CHARACTERIZATION OF QUANTUM STATES: ADVANCES IN QUANTUM TOMOGRAPHY AND TESTS OF NONLOCALITY

BY

CHRIS ZEITLER

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois Urbana-Champaign, 2021

Urbana, Illinois

Doctoral Committee:

      Associate Professor Virginia Lorenz, Chair
      Professor Paul G. Kwiat, Director of Research
      Associate Professor Eric Chitambar
      Professor Jen-Chieh Peng

Abstract

Quantum communication protocols rely on the precise generation of target quantum states. In order to implement these algorithms efficiently, characterization of the produced states is necessary to ensure that the generated state accurately matches the target. In practical communication tasks, there is a tradeoff between the amount of time spent monitoring the source state, and performing the desired task. In this thesis, we present work on the implementation of a Bayesian tomography algorithm that provides improved performance in this low-data regime relative to the standard maximum likelihood approach. In addition, we investigate the entanglement properties of hyperentangled states that display entanglement in multiple degrees of freedom simultaneously. We find that Bell tests beyond the standard CHSH inequality can reveal the higher-dimensional nature of these states.

To my parents.

Acknowledgements

I first want to thank my parents, who taught me the value of curiosity, patience, and dedication. A joy of my life has been having parents as supportive as you. Thank you Lynne and Peter for everything you have done for me.

This research group has fostered a welcoming, collaborative, and inquisitive work space, and I want to thank all the members of the group, both long gone and just beginning their journey here, for the part they play in making it that way. A number have had an outsize influence on me, and I want to thank them for that. Thank you to Trent Graham, who mentored me as a young graduate student and taught me much of what I know about optics. He originally designed and built the hyperentanglement source used in this work. Thank you to Joe Chapman, whose exceptional skills in the lab made the experiments in Chapter 4 possible. He treated me with patience when he was able to record data before I had written tools to process it. Thank you to Courtney Krafczyk for being my canvassing buddy and for always bringing kindness and compassion to our research group. Thank you especially to Michelle Victora, who has been a true confidante and friend, both in and out of the lab. Michelle has helped me feel confident about my work and has brought an unquenchable energy to all the things we have done together.

The friends I have met in Illinois have made my life here rich and fulfilling. Thank you to Cristina, Tom, Michelle, Allycia, Matt, and Carolee for putting up with my desire to read fiction together for the last 6 years. Thank you Allycia for sharing a love of cooking with me and for the countless, impeccable meals you shared with me. Thank you Cristina for showing me the value in being unapologetically yourself. Thank you to Brian and Lily for showing me companionship when I was feeling isolated. Thank you to Mazin for your empathy and friendship, for being a squash partner, and for sharing so many of your passions with me.

Thank you finally to Erin, who has helped me find joy this last year in a time when that was all too rare.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

APD	Avalanche photodiode
BME	Bayesian mean estimation
BSM	Bell state measurement
BS	Beam splitter
BQC	Blind quantum computing
DOF	Degree of freedom
eig	eigenvalue
EPS	Entangled photon source
F	Fidelity
HWP	Half-wave plate
LC	Liquid crystal
MLE	Maximum likelihood estimation
MZI	Mach-Zehnder interferometer
PBS	Polarizing beam splitter
PC	Pockels cell
PPLN	Periodically-poled lithium niobate
QKD	Quantum key distribution
QWP	Quarter-wave plate
SNSPD	Superconducting nanowire single photon detector
SPDC	Spontaneous parametric downconversion
Tr	Trace

Chapter 1

Introduction

In this chapter, we provide an introduction to the field of quantum information through the lens of quantum optics. In Section 1.1, we will discuss the representation of quantum states which form the foundation of the rest of this work and discuss some of their properties. In Section 1.2, we will discuss the tools for experimentally generating, manipulating, and measuring quantum states of light. In Section 1.3, we present an overview of a few quantum communication and quantum networking tasks.

1.1 Representing Quantum States

1.1.1 Pure States

A quantum state is said to be pure if there exists a measurement basis in which a measurement of the state has a deterministic outcome. For example, the state $|0\rangle$ can be seen to be pure because when it is measured in the computational basis, it will always give a measurement outcome of 0. The state $\frac{|0\rangle+|1\rangle}{\sqrt{2}}$ does not have a deterministic result in the computational basis, but is still pure because this state will give a deterministic result in the diagonal basis. Pure states can be represented as a complex vector. For example, we might define the polarization states of a photon as

$$|H\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, |V\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{1.1}$$

Photon polarization will be further discussed in Section 1.2.2. Pure states are constrained by a normalization condition:

$$|\langle\psi|\psi\rangle|^2 = 1, \tag{1.2}$$

and if $|\psi\rangle$ is not normalized, it can be converted to a normalized state by computing

$$|\widetilde{\psi}\rangle = \frac{|\psi\rangle}{|\langle\psi|\psi\rangle|^2}. \tag{1.3}$$

1

In general, attempting to determine a state's purity by finding a deterministic basis is a challenging task, and so other metrics are used to ascertain the purity of a state. Understanding these techniques, however, requires us to consider states that are *not* pure.

1.1.2 Mixed States

Experimentally, all observed quantum states are found to not be pure, that is, regardless of the basis of measurement, the measurement outcome is probabilistic. We can begin to describe these states by considering them as a sum of pure states. However, if the states are simply added, this will result to another pure state. This leads us to view such states as incoherent mixtures of pure states:

$$\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|, \tag{1.4}$$

where ρ is referred to as the state's density matrix. One might wonder whether if such an extension of pure states is sufficient to describe all possible states, but it can be shown that such a description completely specifies the quantum state [1]. Note, however, that the form given in Equation 1.4 is not unique. For example, a state could be decomposed in two ways:

$$\begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} = 0.5|H\rangle\langle H| + 0.5|V\rangle\langle V| = 0.5|D\rangle\langle D| + 0.5|A\rangle\langle A|. \tag{1.5}$$

Thus, while it may be appealing to see the first description of the system as producing horizontally and vertically polarized photons with equal probability, it is equally valid to describe the system as producing diagonally and antidiagonally polarized photons with equal probability (as well as infinitely many other descriptions).

Like pure states, density matrices must observe a normalization condition, given by $\text{Tr}(\rho)=1$, and similarly, a density matrix can be converted to a normalized form by calculating:

$$\widetilde{\rho} = \frac{\rho}{\text{Tr}(\rho)}. \tag{1.6}$$

In addition, the density matrix must be Hermitian ($\rho = \rho^\dagger$) and positive semidefinite ($\text{eig}(\rho) \geq 0$). To understand why these restrictions are necessary, consider expanding ρ in its eigenbasis according to Equation 1.4. In this basis, the values of p_j are the eigenvalues of ρ and correspond to the probability of measuring ρ in the state $|\psi_j\rangle$. Because this is an orthonormal basis, $\sum_j p_j=1$, which simply means that the sum of

all possible measurement outcomes is 1. If ρ does not obey the normalization condition, then this is not true. Similarly, the Hermiticity condition ensures that the probabilities p_j are all real, and the positive semidefinite condition ensures that none of p_j are less than zero.

After expanding our description of quantum states to allow mixed states, we can now describe a mathematical test of to determine if a state is pure. We can conclude that a state is pure if

$$\mathrm{Tr}(\rho^2) = 1, \tag{1.7}$$

and so the quantity $\mathrm{Tr}(\rho^2)$ is referred to as a state's purity. To see why the purity must be 1 for a pure state, recall that for a pure state, there must exist a measurement basis in which the results are deterministic. If we expand ρ in this basis, we then know that exactly one p_j will be 1, and the rest zero. This means that all pure states can be represented as $\rho = |\psi\rangle\langle\psi|$. Putting this into our definition of the purity,

$$\mathrm{Tr}(\rho^2) = \mathrm{Tr}(|\psi\rangle\langle\psi|\psi\rangle\langle\psi|) = \mathrm{Tr}(|\psi\rangle\langle\psi|) = 1. \tag{1.8}$$

For a mixed state,

$$\begin{aligned}
\mathrm{Tr}(\rho^2) &= \mathrm{Tr}(\sum_{jk}(p_j|\psi_j\rangle\langle\psi_j|)(p_k|\psi_k\rangle\langle\psi_k|)) \\
&= \mathrm{Tr}(\sum_j p_j^2|\psi_j\rangle\langle\psi_j|) \\
&= \sum_j p_j^2
\end{aligned} \tag{1.9}$$

Note that this derivation relied on orthonormality of the $|\psi_j\rangle$ basis; if ρ is expressed in a basis that is not orthonormal, Equation 1.9 will not apply. If ρ is mixed, then we know that $0 \le p_j < 1$, and so $p_j^2 \le p_j$. By the normalization condition, $\sum_j p_j = 1$, and so $\sum_j p_j^2 < 1$. This shows that the purity can be used to distinguish if a state is pure or not. If the purity is equal to 1, then the state is pure, and if it is less than 1, the state is mixed. The "most mixed" a state can be is for the probability of all measurement outcomes to be equal. The purity of such a state will then be

$$\mathrm{Tr}(\rho^2) = \mathrm{Tr}(\frac{\mathbb{I}_d\mathbb{I}_d}{d^2}) = \frac{1}{d}, \tag{1.10}$$

where d is the dimension of ρ, and $\mathbb{I}_d$ is the identity matrix of dimension d. Note that the minimum possible purity of a state depends on the state's dimension; while a state with purity 0.5 will be completely mixed if d=2, it will have twice the purity of a completely mixed state if d=4.

1.1.3 Entanglement

One of the fundamental phenomena of quantum states that distinguishes them from their classic analog is that of entanglement. Entanglement is a type of correlation between quantum systems. If two systems can be described as

$$\rho_{AB} = \rho_A \otimes \rho_B, \tag{1.11}$$

then the state ρ_{AB} is said to be *separable* and is not entangled. If ρ_{AB} cannot be expressed in the form of Equation 1.11, then it is said to be entangled. This negative definition of entanglement is cumbersome to use, especially in experimental settings where ρ_{AB} itself may be unknown. Techniques for detecting the presence of entanglement are further discussed in Chapter 4.

Entanglement has a number of features that distinguish it from classical correlations. First, entanglement is basis-independent, and the correlations from entanglement will be present regardless of the basis the state is measured in. This is most easily seen in the state

$$|\Psi^-\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}, \tag{1.12}$$

the antisymmetric singlet state. Because this is the only antisymmetric state of two qubits, it must have the form of Equation 1.12 regardless of what basis it is expressed in [2]. Similarly, it must exhibit the same correlations regardless of the basis of measurement. Compare this to the separable state $|00\rangle$, which has perfect correlations in the computational basis but becomes less correlated in other bases.

Second, the local state of entangled subsystems will be mixed. Consider again the state $\rho_{AB} = \Psi^-$. We can determine the local state of subsystem A by taking the partial trace of ρ_{AB} over subsystem B:

$$\rho_A = Tr_B(\rho_{AB}) = \frac{|0\rangle\langle0| + |1\rangle\langle1|}{2}. \tag{1.13}$$

ρ_A is completely mixed, as is ρ_B. This fact is exploited in quantum steering [3], in which one party's ability to leave the one subsystem in a pure state by measuring the other subsystem is used as a signal of entanglement. For more information on quantum steering, see Section 4.3. In comparison, if we again consider the separable state $|00\rangle$, we can see that each subsystem is in a pure state.

This fact is especially important when designing entanglement sources because correlations to unintended degrees of freedom can lead to a reduction in the degree of entanglement in the target degree of freedom.

For example, a source may be designed to produce the state

$$\rho_{target} = \frac{|HH\rangle + |VV\rangle}{\sqrt{2}} \qquad (1.14)$$

but may develop correlations with the frequency of the photons produced, such that the H and V components have slightly different spectra, such as

$$\rho_{exp} = \int d\delta\omega \frac{|HH\rangle|\omega_1, \omega_2\rangle + e^{i\phi(\omega_1, \omega_2)}|VV\rangle|\omega_1 + \delta\omega, \omega_2 - \delta\omega\rangle}{\sqrt{2}}. \qquad (1.15)$$

These correlations mean that the polarization subsystem will always be partially mixed and not maximally entangled. Fidelity is a way to quantify the degree to which a state is similar to a target state. For pure states, the fidelity has a simple form:

$$F(|\psi\rangle, |\phi\rangle) = |\langle\psi|\phi\rangle|^2. \qquad (1.16)$$

For mixed states, the fidelity takes a more complex form [1]:

$$F(\rho, \sigma) = \mathrm{Tr}(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}})^2. \qquad (1.17)$$

If $F(\rho, \sigma) = 1$, then $\rho = \sigma$, and if $F(\rho, \sigma) = 0$, then ρ and σ are orthogonal.

1.1.4   Measurements

A distinctive feature of quantum systems is that they are almost always modified by the measurement process. The most basic measurement of a quantum system is a projective measurement. After a projective measurement, the measured system is left in the state $|i\rangle$ when the measurement outcome i is observed, regardless of the systems state before measurement. Here, m is an eigenvalue of the measurement operator with associated eigenvector $|m\rangle$. For projective measurements, the state associated with each measurement is pure. By expanding our description of measurements to include mixed-state outcomes, we arrive at the most general form of measurement on quantum systems, called positive operator-valued measures (POVMs) [1]. A POVM consists of a number of measurement operators M_i, constrained such that

$$\sum_i M_i = \mathbb{I}. \qquad (1.18)$$

In this framework, projective measurements are simply a subset of POVMs with $M_i = |i\rangle\langle i|$, where the set of states $|i\rangle$ form an orthonormal basis. The probability that measuring a state with a POVM will yield the

5

ith outcome can be calculated as

$$Pr(i|\rho) = \text{Tr}(M_i\rho). \tag{1.19}$$

It is worth mentioning a few POVMs due to their ubiquity. When measuring photon polarization, three measurement bases are often used. The first is the H/V basis, where H corresponds to a photon with polarization aligned horizontally, and V corresponds to a photon with polarization aligned vertically[1]:

$$M_1 = |H\rangle\langle H|$$
$$M_2 = |V\rangle\langle V|. \tag{1.20}$$

The next is the D/A basis, where the photon polarization is diagonally or anti-diagonally aligned:

$$M_1 = |D\rangle\langle D|, \ |D\rangle = \frac{|H\rangle + |V\rangle}{\sqrt{2}}$$
$$M_2 = |A\rangle\langle A|, \ |A\rangle = \frac{|H\rangle - |V\rangle}{\sqrt{2}}. \tag{1.21}$$

Finally, in the R/L basis, R corresponds to a photon with right-handed circular polarization, and L corresponds to a photon with left-handed circular polarization:

$$M_1 = |R\rangle\langle R|, \ |R\rangle = \frac{|H\rangle + i|V\rangle}{\sqrt{2}}$$
$$M_2 = |L\rangle\langle L|, \ |L\rangle = \frac{|H\rangle - i|V\rangle}{\sqrt{2}}. \tag{1.22}$$

These three bases (or their analogs for other systems) are important because they comprise the over-complete set of measurements used in quantum tomography. The density matrix representing a state of dimension d has $d^2 - 1$ independent parameters, so $d^2 - 1$ linearly independent measurements are necessary to specify a state. For a polarization qubit, these measurements could be $|H\rangle$, $|D\rangle$, and $R\rangle$. Often, the number of incident photons is unknown, so it is necessary to measure in one complete basis for normalization purposes, which could be achieved by adding $|V\rangle$, for example. This gives some preference to the H/V basis, decreasing uncertainty in this basis relative to the others. For symmetry (and because many experimental setups are conducive to measuring complete bases anyway), the H/V, D/A, and R/L bases comprise the standard set of tomographic measurements. For measurements on two-qubit states, the Bell basis is often used. This

---

[1] While this picture corresponds to our normal notion of horizontal and vertical relative to the table that an optics experiment is carried out on, these definitions should usually be calibrated to a specific polarizing optic element in the experiment, such as a polarizing beam splitter.

basis consists of the four maximally-entangled two-qubit states referred to as Bell states $|\Phi^{\pm}\rangle$ and $|\Psi^{\pm}\rangle$:

$$
\begin{aligned}
M_1 = |\Phi^+\rangle\langle\Phi^+|, \ |\Phi^+\rangle &= \frac{|00\rangle + |11\rangle}{\sqrt{2}} \\
M_2 = |\Phi^-\rangle\langle\Phi^-|, \ |\Phi^-\rangle &= \frac{|00\rangle - |11\rangle}{\sqrt{2}} \\
M_3 = |\Psi^+\rangle\langle\Psi^+|, \ |\Psi^+\rangle &= \frac{|01\rangle + |10\rangle}{\sqrt{2}} \\
M_4 = |\Psi^-\rangle\langle\Psi^-|, \ |\Psi^-\rangle &= \frac{|01\rangle - |10\rangle}{\sqrt{2}}.
\end{aligned}
\tag{1.23}
$$

1.2 Creating Photonic Qubits

In this work, we will use photons as the physical system for our qubits. Photons offer a number of degrees of freedom in which to encode qubits, and we will discuss the strategies to generate, manipulate, and measure qubits encoded in those degrees of freedom. In Section 1.2.1, we will discuss the use of spontaneous parametric downconversion (SPDC) to generate photon pairs. In Section 1.2.2, we will discuss techniques to use SPDC sources to generate pairs that are entangled in polarization, as well as ways to manipulate and measure polarization states. In Section 1.2.3, we will do the same for time-bin qubits.

