We identified ρ to within an uncertainty which is at most $O(1/\sqrt{\tilde{N}})$ using one very long ($\tilde{N} = 10^7$) static tomography experiment, whose overwhelming size ensures accuracy sufficient to calibrate the other experiments, all of which involve $N \leq 3 \times 10^4$ photons.

Our "standard" (static) protocol involved repeatedly preparing our target state, collecting $N/3$ photons at each of the three measurement settings corresponding to σ_x, σ_y, and σ_z, and computing $\widehat{\rho}_{\rm MLE}$ as outlined in [29]. Each data point in figure 5.4a represents an average over many ($\sim$150) repetitions.

[4] All quoted uncertainties herein are 1σ, or 68% confidence intervals. Therefore, we don't expect the the "true" value to lie within the error bars more than 68% of the time. Most of the results given here agree with theoretical predictions to within 2σ (95% confidence intervals), a common criterion for consistency between data and theory.

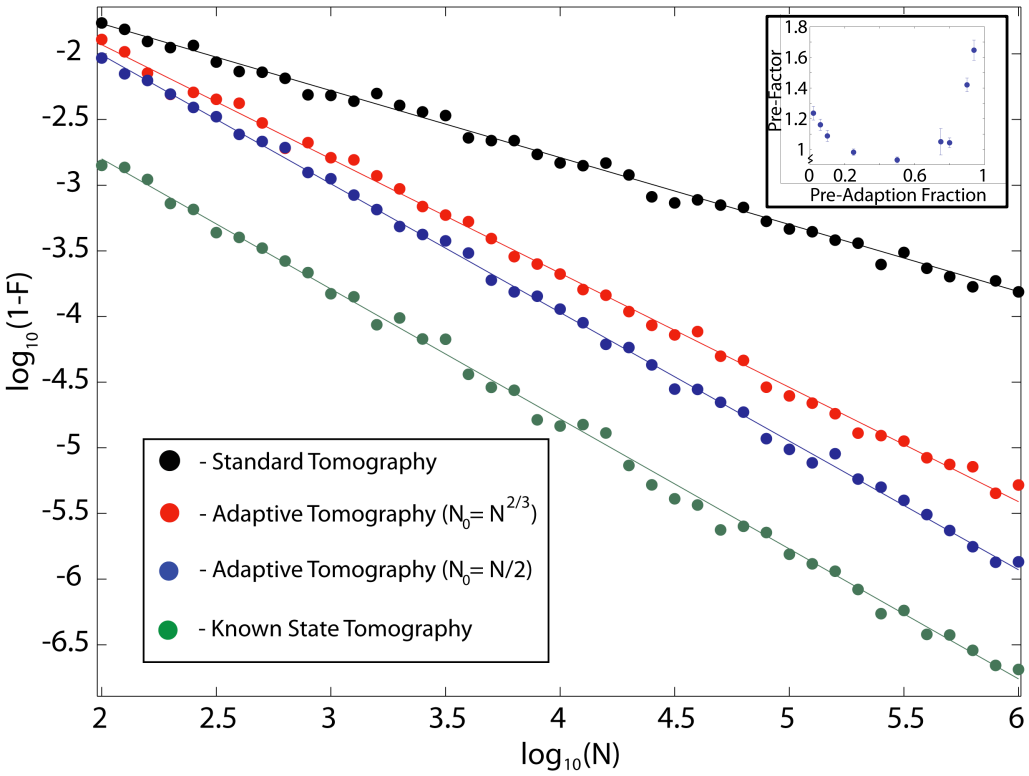

Figure 5.2. Average infidelity $1 - F(\widehat{\rho}, \rho)$ vs. sample size N for Monte Carlo simulations of four different tomographic procotocols: standard tomography (black), the procedure proposed in [1] using $N_0 = N^{2/3}$ (red), our procedure using $N_0 = N/2$ (blue), and "known basis" tomography (green). Both adaptive procedures clearly outperform static tomography, but our procedure clearly outperforms the $N_0 = N^{2/3}$ approach, and matches the asymptotic scaling of known-basis tomography. The inset shows the dependence of the prefactor (β) on $\alpha = N_0/N$.

To do adaptive tomography, we measured $N_0 = N/2$ photons, used the data to generate an ML estimate $\widehat{\rho}_0$, then rotated the measurement bases so that one diagonalized $\widehat{\rho}_0$. So, if the preliminary estimate is

$$\widehat{\rho}_0 = \lambda_1 \left|\psi_1\right\rangle\!\left\langle\psi_1\right| + \lambda_2 \left|\psi_2\right\rangle\!\left\langle\psi_2\right|,$$

we define $|\psi_{3/4}\rangle = (1/2)(|\psi_1\rangle \pm |\psi_2\rangle)$ and $|\psi_{5/6}\rangle = (1/2)(|\psi_1\rangle \pm i|\psi_2\rangle)$, and then measure the bases $\{\{|\psi_1\rangle, |\psi_2\rangle\}, \{|\psi_3\rangle, |\psi_4\rangle\}, \{|\psi_5\rangle, |\psi_6\rangle\}\}$. We measured the remaining $N - N_0$ photons

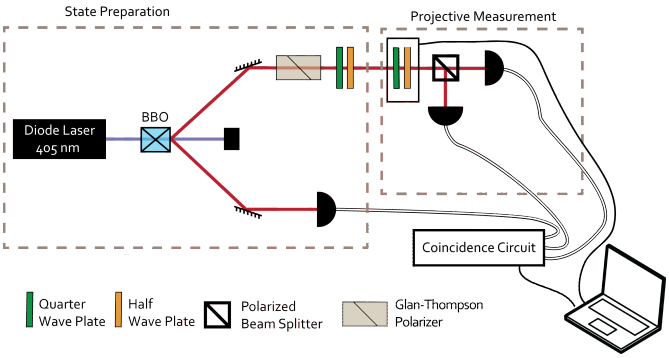

Figure 5.3. Spontaneous parametric downconversion is performed by pumping a nonlinear BBO crystal with linearly polarized light. One photon is sent directly to a detector as a trigger. A rotation using a quarter-half waveplate combination prepares the other photon in any desired polarization state. Finally, a projective measurement onto any axis of the Bloch sphere is performed by a quarter-half waveplate combination followed by a polarizing beamsplitter. The measurement waveplates are connected to a computer to enable adaptation.

in these new bases and constructed a final ML estimate using the data from both phases.

We fit a power law ($1 - F = \beta N^p$) to the average infidelity of each protocol (Fig. 5.4a), and found $p = -0.51 \pm 0.02$ for standard tomography, $p = -0.71 \pm 0.04$ for the procedure of Ref. [1], and $p = -0.90 \pm 0.04$ for our adaptive procedure.

Our data generally match the theory; adaptive tomography outperforms standard tomography by an order of magnitude even for modest ($\sim 10^4$) N. Experiments that achieve very low infidelities ($\sim 10^{-4}$) show small but statistically significant deviations from theory, which we believe can be explained by waveplate misalignment – fluctuations on the order of 10^{-3} radians are sufficient to reproduce the observed deviations in simulations. For a detailed discussion of systematic error and how it affects our results please see the supplementary material.

There is an even simpler adaptive procedure. After obtaining a preliminary estimate $\widehat{\rho}_0$, we measured *all* of the remaining $N/2$ samples in the diagonal basis of $\widehat{\rho}_0$, neglecting the second and third bases presented in the previous section's protocol. This *reduced adaptive tomography* procedure requires just one extra measurement setting (full adaptive tomography requires three), but achieves the same $O(\frac{1}{N})$ infidelity (Fig. 5.4b). The best fits to the exponent p in $1 - F = \beta N^p$ are $p = -0.51 \pm 0.02$ for standard tomography and $p = -0.88 \pm 0.05$ for reduced adaptive tomography (not significantly different from the results shown in Fig. 5.4a). In higher dimensional systems, reduced adaptive tomography should provide even greater efficiency advantages.

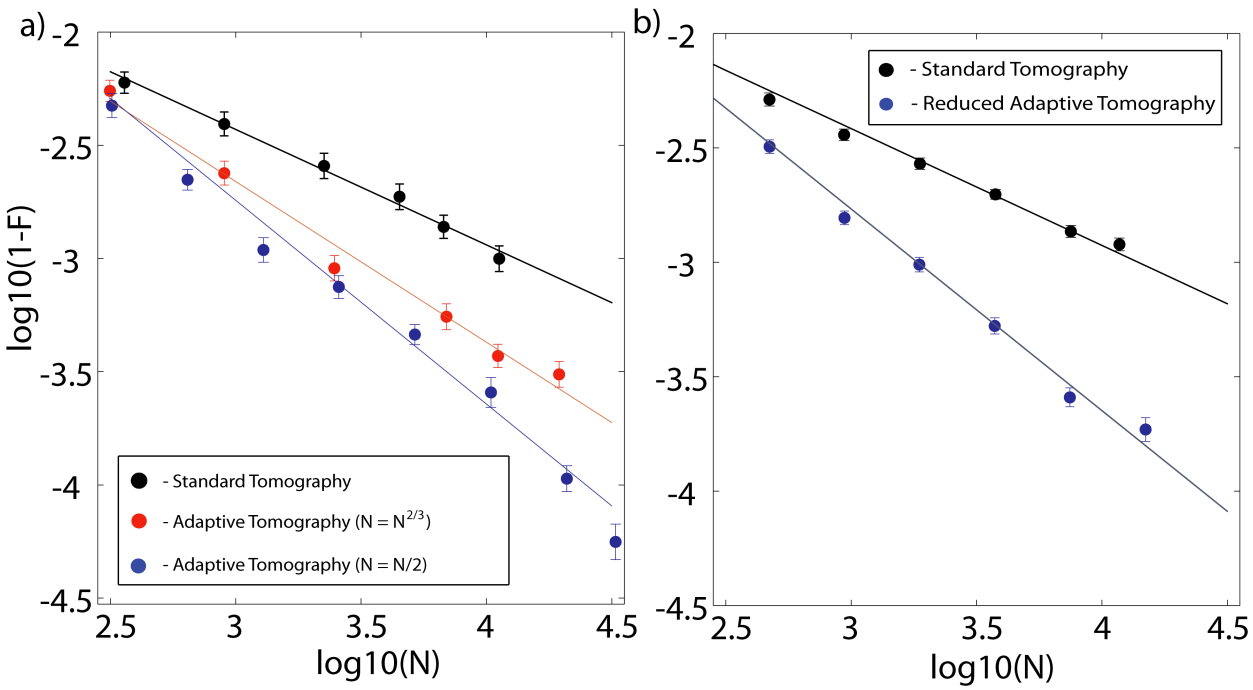

Figure 5.4. Experimental data: a) The average infidelity $1 - F(\widehat{\rho}, \rho)$ for the three tomographic protocols shown in Fig. 5.2 vs. the number of samples N. Each average is over 150 different realizations of the experiment. b) Average infidelity $1 - F(\widehat{\rho}, \rho)$ for standard tomography (black) and *reduced* adaptive tomography (blue) is plotted versus N. Each average is over 200 different realizations of the experiment; error bars are standard deviation of the mean of these samples. Error bars are standard deviation of the mean of these samples.