1.2.1 Generating Photon Pairs

In order to encode qubits with photons, photons must first be created. One approach to generate photons at the single-photon level is through the use of spontaneous parametric downconversion (SPDC), a process in which a high-energy pump photon is converted into two, lower-energy photons. Such a process is forbidden in vacuum, so SPDC must take place in certain materials that have a nonzero $\chi^{(2)}$ nonlinearity [4]. $\chi^{(2)}$ refers to the second-order correction to the susceptibility tensor, which describes the induced dipole moment per unit volume of a material in the presence of an electric field. While a linear approximation is often a sufficient description, for some materials (for example, crystals that lack inversion symmetry), there are relevant higher-order corrections. For materials with a $\chi^{(2)}$ nonlinearity, there will be a coupling between the incident pump electric field and the electric field modes of the daughter photons [5]. For this coupling to lead to an appreciable probability for downconversion, this process must conserve energy and momentum:

$$
\begin{aligned}
\omega_p &= \omega_1 + \omega_2 \\
\mathbf{k_p} &= \mathbf{k}_1 + \mathbf{k}_2,
\end{aligned}
\tag{1.24}
$$

where the conservation of momentum is generally referred to as phase matching. The source used in this work uses a 532-nm pump to generate SPDC photons at 810 and 1550 nm, satisfying energy conservation
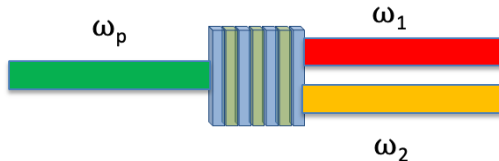
Figure 1.1: Diagram of the downconversion process. A high-energy pump enters a nonlinear crystal, in which some of the pump undergoes downconversion into two lower-energy photons, subject to energy and momentum conservation.
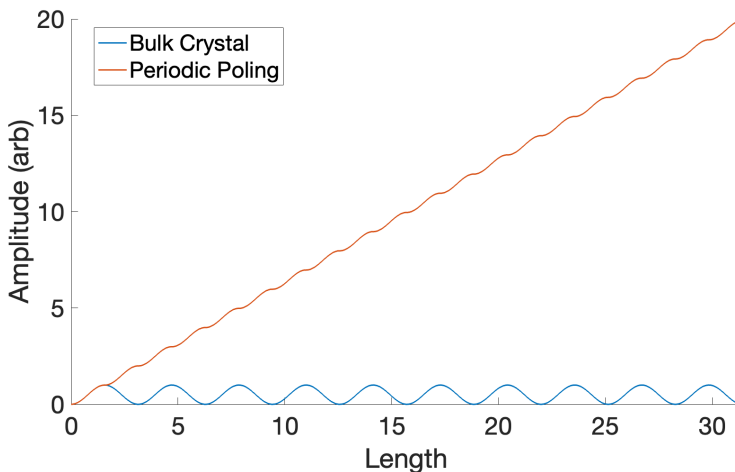
Figure 1.2: Effect of periodic poling on downconversion amplitude. With a bulk crystal (blue), the downconversion amplitude does not grow over the length of the crystal due to the momentum mismatch. By flipping the sign of the nonlinear coefficient every quarter period by periodically poling the crystal, the downconversion amplitude steadily increases (red).

because $\frac{1}{810} + \frac{1}{1550} = \frac{1}{532}$. In the lithium niobate crystal that facilitates this interaction, there is a mismatch between the pump momentum and the downconversion momenta

$$\Delta \mathbf{k} = \mathbf{k}_p - \mathbf{k}_1 - \mathbf{k}_2, \tag{1.25}$$

leading to rapid oscillations of energy between the pump and downconversion modes but no appreciable increase in the downconversion probability over the length of the crystal. To overcome this, the crystal is engineered to be periodically poled, causing the sign of the nonlinear coefficient to swap at regular intervals [6]. If this poling period is properly selected so that the nonlinear coefficient switches sign just as energy would begin to oscillate back into the pump beam, then there will be a build up of amplitude over the length of the crystal. This periodic poling technique is called quasi-phase matching and is Type-0 for this crystal,

meaning that a horizontally polarized 532-nm pump photon will downconvert into 810-nm and 1550-nm photons that are both horizontally polarized.

1.2.2 Polarization Qubits

Polarization qubits are ubiquitous in quantum optics, and the technology to manipulate them is highly developed.
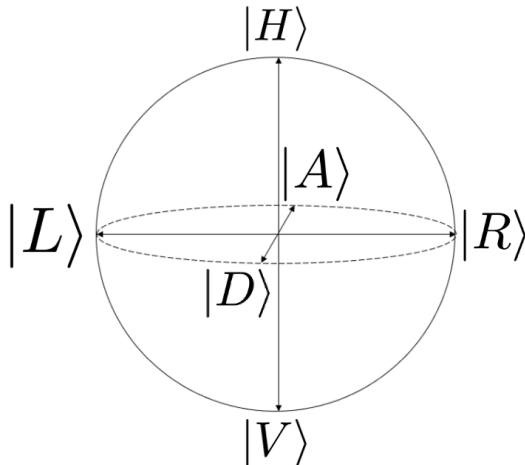
Figure 1.3: Poincaré sphere for polarization qubits. Pure polarization states reside on the surface of the sphere, and states become more mixed towards the origin.

There are a number of techniques for designing an SPDC source that creates polarization-entangled pairs, such as using a pair of crystals [7], splitting the pump with a beam displacer [8], and carrying out the SPDC process in an interferometer [9]. Fundamentally, all these techniques rely on the same principle, in which two processes that generate orthogonally polarized pairs are combined coherently. In Chapter 4, we present a source that utilizes the interferometric technique, and a schematic for this method is shown in Fig. 1.4. In this source, the counterpropagating paths of a polarizing Sagnac interferometer that contains a halfwave plate and a nonlinear crystal give rise to these two processes.

Polarization qubits can be manipulated through the use of properly designed birefringent materials. A birefringent material is one that has different indices of refraction for different polarizations. The axis along which the index of refraction is lowest is referred to as the fast axis, and the direction perpendicular to this axis is referred to as the slow axis. The two polarization modes that correspond to these axes are then the eigenmodes of the material. When traveling through a material, the phase of light changes by an amount

$$\phi = \frac{2\pi n L}{\lambda},$$
(1.26)

9

$$|V\rangle_{pump} \longrightarrow |HH\rangle \qquad |H\rangle_{pump} \longrightarrow |VV\rangle$$

CW   CCW

+
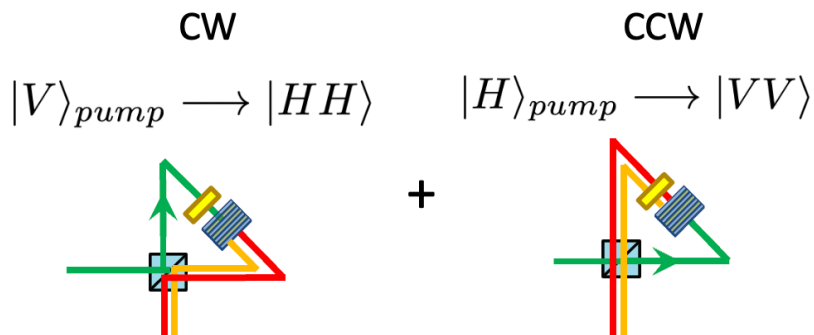
Figure 1.4: Diagram of a Sagnac polarization entanglement source. A diagonally polarized pump is incident on the interferometer's beamsplitter. The vertical (horizontal) portion of the pump circulates clockwise (counterclockwise) leading to a pair of horizontal (vertical) SPDC photons due to the halfwave plate in the interferometer. These processes add coherently, leading to a polarization-entangled state.

where n is the index of refraction, L is the length through which the light propagates, and λ is the wavelength of the light in vacuum. Because n is different for different polarizations, passing through a birefringent material leads to a phase shift between the two polarization eigenmodes. By carefully engineering the thickness of these birefringent materials, a specific phase shift can be applied (note from Equation 1.26 that this phase shift does depend on wavelength, however). In this way, birefringent materials can be used to enact unitary operations on polarization qubits. Usually, L is set so that ϕ is either π (HWP, a halfwave plate) or $\frac{\pi}{2}$ (QWP, a quarterwave plate). These devices are useful because the composition of a QWP-QWP-HWP can be used to generate an arbitrary unitary operation by properly setting the orientations of the waveplates (see Section 2.1). In addition, the combination of a QWP and a HWP can be used to change a specific pure state to any other pure state.

One other useful device for manipulating polarizations is the liquid crystal (LC). The birefringence of a liquid crystal depends on voltage applied across the crystal. This allows the LC to act as a halfwave plate, a quarter waveplate, or other phase retarder, depending on the applied voltage, or to apply a tunable phase shift to a certain polarization.

Pockels cells (PC) are another electro-optic device that displays a voltage-dependent birefringence [10]. In contrast to liquid crystals, Pockels cells are designed so that their birefringence can be manipulated very rapidly, at a time scale faster than the nanosecond-scale separation of the time-bins discussed in the following section. This allows Pockels cells to selectively manipulate the polarization of one time-bin while leaving the other unchanged.

Projective measurements on polarization can be carried out through the use of a polarizing beam splitter (PBS) or polarizer. PBSs transmit one polarization and reflect the orthogonal polarization. By placing a

QWP and a HWP before the PBS, the basis of measurement of the PBS can be modified to allow arbitrary projective measurements. Polarizers work in a similar fashion, transmitting only one polarization and absorbing the other. Combining a polarizer with a QWP and HWP enables arbitrary projections. Usually, PBSs are preferable because they allow measurements of the two orthogonal polarizations simultaneously.
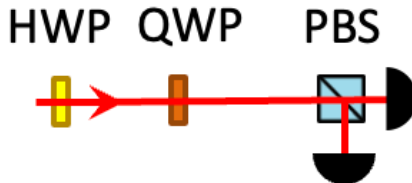
Figure 1.5: Diagram of a polarization analyzer. By modifying the orientation angle of the HWP and QWP, the PBS can be put into an arbitrary measurement basis.

1.2.3 Time-Bin Qubits

Qubits can also be encoded in the creation time of the photons. Although this is a continuous degree of freedom, it can be converted to a system with discrete encoding (e.g., qubits, or more generally, qudits) through the use of binning—for example, by setting $|t_1\rangle \equiv [t_1, t_1 + \Delta t]$ and $|t_2\rangle \equiv [t_2, t_2 + \Delta t]$, where the intervals indicate the time during which the photon was created.

SPDC sources can be modified to produce time-bin entangled photons by placing the pump in a coherent superposition of timing modes before sending it through the nonlinear crystal. The pump can be prepared in such a way by sending it through an unbalanced interferometer, so that one path has a longer propagation time than the other. In order to interfere subsequent time-bins, the early time-bin must be delayed to overlap with the later time-bin. The precision with which this delay must be carried out is determined by the width of the pulse, given by

$$t_{pulse} \approx \frac{\lambda^2}{c\Delta\lambda}, \tag{1.27}$$

where λ is the pump's wavelength and $\Delta\lambda$ is its frequency bandwidth. Fig. 1.6 provides a schematic for a time-bin entangled downconversion source.

Measuring time-bin qubits in the t_1/t_2 basis is simple, as the time of the detection event can be used to determine the measurement outcome. This does necessitate having detection electronics that can resolve the two time-bins, as well as a shared master clock. Measuring in superposition bases $\frac{|t_1\rangle \pm e^{i\phi}|t_2\rangle}{\sqrt{2}}$ is more challenging and necessitates the use of unbalanced interferometers. Because these measurements are fundamentally phase-sensitive, these measurements generally require active stabilization to ensure that the
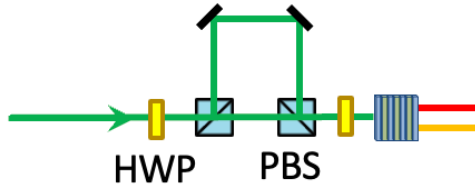
Figure 1.6: Diagram of a time-bin entangled SPDC source. The pump travels through an unbalanced interferometer to prepare it in a superposition of time-bins before passing through the nonlinear crystal. The HWP before the interferometer allows control of the relative amplitude of the two time-bins, while the HWP after the interferometer allows the pump's polarization state to be set before the downconversion process.
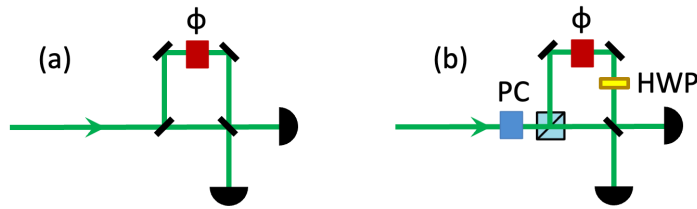
Figure 1.7: Diagram of a time-bin measurement system. (a) Passive system. Because this system does not actively sort the time-bins, there are three possible measurement times at the detector, arising from the combination of photon creation times and interferometer paths. For the middle measurement time, this system measures in the superposition basis determined by ϕ. Detection events on both detectors for the early measurement time can be combined with late measurement times to measure in the t_1/t_2 basis. (b) Active system. Here, the first beamsplitter is replaced by an active switch. To measure in the t_1/t_2 basis, the switch simply transmits the time-bins; detector timing then projects onto the t_1/t_2 states. To measure in the superposition basis $t_1 \pm e^{i\phi}t_2$, the switch routes the early time-bin to the long path, and the late time-bin to the short path. Detectors after the final recombining beamsplitter then project onto $\frac{|t_1\rangle \pm e^{i\phi}t_2\rangle}{\sqrt{2}}$.

measurement interferometer matches the time-bin separation and is phase-stable. An example time-bin measurement system is shown in Fig. 1.7.

1.3 Quantum communication and quantum networks

Having described how to both represent and generate entangled quantum states, we can now move to providing a number of applications that these states can help enable. We will focus on three applications: quantum key distribution (QKD), quantum repeaters, and blind quantum computing. This by no means constitutes an exhaustive list of applications in the quantum communication and quantum information science field, but these protocols should be illustrative of the types of advantages and challenges that arise in this field. These three applications also span a wide range of technological maturity and can, therefore,

provide context as to what advances in quantum information science can enable. The work presented in Chapters 3 and 4 on quantum tomography and verification of entanglement can be seen as providing an improvement on the tools and techniques available when implementing quantum protocols such as these, and this discussion can provide context for the work presented in these chapters.

1.3.1 Quantum Key Distribution

Quantum Key Distribution (QKD) was first proposed in 1984 by Bennett and Brassard [11]. QKD is a cryptography scheme that allows for provably secure communication. To begin one version of the protocol, Alice and Bob are each sent one photon from an entangled pair source. They then independently and randomly choose to measure in one of two predetermined bases. Because entangled states display correlations in all bases, whenever they both measure in the same basis, their measurements are correlated, and they can use these correlations as the basis for a one-time pad to encrypt and decrypt their communication. If an eavesdropper were to make measurements on photons before they reached Alice and Bob, the degree of entanglement will be reduced, leading to errors in the correlations between Alice and Bob's measurements. By comparing a subset of their measurements, it is possible to deduce the maximum amount of information that an eavesdropper could have about the remaining data, and subsequent processing can be employed to create a completely secure key. The security of this protocol is based on two features of quantum mechanics. The first is that measurement of a quantum state in general causes changes to the state itself. Because of this, an eavesdropper who measures the transmitted quantum states is detectable because, some fraction of the time, she will necessarily cause measurable changes to the states that she intercepts. This provides security when combined with the second feature: the no-cloning theorem [12] . This theorem states that an arbitrary qubit cannot be copied[2]. This limits the capability of an eavesdropper; unlike in a classical scenario, the eavesdropper cannot copy the signal and measure the copy, leaving the original unchanged. The security of the communication is judged by the rate at which measurements differ from their expected outcomes, called the qubit error rate. By attributing these errors to an eavesdropper, it is possible to bound the amount of information that a third party could have. While the protocol is completely secure when implemented perfectly, in principle it is challenging to completely remove "side-channels" that can leak information to an eavesdropper [13, 14, 15]. One approach to remove many of these side channels is to implement the protocol in a "device-independent" way [16, 17]. In such a scheme, the security is ensured by the violation of a Bell inequality, which itself does not rely on devices working as intended to indicate the presence of entanglement.

---

[2]A common proof for this theorem considers a cloning unitary operator U that copies one state onto another: $U|\psi 0\rangle = |\psi\psi\rangle$. If U is applied to two arbitrary states $|\psi\rangle$ and $|\theta\rangle$, then $U|\psi 0\rangle = |\psi\psi\rangle$ and $U|\theta 0\rangle = |\theta\theta\rangle$. This implies that $\langle\theta|U^\dagger U|\psi\rangle = \langle\theta\theta|\psi\psi\rangle$, which means that $\langle\theta|\psi\rangle = \langle\theta|\psi\rangle^2$, so either $|\psi\rangle = |\theta\rangle$ or the two are orthogonal. This contradicts the states being arbitrary, so U must not exist.

Based on the degree of violation of the Bell inequality, a bound can be set on the amount of information leaked to a third party.

1.3.2 Quantum Repeaters

While QKD is a robust technology with many experimental implementations, quantum repeaters are an enabling technology that still lacks effective experimental implementation. Due to the no-cloning theorem, it is not possible to amplify quantum signals. Because attenuation is inevitable when photonically-encoded quantum states are transmitted over significant distances, this places a limit on the internode separation in a quantum network.. Quantum repeaters attempt to overcome this challenge through the distribution of quantum states through entanglement swapping [18]. In the process of entanglement swapping [19], a Bell measurement is made on the two halves of two entangled pairs. This projects the other two photons into a Bell pair as well, effectively swapping which photons are entangled. A schematic for such a device is shown in Fig. 1.8. This simple version obscures some of the largest technical challenges, most importantly, that it requires synchronous success of all elements to distribute the entanglement. In practice, both the generation of entangled pairs and the projection onto a Bell state are probabilistic events. In order to realize an improvement rate of entanglement distribution, quantum memories [20] must be used to allow asynchronously successful entanglement links to be swapped.
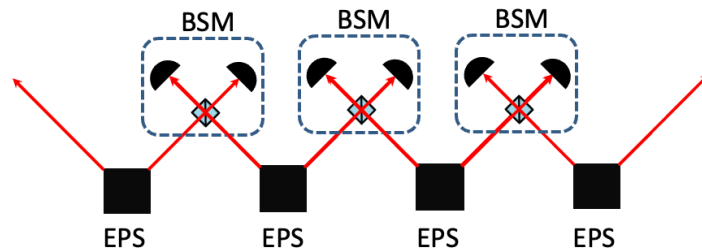
Figure 1.8: Schematic of a quantum repeater. By repeatedly carrying out Bell state measurements (BSM) between photons from adjacent entangled photon sources (EPS), the entanglement can be distributed across the network.

1.3.3 Blind Quantum Computing

Blind quantum computing (BQC) is a client-server model of quantum computing whereby the server is blinded from the details of the computation being performed [21]. In a scenario where quantum computing

is a scarce resource, many parties may wish to utilize a central quantum computer. Directly programming the computer in a straightforward way would lead those running the server quantum computer to be able to deduce the computation and its results, which may be against the interests of the client. BQC solves this problem by providing a means for a client with limited quantum resources to direct quantum computation on a remote server without revealing the computation or its results. BQC works in a cluster state, measurement-based quantum computing architecture. In this framework, the server generates a cluster state by entangling qubits that are prepared and sent to it by the client. The computation then proceeds in an iterative way by the repeated measurement of qubits in the cluster state. The measurement results are used by the client to determine the basis of future measurements and also constitute the outcome of the computation. In order for this process to be blind, the server must not be able to determine the input state, the measurement performed, or the result. In order to blind the input state, the client applies random phases to the supplied input state before transmission, so that only they know what these phases are. The computation and the measurement result can be simultaneously blinded by adding random π phase shifts to the basis that each measurement is carried out in; in this way, the client knows which measurement results contain bit flips and which do not. It is not possible to fully blind the computation itself because the minimum complexity of a calculation depends on the number of qubits in the cluster state, so the server can always put an upper bound on the complexity of the calculation being performed. However, by sending extra unused input qubits, the client can effectively mask the true size of the computation.