5.4 Discussion

We demonstrated two easily implemented adaptive tomography procedures that achieve $1 - F(\widehat{\rho}, \rho) = O(1/N)$ for *every* qubit state. In contrast, any static tomography protocol will yield infidelity $O(1/\sqrt{N})$ for most nearly-pure states. Our simplest procedure requires only one additional measurement setting than standard tomography. We see almost no reason not to use reduced adaptive tomography in future experiments.

Previous work [1] optimized average fidelity over Bures measure, a very respectable choice [30, 31, 32]. Unfortunately, the "hard-to-estimate" states lie in a thin shell at the surface of the Bloch sphere, whose Bures measure vanishes as $N \to \infty$. So although the scheme with $N_0 \propto N^{2/3}$ proposed in [1] achieves Bures-average infidelity $O(1/N)$, it achieves only

$O(1/N^{5/6})$ infidelity for nearly all of the (important) nearly-pure states [5]

The $O(1/N)$ infidelity scaling achieved by our scheme is optimal, but the constant can surely be improved – i.e., if our scheme has asymptotic error α/N, a more sophisticated scheme can achieve α'/N with $\alpha' < \alpha$. The absolutely optimal protocol requires joint measurements on all N samples [33], and will outperform any local measurement. There is undoubtedly some marginal benefit to adapting more than once, but we have shown that a single adaptation is sufficient to achieve $O(1/N)$ scaling.

We conclude with an observation that may surprise: adaptivity provides no advantage at all if inaccuracy is measured by trace-norm or 2-norm, which aren't hypersensitive to small variations in small eigenvalues. This does *not* undermine our result – it has a simple explanation. Trace-norm ($|\widehat{\rho} - \rho|_1$) quantifies *single-shot* distinguishability. When $N \gg 1$ samples are available, it becomes irrelevant. The relevant quantity is $|\widehat{\rho}^{\otimes N} - \rho^{\otimes N}|_1$, whose behavior is defined by the Chernoff bound [18], which in turn is well approximated by infidelity. So infidelity is a measure of *many-copy* distinguishability. Since tomography is necessarily concerned with $N \gg 1$ copies, the advantages of adaptivity *are* real, and hold for all many-copy metrics (e.g., relative entropy, Chernoff bound, etc.)

5.5 Sources of systematic error

Above, we attributed certain properties of our experimental data to "systematic errors". Here, provide more detail on the role that systematic errors play in our experimental results, and their interplay with static and adaptive tomography. "Systematic errors" is a broad term, incorporating almost everything that can go wrong with an experiment, so we consider several forms of it. We begin by briefly discussing *frame misalignment*, where adaptive tomography yields no advantage, but almost nothing else can either. We then consider some systematic errors in measurement that can be detected and mitigated, and demonstrate through simulations that they affect static tomography and adaptive tomography differently. Using a one-parameter fit and a model of our experiment, we show that waveplate-alignment errors of around 1.5×10^{-3} radians reproduce our experimental results remarkably well. To wrap up, we examine the asymptotic scaling of tomographic infidelity for three different models of systematic error, and conclude that adaptive tomography mitigates these forms of systematic error much better than static tomography can.

Frame Misalignment: The most systematic of errors is a fixed misalignment of reference frames. In our linear optics experiment, where a set of waveplates and a polarizing beamsplitter are used to measure the polarization state of light, this means that all the optical elements are misaligned by the same amount. Varying tomographic strategies can have no effect on this sort of error. In fact, this kind of frame misalignment cannot be de-

[5] Ironically, restricting the problem to pure states falsely trivializes it – the average *and* worst-case infidelity is $O(1/N)$ even for static tomography! The difficulty is not in estimating which pure state we have, but in distinguishing between small eigenvalues ($\lambda = 0$ vs $\lambda = 1/\sqrt{N}$).

tected at all within the experiment. It is equivalent to a change of gauge, has no operational consequences in this context, and is not interesting.

Instead, let us consider some errors that, while less "systematic" than frame misalignment, are also more detectable – and therefore potentially sensitive to different tomographic strategies.

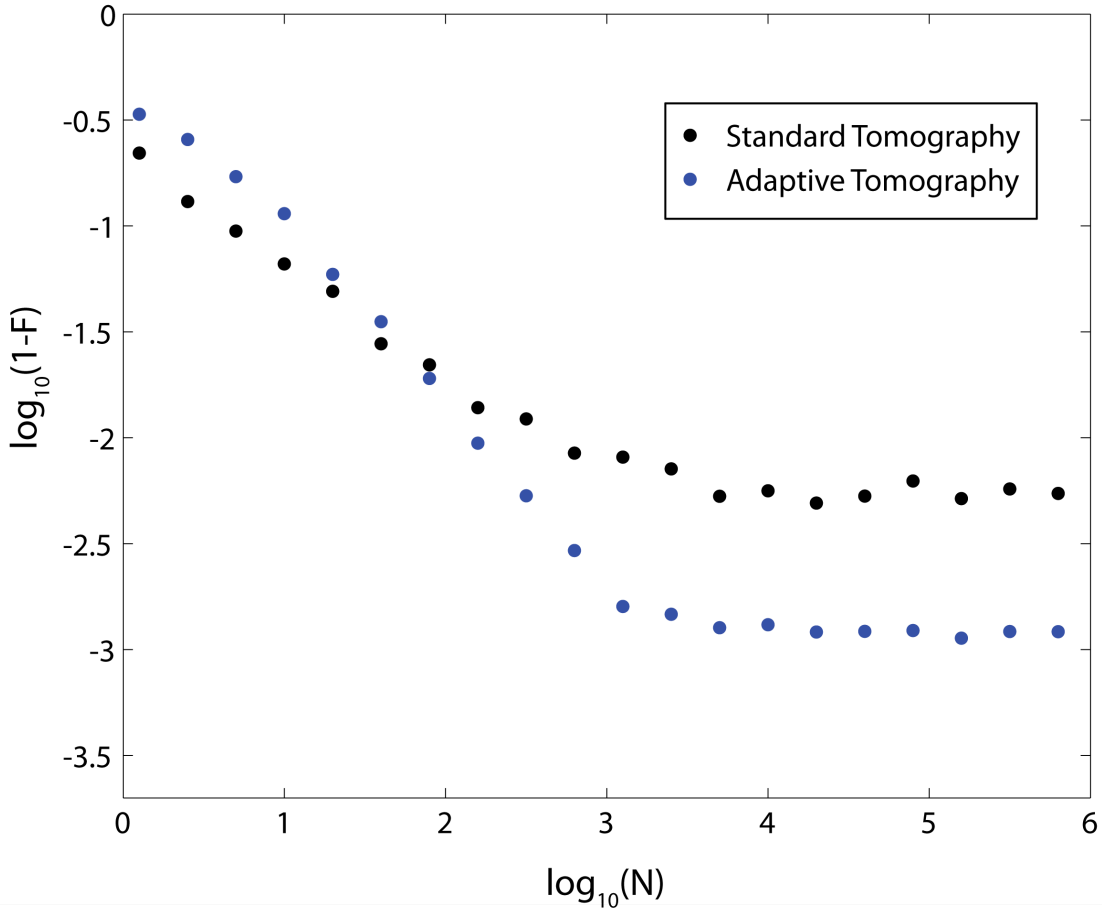

Figure 5.5. Average infidelity vs. sample size (N) for simulations with systematic errors (of Model 1 type; see text) on the order of $E = 10^{-2}$. The infidelity decreases with increasing N up to a point, after which it flattens out after hitting a 'noise floor'. The noise floor occurs at a lower average infidelity for adaptive tomography than for static tomography.

Waveplate Misalignment: A very important source of errors for our experiment is the alignment angle of the waveplates that measure photon polarization states. Our experiment's

reference frame is defined by the polarizing beam-splitter used to make the final projective measurement. To change the basis of the measurement, waveplates are mechanically rotated by a motor with good but finite accuracy. Every time we change the measurement basis, the motor's finite accuracy causes a slight misalignment of the waveplates – and the ensuing photodetections correspond to a measurement of a basis slightly different from the one we intended to measure.

Our terminology for discussing these errors is as follows. We performed (and, in this Supplement, we simulate) a large number of *experiments*. A single experiment (or *experimental run*) comprises the production of N identically prepared photons. Within an experiment, we implement several (3 to 6) *measurement settings*. Each measurement setting corresponds to (i) adjusting the waveplate, then (ii) measuring a large number of photons without any adjustments. In static tomography a single experiment includes three measurement settings (projections onto X,Y, and Z axes of the Bloch sphere), in each of which $N/3$ samples are measured. In adaptive tomography, a single experiment includes six measurement settings (the same initial set of three, and then three more in a rotated frame), each applied to $N/6$ states.

We consider three different models of systematic error in waveplate alignment.

1. **Model 1**. Each time a waveplate (whose purpose is to make a measurement) is moved to a new angle θ, it ends up instead aligned at angle $\theta + \delta\theta$, where $\delta\theta$ is a Gaussian random variable with zero mean and standard deviation E. Thus, in each experimental run, each measurement setting is misaligned by an independent random angle. This angle persists over many samples in the same experimental run, but not across multiple experimental runs.

2. **Model 2**. Each time a waveplate is moved, it misses its target θ by a random angle $\delta\theta$ that is fixed for each *experiment*, rather than for each measurement setting within the experiment. This model is mathematically equivalent to (and can be taken to represent) a small misalignment of the polarizing beam splitter. We take $\delta\theta$, which is fixed for each individual experiment, to be a Gaussian random variable with zero mean and standard deviation E.

3. **Model 3**. The waveplates are misaligned by an angle $\delta\theta$ that is fixed for each experiment (as in Model 2), but each experiment has the same fixed misalignment. In this model, $\delta\theta = E$ is not a random variable.

We believe that Model 1 best represents our experiment. The waveplate motors used in our experiment have an finite precision, and every time we change their angle, they return to a factory-set "home" position before realigning (which eliminates or at least minimizes correlation between successive alignment errors). Thus, the waveplate angle picks up a different random error each time it is moved to a new measurement setting.

We simulated the effect of Models 1-3 on adaptive and static tomography. In each simulation, 200 independent experimental runs were generated (each involving at least 3

measurement settings, with many identically prepared photons measured at each setting). We averaged the tomographic infidelity of these 200 runs to characterize the effect of random waveplate errors in each model.

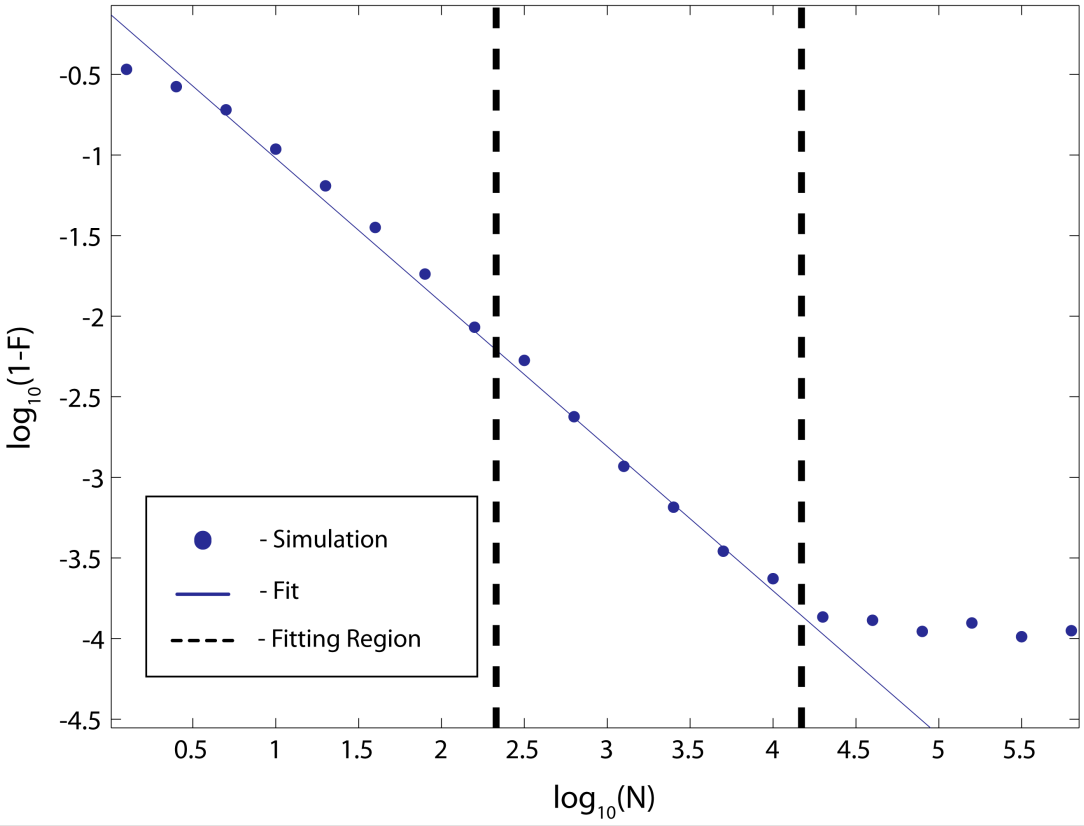

Figure 5.6. Average infidelity vs. sample size (N) for simulations of adaptive tomography with systematic errors (of Model 1 type; see text) on the order of $E = 10^{-3}$. Also plotted is the region over which experimental data was taken (see main body of text) and a line of best fit for this region.

Results: Figure 5.5 shows error (average infidelity) versus sample size (N) for Model 1 with $E = 0.5$ degrees ($\sim 10^{-2}$ radians). As the sample size increases, statistical errors decrease, and so average infidelity decreases. However, the error reaches a clear noise floor as systematic errors begin to dominate. It is higher for standard tomography than for adaptive tomography. We conclude that adaptive tomography is less sensitive than standard tomography to systematic errors.

Since the alignment errors vary randomly from experiment to experiment, an astute experimentalist might achieve higher accuracy by repeating each measurement setting many

(M) times, resetting the waveplate each time. This would work, reducing the infidelity by a factor of $1/\sqrt{M}$, but it adds significantly to the experimental difficulty and complexity. For example, if the waveplates have a precision of ~ 0.5 degrees, then when these systematic errors dominate over statistical errors (see Figure 5.5), the infidelity of static tomography saturates at 10^{-2}, while for adaptive tomography it saturates at 10^{-3}. Achieving the same 10^{-3} accuracy with static tomography would require repeating each measurement $M = 100$ times. Or, the experimentalist could just use adaptive tomography, and achieve it with only a single extra waveplate setting.

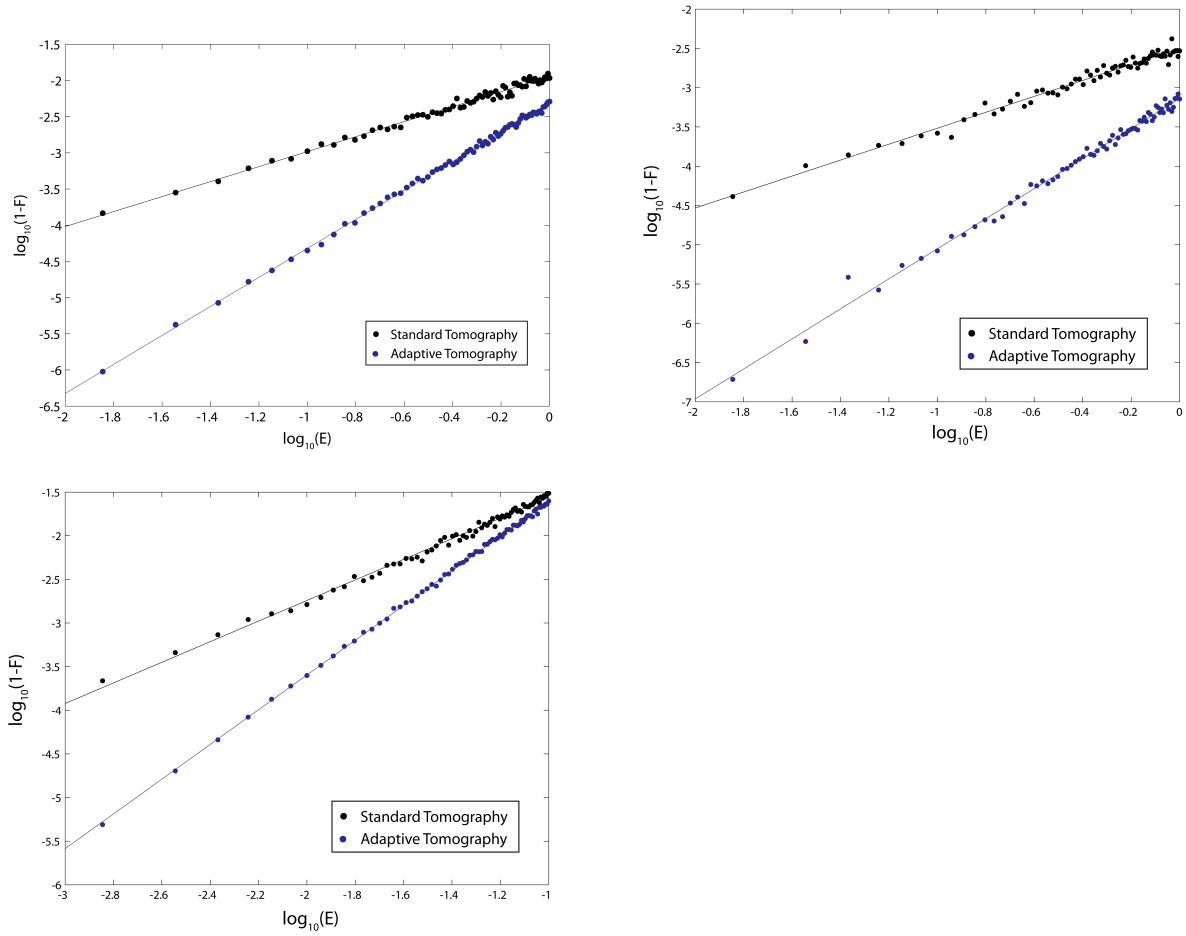

Figure 5.7. Location of the noise floor $(1 - F)$ for Models 1,2,3 (from top to bottom). For each of the three Models, we plot the average infidelity as $N \to \infty$ (to ensure that systematic errors dominate) vs. E (the magnitude of systematic error). Data points are results of simulation, and lines are lines of best fit.

We then performed several simulations of Model 1, in which we varied E (the magnitude of systematic error). For each value of E, we fit a line to the $1-F$ vs. N curve *over the same*

range of N that we observed in our experiment. This is shown in Figure 5.2. A simulation with $E = 0.15$ degrees yielded results almost indistinguishable from our experimental data. In the experimentally observed region, the line of best fit has a slope of -0.895 ± 0.023,

$$1 - F \propto N^{-0.895 \pm 0.023},$$

which matches our experimental data very well.

Finally, we investigated the value of the noise floor for Models 1-3 (Figure 5.7). Each plot in Figure 5.7 shows $1 - F$ (average infidelity) vs. E (magnitude of systematic errors) on a log-log plot, for both standard and adaptive tomography. We examined sufficiently high N to guarantee that systematic errors dominate.

1. For Model 1, the line of best fit (on a log-log plot) for standard tomography has a slope of 1.03 ± 0.01 and the line of best fit for adaptive tomography has a slope of 2.00 ± 0.01.
2. For Model 2, the slopes of the lines of best fit are 1.01 ± 0.02 and 1.91 ± 0.02.
3. For Model #3, the slopes of the lines of best fit are 1.18 ± 0.01 and 1.99 ± 0.01.

We conclude that in all three models of systematica error that we considered here, it's fair to say that average infidelity scales *linearly* with E for standard tomography, and *quadratically* with E for adaptive tomography. Adaptive tomography is substantially more robust to systematic errors than standard tomography – not just by a constant factor, but qualitatively so.

Conclusion: We have shown that for three reasonable models of systematic errors, the average infidelity of adaptive tomography scales with E^2 and the average infidelity of static tomography scales with E, where E is the magnitude of these errors. Infidelity is very sensitive to spectral errors (i.e., changes in the eigenvalues of the estimated density matrix), but not to unitary errors (changes in the eigenvectors). The primary result of systematic errors in the measurement basis – i.e., measuring the wrong basis by an angle E – is a unitary error in the estimate by $O(E)$. As we have shown in the main text, adaptive tomography measures the eigenvalues of the density matrix to much higher precision than static tomography. Furthermore, adaptive tomography (because it specifically seeks to measure the diagonal basis of ρ) still obtains an accurate estimate for the eigenvalues even in the presence of systematic error. Even if we get the basis wrong by an angle of $O(E)$, this only affects the measurement probabilities (and therefore the estimated spectrum) by $O(E^2)$.