Chapter 2

Quantum Gates

A common framework for discussing manipulations of quantum systems is through the use of quantum gates. Each quantum gate is a unitary operation, and the composition of certain gates can allow for universal quantum computation, just as the composition of simple logic gates can lead to universal classical computation. This can simplify the demands on experimental systems because only a limited number of types of gates need to be implemented. Universal computation can be achieved through the combination of a minimal set of single-qubit gates with an entangling two-qubit gate. A common set of single-qubit gates that can be used to achieve universality are the Pauli operators in conjunction with the Hadamard gate, the S gate, and the T gate. Matrix representations of these gates are shown in Equation 2.1:

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, S = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}, T = \begin{bmatrix} 1 & 0 \\ 0 & e^{\frac{i\pi}{4}} \end{bmatrix}. \quad (2.1)$$

A commonly used class of entangling gates are called controlled gates, which conditionally apply a unitary to one qubit depending on the state of the other; such gates have the form

$$CU = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & u_{11} & u_{12} \\ 0 & 0 & u_{21} & u_{22} \end{bmatrix}. \quad (2.2)$$

Here, U could be any single-qubit gate. For example, the controlled-Y gate would have a matrix representation

$$CY = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \end{bmatrix} \quad (2.3)$$

16

and would apply a Y gate to the second qubit if the first were in state $|1\rangle$, and leave the state unchanged if the first qubit were in state $|0\rangle$. In the following sections, we will discuss implementations of these gates for polarization and time-bin qubits.

2.1 Polarization Gates

For polarization qubits, we set the computational basis to be

$$|0\rangle \equiv |H\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, |1\rangle \equiv |V\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{2.4}$$

Implementing all unitary operations on polarization qubits is possible through the use of waveplates. An arbitrary unitary transformation on polarization states can be generated from three waveplates, so all gates take, at most, three waveplates to implement, but some can be achieved using only a single waveplate. Implementing these gates requires the waveplate to be rotated. These rotations are defined by the angle of the waveplate's fast optic axis (i.e., the axis of the waveplate with the lowest index of refraction) with respect to vertical, and are given here in degrees. Note that these gates are implemented up to a global phase. For example, the series of waveplates in Fig. 2.1 implementing a Y gate actually apply the unitary $e^{\frac{i\pi}{2}}Y$. To implement the remaining single-qubit gates, note that the Z, H, S, and T gates are all diagonal
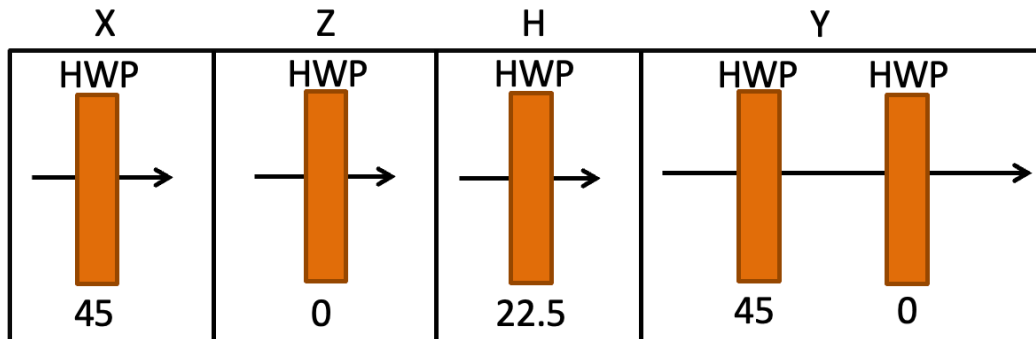
Figure 2.1: Schematic for implementing polarization gates with waveplates. The X, Z, and H gates can be implemented with a single waveplate, but the Y gate is more complicated. The values under each waveplate indicate the orientation of the waveplate's fast axis in degrees.

with a different phase applied to the $|1\rangle$ state. These gates can be generalized to an arbitrary phase gate:

$$P(\phi) = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{bmatrix}. \tag{2.5}$$

By implementing this arbitrary phase gate, all the other necessary gates will be implemented. Fig. 2.2 shows a technique for implementing an arbitrary phase gate using two QWPs and a HWP.
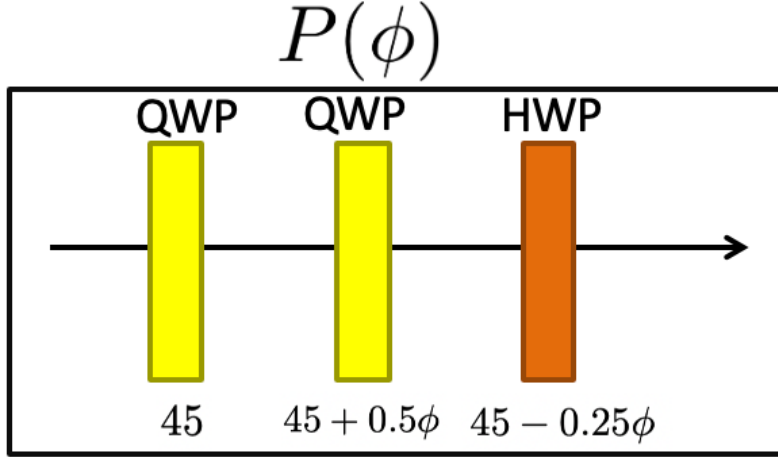
Figure 2.2: Schematic for implementing a polarization phase gate with waveplates. The values under each waveplate indicate the orientation of the waveplate's fast axis, in degrees.

2.2   Timing Gates

For time-bin gates, we set the computational basis to be

$$|0\rangle \equiv |t_1\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, |1\rangle \equiv |t_2\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{2.6}$$

Direct manipulation of timing-mode qubits is challenging because arbitrary unitary operations, in general, require moving state amplitude from one time-bin to the other. Fig. 2.3 shows ways to implement single-qubit gates on time-bin qubits. These gates are presented in a polarization-independent way. In general, carrying out these gates requires the use of an ancilla degree of freedom (DOF) to control the two time-bins independently, and polarization is an effective tool for this because of the maturity of the technology for manipulating polarization[1]. This presents a problem if the polarization is already being used as another qubit on the photon. To overcome this problem, beam displacers are used to map the polarization qubit onto a pair of spatial modes. After this mapping has occurred, the polarization can be used as an ancilla

---

[1]Here we use polarization to control the effective splitting ratio of the beamsplitter—ranging from complete transmission to complete reflection. One could achieve similar control by replacing each beamsplitter with a tunable Mach-Zehnder interferometer (MZI); by changing the phases inside and outside the MZI, one can similarly control the transmission and reflection amplitudes of this effectively tunable beamsplitter. Such methods have been employed in recent integrated optic qubit-processing chips [22], where the integration enables many interferometers on a single chip. The disadvantage of such an approach is that any phase drifts in the interferometers change the effective "reflectivity" of the "beamsplitters" they are implementing.

DOF to facilitate the timing gate, and a final beam displacer can be used to transduce the spatial modes back into a polarization qubit.
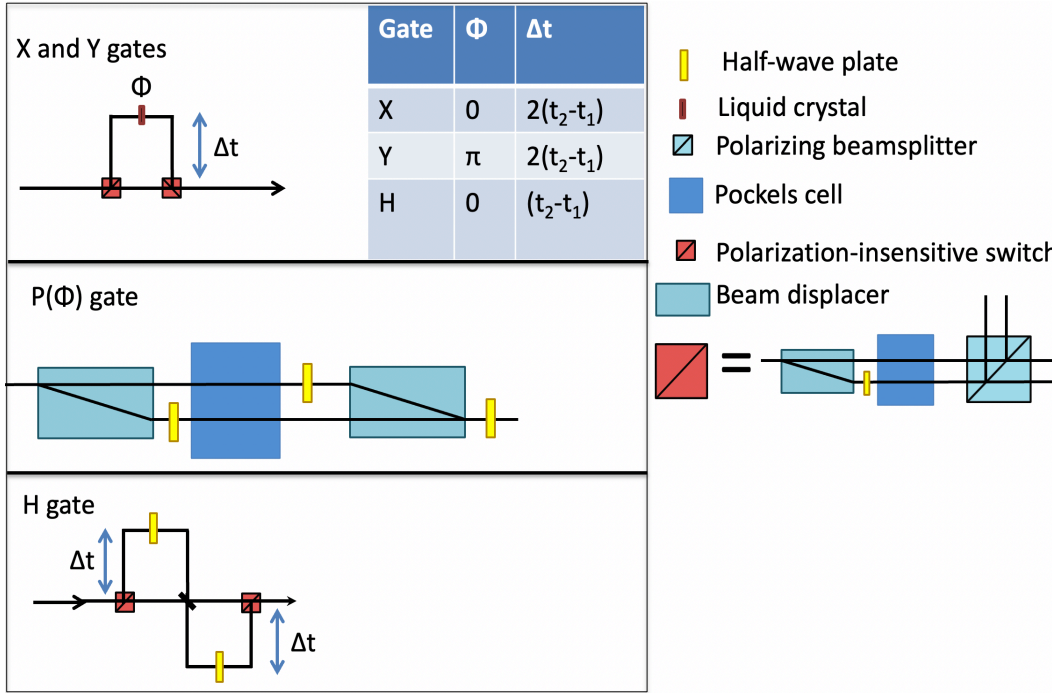
Figure 2.3: Schematic for implementing gates on time-bin qubits. Because the X,Y, and P(ϕ) gates do not require mixing time-bins, they can be implemented using only one interferometer. To apply an arbitrary phase-gate, the Pockels cell voltage should be set so it applies a phase of ϕ and should be on for the late time-bin and off for the early time-bin. The gates shown here are constructed in a polarization-independent way using beam displacers to map the polarization modes to spatial modes before using the polarization as an ancilla to assist in the operation of the time-bin gate. Implementing an H gate requires mixing time-bins, so two unbalanced interferometers are necessary. In the polarization-insensitive switches, the Pockels cells are set to the halfwave voltage with their axis oriented at 45°, so that they rotate $|H\rangle$ to $|V\rangle$. The Pockels cell fires for the early time-bin so that this time-bin is routed into the PBS's reflected port and is vertically polarized, while the late time-bin is unaffected by the Pockels cell and is routed through the PBS's transmitted port and is horizontally polarized. All HWPs are also set at 45° to swap $|H\rangle$ and $|V\rangle$.

Implementing X and Y gates requires delaying the early time-bin relative to the late one. By delaying the early time-bin by twice the time-bin separation, the amplitudes in the early and late time-bins can be swapped. By applying a phase shift to one path of the delay interferometer, this technique can implement either an X or Y gate.

Implementing arbitrary phase gates can be carried out using a pair of beam displacers and a Pockels cell. The beam displacers allow the use of a Pockels cell to apply the necessary phase without disturbing the polarization state of the photon. The Pockels cell is set to apply a phase ϕ to the second time-bin.

Implementing a Hadamard gate is more challenging than the other gates because it requires the creation

of superpositions of time-bins, so two unbalanced interferometers are necessary. The first delays the early time-bin so that it arrives at the beamsplitter at the same time as the late time-bin. This causes the output modes of the beamsplitter to be coupled to superpositions of the time-bins (See Fig. 2.3). Finally, a second unbalanced interferometer is used to map the two spatial-modes back onto a single spatial-mode with two time-bins. Because the polarization-insensitive switch causes the two spatial modes to have different polarizations, a halfwave plate is inserted in the reflected port of the PBS so that the two spatial modes have the same polarization when they interfere at the intermediate beamsplitter. After the spatial modes mix at the beamsplitter, the polarization of one mode must again be modified so that the second polarization-insensitive switch can properly route both time-bins onto a single spatial mode. Schematics for these time-bin gates are shown in Fig. 2.3.

$$|H\rangle \otimes \frac{\alpha|t_1\rangle + \beta|t_2\rangle}{\sqrt{|\alpha|^2 + |\beta|^2}} \rightarrow \frac{\alpha|V\rangle + \beta|H\rangle}{\sqrt{|\alpha|^2 + |\beta|^2}} \otimes |t_2\rangle$$
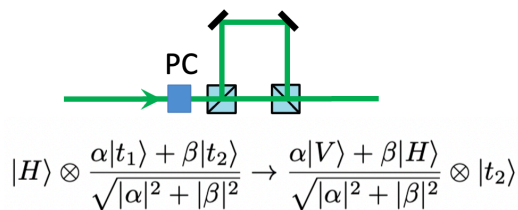
Figure 2.4: Diagram of a time-bin-polarization transducer. The Pockels cell fires to direct the early time-bin into the long path, delaying it and mapping it onto vertical polarization. The late time-bin remains horizontally polarized, and the qubit is recombined on the second polarizing beamsplitter; the relative path-length difference is set to match the time-bin qubit difference $t_2 - t_1$. This device can also be run in reverse to convert a polarization qubit into a time-bin qubit.

These techniques can be cumbersome because of the number of optical elements (e.g., $\sim$10 elements to realize a single X gate) necessary to create time-bin gates. For this reason, it is often more effective to transduce a timing qubit into a qubit in another degree of freedom. Such a device is shown in Fig. 2.4; this time-to-polarization transducer converts a time-bin qubit into a polarization qubit through the use of a Pockels cell and an unbalanced interferometer. Such a device can also be used to transduce a polarization qubit into a timing qubit. After converting the timing qubit into a polarization qubit, the polarization manipulation techniques in Section 2.1 can be used to efficiently modify the qubit's state before it is transduced back into a time-bin qubit.

2.3 Two-qubit gates between polarization and time-bin

While the single-qubit timing gates can be complicated, carrying out two-qubit gates using polarization and timing is relatively simple. If the time-bin is used as the control qubit, then an arbitrary control gate can
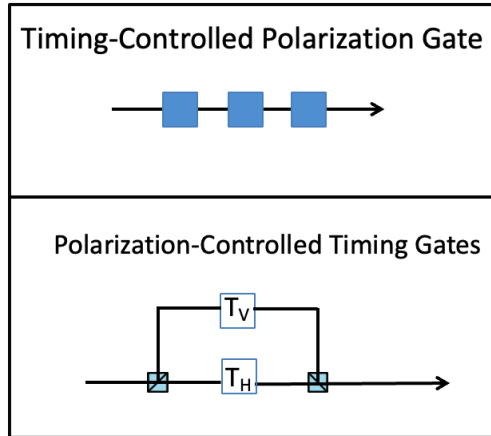
Figure 2.5: Schematic of two-qubit gates using polarization and time-bin. To use the time-bin qubit as the control, a series of three Pockels cells are used to apply the desired unitary to one time-bin and not the other. To use the polarization qubit as a control, a PBS can be used to direct the horizontal polarization to a spatial mode that contains a time-bin gate T_H and the vertical polarization to a spatial mode that contains a time-bin gate T_V. These two spatial modes can then be recombined using another PBS.

Gate	PC 1 Phase	PC 2 Phase	PC 3 Phase
Z	0	π	0
S	0	$\frac{\pi}{2}$	0
T	0	$\frac{\pi}{4}$	0
H	$\frac{3\pi}{2}$	$\frac{\pi}{2}$	$\frac{\pi}{2}$
X	0	0	π
Y	0	π	π

Table 2.1: Table of Pockels cell phases for implementing arbitrary unitary operations with Pockels cell orientations fixed. The first Pockels cell is oriented at $45°$, the second at $0°$, and the third at $-45°$.

be carried out using a series of three Pockels cells (each acting like a waveplate, as discussed in Section 2.1). This allows one unitary to be applied to the first time-bin and a different unitary to be applied to the second time-bin. To generate any desired controlled gate, the first two Pockels cells can be set to the quarter-wave voltage and the last to a half-wave voltage. Then, their axes can be aligned to the angles given in Fig. 2.1 or Fig. 2.2, depending on the desired unitary. The Pockels cells must then be driven to apply this unitary to the second time-bin. However, it is experimentally simpler to modify a Pockels cells phase than to rotate its axis. It is also possible to apply an arbitrary unitary using a series of three Pockels cells with fixed orientations by varying the phase applied by each device. We will consider the case where the first Pockels cell is oriented at $45°$, the second at $0°$, and the third at $-45°$, although other configurations can also be used. This configuration can easily create a $P(\phi)$ gate by applying a phase shift of ϕ on the second Pockels cell, and zero phase on the others. Phase settings for the gates in Equation 2.1 are shown in Table 2.1.

If, instead, the polarization is used as the control qubit, then a PBS can be used to couple the two polarization modes to separate spatial modes. Then, a different time-bin gate (such as those described in Fig. 2.3) can be applied to each spatial mode before the spatial modes are recombined using a final PBS. A diagram of these two-qubit gates is shown in Fig. 2.5.

Chapter 3

Quantum Tomography

A ubiquitous feature in the practical application of quantum tasks is the need to verify that the states being generated are the ones desired. Whether the goal is to generate specific qubit states to transfer to a quantum server in the case of blind quantum computing [21] or to create Bell pairs in order to distribute entanglement across a quantum network, success of a protocol is dependent on the ability to produce specific states. It is therefore critical to be able to measure the quality of the states being generated. In the case of entanglement distribution across a quantum network, it has been shown that the optimal strategy can depend on the quality of the entanglement being generated [23]. Quantum tomography is the process through which the state of a quantum object can be determined. Because measurements of quantum objects are projective and thus modify the state, tomography can only be carried out with a high degree of accuracy when many identical copies of the state are available for measurement.[1] If only one copy of the state is available, then the task of optimally determining its state is referred to as quantum orienteering [24], for which techniques such as entanglement-assisted measurements can boost the quality of the result; in general the infinite amount of information conveyed by a quantum state (e.g., of an arbitrary qubit) cannot of course be perfectly determined by a finite number of measurements. Usually, multiple copies of the state of interest are available and the precision requirements of the state reconstruction are higher than orienteering allows, so the more general technique of quantum tomography is employed. Tomographies consist of two main steps: data acquisition and algorithmic state reconstruction. In the data acquisition step, a number of observables O_i are measured on (hopefully) identical copies of the state to be reconstructed in order to determine the probabilities of the various outcomes for these observables. In many cases, the exact number of incoming copies of the state is not known, and the measurement consists of the number of observed events M_i for each outcome. By measuring a full basis, these frequencies can be converted to probabilities; these probabilities are then processed by a tomography algorithm to produce a quantum state that is reflective of those measurements.