Chapter 6

$SU(2)$-covariant probe states for quantum process characterization

The goal of quantum metrology is to measure or detect physical phenomena with surprising precision by exploiting quantum resources. Often, this means using entangled states to achieve greater resolution or sensitivity. For example, squeezed light[34] and N00N states [35, 36, 37] have been used in interferometers to achieve higher precision in *single parameter* estimation. N00N states are maximally sensitive to small $U(1)$ phase shifts [38], but they are fragile. Other parameters might be best detected or estimated by a different optimal state [39, 40], and for estimating a even a simple three-parameter SU(2) process the optimal state is unknown [41]. At the opposite extreme of metrology is *quantum process tomography* (QPT) [42, 43]. Here, the goal is to learn *every* parameter of an unknown process. QPT requires a diverse set of probe states, and the overall accuracy of estimation depends on the properties of the entire set. For process tomography on a single quantum optical mode, Lobino *et al.* [44] showed that it is sufficient to (1) prepare a single Glauber coherent state and (2) displace it by a variety of phase space translations. This approach, in which a single "fiducial" state is multiplied into a complete set of probe states by easily implemented group transformations, has the great merit of experimental ease. But while sufficiently large coherent-state ensembles are sufficient for process tomography, they are not efficient. Coherent states are very "classical" [45], and provide exponentially little information about parameters of some quantum processes, motivating a search for set of states for tomography which provide equal information about all possible processes (see [46] for a precise statement of this problem).

In this chapter, we examine a closely related question for 2-photon polarization ("biphoton") states. Like an optical mode, this system admits [spin]-coherent states (as well as others). The corresponding symmetry group, $SU(2)$, is transitive on the set of coherent states, i.e. a spin-coherent state can be transformed into any other spin-coherent state by applying a polarization (SU(2)) rotation. We prepare a wide range of probe states, and quantify their performance at two opposite extremes of the metrology spectrum: (1) their ability to *detect* random $SU(2)$ phase shifts, and (2) their ability to *characterize* an unknown process, when displaced by a variety of $SU(2)$ rotations and used for QPT. Remarkably, the most sensitive detector states (N00N states) are also among the *least* effective for QPT! The optimal $SU(2)$-covariant set (i.e. a set generated by applying SU(2) operations to a fiducial state) for QPT is generated by a state that is neither spin-coherent nor N00N, but outperforms both of them. When displaced by uniformly random $SU(2)$ operations, it generates a 2-design [47, 48], confirming the theoretical prediction that 2-designs should be optimal for process tomography.

We focus on a particular, important family of processes that we call $SU(2)$ *jitter*. In $SU(2)$ jitter, an N-photon state experiences a small random collective $SU(2)$ rotation, whose magnitude is Gaussian-distributed. Detecting and characterizing $SU(2)$ jitter is important because it is a common model for *decoherence*[49], the primary enemy of quantum information and computation [50]. Noiseless subsystems [51, 52] were designed against this noise model.

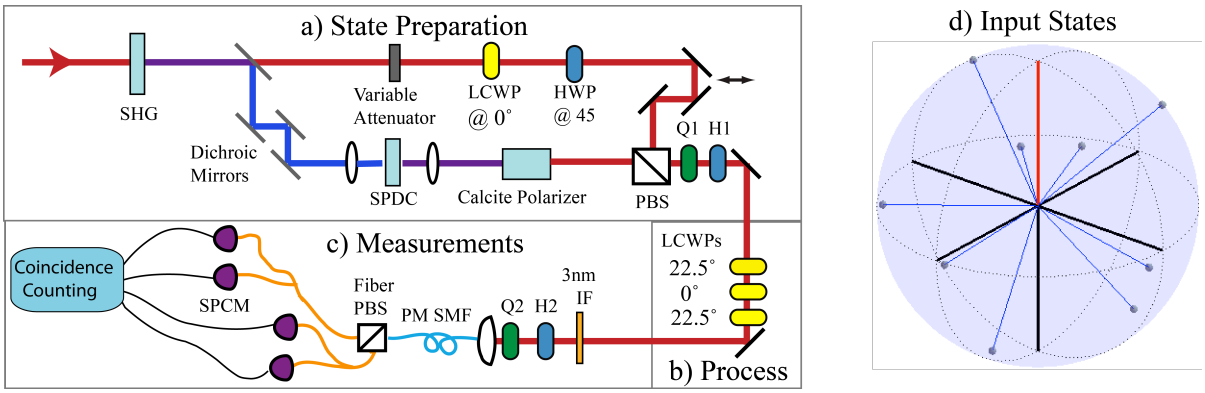

Figure 6.1. Schematic diagram of the experimental apparatus used to generate and measure different biphoton states. **a) State Preparation** – We can prepare any biphoton state in two steps. First, we prepare a state of the form $\sqrt{x}\,|2,0\rangle_{H,V} + e^{i\phi}\sqrt{1-x}\,|0,2\rangle_{H,V}$ by using a polarizing beam splitter (PBS) to combine a vertically polarized weak coherent state with the output of horizontally polarized type-I collinear down-conversion. The amplitude x is set by attenuating the laser intensity relative to the down-converted intensity. The angle ϕ is set by the relative phase between the two paths, which we control using a liquid-crystal wave plate (LCWP). Then, we apply any desired polarization rotation in $SU(2)$, using quarter- and half-waveplates, to produce any desired biphoton state. **b) Process** – 3 LCWPs, oriented as shown, are used to perform arbitrary polarization rotations. To implement depolarization, the retardances of the LCWPs are made to fluctuate during each measurement. **c) State Measurement** – The biphoton states pass through wave plates, are coupled into a polarization-maintaining fiber and sent to a polarizing beam splitter. The output of each port of the beam splitter is probabilistically split, using 50:50 fiber beam splitters, and sent to single-photon counting modules. **d) Input States** – A graphical representation of the set of states used for process tomography. A fiducial state is prepared and rotated to nine other states. These states (including the unrotated fiducial state) make up the set of ten input states used for process tomography. The fiducial state is represented by the red line, and it is rotated to each of the nine other points on the sphere. The rotations are chosen to be (approximately) uniformly distributed on the surface of the sphere.

6.1 Experimental Methods and Preliminary Validations

A *biphoton* is a system of two photons in the same spatial and temporal modes[53, 54]. Basis states can be described by specifying the number of photons polarized horizontally (H) and vertically (V), e.g.

$$C_0 \left|2,0\right\rangle_{H,V} + C_1 \left|1,1\right\rangle_{H,V} + C_2 \left|0,2\right\rangle_{H,V}. \tag{6.1}$$

A single-photon polarization state is a two-level system and it is isomorphic to a spin-1/2 particle. We can define a single horizontally-polarized photon to be a spin-up spin-1/2 particle and a vertically-polarized photon to be a spin-down spin-1/2 particle. Similarly, the polarization state of a biphoton is isomorphic to a spin-1 particle [55]. Hence we can write a general biphoton state in the angular-momentum basis $\{\left|J=1,m=+1\right\rangle, \left|J=1,m=0\right\rangle, \left|J=1,m=-1\right\rangle\}$ as

$$C_0 \left|1,+1\right\rangle + C_1 \left|1,0\right\rangle + C_2 \left|1,-1\right\rangle, \tag{6.2}$$

where the equivalence between the two biphoton bases can be seen by comparing equations 6.1 and 6.2. A collective polarization rotation corresponds to an $SU(2)$ rotation of the effective spin-1 particle about some axis $\vec{r}$ by an angle θ.

Two photons polarized in the same direction form a *spin-coherent* state, and these spin-coherent states are analogous to the Glauber coherent states of an optical mode [56]. Just as [44] used displaced Glauber coherent states as input states for QPT, a set of at least nine distinct spin-coherent states can form a complete probe set for QPT, and can be generated by applying various $SU(2)$ rotations to a single fiducial spin-coherent state. Here, we generalize this procedure in a simple way: we prepare a fiducial state that is *not* necessarily a spin-coherent state, and generate candidate probe sets for QPT by applying 10 distinct $SU(2)$ rotations to it. (We prepare 10 fiducial states instead of 9 because it is experimentally convenient, and provides a small amount of useful redundancy). Our fiducial states take the form

$$\left|\psi_x\right\rangle = \sqrt{x}\left|2,0\right\rangle_{H,V} + \sqrt{1-x}\left|0,2\right\rangle_{H,V}, \tag{6.3}$$

and are prepared using the apparatus sketched in Fig. 6.1a and the methods described in [57, 58, 59, 60, 61, 62].

In brief, we prepare our biphoton states by combining vertically-polarized laser light with horizontally-polarized collinear down-converted light into the same spatial mode by overlapping them at a polarizing beamplitter. We then post-select on two-photon events in this mode. Since down conversion only produces photon in pairs we can only ever detect two horizontally-polarized down-converted photons or two vertically-polarized laser photons. If the photons from both sources are indistinguishable this process will result in the superposition state of Eq. 6.3. To set x we tune the amplitude of the laser light while keeping the amplitude of the the down-converted light fixed. We actively stabilize the phase between the laser light and the down-converted light, allowing for data runs lasting several hours [62]. The class of biphoton states which we can prepare (Eq. 6.3) includes spin-coherent states ($x = 0, 1$) and the two-photon N00N state ($x = \frac{1}{2}$). In fact, *any* biphoton state can be prepared by choosing some value of x and then applying some $SU(2)$ rotation using wave

plates. (N-photon states have $N-1$ parameters that are $SU(2)$-invariant; for the biphoton, x is the only parameter.)

The horizontally- and vertically-polarized photons are not perfectly mode-matched when generated. We remedy this by passing them through a 3nm filter and coupling them into a single-mode fiber, which discards any photons not in the desired mode. This procedure results in near-perfect biphotons, as quantified by the [very small, $< 2\%$] amount of population in the anti-symmetric subspace. We characterize our state preparation by doing quantum state tomography using the apparatus of Fig 6.1c as described in Ref. [54, 55]. To visualize quantum polarization states on the Poincare sphere we must use a quasi-probability distribution because arbitrary N-photon polarization states cannot be represented as a point on the Poincare sphere (as can be done with classical light). We use Wigner distributions as defined in Ref. [63]. Our resulting experimentally-measured biphoton states, after numerically filtering out the anti-symmetric subspace, are depicted in the first column of Fig. 6.2 as Wigner distributions plotted on the Poincare sphere [64]. As x is increased, the states become more "nonclassical", with the most nonclassical state being the N00N state with $x = 0.5$ (See row (c), column 1 of Fig. 6.2). This procedure lets us prepare any desired state with fidelity $\geq 93\%$.

The process we study, depolarization due to fluctuating $SU(2)$ rotations, manifests itself in many experimental systems – e.g., a spin in a fluctuating magnetic field, or a polarization state propagating through an optical fiber which thermally fluctuates with time. In both cases, the state (the spin state or the polarization state) couples to an external system (a time-varying field or a time-varying birefringence), causing depolarization. For example, a spin is coupled to a magnetic field via the Hamiltonian $\hat{H} = -\mu \vec{J} \cdot \vec{B}$, where $\vec{B} = B_o \vec{r}$ (the direction of the magnetic field is $\vec{r}$ and the strength of the magnetic field is B_o). This has the effect of rotating the spin about an axis $\vec{r}$ by an angle $\theta = \mu B_o$, as described by the unitary operator $\hat{U} = e^{-i\theta \frac{\vec{r}\cdot\vec{J}}{\hbar}}$. If $\vec{B}$ changes with time, then the output state of this process can be described by properly averaging over $\vec{r}$ and θ [65]. (Polarization states are similarly rotated by a birefringent refractive index, where θ is related to the strength of the birefringence and $\vec{r}$ to the orientation of optical axis.) In $SU(2)$ jitter we consider isotropic depolarization (by which we mean that there is no preferred rotational axis) and we assume that $\vec{r}$ and θ are uncorrelated. In this case, the quantum process is:

$$D_\gamma[|\psi\rangle\langle\psi|] = \int d\vec{r} \int d\theta P(\theta) e^{-i\theta \frac{\vec{r}\cdot\vec{J}}{\hbar}} |\psi\rangle\langle\psi| e^{+i\theta \frac{\vec{r}\cdot\vec{J}}{\hbar}}, \tag{6.4}$$

where the rotation axis $\vec{r}$ is uniformly random and the angle θ has a distribution, $P(\theta)$, which we take to be Gaussian:

$$P(\theta) \propto e^{-\frac{\theta^2}{2\gamma^2}}. \tag{6.5}$$

The overall strength of the depolarization process is quantified by γ, the width of the distribution of θ.

We implement the depolarization process described by Eq. 6.4 using three liquid crystal wave plates (LCWP) as shown in Fig. 6.1b. Each LCWP applies an adjustable polarization rotation, and the rotation angle can be changed very rapidly. This allows us to easily apply

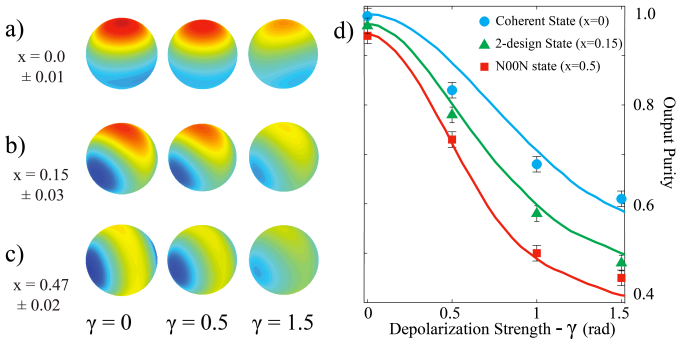

Figure 6.2. Typical results of state tomography, and the effect of $SU(2)$-jitter depolarization. In plots **(a)-(c)**, we show experimentally reconstructed Wigner functions (plotted on the Poincare sphere) for three different states after they have been decohered by three different amounts of $SU(2)$ jitter. Row (a) shows spin-coherent states with x=0, row (b) shows "2-design" states with x=0.15, and row (c) shows nearly-N00N states with x=0.47. Each row shows the effect of applying depolarization with strength (see Eq. 6.4) $\gamma = 0, 0.5, 1.5$. In **(d)**, we plot the *purity* of the same reconstructed states shown in (a-c). The solid lines are the theoretical predictions given by simulations of the process (Eq. 6.4).

50 different (randomly selected) rotations to implement a single process, closely approximating the ideal process of equation 6.4 as a statistical mixture of these 50 random unitary processes as:

$$\tilde{D}_\gamma[|\psi\rangle\langle\psi|] = \sum_{k=1}^{50} e^{-i\theta_k \frac{\vec{r}_k\cdot\vec{J}}{\hbar}} |\psi\rangle\langle\psi| e^{+i\theta_k \frac{\vec{r}_k\cdot\vec{J}}{\hbar}}. \tag{6.6}$$

Using this implementation we can apply precisely calibrated $SU(2)$ jitter. We verified both our state preparation *and* our implementation of depolarization by performing state tomography on $\rho = \tilde{D}_\gamma[|\psi_x\rangle\langle\psi_x|]$ for several values of x and γ. Figure 6.2a-c shows the Wigner functions of the resulting reconstructed states. Plots in column 1 are for undecohered states, while columns 2-3 show the effects of $SU(2)$ jitter with strengths of $\gamma = 0.5$ and $\gamma = 1.5$ rad (respectively). Increasing γ blurs the Wigner function. This is captured quantitatively by the state's *purity*, a reasonable proxy for the amount of depolarization suffered. Figure 6.2d plots the output purity (computed from the tomographic estimate) versus γ, and compares it to the prediction of numerical simulations of $\tilde{D}_\gamma$ (solid lines), for three different input states with $x = 0$ (blue), $x = 0.15$ (green) and $x = 0.47$ (red). The only inputs to our simulation are the experimentally measured purities of the input states when $\gamma = 0$, which in a perfect experiment would be 1, but are slightly degraded by experimental noise. We observe excellent agreement between simulation and experiment, confirming that our process performs as expected. In particular, the N00N state loses purity more rapidly than any other state (as γ is increased), indicating that N00N states are indeed the most fragile (and thus potentially sensitive) to $SU(2)$ jitter. The fragility of N00N states to a similar model of SU(2) depolarization was also pointed out in [49]. In the next section we discuss an experiment exploiting this fragility to detect depolarization.

6.2 Detecting depolarization

We examined different probe states effectiveness at *detecting* $SU(2)$ jitter. This corresponds to distinguishing between two processes: $\mathbb{1}$ (no depolarization) or D_γ ($SU(2)$ jitter). Acting on the probe state $|\psi_x\rangle\langle\psi_x|$, these alternatives produce either $|\psi_x\rangle\langle\psi_x|$ or $\rho_D = D_\gamma[|\psi_x\rangle\langle\psi_x|]$, and to distinguish these alternatives we simply perform a POVM measurement with two outcomes,

$$\mathcal{M} = \{|\psi_x\rangle\langle\psi_x|, \mathbb{1} - |\psi_x\rangle\langle\psi_x|\}. \tag{6.7}$$

In simple terms, we are checking to see whether the probe state changed at all. If done perfectly, this protocol has one-sided error; it may fail to detect D, but will never detect it in error.

To implement this measurement experimentally, we recall that if $|\psi_x\rangle$ is a spin-coherent state then it can be written as $\hat{U}|2,0\rangle_{H,V}$ for some $\hat{U} \in SU(2)$. We can implement $\hat{U}^\dagger$ using the quarter- and half-wave plates labeled Q2 and H2 in Fig. 6.1c, and after performing this inverse rotation on the output state, detection of two photons at the H-port of the PBS corresponds to the $|\psi_x\rangle\langle\psi_x|$ outcome of $\mathcal{M}$. Similarly, if $|\psi_x\rangle$ is a N00N state, it can be

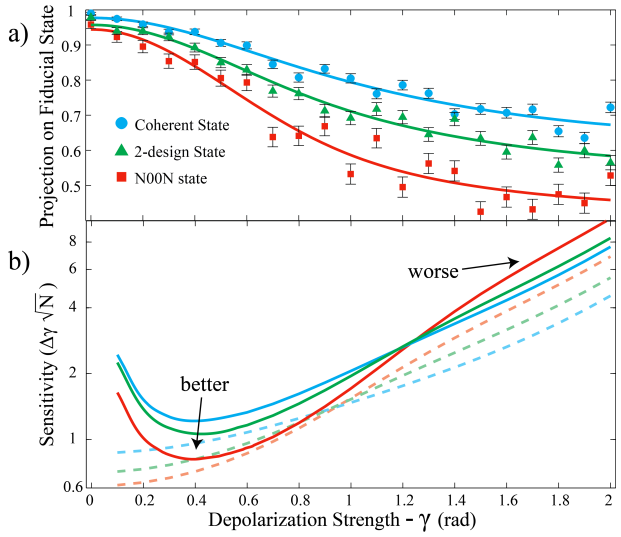

Figure 6.3. Sensitivity of different states to depolarization: This figure shows two different measures of the probe state's ability to detect depolarization, for three different probe states, and compares theoretical predictions to experimental data. Plot **(a)** shows the probability that depolarization is *not* detected, which is simply the probability of finding the system in the same state in which was prepared despite depolarization having happened. The theory (solid line) is simply the projection $\langle\psi_x|\,\rho\,|\psi_x\rangle$, where $|\psi\rangle_x$ is the probe state and ρ is the decohered probe state. Experimental data points are empirical probabilities of nondetection. Plot **(b)** shows the *sensitivity* (Eq. 6.12) scaled by $\sqrt{N}$ (where N is number of two-photon counts detected at each point) of each state to small changes in depolarization, the solid lines are calculated from the slopes of the fit to the detection data shown in (a), and the dashed lines are calculated for ideal input states.

written as

$$\begin{aligned} |\psi\rangle_x &= \frac{1}{\sqrt{2}}\left(|2,0\rangle_{H,V} + |0,2\rangle_{H,V}\right) \\ &= U\,|1,1\rangle_{H,V} \end{aligned}$$

where $U^\dagger$ can be implemented by a HWP at 22.5°. Thus, after performing this inverse rotation on the output state, detection of 2 coincident photons at the H- and V-ports of the PBS corresponds to the $|\psi_x\rangle\langle\psi_x|$ outcome of $\mathcal{M}$.

In our experiment, for states with $x \neq 0$ or $x \neq 0.5$, we estimated the value $\langle|\psi_x\rangle\langle\psi_x|\rangle$ indirectly from two measurements. In general the density matrix describing the system at the output will be:

$$\rho = \begin{pmatrix} a & f & d \\ f^* & b & g \\ d^* & g^* & c \end{pmatrix}, \tag{6.8}$$

where ρ is written in the same basis as the state in equation 6.3. If the process is pure SU(2) depolarization, and the input state is given by equation 6.3 then d will be real. (We check that this is the case by performing quantum state tomography on ρ for several depolarization strengths.) In this case, the expectation value of the projection onto $|\psi_x\rangle$ is

$$\langle|\psi_x\rangle\langle\psi_x|\rangle = \mathrm{Tr}(|\psi_x\rangle\langle\psi_x|\,\rho) = ax + c(1-x) + 2d\sqrt{x(1-x)}. \tag{6.9}$$

So we can estimate $\langle|\psi_x\rangle\langle\psi_x|\rangle$, for any value of x, by measuring a, c, and d. If ρ is sent directly to a PBS both photons will be transmitted with probability a, both will be reflected with probability c, and one will exit each port with probability b. All of these probabilities can readily be measured via coincident detection between different combinations of the four detectors in fig 6.1c. To measure d, a half-waveplate at 22.5° is inserted before the PBS. Now one photon will exit each port of the PBS with probability

$$P_{HV} = \frac{1}{2} - d - \frac{b}{2}. \tag{6.10}$$

Since b is already known, measuring P_{HV} gives us an estimate of d, which gives us enough information to reconstruct $\langle|\psi_x\rangle\langle\psi_x|\rangle$ for any value of x. While our "two-step" method works, it is possible to directly perform this projection for any biphoton state. One way to do this would by time reversing the state preparation techniques of [58, 66, 36]), where it is shown how arbitrary N-photon polarization states (occupying the a single spatial mode) can be built up by combining N photons one-at-a-time into the same mode. The polarization of each individual photon sets the N-photon polarization state. The time reverse of this process would be to split a single-mode N-photon polarization state into N separate spatial modes (probabilistically), and to detect a single photon with a specific polarization in each mode. Thus this would correspond to projecting onto a state set by the N different polarization measurements.