---

[1]Tomography using "weak" measurements has also been investigated [25], but here a much greater number of copies of the state are needed, which likely makes it unsuitable for practical applications.

3.1 Maximum Likelihood Estimation

3.1.1 Constructing a Likelihood Function

One way of choosing a representative quantum state to reflect measurements is to select the state that was most likely to have produced those measurements. Indeed, this strategy is the most pervasive tomography technique [26]. If we can construct a function $\mathcal{L}(\rho)$ that is proportional to the likelihood of a given state ρ, then we can cast the task of tomography as an optimization problem over all possible density matrices. In quantum optics, measured photon counts obey Poissonian statistics, so we can use a Poissonian distribution as the basis for our likelihood function. For mean values greater than a few counts, we can approximate a Poissonian distribution as a Gaussian distribution, which will be a useful approximation to make for computational reasons (that the approximation begins to fail near mean values of zero can be easily seen due to the symmetric nature of the Gaussian distribution; near zero, it will predict nonzero probabilities for negative count values, whereas the Poissonian distribution admits nonzero probabilities only for positive values). The likelihood will have the form

$$\mathcal{L}(\rho) = \prod_i e^{-\frac{(M_i - M_{expected}(\rho))^2}{2\sigma^2}}, \tag{3.1}$$

where M_i is the number of counts observed when the measurement observable was O_i, $M_{expected}(\rho)$ is the number of expected counts when the observable O_i is measured for the state ρ, and σ is the standard deviation of $M_{expected}$. For a given measurement and quantum state, we would predict the probability of observing a count as $|\langle O|\rho\rangle|^2$; then, the number of expected counts will be $\mathcal{N}|\langle O|\rho\rangle|^2$, where $\mathcal{N}$ is the number of incident photons. We will henceforth assume these are photons, but of course the arguments apply to any physical system (e.g., atomic states, spin states, etc.) In most photon counting experiments, we do not know how many photons were emitted, so the value of $\mathcal{N}$ is unknown. In maximum likelihood estimation (MLE) techniques for which the amplitude is unknown, this becomes another free parameter in the optimization, and we can include this factor in ρ if we do not enforce a normalization constraint on ρ. Poissonian distributions have the property that the standard deviation σ is equal to the mean, so we can determine σ^2 to be $\mathcal{N}^2|\langle O|\rho\rangle|^4$. The likelihood will then be

$$\mathcal{L}(\rho) = \prod_i e^{-\frac{(M_i - |\langle O_i|\rho\rangle|^2)^2}{2|\langle O_i|\rho\rangle|^4}}. \tag{3.2}$$

At this point, it is worthwhile to have a brief discussion of what this likelihood represents. The quantity calculated is the likelihood of observing a set of measurements M_i (with associated measurement operators

O_i), if the input state was the given density matrix ρ; in other words, L is proportional to the probability of observing the counts M_i, conditioned on the state entering the measurement system being ρ. This explains why we estimate σ as $M_{expected}$ and not as M_i; although this distinction has minor import because $M_i \approx |\langle O|\rho\rangle|^2$ for the most likely state, it helps to produce meaningful results in the edge case of measuring zero counts. If we estimated σ as M_i, then we would erroneously conclude that the likelihood was zero for any state for which $M_i \neq |\langle O_i|\rho\rangle|^2$. In contrast, if $|\langle O|\rho\rangle|^2 = 0$, then ρ must be a pure state, so if $M_i \neq 0$, then there is no possibility that the measured state was ρ and the likelihood should be zero.

3.1.2   Determining the Most Likely State

Using the likelihood function $\mathcal{L}$, the density matrix with the highest likelihood of producing the observed set of measurements is then chosen as the estimator of the state. Computing a maximum likelihood tomography is therefore reduced to an optimization problem: finding $\rho(\mathbf{t})$ such that $L(\rho(\mathbf{t}), M)$ is maximized. Here, $\rho(\mathbf{t})$ denotes a certain parameterization of the density matrix, where $\{\mathbf{t}\}$ are the parameters being optimized. If a naive parameterization is used, it is possible to produce invalid density matrices with negative eigenvalues; for this reason the Cholesky decomposition is used [27]: $\rho(\mathbf{t}){=}TT^\dagger$, where

$$T = \begin{bmatrix} t_1 & 0 & \ldots & 0 \\ t_{2^n+1} + it_{2^n+2} & t_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & 0 \\ t_{4^n-1} + it_{4^n} & t_{4^n-3} + it_{4^n-2} & t_{4^n-5} + it_{4^n-4} & t_{2n} \end{bmatrix}. \tag{3.3}$$

This parameterization ensures that for any set of real parameters t, $\rho(\mathbf{t})$ will be a valid density matrix.

3.1.3   Drawback to Maximum Likelihood Tomography

Despite the prevalence of its use in the quantum information community, there are a number of drawbacks to this technique that motivate the need to search for better alternatives. The foremost drawback is that maximum likelihood estimation has a systematic bias towards pure states [28]. This bias is more pronounced at low measurement amplitudes, and the presence of this bias can be observed by analyzing the algorithm's low-count behavior. For the case in which no measurements are recorded, the best estimate is that all states are equally likely, so the average state would be the completely mixed state. Alternatively, the best estimate could be chosen as the state with the highest average fidelity with all other possible states. In either case, without outside information about what the state should be, the best initial guess is the completely mixed state. Fig. 3.1 shows that maximum likelihood tomography does not reproduce this desired low-amplitude

behavior: even for the smallest number of measurements, it can still report a completely pure state; this is a manifestation of the bias inherent in maximum likelihood tomography. This bias is unavoidable because the likelihood function being optimized cannot be symmetric due to the boundary imposed by the condition that $Tr(\rho) \leq 1$. This condition leads the likelihood function to always be skewed towards pure states (except for the case in which the most likely state is the completely mixed state).
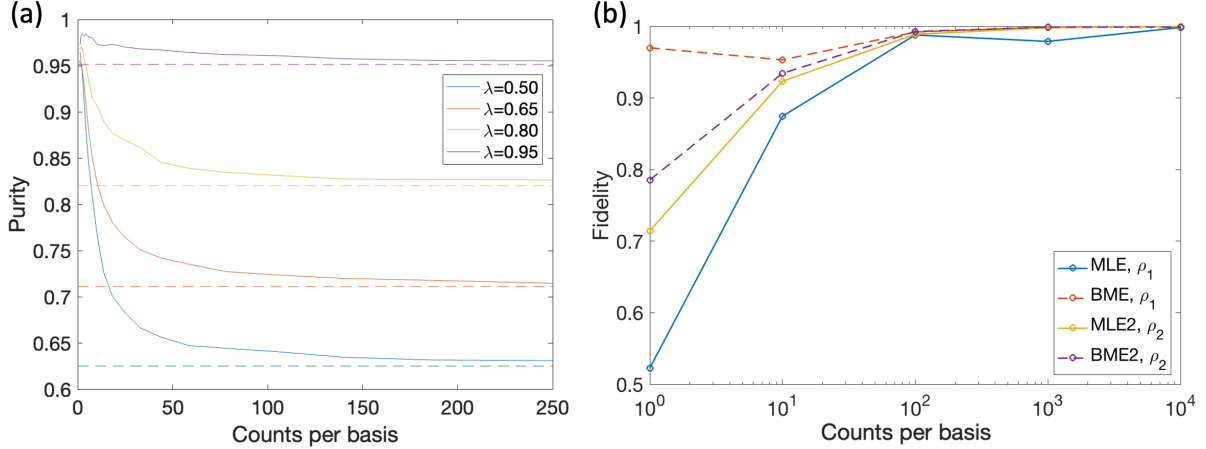
Figure 3.1: (a) Example maximum likelihood tomographies, demonstrating that the reconstructed state has increasingly high purity as the number of measured counts decreases. Analyzed states have the form $\lambda|H\rangle\langle H| + \frac{1-\lambda}{2}\mathbb{I}_2$. Solid lines show the purity of the tomographically reconstructed state, while dashed lines show the actual purity of the state being measured. (b) Sample tomographies for two qubit states of varying purity, showing the advantage of the BME method at low count rates. For ρ_1, λ=0.1 and for ρ_2, $\lambda = 0.9$.

3.2 Bayesian Mean Estimation

The origin of the bias in maximum likelihood estimation arises from the fact that the estimator being used is the maximum, which can be a biased estimator. This bias can be avoided by representing the data with an unbiased estimator, such as the mean. This is the heart of the tomographic approach called Bayesian mean estimation (BME) [29]. In this framework, the primary task is determining the posterior probability distribution, which here consists of the probability that each possible density matrix ρ produced our measurements— that is, the probability of a density matrix conditioned on our observed measurements $Pr(\rho|M)$. The likelihood function in the previous section was proportional to a related quantity: the probability of a given measurement outcome occurring, conditioned on ρ being the state measured: $Pr(M|\rho)$. Bayes's rule relates these two quantities together:

$$Pr(\rho|M) = Pr(M|\rho)\frac{Pr(\rho)}{Pr(M)}. \tag{3.4}$$

$Pr(\rho|M)$ is referred to as the posterior, while $Pr(\rho)$ is called the prior. $Pr(M|\rho)$ is simply the likelihood function discussed earlier, while $Pr(M)$ corresponds to a normalization factor that makes the posterior a proper probability distribution. With the posterior, it is possible to calculate the mean state by computing the integral

$$\rho_{BME} = \int_{\rho} \rho Pr(\rho|M)d\rho. \tag{3.5}$$

Other statistical properties of the distribution can be calculated in an analogous way. In general, this integral does not have an analytic solution, and numerical methods must be used to calculate ρ_{BME}. In the coming sections, we describe our application of Monte Carlo techniques to estimate the integral in Equation 3.5. Fig. 3.2 presents an overview of the algorithm, which will be examined in-depth in the following sections. The algorithm proceeds as follows. Initially, sample states are drawn from an unbiased prior distribution, and their likelihood is calculated. Details on the choice of a prior distribution are discussed in Section 3.3. When enough states from this prior have been sampled, the sampled states and associated likelihoods are used to estimate the posterior distribution. The approach to estimating the posterior is discussed in Section 3.3.1. The estimate of ρ_{BME} before sampling the current estimate of the posterior is recorded as $\rho_{previous}$. Next, states are repeatedly sampled from the estimate of the posterior distribution, and their likelihoods are calculated and used to update the value of ρ_{BME}. After a set number of states are sampled from this posterior estimate, the stability of ρ_{BME} is calculated by comparing its current value to the previous value $\rho_{previous}$. If the stability is sufficiently high, then the algorithm terminates, returning ρ_{BME} as the estimate of the incoming state. A description of how a numeric measure of this stability was constructed and how a threshold was developed for when this quantity indicates stability is covered in Section 3.4. If the stability is not high enough, the new state-likelihood pairs that were sampled from the previous posterior estimate are used to update the posterior estimate. This process of repeatedly sampling one posterior estimate, evaluating the stability of ρ_{BME}, then updating the estimate of the posterior comprises one *iteration* of the algorithm, which consists of many *samples* of one posterior estimate. For example, if the algorithm completed in two iterations, then the estimate of ρ_{BME} would consist of samples from three sources: 1) 300 samples of an unbiased prior, 2) 1000 samples of the first estimate of the posterior, and 3) 1000 samples from the second estimate of the posterior. How many samples of the initial prior are used and how many samples are used in each iteration of the algorithm are set as inputs to the algorithm, and the values given here are examples. Benchmarks for the performance of this algorithm are given in Section 3.5, and its potential applications are discussed in Section 3.6. Those familiar with Bayesian estimation approaches may wonder at what point in the measurement process this technique can be applied, with the idea that a new estimate of the tomography could be calculated after each measurement. While the BME algorithm could be applied in this way, so

too could the MLE algorithm. The two differ slightly in the types of results they return if they are applied to data collected from measurements that do not span the space of density matrices, as may be possible in the middle of a set of measurements. In such a situation, the BME algorithm will return a state that is completely mixed (up to statistical noise), while the MLE algorithm will return a state that depends on the initial conditions of the optimization because an undercomplete set of measurements leads to the most likely state not being unique. In any case, these algorithms are not designed to produce meaningful results in that scenario, and if an undercomplete set of measurements can determine the feature of interest in the state, then a more specialized data analysis (e.g., measurements to determine a certain phase in the density matrix), and not a tomographic reconstruction, will be more applicable. For all the results presented here, the assumption is made that all the data is collected first, and afterwards, the algorithms are applied to the completed data sets.
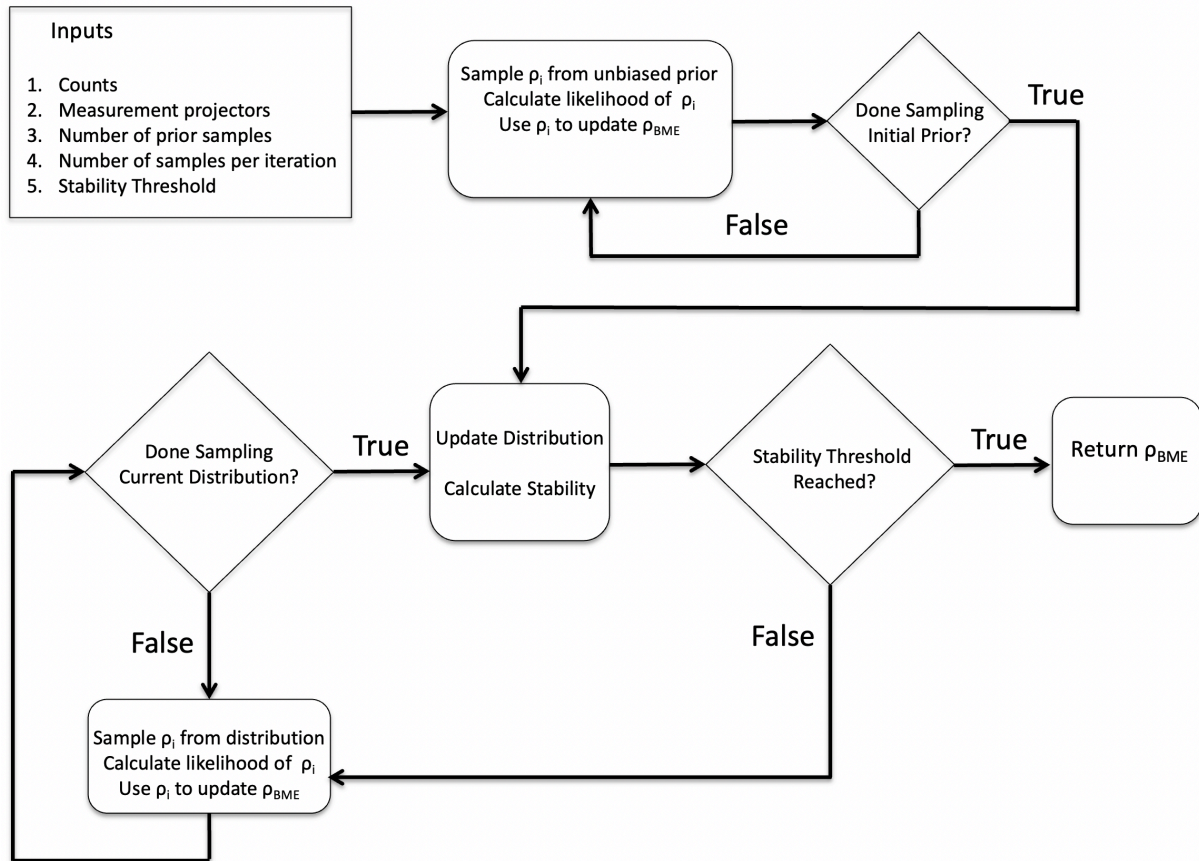
Figure 3.2: Schematic for the Bayesian mean estimation algorithm.

3.3 Prior Distributions

In order to make progress towards determining ρ_{BME}, a prior distribution $Pr(\rho)$ must first be selected. It is necessary that this prior distribution has a nonzero probability for all density matrices, or certain states cannot be accurately reconstructed. Other than this restriction, the choice of a prior distribution is unimportant in terms of the correctness of the final result, but it does have a large impact on the computation time required to arrive at the final answer. Broadly, there are two types of priors: "unbiased" and "insightful." The mean state of an unbiased prior is the completely mixed state. Using an unbiased prior is the most conservative approach, because it will have the lowest worst-case convergence time. When sampling from an unbiased prior, we randomly sample a matrix A from the Ginibre ensemble (random matrices of Gaussian distributed values), where the A_{ij} entry is $c + d\mathbf{i}$, and c and d are normally distributed random variables.[2] The density matrix is then given by

$$\rho_{sample} = \frac{AA^\dagger}{Tr(AA^\dagger)}. \tag{3.6}$$

.

Like in the Cholesky decomposition, constructing a density matrix in this fashion ensures that ρ_{sample} is physical for any values of A_{ij}. This distribution produces states that are on average less pure than a uniform distribution. In this case, a uniform distribution would be one that produces states that are evenly spaced within the Bloch sphere. Fig. 3.3 (a) shows the probability of sampling a qubit with a given purity from both a uniform distribution and the Ginibre distribution. Results from the Ginibre distribution were produced with random sampling, while the uniform distribution probabilities were calculated directly from the geometric properties of the Bloch sphere. In the Bloch sphere representation, a state's purity is related to its distance from the center of the sphere, so the density of high purity states is larger because they reside on a spherical shell with a larger radius. For higher dimensions, the geometry is more complicated, and this simple geometric argument to determine the density of states of a given purity does not apply. Fig. 3.3 (b) shows the distribution of purities for states of dimension 3 (e.g., a single qutrit) sampled from the Ginibre ensemble, while 3.3 (c) shows this for states of dimension 4 (e.g., two qubits or one d=4 qudit) . The results in Fig. 3.3 indicate that, if a constant number of prior samples are used in a BME tomography, the value of this constant should be determined by the number of samples needed to reliably reconstruct *highly-pure* states because these states will take longer due to their under-representation in the Ginibre distribution. Generating random quantum states is a critical element of the entire BME algorithm, and is an active area of research [30]. Increased sophistication in the generation of random quantum states could lead to increased

---

[2] Specifically, we sample a normal distribution with a mean of 0 and a standard deviation of 1; because these values are subsequently normalized, the choice of standard deviation is unimportant.

computational efficiency of the BME algorithm.
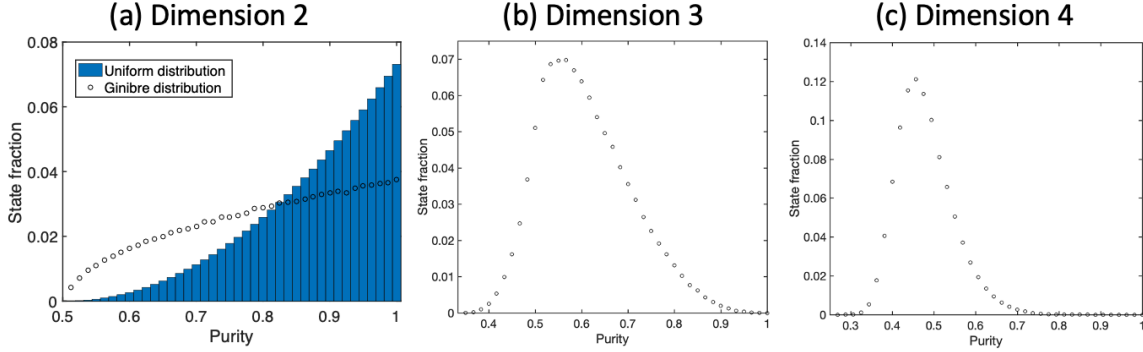
Figure 3.3: Purity Distribution of the Ginibre Ensemble. The Ginibre ensembles oversamples mixed states compared to a uniform distribution for states of dimension 2. For higher dimension states, it samples lower-purity states with increased frequency.

If the state being reconstructed is approximately known, it can be much faster to use an insightful prior. Granade et al. proposed [29] a form for insightful priors, where the distribution is the most entropic distribution for which the mean state is the guessed state; in other words, the distribution is as uniform as possible while still ensuring that the distribution's mean is the chosen state. Care must be taken in using this definition, however, as using a pure state as the guess to generate the insightful prior leads to the distribution being a delta function at that guessed state. This distribution can be generated from an unbiased prior by shifting the unbiased states a probabilistic amount towards the predicted state:

$$\rho_{sampled} = (1-e)\rho_{unbiased} + e(\rho_{fixed}), \tag{3.7}$$

where $\rho_{unbiased}$ is a state randomly sampled from an unbiased distribution, e is a real number sampled from a beta distribution[3] given in Equation 3.8 with parameters $\alpha = 1$ and $\beta = b$, given in Equation 3.9:

$$Beta(x; 1, b) \equiv \mathcal{N}(1-x)^{b-1}, x \in [0,1], \tag{3.8}$$

$$b = \frac{Vd}{1 - Vd}, \tag{3.9}$$

---

[3]Beta distributions often appear in Bayesian inference because they are the conjugate prior to a variety of likelihood functions. Beta distributions are an effective model for random proportions, like in this case where the parameter e drawn from the beta distribution dictates what proportion of our sample state is determined by our guess and what proportion is determined by the sample of an unbiased prior.

where $\mathcal{N}$ is a normalization factor, d is the dimension of ρ, V is the smallest eigenvalue of ρ, and

$$\rho_{fixed} = (b+1)\rho_{guess} - \frac{b}{d(1+b)}\mathbb{I}_d.$$

(3.10)

This construction creates a distribution that has a mean of ρ_{guess} with maximum entropy [29]. The way these insightful priors function can most easily be seen in the qubit case. For qubits, the lowest eigenvalue V determines a state's purity: for pure states, the lowest eigenvalue will be 0, and for completely mixed states, the lowest eigenvalue will be 0.5. This in turn leads to a beta distribution with a bias towards 0 for mixed states and a bias towards 1 for pure states. This makes sense because, if our guess is very mixed, then need not change the unbiased prior very much, but if our guess is a nearly pure state, the unbiased prior must be modified more so that the mean state is our guess. While using an insightful prior of this form can speed up state reconstructions when the guess is close to the result, it is fundamentally limited by the choice to make the distribution as entropic as possible. This choice means that the variance of the distribution is determined by its mean and not an independently tunable parameter. Practically, it is a very common occurrence that a posterior distribution has very low variance because, for a state that is not drifting, the variance decreases as a function of the number of counts measured. For that reason, while insightful priors of this form can be superior to an unbiased prior, they are still extremely ill-suited for sampling the posterior, with most samples being made of states with very low likelihood and very little impact on the mean.

3.3.1 Adaptive Sampling

The simplest approach to calculating the tomographic mean through Bayes's rule would be to repeatedly sample the prior distribution and calculate the associated likelihood until enough samples have been taken to reasonably estimate ρ_{BME} using Equations 3.4 and 3.5. In practice, this method is not feasible in almost all situations. This is due to the way the likelihood function given in Equation 3.2 changes as more counts are measured. As counts increase, the likelihood of all but the most likely states decreases exponentially, while the rate at which they are sampled remains the same. This means that more and more time is spent sampling states that do not end up contributing to the final mean state at all. To illustrate this, consider the quantity k given by

$$k(\rho) \equiv -log(\mathcal{L}(\rho)) = \sum_i \frac{(M_i - |\langle O|\rho\rangle|^2)^2}{2|\langle O|\rho\rangle|^4}.$$

(3.11)

Fig. 3.4 shows histograms of the values of k for 50 thousand samples of an unbiased prior distribution. In Fig. 3.4 (a), the tomography is calculated from a total of 50 counts, while in Fig. 3.4 (b), the tomography was calculated from 5000 counts. In order to normalize the results between these two scenarios, the minimum
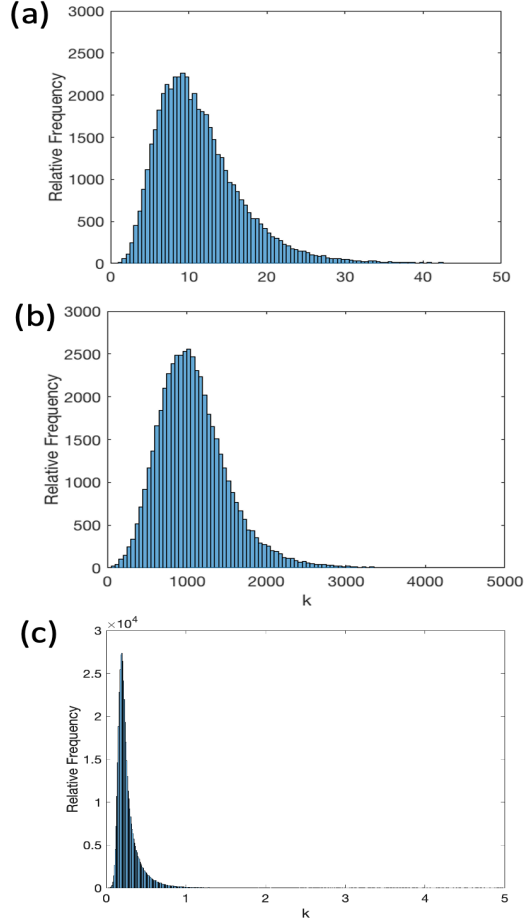
Figure 3.4: Comparison of the distribution in k values for a tomography of a ququart. (a) 50 total counts. (b) 5000 total counts The tomography with 5000 counts displays a far lower fraction of states with k≤10, indicating that almost none of the sampled states contribute to mean.(c) Shows the distribution of k values when the tomography from (b) is carried out using the adaptive prior method described later in Section 3.3.1. Using this method, almost all samples are of high-likelihood states.

of k in each case was subtracted from each value, so that the most likely state has $k = 0$ and $\mathcal{L} = 1$. From this, it can be seen that the vast majority of samples in the high count case have so little likelihood that they do not contribute at all to ρ_{BME}. While, in the low count regime, the most likely state represents only 0.008 of the total likelihood, in the high count case, the most likely state makes up almost all of it, with the second most likely state representing only 5×10^{-9} of the total. For almost all tomographies, sampling only from the prior is not a feasible computation method for determining the mean because high likelihood states are sampled so infrequently.

If unbiased priors are insufficient for efficient sampling, then the next method to investigate would be insightful priors. However, the previously discussed insightful priors are also insufficient for efficient sampling. Because the insightful prior distributions are designed to be maximally entropic, they are parameterized only

32

by their mean, and thus the standard deviation will always be fixed once the insightful guess is made. This is not the case for the posterior distributions that we want to sample from. Consider taking tomographic measurements of a stable source emitting some state ρ. When only a few counts are recorded, the posterior distribution should be relatively wide with its mean centered at ρ (or near ρ, if Poissonian statistics and environmental noise are considered). As more and more measurements are made, the mean remains the same, while the distribution becomes more narrow, reflecting the increased certainty that ρ is the true mean. This example illustrates why maximally entropic insightful priors are insufficient to represent posterior distributions in BME tomography. This difference is significant because the efficiency of the tomographic algorithm relies on our ability to efficiently sample the posterior distribution. The most efficient possible way to do this would be to sample the posterior directly; however, the original task is equivalent to calculating the (currently unknown) posterior distribution, so this is not an option.

The next best option is to estimate the posterior distribution as well as possible, given the currently available information. At the beginning of the algorithm, with no likelihoods sampled, that best guess is the prior, be it an unbiased or insightful distribution. As more states are sampled, information about the posterior is revealed in the form of the likelihoods of the sampled states. From this, an estimate of the posterior can be made and from it, subsequent states can be sampled. To the extent that this distribution is a better estimate than the prior, sampling from this distribution will increase the convergence speed of the algorithm by sampling high likelihood states more often.

For high count rate tomographies, this approach is necessary in order to lower the computation time to tractable levels. To see why this must be the case, consider the fraction of quantum states that contribute significantly to the final mean. These states represent the ones that must be sampled to effectively estimate the mean of the distribution. For the qubit case, an approximation for this fraction can be made by finding the surface of states with likelihood equal to half of the maximum. Fig. 3.5 shows how the fraction of states inside this surface changes as a function of count rate. This example was calculated for the state $0.95|H\rangle\langle H| + \frac{0.05}{2}\mathbb{I}_2$, and the surface was estimated as a sphere. The radius of the sphere was determined by finding the state of the form $\lambda|H\rangle\langle H| + \frac{1-\lambda}{2}\mathbb{I}_2$ with likelihood equal to half the likelihood of the most likely state. There are two such states that meet this criteria; for simplicity, the more mixed state was chosen, which leads this estimate to be an upper bound on the true value because the skewed likelihood distribution has a longer tail for more mixed states. Fig. 3.5 (a) shows the shape of the likelihood distribution for one count rate.

If the mean is calculated by sampling from a uniform prior, then the necessary computation time is proportional to the inverse of the volume fraction given in Fig. 3.5 because this fraction is equal to the
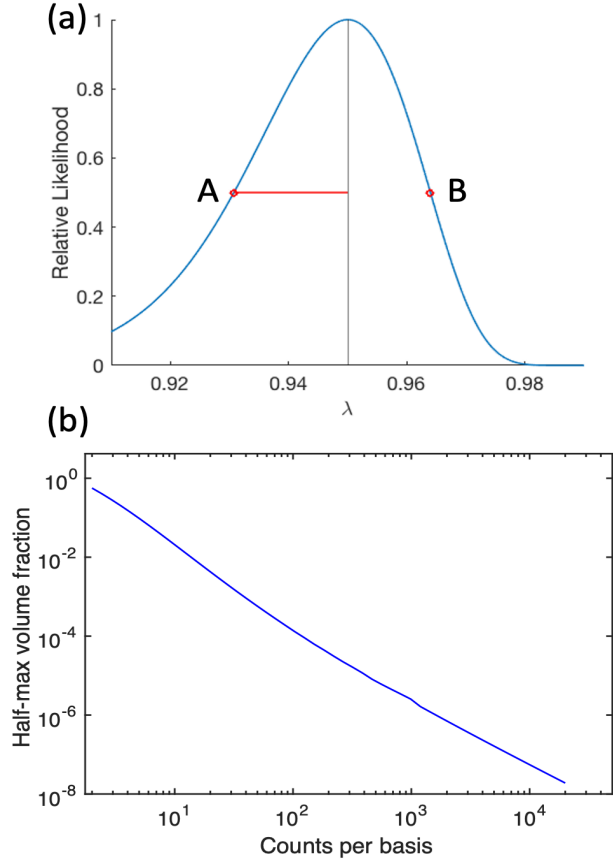
Figure 3.5: Estimate of the fraction of states contributing to the Bayesian mean. The estimate is made for the state $\rho = 0.95|H\rangle\langle H| + \frac{0.05}{2}\mathbb{I}_2$. (a) Relative likelihood of states near ρ, showing the locations (A and B) of the two states with half the likelihood of ρ. A is the more mixed of the two states, and is used in calculating the state fraction in (b). The distance between ρ and A is used as an estimate for the radius for the sphere that contains the states with likelihood more than half that of ρ. Results are the upper-bound for the fraction of states with likelihood at least half of the most likely state because of the asymmetry of the likelihood function.

probability that state will be sampled from that region. As the fraction of states that are in this region decreases, more and more samples are necessary to ensure the states relevant to the mean are sampled. While this means that sampling from a uniform distribution is feasible at low count rates when a high fraction of states have relatively high likelihoods, computation times become intractable for high-count tomographies. This issue is further exacerbated for higher-dimensional states, as increased state-space volume and a more sharply-peaked likelihood function due to an increased number of measurements lead to an even smaller relevant fraction of states.

To implement an adaptive estimation approach, a model for the form of the distribution must be chosen, in addition to a parameterization for that distribution. While an infinite number of parameterizations are

34

possible, because a normal distribution does not have compact support, some choices (e.g., the density matrix parameters ρ_{ij} themselves) of parameterizations of a normally distributed model will lead to sampling of nonphysical density matrices, which can in turn lead to a nonphysical mean. In order to ensure that only physical density matrices are produced, the Cholesky decomposition's $\mathbf{t}$ parameters were chosen. This allows the use of a multivariate normal distribution to model the posterior. A multivariate normal distribution has the form

$$p(\mathbf{t}; \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2}(\mathbf{t} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{t} - \boldsymbol{\mu})), \quad (3.12)$$

where $\boldsymbol{\mu}$ is a vector of the mean values of $\mathbf{t}$, Σ is the covariance matrix for $\mathbf{t}$, and n is the length of $\mathbf{t}$. It is necessary to use a multivariate normal distribution, as opposed to a product of univariate normal distributions, because the $\mathbf{t}$ parameters exhibit strong correlations. A product of univariate normals is unable to describe these correlations, so such a model still samples low-likelihood states very frequently.
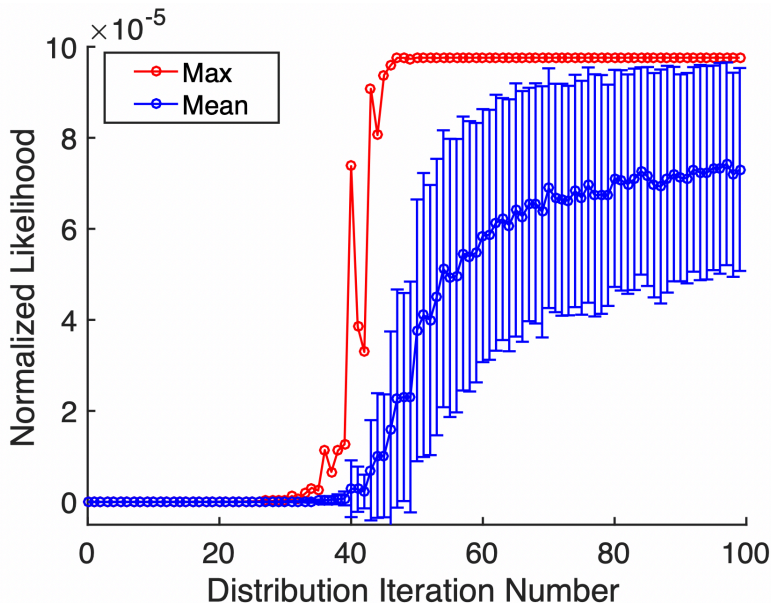
Figure 3.6: Example progression of adaptive distributions over 100 iterations. Each iteration consists of 300 samples from the current estimate of the posterior distribution. The likelihood of the sample with the highest likelihood is shown in red, while the mean likelihood of the samples in each iteration is shown in blue. Error bars for the mean are calculated from the standard deviation of the likelihood in each iteration. The mean likelihood continues to rise after the maximum likelihood plateaus, showing that the distribution continues to adapt to more closely represent the posterior even after it begins to sample the most likely state. Iteration zero consists of the states sampled from the initial prior distribution. The likelihoods are normalized so that $\sum_i \mathcal{L}_i = 1$.

3.4 Stopping Condition

A necessary component in the Monte Carlo estimation used in Bayesian tomography is a criteria for determining when a sufficient number of states have been sampled. The simplest option would be to sample a predetermined number of states. However, because the algorithm relies on random sampling of states, it would be necessary to always sample enough states that, even in the worst case, a good estimate of the mean would be reached. This means that, for the average tomography, (many) more states than necessary would be sampled, increasing computation time. A more robust strategy is to measure a figure of merit during the computation and stop the computation when this value reaches a certain threshold. This figure of merit should reflect the stability of the computation, in that a point is reached when additional state sampling no longer causes the mean to change, or at least is very unlikely to.

In order to detect that a stable mean has been reached, a numeric representation of stability is necessary. As a base for this measure of stability, we use the infidelity I between the current mean state and the previous mean state before sampling the current estimate of the posterior distribution:

$$I_n \equiv 1 - F(\rho_n, \rho_{n-1}), \tag{3.13}$$

where ρ_i indicates the estimate of ρ after the ith iteration of the algorithm. The fidelity between these states can be seen as a measure of how similar the two states are, while the infidelity shows the extent to which the states are different. By using the infidelity as the basis of the metric, the stability value can been seen as related to the change in the mean over time; just as a value would be deemed stable if the change between iterations was low, ρ_{BME} can be seen as stable if the infidelity between subsequent iterations is low. In general, each iteration contains many samples (for example, the simulations in Section 3.5 use 1000 samples per iteration), so a low infidelity between iterations indicates that, even after many new samples are included in the estimate of ρ, the estimate is largely unchanged.