We note in passing that this protocol is reminiscent of atomic interferometry. There, too, a probe state is prepared and then measured later. The probability (and statistics) of the

results typically oscillate over time, because different atomic states have different energies and accumulate quantum phases that beat against one another. Depolarization makes these oscillations decay and eventually disappear, and this decay is often used to estimate the depolarization strength in the system. Thus, in both interferometry and our experiment, greater fragility to depolarization enables detection of weaker depolarization.

We created a range of probe states parameterized by $x \in [0, \frac{1}{2}]$, varying from a spin-coherent state at $x = 0$ to a N00N state at $x = \frac{1}{2}$ (Fig. 6.1a), subjected them to depolarization (Fig. 6.1b), and performed the measurement $\mathcal{M}$ given in Eq. 6.7 (Fig. 6.1c).

Figure 6.3a shows the experimentally observed probability of failing to detecting $SU(2)$ jitter as a function of the jitter's strength (γ), for three representative probe states: a N00N state (red), a spin-coherent state (blue) and an intermediate state with $x = 0.15$ (green). The experimental results are in good agreement with the simple theoretical prediction (solid lines) given by

$$\begin{aligned} P(\gamma) &= \langle\psi_x| D\left[|\psi_x\rangle\langle\psi_x|\right] |\psi_x\rangle \\ &= A_x e^{-2\gamma^2} + B_x e^{-\frac{\gamma^2}{2}} + C_x, \end{aligned} \tag{6.11}$$

where A_x, B_x and C_x are straightforward but unwieldy functions of x (see equation 4 in section A of the Supplemental Material for their form). We see that the N00N state is consistently the best detector of $SU(2)$ jitter, and that detection probability for any fixed depolarization strength appears to increase (as expected) monotonically with x.

Metrology is also concerned with *estimating* (rather than just detecting) parameters of a process. In this case, that means estimating γ, and this requires repeating the experiment more than once, since a single experiment can at best detect that $\gamma > 0$. We can then estimate the probability plotted in Fig. 6.3a, e.g. as $\hat{P} = n/N$, where the experiment was repeated N times and depolarization was detected in n of them. Armed with our knowledge of the initial state and our estimate $\hat{P}$ of the nondetection probability, we can then estimate γ (e.g., by simply inverting the appropriate theoretical curve shown in Fig. 6.3a. Of course, our estimate ($\hat{\gamma}$) will have some uncertainty:

$$\gamma = \hat{\gamma} \pm \Delta\gamma.$$

$\Delta\gamma$ is the smallest change in γ that can be detected with reasonable probability. We refer to it as *sensitivity* (although it should be noted that smaller $\Delta\gamma$ implies greater sensitivity!), and it is given by [67]

$$\Delta\gamma = \frac{\Delta P(\gamma)}{\frac{dP(\gamma)}{d\gamma}}, \tag{6.12}$$

where $\Delta P(\gamma)$ is the standard deviation of the *estimated* nondetection probability $\hat{P}$. Since our detection protocol is a Bernoulli (coin-flip) process, $\Delta P(\gamma) = \sqrt{\frac{P(\gamma)(1-P(\gamma))}{N}}$. Both $P(\gamma)$ and $\frac{dP(\gamma)}{d\gamma}$ depend only on x (a property of the probe state) and can be computed from Eq. 6.11.

In Fig. 6.3b, we show the dependence of sensitivity, scaled by $\sqrt{N}$ (where N is the number of times the experiment is repeated) on γ and the probe state. We scale the sensitivity by $\sqrt{N}$ that it only depends on the input state and γ. Since sensitivity is not a directly observable quantity, we compare a pure theory prediction to an empirical fit. Dashed lines are pure theoretical predictions, in which both $P(\gamma)$ and $dP(\gamma)/d\gamma$ are calculated using Eq. 6.11 for *ideal* input states. The solid lines are empirical fits: both $P(\gamma)$ and $dP(\gamma)/d\gamma$ are calculated from the smooth empirical fits to data shown as solid lines in Fig. 6.3a.

Whereas N00N states are always the best *detectors* of depolarization, we observe that they are only the best at *measuring* γ for low values of γ. Around $\gamma \approx 1$, they become less sensitive than the other states. This is a direct consequence of their extreme fragility; since almost any amount of depolarization disturbs the N00N state, it does not distinguish well between medium and strong depolarization. We also observe a significant discrepancy between the ideal sensitivity (dashed lines) and the observed value at small γ. This is because our input states are imperfect and not pure – even at $\gamma = 0$, there is a small probability that we will (falsely) detect depolarization! Thus, $P(\gamma)$ is never 1, and $\Delta P(\gamma) \not\to 0$ as $\gamma \to 0$. Still, we find that N00N states are clearly optimal for $\gamma \lesssim 0.9$, beating spin-coherent states by a factor of 1.44 ± 0.05, which agrees well with our theoretical prediction of $\sqrt{2}$ (see section A of the Supplemental Material for derivation).

6.3 Quantum Process Tomography

Using the protocol above, we can *detect* depolarization, and we can even *quantify* its strength. But to do so reliably, we needed to assume that the process is of a specific one-parameter form (pure $SU(2)$ jitter). For example, a consistent and coherent $SU(2)$ rotation would violate this assumption, and might go entirely undetected or be incorrectly diagnosed as jitter (depending on the axis of the measurement and the nature of the probe state). Characterizing general depolarization processes, and correctly diagnosing what is happening, requires quantum process tomography (QPT) [42, 43].

QPT requires not one, but an ensemble of input states – and preparing a suitable ensemble can be quite challenging. We avoid this complexity by using the ideas of Lobino *et al.* [44], and generating diverse input states by applying diverse simple transformations to a single fiducial state (see Fig. 6.1d). Lobino *et al.* prepared optical coherent states by translating the $|0\rangle$ state. The equivalent protocol in our biphoton system would be to prepare the $|2,0\rangle_{H,V}$ state and then generate an ensemble of spin-coherent states by performing different $SU(2)$ rotations on it. However, we go one step further and generalize this process by varying the fiducial state (parameterized, again, by $x \in [0,1]$). We study the dependence of process reconstruction fidelity on x, to determine (in particular) whether the N00N states that best *detect* depolarization are also the most best probe states to *characterize* it.

QPT reconstructs the entire process matrix (or *superoperator*) from the observed measurement statistics. This reconstruction, in essence, involves solving a set of linear equations

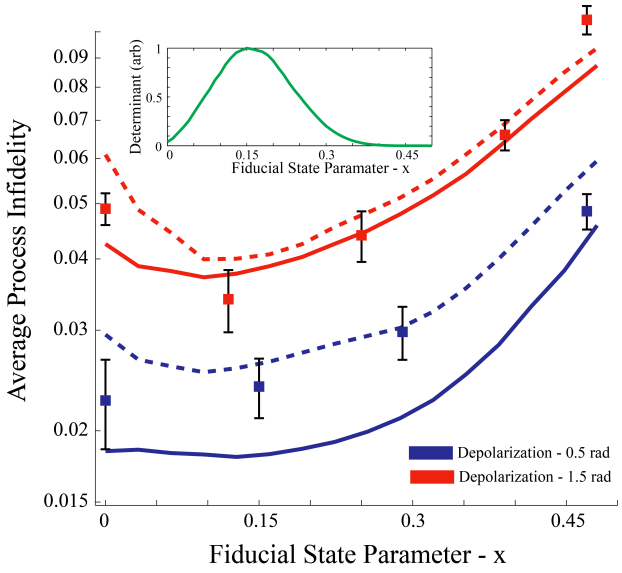

Figure 6.4. Inaccuracy of process reconstruction. This figure shows the *average process infidelity* (API, see Eq. 6.13 – smaller is better) between the true process and its tomographic reconstruction, and its dependence on the fiducial state used to generate the set of probe states for QPT. Solid lines are simulations of the experiment using 10 randomly oriented *pure* input states, while dashed lines are simulations using the 10 experimental input states (as determined using state tomography, and in particular accounting for the decline in their purity as x increases). Squares are experimentally estimated process infidelities. Red and blue represent different depolarization strengths. **INSET: "Completeness" of input sets:** The inset plot shows the determinant of the probe states' Gram matrix (normalized to a maximum of 1), whose inverse appears in the tomographic reconstruction. It depends on x; larger determinants yield a more robust inversion, while zero determinant indicates a tomographically incomplete set. We observe that a N00N fiducial state ($x = \frac{1}{2}$) yields a probe set that is not tomographically complete, while for $x \approx 0.15$ the reconstruction should be as robust and accurate as possible. The main plot confirms this theoretical prediction.

described by a the Gram matrix M of the input states. The reconstructed process is obtained by applying M^{-1} to a vector of observed statistics. M must obviously be full rank, but (moreover) it should not have any small eigenvalues. If M has small eigenvalues, M^{-1} will amplify small statistical fluctuations that result from finite sample size into large errors, and the reconstructed process will have low fidelity with the true process.

This property (the "tomographic power" of a set of probe states) is hard to quantify exactly, but it can be captured approximately by the determinant of the Gram matrix. This is zero if (and only if) the probe ensemble completely fails to probe at least one dimension of state space, and it achieves its maximum value if (and only if) the probe ensemble is a 2-design. Larger determinants are better, implying that the inversion will amplify errors less. In the inset to Fig. 6.4, we plot this determinant as a function of the fiducial state's x-value. For a N00N state ($x = \frac{1}{2}$) the determinant is zero! So, remarkably, the set of states generated by the $SU(2)$ orbit of a N00N state is incomplete – it does not enable QPT at all, and is completely oblivious to at least one parameter of the process (the Gram matrix of the SU(2)-covariant set generated using a N00N state is shown to be rank deficient in section B of the Supplemental Material). At the opposite extreme ($x = 0$), spin-coherent states *do* generate a complete set, suitable for QPT – but, like the coherent states of Lobino *et al.*, they are not optimal for the task. The determinant is small, indicating that at least one parameter of the process is poorly resolved. The maximum value of the determinant is achieved at an intermediate point, $x = \frac{1}{2} - \frac{1}{2\sqrt{2}} \approx 0.15$, which generates a set of probe states that are neither N00N nor spin-coherent. We predicted that this set of states would enable optimally accurate QPT.