Because the mean being calculated is weighted by the likelihood of the sampled states, the stability condition also needs to take into account the relative weight of the states sampled when measuring the stability. If this factor is ignored, there are two ways in which the infidelity could be low. The first possibility is the one we are trying to detect: the mean state remained unchanged even though high-likelihood samples were made. The second possibility is that the sampled states had a low likelihood to begin with, so they would not cause a significant change to the mean in any situation; the infidelity would also be low because $\rho_n \approx \rho_{n-1}$. In order to distinguish the first case from the second, the stability condition must include a factor proportional to the likelihood weighting. One method is to weight the infidelity by the ratio $\mathcal{N}$ between the

expected likelihood if all iterations had a uniform, equal weight and the measured likelihood of the current iteration. If the algorithm is on its nth iteration, and the likelihoods were uniform, then the relative weight for this iteration would be $\frac{1}{n}$. However, the weights are *not* uniform: the current has a weight $\mathcal{L}_n$, and its relative weight is $\frac{\mathcal{L}_n}{\sum_{i=1}^{n} \mathcal{L}_i}$

$$\mathcal{N} = \frac{\sum_i \mathcal{L}_i}{n\mathcal{L}_n}, \tag{3.14}$$

where $\mathcal{L}_i$ is the summed likelihood of the states sampled from the ith estimate of the distribution, and n is the number of the current iteration. Our proposed stopping parameter S is then given by

$$S = \mathcal{N}I_n = \frac{\sum_i \mathcal{L}_i}{n\mathcal{L}_n}\left[1 - F(\rho_n, \rho_{n-1})\right], \tag{3.15}$$

and the algorithm will terminate when

$$S \leq C, \tag{3.16}$$

where C is a parameter which will be determined in Section 3.4.1. To see how this normalizing factor works, consider a case where the current iteration has had a much higher likelihood than the average iteration. In this case, $\mathcal{N}$ will be small, and S will be below the threshold even for larger values of I_n. Similarly, if the current iteration has a much lower likelihood than the average iteration, then $\mathcal{N}$ will be large, and the infidelity must be lower for S to be below the required threshold.

3.4.1 Stopping Parameter Tuning

In order to be useful, the stopping parameter S needs to be paired with a threshold C that is calibrated to some other functional metric of tomographic performance. Ideally, the tomography could be set to some threshold fidelity, and then the tomography would complete when this fidelity was reached. Because the target state may not be known, it may not be possible to directly calculate this metric. Instead, statistical information must be gathered relating the final state fidelities to the value of the stopping parameter. In order to determine this relationship, the tomography algorithm was run for a fixed number of samples expected to be significantly larger than the number required to reach a stable result. Then, the sample number for which $F(\rho_i, \rho_{final}) \geq F_{threshold}$ was recorded, along with the corresponding stability at that sample S_i. This process was carried out for a variety of count magnitudes ranging from 5 to 250 average counts per basis, for states with dimension 2, 3, and 4. Ten states were randomly selected from the Ginibre ensemble for each combination of dimension and count rate. The results did not show a significant dependence on these parameters, so the combined results of this study are shown in Fig. 3.7, with the samples sorted by

stopping parameter. (A version binned by amplitude and dimension can be found in Appendix A.) This data indicates that, while most tomographies converge to ρ_{final} for $S < 10^{-7}$, some remain significantly different from ρ_{final} until $S < 10^{-12}$.
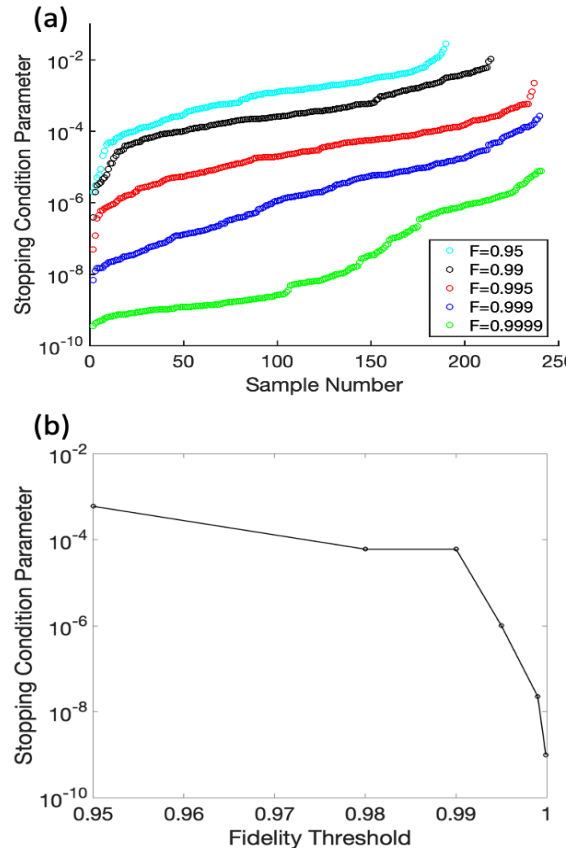
Figure 3.7: Relation between the stopping condition parameter and the fidelity with the final state. When the fidelity $F(\rho_i, \rho_{final})$ exceeded the given thresholds, the value of S_i was recorded. This was carried out for a number of different state dimensions and count amplitudes. (a) The stopping condition parameters values were then sorted so the features of the data can be more easily discerned, so that sample 1 was the sample that required the lowest stopping parameter to reach the necessary fidelity, while sample 240 required the highest stopping parameter. Measurement of the stability began after sampling the initial prior, and some tomographies exceeded the low-fidelity thresholds while sampling the initial prior, so these tomographies do not have an associated stopping parameter for low-fidelity thresholds. (b) Stopping condition values such that 99% of tomographies reach the given fidelity threshold.

3.5 Benchmarking

3.5.1 Simulated Results

Figs. 3.8, 3.9, and 3.10 show simulation results comparing the effectiveness of Bayesian mean estimation and maximum likelihood estimation. In Fig. 3.8, target states were randomly selected from the Ginibre distribution for a variety of count rates. At each count rate, 10 states were reconstructed. The counts were simulated for the canonical over-complete measurement set. The counts were sampled from a Poissonian distribution with expectation value of the measurement as the mean value. The Bayesian mean algorithm was run with a stopping condition of $S \leq 10^{-8}$, 1000 samples per prior update, and 2000 samples using the initial unbiased prior. The computation time and the fidelity of the tomography were measured for both the Bayesian mean estimation and maximum likelihood algorithm.

The results in Fig. 3.9 were generated in the same way as Fig. 3.8, but with two specific target states of differing purities. The states had the form

$$\rho = \lambda \rho^{pure} + \frac{1 - \lambda}{d} \mathbb{I}_d, \tag{3.17}$$

where $\rho_{11}^{pure} = 1$, all other entries are 0, and d is the dimension of the state.

The results in Fig. 3.10 were generated from tomographies with the same parameters as Fig. 3.8 and 3.9, with an entangled target state with varying purity. This state had the form

$$\rho = \lambda \Phi^+ + \frac{1 - \lambda}{4} \mathbb{I}_4. \tag{3.18}$$

Fig. 3.10 (a) presents the results of the MLE algorithm, while Fig. 3.10 (b) shows the results of the BME algorithm, and Fig. 3.10 (c) shows the difference between the two.

These results demonstrate the effectiveness of the Bayesian approach for tomographies with low count rates. For tomographies with 25 or fewer average counts per basis, the Bayesian approach lead to significantly higher fidelities and much lower variance, at the cost of higher computation time. The only exception was for highly pure states, for which the MLE algorithm was slightly better, even for low-count tomographies. For qubit tomographies, the Bayesian algorithm also produced comparable fidelities. In higher-dimensional reconstruction, the Bayesian algorithm did not match the fidelities of the maximum likelihood method. According to [28], this is not the expected behavior of the Bayesian approach, indicating that further refinement in the implementation of the algorithm is necessary to make it an appropriate technique to use at higher count rates for high-dimensional states. Two possible causes are flaws in the stopping condition (e.g.,
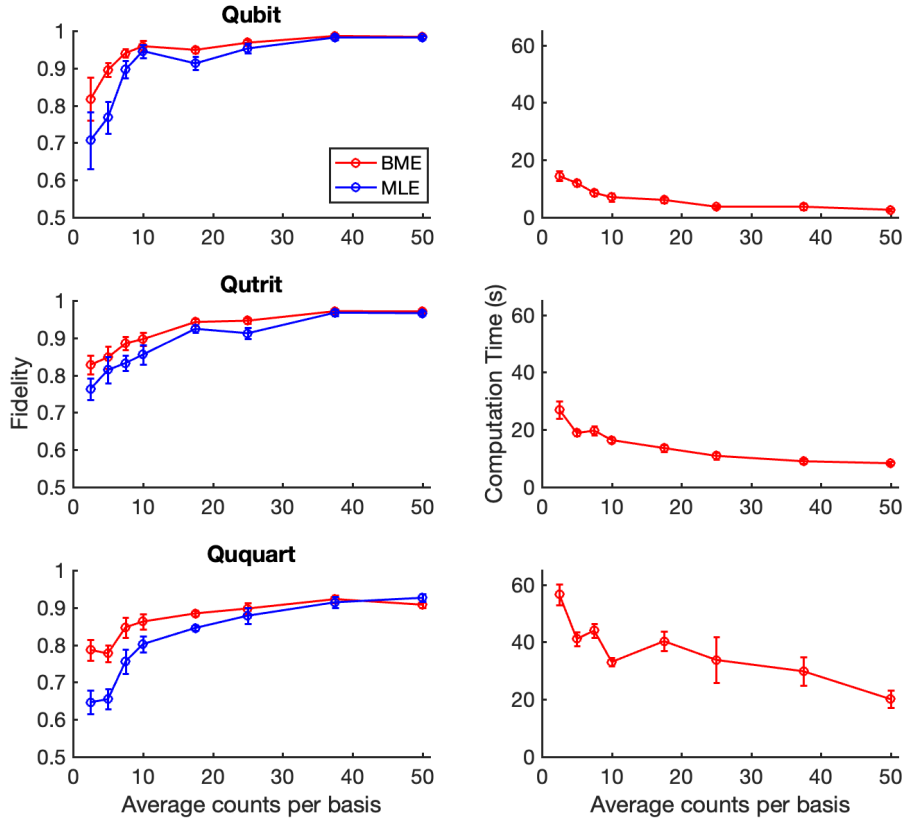
Figure 3.8: Comparison of the accuracy and computational resources of maximum likelihood tomography and Bayesian mean estimation. Each data point consists of ten tomographies of states randomly drawn from the Ginibre ensemble. Error bars represent the standard error of the mean. The plots on the left show the fidelity of the reconstructed states with the measured state, while the plots on the right show the average computation time in seconds. All MLE tomographies took less than 3 seconds to compute and took 0.3 seconds on average.

an insufficient number of samples to reliably estimate the mean) or issues with the adaptive estimation of the distribution; however, Fig. 3.11 provides evidence that it is not likely that the issue arises from insufficient sampling. This figure shows the results of a tomography generated in a similar fashion to Fig. 3.8 for states with dimension 4, except that the algorithm was stopped after a fixed number of samples chosen to be much larger than necessary to reach the stability condition. Despite sampling $\sim$10 million more states than necessary to pass the stability condition, the fidelity did not appreciably increase.

This means that the likely cause of the reduced fidelities, relative to the maximum likelihood approach, arises from the iterative estimation of the posterior distribution. One possible issue is with the functional form of the estimate of the distribution. While the chosen distribution is of the form $e^{f(T^2)}$, where f is a linear function of T, the likelihood function has a form $e^{f(T^4)}$ because $\rho = TT^\dagger$. This mismatch may lead to an
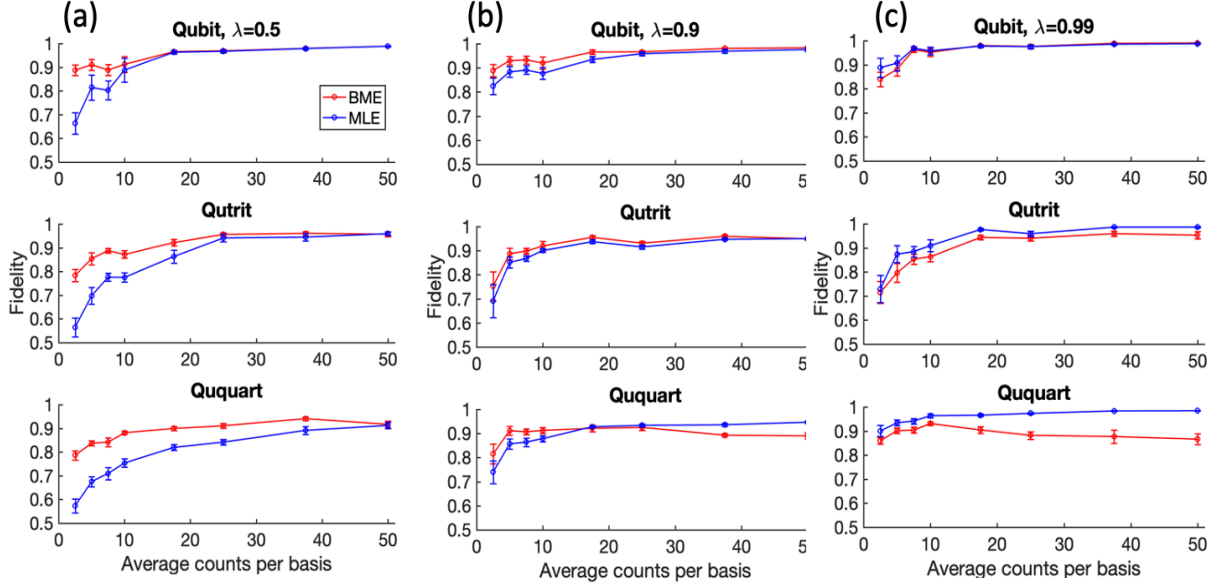
Figure 3.9: Comparison of MLE and BME for states of varying purity with the form $\lambda\rho^{pure} + \frac{1-\lambda}{d}\mathbb{I}_d$, where d is the dimension of the state, and (a) $\lambda = 0.5$, (b) $\lambda = 0.9$, (c) $\lambda = 0.99$. Each data point consists of ten tomographies with different Poissonian noise.

oversampling of some states and a degradation in the estimation of the mean. In order to investigate if this aspect of the adaptive priors led to decreased fidelity for highly pure states, we redesigned the algorithm to compute the priors from the density matrix elements directly, instead of from the parameters of the Cholesky decomposition. For example, for a qubit, the four parameters would be ρ_{11}, ρ_{22}, $\text{Re}(\rho_{12})$, and $\text{Im}(\rho_{12})$. This method can sample nonphysical states. In the cases for which a nonphysical state was sampled, the state was rejected and a new state was sampled until the state was physical. To test the performance of this alternative method, we reran the simulation in Fig. 3.10, using the alternative adaptive sample algorithm. The results are shown in Fig. 3.12. This method displayed worse performance for highly pure states than the original technique, indicating that the original parameterization was not the cause of the reduced fidelity for high-dimensional, highly pure states.

3.5.2 Application to Experimental Data

In addition to simulated data, a comparison between the techniques in an experimental context was also performed [31]. In this experiment, states of the form

$$\Phi = |0\rangle + e^{i\phi_1}|1\rangle + e^{i\phi_2}|2\rangle + e^{i\phi_3}|3\rangle \tag{3.19}$$
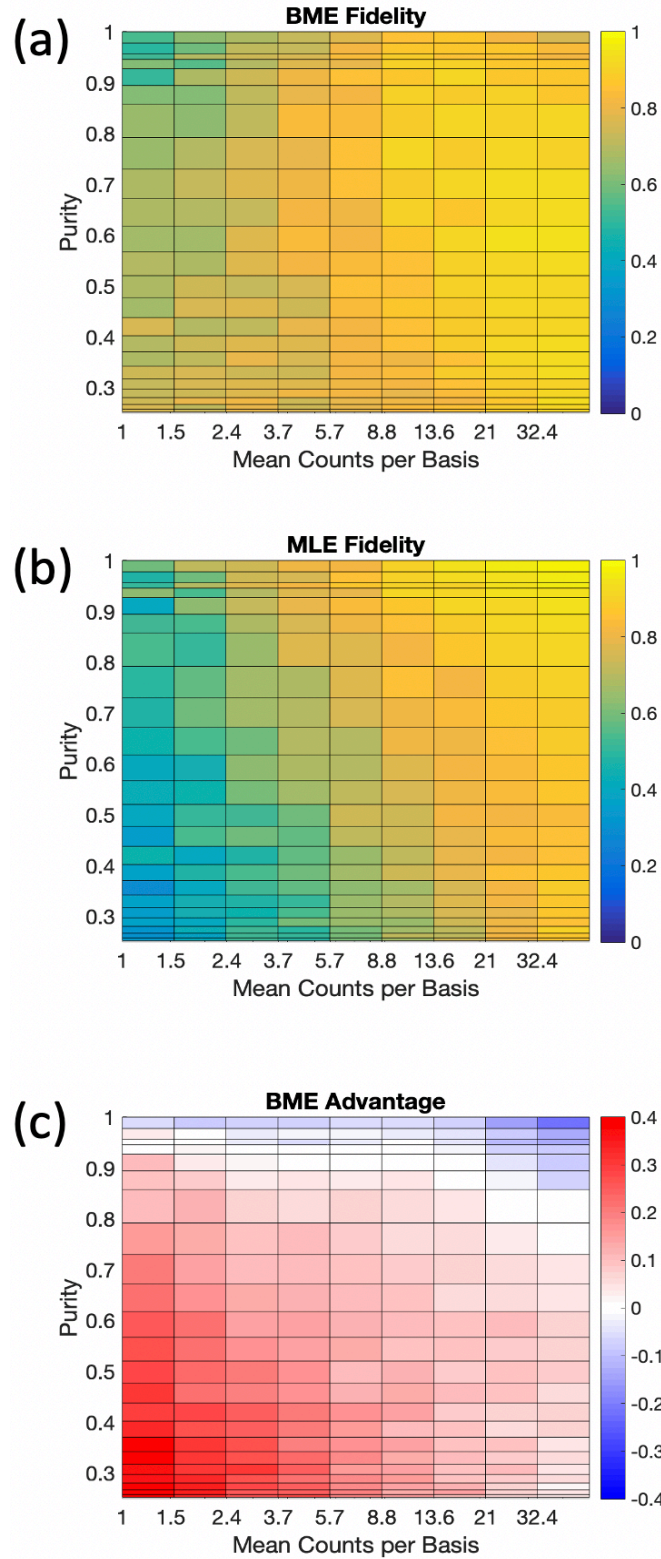
Figure 3.10: Comparison of the accuracy of maximum likelihood tomography and Bayesian mean estimation for an entangled state of the form $\lambda\Phi^+ + \frac{1-\lambda}{4}\mathbb{I}_4$. Each data point is calculated from the mean of ten tomographies with different Poissonian noise. (a) Fidelity of the MLE tomography. (b) Fidelity of the BME tomography. (c) The difference in fidelity between the BME and MLE tomographies
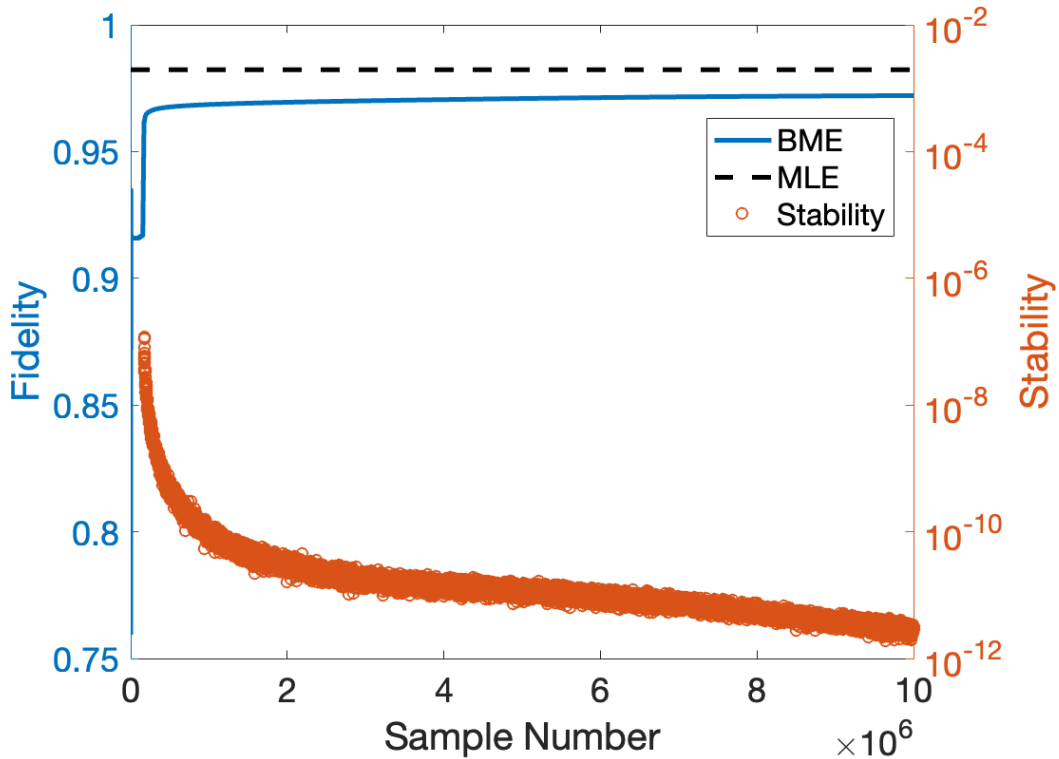
Figure 3.11: Example tomography with sample number far past the stability threshold. Despite sampling far more states than necessary to exceed the stability threshold of $S \leq 10^{-8}$, the fidelity of the BME tomgraphy does not appreciably increase and is still less than that of the MLE tomography, indicating that increased sampling is not a viable strategy to close the gap between the techniques for high count rate tomographies. The tomographies were taken for a random state of dimension 4, purity 0.42, and an average of 250 counts per basis.