The $x \approx 0.15$ ensemble is special and unique in another way; it forms a 2-design (as shown in section B of the Supplemental Material). Informally, 2-designs are sets of quantum states whose projectors span the vector space of operators as uniformly as possible; more precisely, the ensemble's 2nd moments are equal to those of the uniform Haar ensemble over pure states. Common examples of 2-designs include mutually unbiased bases (MUBs) [68] and symmetric informationally complete measurements (SIC-POVMs) [69, 70], and there is strong theoretical [68] and experimental [71] evidence that 2-designs are optimal for state tomography. Our results here are the first experimental evidence that 2-designs are optimal for QPT (see theory in Refs. [47, 48]).

Since theoretical analysis predicts that the $x \approx 0.15$ ensemble should outperform every other $SU(2)$-generated ensemble at QPT (including the nominally more-sensitive N00N ensemble), we did an experiment to test the prediction. We prepared several different $SU(2)$-covariant sets of input states – each generated by applying 10 different collective polarization ($SU(2)$) rotations to a single fiducial state with values of x ranging from 0 to 0.47 as detailed in Fig. 6.1d – and used them to perform QPT on an $SU(2)$-jitter process. **No a priori assumptions were made about the nature or structure of the process**[1]. We performed QPT by (1) preparing many copies of each of the 10 states, (2) sending them through

[1]We also verified experimentally that our process preserves the biphoton structure – i.e., it does not violate permutation symmetry, and therefore does not change the population in the anti-symmetric subspace – and so we can treat it as a quantum process on a 3-dimensional Hilbert space.

the process, (3) performing a tomographically complete set of measurements on each output state, and (4) using maximum-likelihood estimation (MLE) to reconstruct the process.

Evaluating the performance of an experimental tomographic procedure (i.e., to rank our three different input ensembles) is nontrivial. We cannot assume that we know the "true" process, yet the standard metric of tomographic success is "How close is the reconstructed process to the 'true' process?" We circumvent this problem by using a measure of accuracy that can be estimated directly, the *average process infidelity*(API) [72]. The API between two processes $\mathcal{E}$ and $\mathcal{F}$ is the [mixed-state] quantum infidelity $1 - F\left(\mathcal{E}\left[|\psi\rangle\langle\psi|\right], \mathcal{F}\left[|\psi\rangle\langle\psi|\right]\right)$, averaged over all *pure* inputs to the process according to the unitarily invariant Haar measure:

$$\mathrm{API}(\mathcal{E}, \mathcal{F}) = \int_{\mathrm{Haar}} \left[1 - F\left(\mathcal{E}\left[|\psi\rangle\langle\psi|\right], \mathcal{F}\left[|\psi\rangle\langle\psi|\right]\right)\right] d\psi \tag{6.13}$$

The API vanishes for a perfect reconstruction, and increases with errors in tomography.

We estimate the true process using QPT; $\hat{D}_\gamma$ is our estimate. To quantify the quality of $\hat{D}_\gamma$ we empirically measure the API by:

1. Prepare [many copies of] 40 different randomly chosen input states ρ_i $(i = 1 \ldots 40)$.
2. Use state tomography to obtain an estimate $\hat{\rho}_i$ of each input state
3. Apply the depolarization process (D_γ) to each state.
4. Use state tomography to obtain an empirical estimate $\hat{\rho'}_i$ of each *output* state.
5. Compute the quantum fidelity between (a) the empirical output state $\hat{\rho'}_i$ and (b) the output state *predicted* by our QPT estimate, $\hat{D}_\gamma\left[\rho_i\right]$.
6. Averaging this fidelity over all 40 input states.

The resulting number requires no a priori assertion about the "true" process, and it is a good quantifier of how accurately the QPT estimate $\hat{D}_\gamma$ predicts independent experimental results. But it is also an estimate of the theoretical API as defined in [72], and deviates from it only inasmuch as (i) we have approximated the integral in Eq. 6.13 by a sum over 40 random states; (ii) those states are not quite pure; and (iii) state tomography on finitely many samples is never quite perfect $(\hat{\rho} \neq \rho)$.

Figure 6.4 shows the dependence of the empirical API on the fiducial state parameter x, for two different depolarization strengths: $\gamma = 0.5$ rad (blue) and $\gamma = 1.5$ rad (red). Points represent experimentally measured APIs, while solid and dashed lines represent two different simulations of our experiment. The solid lines are generated by simulating process tomography using 10 randomly chosen *pure* input states; the dashed lines are generated by simulating process tomography using the same 10 nearly-pure input states used in the experiment. Both simulations used 40 different random states to estimate the empirical API, just as in the experiment.

We observe a minimum in the API (i.e., optimal reconstruction fidelity) at $x \approx 0.15$ for both of the depolarization strengths – exactly where theory predicted. This minimum API coincides with the maximum value of $\det(M)$ (Fig. 6.4, inset). We also confirm that spin-coherent probe ensembles are not optimal. N00N state ensembles consistently generate the *least* accurate QPT. Ironically, while the N00N ensemble should in theory fail catastrophically, it is (slightly) redeemed by experimental imperfections in state preparation, which result in the N00N ensemble being not quite perfectly incomplete. However, it still achieves a much worse API than any other $SU(2)$-covariant input ensemble.

6.4 Conclusions

It is well known that entanglement can (and usually does) improve metrology. However, it has also been taken for granted that ensembles of coherent states (which, in multi-photon systems, are not entangled at all) are "good enough" for process tomography. We have shown that both of these beliefs should be interpreted cautiously. On one hand, while maximally entangled N00N states are indeed optimal for *detecting* a particularly common and important form of decoherence, they are very bad for *characterizing* it in detail. And while [spin]-coherent states are indeed sufficient for QPT, they are not optimal. Our results can be summarized as showing that the most robust and flexible way to probe decoherence is with "partly entangled" states, intermediate between N00N and coherent states.

Our experimental results show that in the presence of prior information (that the system is undergoing pure $SU(2)$ jitter), the optimal biphoton probe states are N00N states. We expect that this result will be of utility in magnetometry and atomic physics, where interferometry is often used to estimate noise. On the other hand, in the complete absence of prior information, we have shown that an intermediate entangled state is much better at performing QPT. Our method generalizes the technique of Ref. [44] – preparing a single fiducial state and displacing it – to generate a set of states we believe are optimal for performing QPT on *any* process, not just $SU(2)$ jitter. This set of states forms a 2-design, and our work is the first experimental evidence confirming that 2-designs are optimal for QPT. Furthermore, we expect the advantage of 2-designs over other sets to increase with the dimensionality of Hilbert space. It is also worth stating that our analysis could be generalized to other symmetry groups, but doing so is an unsolved problem. Our results imply that one can greatly improve the accuracy of QPT by choosing the right set of input states – but, surprisingly, the "right" states for QPT are not those most sensitive to the process. We conclude that detailed state engineering can be very useful in tailoring probe states or ensembles to specific tasks in the characterization (and ultimately remediation) of depolarization.

References

[1] E Bagan, MA Ballester, RD Gill, R Munoz-Tapia, and O Romero-Isart. Separable measurement estimation of density matrices and its fidelity gap with collective protocols. *Phys. Rev. Lett.*, 97(13):130501, 2006.

[2] SJ Van Enk and Robin Blume-Kohout. When quantum tomography goes wrong: drift of quantum sources and other errors. *New Journal of Physics*, 15(2):025024, 2013.

[3] Robin Blume-Kohout. Optimal, reliable estimation of quantum states. *N. J. Phys.*, 12(4):043034, 2010.

[4] Robin Blume-Kohout. Robust error bars for quantum tomography. *arXiv:1202.5270*, 2012.

[5] John Smolin, Jay Gambetta, and Graeme Smith. Efficient method for computing the maximum-likelihood quantum state from measurements with additive gaussian noise. *Phys. Rev. Lett.*, 108:070502, Feb 2012.

[6] Seth Merkel, Jay Gambetta, John Smolin, Stefano Poletto, Antonio Córcoles, Blake Johnson, Colm Ryan, and Matthias Steffen. Self-consistent quantum process tomography. *Phys. Rev. A*, 87:062119, Jun 2013.

[7] Joseph M. Renes, Robin Blume-Kohout, A. J. Scott, and Carlton M. Caves. Symmetric informationally complete quantum measurements. *Journal of Mathematical Physics*, 45(6):2171–2180, 2004.

[8] Zdenek Hradil. Quantum-state estimation. *Phys. Rev. A*, 55(3):1561–1564, 1997.

[9] Matteo Paris and Jaroslav Rehacek. *Quantum state estimation*, volume 649. Springer, 2004.

[10] Robin Blume-Kohout. Hedged maximum likelihood quantum state estimation. *Phys. Rev. Lett.*, 105(20):200504, 2010.

[11] David Gross, Yi-Kai Liu, Steven T Flammia, Stephen Becker, and Jens Eisert. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.*, 105(15):150401, 2010.

[12] M Christandl and R Renner. Reliable quantum state tomography. *Phys. Rev. Lett.*, 109(12):120403, 2012.

[13] W. K. Wootters. Statistical distance and hilbert space. *Phys. Rev. D*, 23:357–362, Jan 1981.

[14] Carl W Helstrom. *Quantum detection and estimation theory*, volume 84. Academic press New York, 1976.

[15] Christopher A Fuchs. Distinguishability and accessible information in quantum theory. *arXiv preprint quant-ph/9601020*, 1996.

[16] C.A. Fuchs and J. van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *Information Theory, IEEE Transactions on*, 45(4):1216–1227, 1999.

[17] J. Calsamiglia, R. Muñoz Tapia, Ll. Masanes, A. Acin, and E. Bagan. Quantum chernoff bound as a measure of distinguishability between density matrices: Application to qubit and gaussian states. *Phys. Rev. A*, 77:032311, Mar 2008.

[18] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz Tapia, E. Bagan, Ll. Masanes, A. Acin, and F. Verstraete. Discriminating states: The quantum chernoff bound. *Phys. Rev. Lett.*, 98:160501, Apr 2007.

[19] Xing-Can Yao, Tian-Xiong Wang, Ping Xu, He Lu, Ge-Sheng Pan, Xiao-Hu Bao, Cheng-Zh Peng, Chao-Yang Lu, Yu-Ao Chen, and Jian-Wei Pan. Observation of eight-photon entanglement. *Nature Photonics*, 6:225–228, 2011.

[20] Richard D. Gill and Serge Massar. State estimation for large ensembles. *Phys. Rev. A*, 61:042312, Mar 2000.

[21] E Bagan, MA Ballester, Richard D Gill, A Monras, and R Munoz-Tapia. Optimal full estimation of qubit mixed states. *Phys. Rev. A*, 73(3):032301, 2006.

[22] F. Huszár and N. M. T. Houlsby. Adaptive bayesian quantum tomography. *Phys. Rev. A*, 85:052120, May 2012.

[23] Jaroslav Řeháček, Berthold-Georg Englert, and Dagomir Kaszlikowski. Minimal qubit tomography. *Phys. Rev. A*, 70:052321, Nov 2004.

[24] Ryo Okamoto, Minako Iefuji, Satoshi Oyama, Koichi Yamagata, Hiroshi Imai, Akio Fujiwara, and Shigeki Takeuchi. Experimental demonstration of adaptive quantum state estimation. *Phys. Rev. Lett.*, 109(13):130404, 2012.

[25] Mark D. de Burgh, Nathan K. Langford, Andrew C. Doherty, and Alexei Gilchrist. Choice of measurement sets in qubit tomography. *Phys. Rev. A*, 78:052122, Nov 2008.

[26] A J Scott. Tight informationally complete quantum measurements. *J. Phys. A*, 39(43):13507, 2006.

[27] Takanori Sugiyama, Peter S Turner, and Mio Murao. Effect of non-negativity on estimation errors in one-qubit state tomography with finite data. *N. J. Phys.*, 14(8):085005, 2012.

[28] J. Löfberg. Yalmip : A toolbox for modeling and optimization in matlab. *Proc. CACSD (Taipei)*, 2004.

[29] Daniel F. V. James, Paul G. Kwiat, William J. Munro, and Andrew G. White. Measurement of qubits. *Phys. Rev. A*, 64:052312, Oct 2001.

[30] Matthias Hübner. Explicit computation of the bures distance for density matrices. *Phys. Lett. A*, 163(4):239–242, 1992.

[31] D Petz and C Sudar. Geometries of quantum states. *J. Math. Phys*, 37(6):2662–2673, JUN 1996.

[32] Karol Życzkowski and Hans-Jürgen Sommers. Average fidelity between random quantum states. *Phys. Rev. A*, 71:032313, Mar 2005.

[33] S. Massar and S. Popescu. Optimal extraction of information from finite quantum ensembles. *Phys. Rev. Lett.*, 74:1259–1263, Feb 1995.

[34] G. Breitenbach, S. Schiller, and J. Mlynek. Measurement of the quantum states of squeezed light. *Nature*, 387(6632):471–475, May 1997.

[35] Agedi N. Boto, Pieter Kok, Daniel S. Abrams, Samuel L. Braunstein, Colin P. Williams, and Jonathan P. Dowling. Quantum interferometric optical lithography: Exploiting entanglement to beat the diffraction limit. *Physical Review Letters*, 85(13):2733–2736, 2000.

[36] M. W. Mitchell, J. S. Lundeen, and A. M. Steinberg. Super-resolving phase measurements with a multiphoton entangled state. *Nature*, 429(6988):161–164, May 2004.

[37] Philip Walther, Jian-Wei Pan, Markus Aspelmeyer, Rupert Ursin, Sara Gasparoni, and Anton Zeilinger. De broglie wavelength of a non-local four-photon state. *Nature*, 429(6988):158–161, May 2004.

[38] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Advances in quantum metrology. *Nature Photonics*, 5(4):222–229, 2011.

[39] Philip J. D. Crowley, Animesh Datta, Marco Barbieri, and Ian A. Walmsley. Multiparameter quantum metrology. *arXiv:1206.0043*, May 2012.

[40] Peter C. Humphreys, Marco Barbieri, Animesh Datta, and Ian A. Walmsley. Quantum enhanced multiple phase estimation. *Physical Review Letters*, 111(7):070403, August 2013.

[41] Xiao-Qi Zhou, Hugo Cable, Rebecca Whittaker, Peter Shadbolt, Jeremy L. O'Brien, and Jonathan C. F. Matthews. Quantum-enhanced tomography of unitary processes. *arXiv:1402.2897*, 2014.

[42] M. W. Mitchell, C. W. Ellenor, S. Schneider, and A. M. Steinberg. Diagnosis, prescription, and prognosis of a bell-state filter by quantum process tomography. *Physical Review Letters*, 91(12):120402, 2003.

[43] J. L. O'Brien, G. J. Pryde, A. Gilchrist, D. F. V. James, N. K. Langford, T. C. Ralph, and A. G. White. Quantum process tomography of a controlled-NOT gate. *Physical Review Letters*, 93(8):080502, 2004.

[44] Mirko Lobino, Dmitry Korystov, Connor Kupchak, Eden Figueroa, Barry C. Sanders, and A. I. Lvovsky. Complete characterization of quantum-optical processes. *Science*, 322(5901):563–566, October 2008.

[45] Roy J. Glauber. Coherent and incoherent states of the radiation field. *Physical Review*, 131(6):2766–2788, September 1963.

[46] Robin Blume-Kohout and Peter S. Turner. The curious nonexistence of gaussian 2-designs. *Communications in Mathematical Physics*, pages 1–17, 2014.

[47] A J Scott. Optimizing quantum process tomography with unitary 2-designs. *Journal of Physics A: Mathematical and Theoretical*, 41(5):055308, February 2008.

[48] A. Fernández-Pérez, A. B. Klimov, and C. Saavedra. Quantum process reconstruction based on mutually unbiased basis. *Physical Review A*, 83(5):052332, May 2011.

[49] Ángel Rivas and Alfredo Luis. Su(2)-invariant depolarization of quantum states of light. *Phys. Rev. A*, 88:052120, Nov 2013.

[50] D. H. Mahler, L. Rozema, A. Darabi, and A. M. Steinberg. Identification of decoherence-free subspaces without quantum process tomography. *Phys. Rev. A*, 86:052101, Nov 2012.

[51] Lorenza Viola, Emanuel Knill, and Seth Lloyd. Dynamical generation of noiseless quantum subsystems. *Physical Review Letters*, 85(16):3520–3523, October 2000.

[52] Lorenza Viola, Evan M. Fortunato, Marco A. Pravia, Emanuel Knill, Raymond Laflamme, and David G. Cory. Experimental realization of noiseless subsystems for quantum information processing. *Science*, 293(5537):2059–2063, September 2001. PMID: 11557885.

[53] Yu. I. Bogdanov, M. V. Chekhova, S. P. Kulik, G. A. Maslennikov, A. A. Zhukov., C. H. Oh, and M. K. Tey. Qutrit state engineering with biphotons. *Physical Review Letters*, 93(23):230503, December 2004.

[54] R. B. A. Adamson, L. K. Shalm, M. W. Mitchell, and A. M. Steinberg. Multiparticle state tomography: Hidden differences. *Physical Review Letters*, 98(4):043601, January 2007.

[55] R. B. A. Adamson, P. S. Turner, M. W. Mitchell, and A. M. Steinberg. Detecting hidden differences via permutation symmetries. *Physical Review A*, 78(3):033832, 2008.

[56] F. T. Arecchi, Eric Courtens, Robert Gilmore, and Harry Thomas. Atomic coherent states in quantum optics. *Physical Review A*, 6(6):2211–2237, December 1972.

[57] Holger F. Hofmann and Takafumi Ono. High-photon-number path entanglement in the interference of spontaneously down-converted photon pairs with coherent laser light. *Physical Review A*, 76(3):031806, 2007.

[58] Jaromír Fiurášek. Conditional generation of n-photon entangled states of light. *Physical Review A*, 65(5):053818, May 2002.

[59] Itai Afek, Oron Ambar, and Yaron Silberberg. High-NOON states by mixing quantum and classical light. *Science*, 328(5980):879–881, May 2010.

[60] Shamir Rosen, Itai Afek, Yonatan Israel, Oron Ambar, and Yaron Silberberg. Sub-rayleigh lithography using high flux loss-resistant entangled states of light. *Physical Review Letters*, 109(10):103602, September 2012.

[61] Y. Israel, I. Afek, S. Rosen, O. Ambar, and Y. Silberberg. Experimental tomography of NOON states with large photon numbers. *Physical Review A*, 85(2):022115, February 2012.

[62] Lee A. Rozema, James D. Bateman, Dylan H. Mahler, Ryo Okamoto, Amir Feizpour, Alex Hayat, and Aephraim M. Steinberg. Scalable spatial superresolution using entangled photons. *Phys. Rev. Lett.*, 112:223602, 2014.

[63] G. Agarwal. Relation between atomic coherent-state representation, state multipoles, and generalized phase-space distributions. *Phys. Rev. A*, 24(6):2889–2896, 1981.

[64] L. K. Shalm, R. B. A. Adamson, and A. M. Steinberg. Squeezing and over-squeezing of triphotons. *Nature*, 457(7225):67–70, January 2009.

[65] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 1 edition, 2000.

[66] Holger F Hofmann. Generation of highly nonclassical n-photon polarization states by superbunching at a photon bottleneck. *Physical Review A*, 70(2):023812, 2004.

[67] Ryo Okamoto, Holger F Hofmann, Tomohisa Nagata, Jeremy L O'Brien, Keiji Sasaki, and Shigeki Takeuchi. Beating the standard quantum limit: phase super-sensitivity of n-photon interferometers. *New Journal of Physics*, 10(7):073033, July 2008.

[68] William K Wootters and Brian D Fields. Optimal state-determination by mutually unbiased measurements. *Annals of Physics*, 191(2):363–381, May 1989.

[69] Joseph M. Renes, Robin Blume-Kohout, A. J. Scott, and Carlton M. Caves. Symmetric informationally complete quantum measurements. *Journal of Mathematical Physics*, 45(6):2171–2180, May 2004.

[70] Z. E. D. Medendorp, F. A. Torres-Ruiz, L. K. Shalm, G. N. M. Tabia, C. A. Fuchs, and A. M. Steinberg. Experimental characterization of qutrits using symmetric informationally complete positive operator-valued measurements. *Physical Review A*, 83(5):051801, May 2011.

[71] R. B. A. Adamson and A. M. Steinberg. Improving quantum state estimation with mutually unbiased bases. *Physical Review Letters*, 105(3):030406, July 2010.

[72] Alexei Gilchrist, Nathan K. Langford, and Michael A. Nielsen. Distance measures to compare real and ideal quantum processes. *Physical Review A*, 71(6):062310, June 2005.

DISTRIBUTION:

1	MS 0359	D. Chavez, LDRD Office, 1911
1	MS 1248	Steve Rinaldi, 5643
1	MS 1322	Robin Blume-Kohout, 1425
1	MS 1322	John Aidun, 1425
1	MS 0899	Technical Library, 9536 (electronic copy)

Sandia National Laboratories