were generated and reconstructed with both a maximum likelihood and Bayesian mean approach. The source was attenuated to produce tomographic results over a range of count rates. Because this analysis was completed early in the implementation of the Bayesian mean algorithm before a majority of the previously discussed features were developed, these results were computed using a fixed number of samples, with no adaptive estimation of the posterior. It was possible to make these calculations with no adaptive estimations because the count rates observed were low, so an unbiased prior still sampled high likelihood states with reasonable frequency. In addition to computing the fidelity with the target state, the accuracy of the reconstructed phases ϕ_i was also measured by calculating the average phase error for each of the three phases ϕ_i. Each data point in this analysis consists of 16 tomographies of 10 different target states for a total of 160 tomographies per data point. This data provides a similar picture to the simulated results presented above: for low count-rate tomographies, the Bayesian approach provides more accurate results.

This analysis also shows that this increased accuracy applies not only to the whole density matrix but also quantities calculated from subsets of the density matrix.
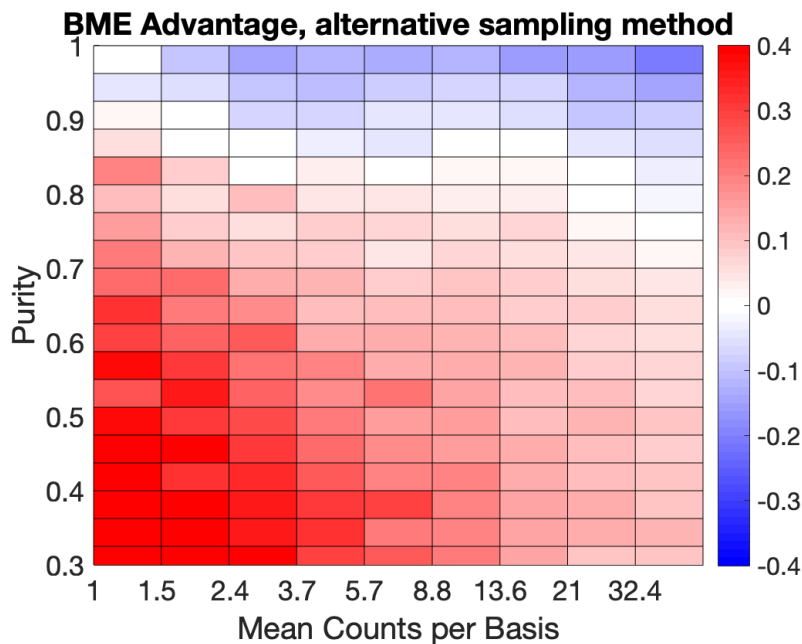
Figure 3.12: Comparison of MLE and BME using an alternative adaptive prior algorithm. The simulation was identical to the one used in Fig. 3.10, except that the alternative adaptive prior algorithm was used. This technique displayed worse performance for highly-pure states than the original algorithm.

3.6   Applications

The benchmarks in Section 3.5 indicate that there is significant benefit to analyzing tomographic data with the Bayesian mean algorithm in the low-count regime. In many quantum optics experiments, care is taken to ensure that data does not fall into this regime because of its increased statistical noise, although this is not possible in all cases. However, the best realm of application to this technique may not be in experimental quantum optics but in profiling a future quantum internet. Consider a quantum communication task where Alice is trying to communicate to Bob over some quantum channel but the success of the task depends on Alice reliably preparing desired quantum states. If either her source or the quantum channel experiences drift, then the states she sends to Bob will shift over time. Because of this, Alice and Bob will need to periodically perform verifications on the transmitted states in order to tune Alice's source or compensate for changes in the quantum channel, possibly by performing tomographies. In this scenario, Alice and Bob want to minimize the amount of time they spend analyzing their source and channel because these analyses
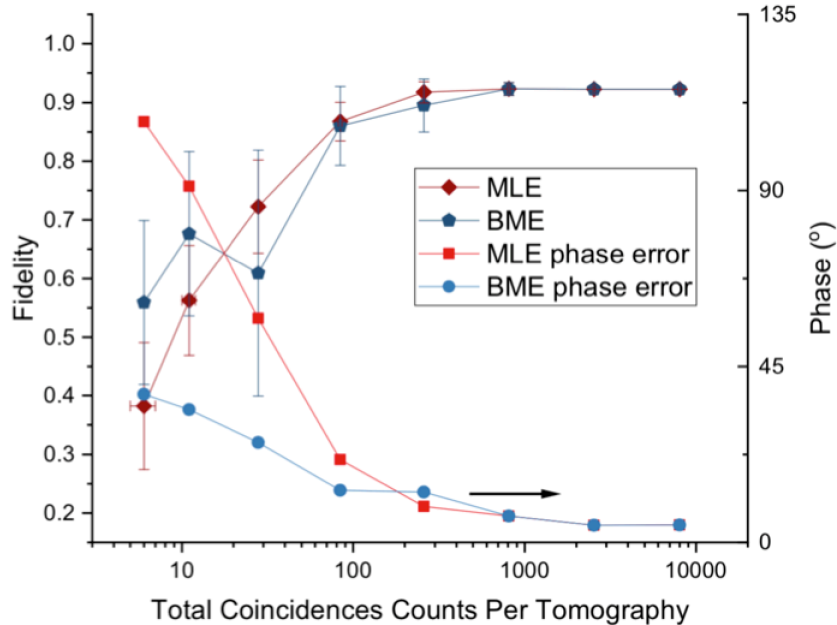
Figure 3.13: Results from a quantum optics experiment comparing the effectiveness of tomography techniques in reconstructing ququarts. These results are shown against the total number of measured counts over all nine measurement bases.

directly decrease the fraction of time spent on the true communication task they are trying to accomplish. In addition, if the characterization takes too long, the results will be the average over the (changing) state over time. These constraints motivate Alice and Bob to perform their tomographies in the fastest time possible in order to actively correct for changes in the source and channel. This is precisely the scenario in which the Bayesian approach can provide benefits by increasing the quality of the state reconstruction without necessitating the use of more experimental resources. One important difference between these types of applications and the simulations presented earlier is that the previous simulations assumed ρ is unknown. Consider the case for which repeated tomographies of a slowly drifting state are made. In such a case, we can use our previous estimate of the state ρ_{prev} to provide an estimate of the current state ρ_{curr}. If BME is employed in such a scenario, then the final estimate of the posterior in the previous tomography can be used as the initial prior for the current tomography, as long as the state drifts slowly enough that ρ_{curr} has a high likelihood of being sampled from the posterior for ρ_{prev}. This will be true when the change in counts between subsequent tomographies due to drift is comparable to or lower than the Poissonian statistical noise in the measurements. Applying this technique can reduce computation time and increase state reconstruction fidelity. Results of an example application are shown in Fig. 3.14. In this scenario, a state with the form given in Equation 3.18 is repeatedly measured. Between measurements, the state is

45

allowed to drift by resampling its Cholesky parameters from a normal distribution such that the fidelity between subsequent states is ∼0.998. This technique shows a marked benefit, especially at low counts rates for pure states, where the new BME technique yielded an average fidelity of 0.96±0.02, compared with an average fidelity of 0.83±0.05 for the standard BME algorithm, and 0.93±0.03 for the MLE algorithm. In addition, the new BME technique was 2.25 times faster than the standard BME technique.
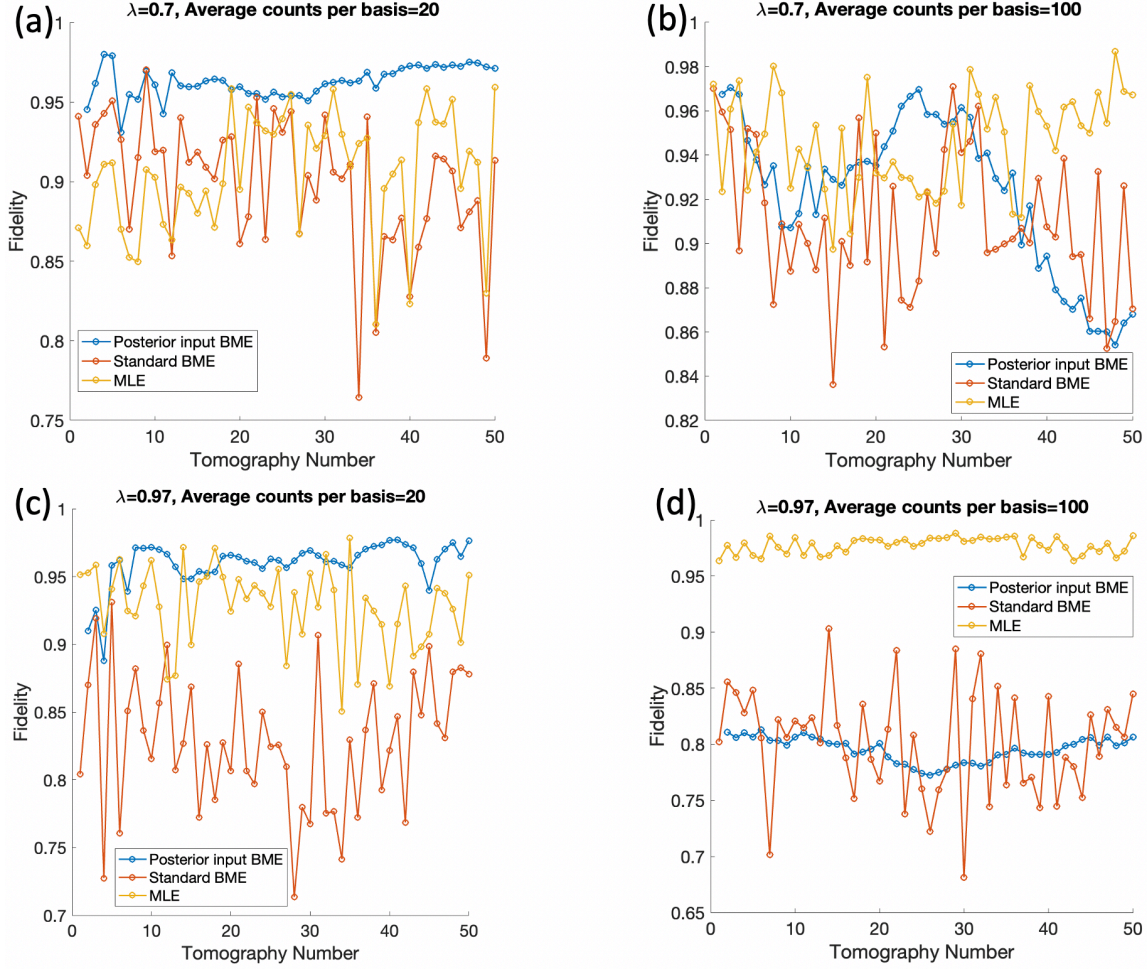
Figure 3.14: Repeated tomographies of a slowly drifting state. For this simulation, an entangled state with the form given in Equation 3.18, with either λ=0.7 ((a) and (b)) or λ=0.97 ((c) and (d)), was allowed to drift randomly between repeated tomographies. The magnitude of this drift was set so that the average fidelity between subsequent states was 0.998. In (a) and (c), the average number of counts per basis is 20 (∼180 total counts), while in (b) and (d), the average is 100 (∼900 total counts). The average tomography fidelities are (a) $F_{Posterior}$=0.96±0.01, F_{BME}=0.90±0.05, and F_{MLE}=0.91±0.04; (b) $F_{Posterior}$=0.92±0.04, F_{BME}=0.91±0.03, and F_{MLE}=0.95±0.02; (c) $F_{Posterior}$=0.96±0.02, F_{BME}=0.83±0.05, and F_{MLE}=0.93±0.03; (d) $F_{Posterior}$=0.80±0.01, F_{BME}=0.80±0.05, and F_{MLE}=0.98±0.01.

46

Chapter 4

Tests of Entanglement

4.1 Introduction

Entanglement, now seen as a fundamental feature of quantum mechanics, was first discussed as an obvious sign of the theory's shortcoming in Einstein, Podolsky, and Rosen's 1935 "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" [32]. In this work, they present the "EPR paradox," the thought experiment wherein interactions with one half of an entangled system can cause instantaneous changes in the other half of the system, regardless of their physical separation. This work motivated the development of "hidden-variable models," where an undetectable hidden variable, in addition to the standard quantum mechanical description of the state, allowed for deterministic measurement outcomes without the reliance on instantaneous effects from distant particles. In the 1970s, John Bell developed a test to distinguish these hidden-variable theories from quantum mechanical ones [33] by specifying a quantity that had different maximal bounds in the two models. Since their advent, Bell tests [34] have been a focus of foundational research in quantum mechanics, providing a means to demonstrate the nonlocal effects present in quantum mechanics and the presence of entanglement. Other techniques, such as quantum steering and entanglement witnesses, expand the applicability of entanglement verification to a wider set of scenarios with differing assumptions. Initially, these tests were conceived of as thought experiments that revealed unexpected (or to some, illogical) features of quantum mechanics; however, repeated experimental verification of the correlations that are the hallmark of entangled states has left little doubt that "spooky action at a distance" is a part of reality. The refinement of these measurement techniques culminated in a trio of "loophole-free" tests of nonlocality using Bell inequalities, providing compelling evidence that nature is truly nonlocal [35, 36, 37].

Now that the original purpose of Bell tests, providing a measurable criteria for separating local and nonlocal theories, has been largely fulfilled, a new era for Bell tests has been entered when they are used as a tool for probing and verifying the proprieties of quantum states, most recently as a means to ensure cryptographic security in the "device-independent" QKD protocols without a reliance on the trust of mea-
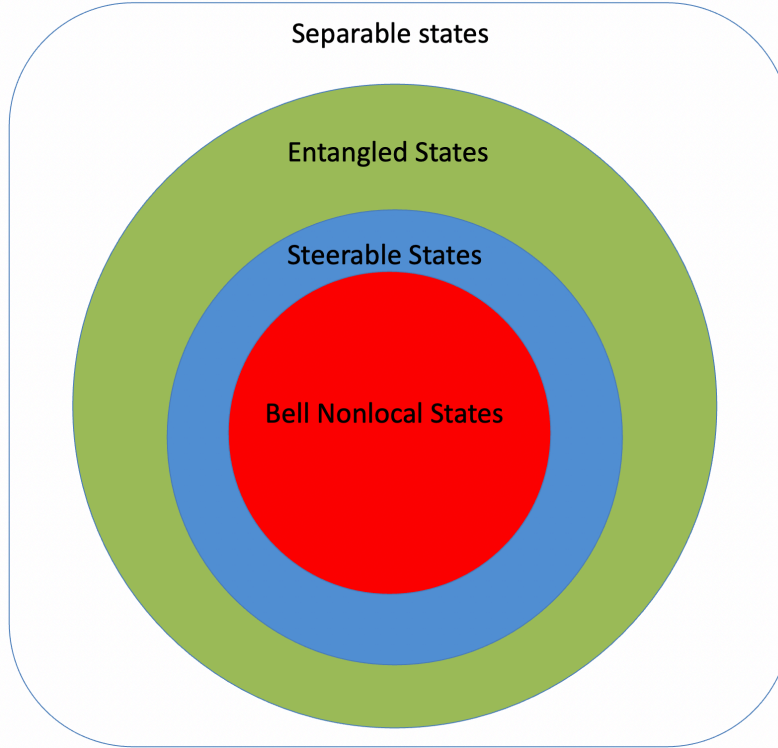
Figure 4.1: Hierarchy of entangled states. All Bell nonlocal states are steerable, and all steerable states are entangled, but there are Bell local steerable states and unsteerable entangled states.

surement devices [17]. The ability to draw conclusions about the measured quantum state without the need to trust the devices used to make the measurement is a defining feature of Bell tests and can be used to distinguish them from other methods of verifying entanglement. In contrast, entanglement witnesses require trusting all measurement devices [38]. Quantum steering represents an intermediate case, in which one of the measurement devices must be trusted, while no assumptions are made about the second measurement device.

While entanglement is a prerequisite for both nonlocality and steerability, the three concepts are distinct. There are local steerable states (that is, states that can be steered but cannot violate a Bell inequality) and unsteerable entangled states. Understanding the nuances between these classifications is still an active area of research.

4.2 Bell Tests

Bell tests were first conceived of in 1965 [33] as a way to determine if reality could have a local realistic description and to show that the predictions of quantum mechanics were incompatible with such a description.

Local realism is a concept initially developed by Einstein, Podolsky, and Rosen [32]. For local realistic theories, the values of properties of a physical system that can be predicted are viewed as "elements of reality," even if no measurement is made on the system to determine what this value is[1]. Such a definition is incompatible with quantum mechanics, so Einstein believed that quantum mechanics must be incomplete in order to accommodate local realism. Bell tests provide an experimental technique to verify if local realism holds. Despite their historical motivation in verifying the existence of local realism, i.e., locality *and* realism, violations of Bell inequalities are generally attributed to the measured quantum states possessing nonlocal features, and the framework for deriving Bell inequalities can be constructed with a focus only on locality [39].
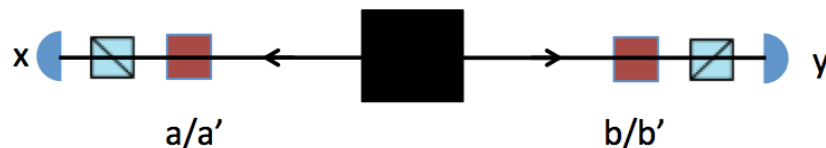
Figure 4.2: Schematic for a typical Bell test. Photons are distributed to Alice and Bob from a (possibly entangled) source. They randomly choose between measurements settings and record the outcome of that measurement. The Bell parameter is then calculated from the settings (a,b) and the measurement outcomes (x,y).

Fig. 4.2 shows the basic structure of a Bell test. Two parties, Alice and Bob, each receive one photon from a (possibly entangled) pair. They each randomly choose what basis to measure in from a predetermined set, which is in general not the same for Alice and Bob. After the measurement, they classically share their measurement settings and results for analysis. The Bell parameter S can then be computed based on this data by computing the conditional probabilities $p(a, b|x, y)$ of the outcomes given the measurements, where a (b) indicates which measurement setting Alice (Bob) used and x (y) indicates the measurement outcome. Depending on the specific form of the Bell inequality, S may also depend on local terms of the form $p(a|x)$ and $p(b|y)$. These forms can always be converted to a related Bell inequality that relies only on the joint probability $p(a, b|x, y)$. The Bell inequality then takes the form

$$S_{Bell} = \sum_{a,b,x,y} \beta_{a,b,x,y} p(a, b|x, y)$$

$$|S_{Bell}| \leq L,$$

(4.1)

---

[1]Einstein, Podolsky, and Rosen described "elements of reality" as follows: "If, without in any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity."

where $\beta_{a,b,x,y}$ are integer coefficients that depend on the Bell inequality being measured and the inequality bound L applies to local realistic models.

4.3   Quantum Steering

Quantum steering was first conceived of in 1936 by Schrödinger in a paper considering the local state for a pair of entangled particles [3]. He observed that, while the particles are initially in a locally mixed state, careful measurements of one half of the pair would cause the other half to be left in a definite pure state, so that certain measurements on one half would "steer" the other half into a predictable state. At the time, Schrödinger was dubious that this result could actually reflect reality, stating, "It is suggested that these conclusions, unavoidable within the present theory but repugnant to some physicists including the author, are caused by applying non-relativistic quantum mechanics beyond its legitimate range."
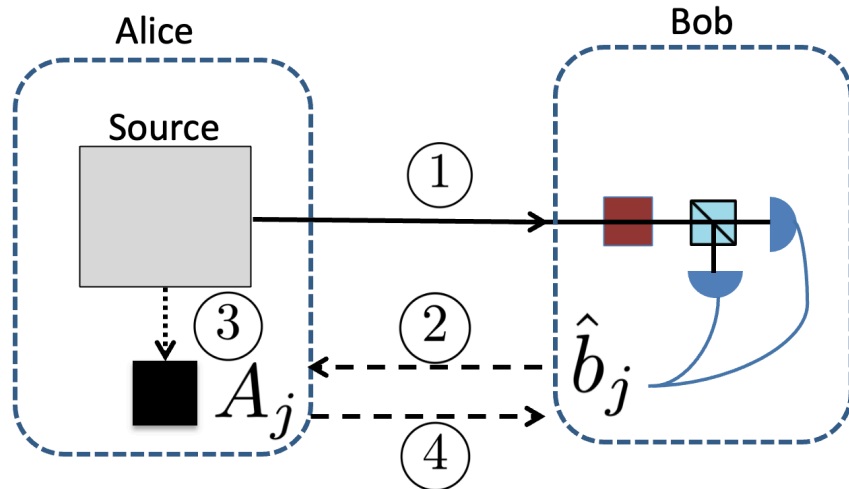
Figure 4.3: Schematic for a typical steering measurement. (1) Alice sends a photon to Bob, who informs her (2) of which randomly chosen measurement he will make. Alice then selects a measurement to match Bob's that will maximize the correlations between their results (3). She sends her measurement outcome A_j to Bob (4), who concludes whether her source displays genuine entanglement by analyzing the correlations between his measurement and the results Alice sends him.

Fig. 4.3 shows the basic structure of a steering experiment. Alice wants to prove to Bob that she has an entangled state. In order to do this, Alice transmits half of the entangled pair to Bob. After receiving the photon, Bob informs Alice of the measurement basis he will be using. In contrast to the type of communication in a Bell test, this information is not just a label but also indicates exactly what basis Bob will measure in. This information allows Alice to choose her measurement in order to steer Bob's half of the state correctly. If the shared state is maximally entangled, then Alice can select a measurement so that

Bob's half is left in an eigenstate of his measurement operator, and Alice can predict what this state is, based on her own measurement outcome. Alice then sends Bob the outcome of her measurement, allowing him to compute the correlation between the results. Bob then repeats this process using at least one other measurement basis to determine the steering parameter, which he computes as

$$S_{Steer} = \frac{1}{N} \sum_j^N \langle A_j \hat{b}_j \rangle, \tag{4.2}$$

where A_j is the outcome of Alice's measurement, $\hat{b}_j$ is the measurement operator for the current basis, and N is the number of measurement settings that Bob chooses between. This formulation shows that quantum steering is a test of a state's ability to display correlations in multiple bases; while separable states can display correlations, these correlations are always basis dependent, and so by making measurements in multiple bases, the presence of correlations can be used to distinguish separable and entangled states. For example, consider the case in which Alice and Bob share the maximally entangled state

$$|\Phi^+\rangle \equiv \frac{1}{\sqrt{2}}(|HH\rangle + |VV\rangle) = \frac{1}{\sqrt{2}}(|DD\rangle + |AA\rangle) \tag{4.3}$$

and Bob chooses to measure in the H/V and D/A bases. This state displays perfect correlations in both bases, so the steering parameter will be 1. In contrast, if they shared the state

$$\rho_{mixed} = \frac{1}{2}(|HH\rangle\langle HH| + |VV\rangle\langle VV|) =$$
$$\frac{1}{4}(|DD\rangle\langle DD| + |AA\rangle\langle AA| + |DA\rangle\langle DA| + |AD\rangle\langle AD| + \tag{4.4}$$
$$|AA\rangle\langle DD| + |DD\rangle\langle AA| + |AD\rangle\langle DA| + |DA\rangle\langle AD|),$$

then their measurements would still be correlated in the H/V basis; however, in the A/D basis, their measurements are totally uncorrelated, so the final steering parameter will be 0.5.

The value of the bound on S_{Steer} can be determined by finding the separable state that maximizes S_{Steer}. This bound depends on the choice of measurements made by Bob. For example, if two measurements are used and they are nearly the same basis, then the bound will be close to 1, while if they are mutually unbiased, then the standard bound of $\frac{1}{\sqrt{2}}$ will apply. As shown in Fig. 4.1, there are steerable states that are Bell local, as well as unsteerable entangled states. For example, the Werner state $\lambda|\Phi^+\rangle\langle\Phi^+| + \frac{1-\lambda}{4}\mathbb{I}_4$ is Bell nonlocal and steerable for $\lambda = [\frac{1}{\sqrt{2}}, 1]$, Bell local but steerable for $\lambda = [\frac{1}{2}, \frac{1}{\sqrt{2}}]$, unsteerable but entangled for $\lambda = [\frac{1}{3}, \frac{1}{2}]$, and unentangled for $\lambda < \frac{1}{3}$ [40].

4.4   Trust Comparison

One framework for understanding the difference between Bell tests and quantum steering experiments is that of measurement device trust, the concept that only certain elements of an experimental system must work in the expected way for reliable conclusions to be drawn from the results. Bell tests do not require trusted measurement devices because the measurements themselves are treated as a black box, and the final Bell parameter is calculated from only the measurement outcomes, with no reliance on the nature of the measurements made. In addition, the bound of the inequality does not depend on the measurements made, so the veracity of conclusions drawn from a Bell inequality do not depend on the measurements being executed as expected, e.g., if Bob accidentally or intentionally measures in the non-optimal basis, the effect will be to reduce the amount by which the Bell inequality is violated— in no case does one falsely conclude a nonlocal correlation when there was none.

In contrast, tests of quantum steering *do* rely on one party accurately carrying out the proper measurements. While the party steering the states is treated as a black box, the party verifying that steering has occurred must carry out measurements that are accurate in order for a valid conclusion to be drawn. This is because, while the steering parameter itself is constructed from the correlations of measurement outcomes and is thus measurement-agnostic, the steering *bound* depends on the measurements made (while the presence of the Bob's measurement choice $\hat{b}_j$ in Equation 4.2 may make it seem as if the maximum steering parameter depends on this choice, Alice can always change her measurement based on Bob's choice to maximize S_{Steer}.) This is easiest to see by observing that the bound depends on the number of unique measurements made; if some experimental error causes fewer unique measurements to be made than assumed in the data analysis, separable states can appear steerable.

To illustrate the sensitivity of these entanglement certification methods to errors in the measurement devices, consider the following scenario in which one party's measurement system does not act as expected. For simplicity in this example, we will limit the measurement settings to linear polarizations and include two types of measurement device error. In the first type, Bob's measurements are not calibrated relative to each other properly, so instead of measuring settings b and b', he will measure b and $b' + \delta b'$. This error could arise from a miscalibrated polarizer or halfwave plate, and for low values of $\delta b'$, is representative of the standard imprecision in setting these values. In the second type of error, Alice and Bob's measurement bases are rotated relative to each other (e.g., if one is on a satellite), so that both of Bob's measurement bases are off by an amount δab. Bob then measures in the basis $b + \delta ab$ and $b' + \delta ab$. While this type of error is likely small in tabletop experiments where Alice's and Bob's measurements can be calibrated to the same polarization reference, in a more general networking scenario, Alice and Bob may be rotated relative to each
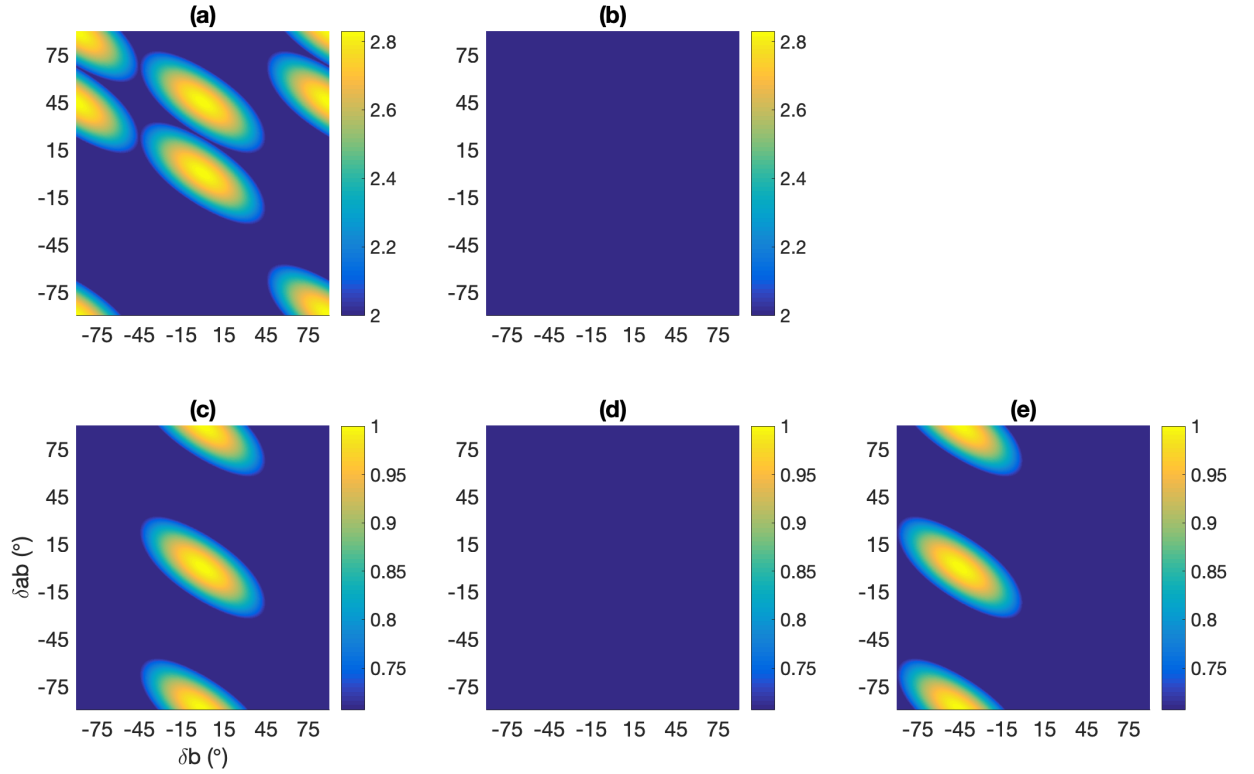
Figure 4.4: Dependence of Bell and steering parameters to measurement errors. The dark blue regions are areas where the bound is not exceeded, while the colored regions correspond to areas where the parameter exceeds the relevant bound. (a) Bell parameter for an entangled state. (b) Bell parameter with a separable state. (c) Steering parameter with an entangled state. (d) Steering parameter with a separable state and Bob making standard measurements. (e) Steering parameter with a separable state and Bob trying to maximize correlations.

other, which will introduce errors of this type in the absence of compensation. If both of these errors occur simultaneously, then Bob's measurement bases will be $b + \delta ab$ and $b' + \delta b' + \delta ab$. For Bell measurements, we set $a = 0°$, $a' = 45°$, $b = 22.5°$, and $b' = 67.5°$. For steering measurements, we set $b = 0°$ and $b' = 45°$ so that b corresponds to the H/V basis and b' to the D/A basis. In the first case, we will investigate the effects of these errors when Alice and Bob share a maximally entangled state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|HH\rangle + |VV\rangle)$ and carry out a CHSH Bell test on the shared state, while in the second case, they perform a two-basis steering measurement. The results are shown in Fig. 4.4 (a) and (c). Both techniques display maximum violations when the errors are zero. The existence of regions in Fig. 4.4 (a) that do not violate a Bell inequality illustrates that while Bell tests are measurement-device independent, errors in measurement devices can still hinder identification of truly nonlocal states. In the next case, a Bell test is carried on the separable state $|HH\rangle$. This result is shown in Fig. 4.4 (b) and shows that, even with measurement errors, a separable state

is never identified as nonlocal. In Fig. 4.4 (d), Alice and Bob carry out the same steering measurement as before, but the shared state is $|HH\rangle$. In this situation, Alice does not modify her strategy to try to take advantage of the lack of shared entanglement. However, if Alice were to change her strategy to try to maximize the observed steering parameter, she can trick Bob into thinking the state they share is steerable. Results for one such strategy are shown in Fig. 4.4 (e), where Alice picks a strategy that maximizes the steering parameter for $\delta ab = 0°$ and $\delta b' = -45°$. For these values, both of Bob's measurement bases are the same, and the state she sends him is an eigenstate of this basis, so she can perfectly predict his measurement outcomes. While this simplistic strategy may be noticeable because of the lack of randomness in outcomes, a more sophisticated actor could reproduce the statistics of an entangled state by changing the state sent to Bob— for example, randomly choosing to send either $|H\rangle$ or $|V\rangle$. Fig. 4.4 (d) and (e) show that, for a steering experiment, measurement-device errors can be exploited by Alice to mislead Bob into concluding they share an entangled state.

4.5 Entanglement Verification of Hyperentangled States

Most Bell tests to date have used some version of the CHSH inequality applied to qubits, but the space of possible tests and states to test is much larger. We investigated the application of tests of nonlocality on hyperentangled quantum states—that is, photons which are entangled simultaneously in multiple degrees of freedom. In this section, we discuss the experimental system used to create and measure states that are simultaneously entangled in polarization and time-bin modes. In Section 4.6.1, we present results of CHSH Bell inequalities using this hyperentangled state and some related shortcomings in using this method to characterize the state. In Section 4.6.2, we discuss how Bell tests can be extended, focusing specifically on the case where more than two measurement bases are used and report on the results of making these measurements on hyperentangled states. Finally, in Section 4.6.3 we investigate the application of quantum steering measurements to these hyperentangled states.

4.5.1 Hyperentangled Pair Source

The hyperentanglement source is driven by a 532-nm pulsed laser (Spectra Physics Vanguard 2.5W 355 laser, frequency doubled from 1064-nm) with an 80-MHz repetition rate. The pump laser is sent through an unbalanced Mach-Zehnder interferometer to put each pulse into a superposition of an early and a late time bin, separated by 2.4 ns, large compared to the 7-ps pulse duration.. After the interferometer, the polarization of the pump beam is prepared using a quarter and half waveplate. In an ideal system, the pump

polarization would be diagonal at this point; however, the polarizing Sagnac interferometer that produces polarization-entangled photon pairs contains a broadband halfwave plate that causes high loss ($\sim 30\%$) for the pump wavelength. Because of the geometry of the interferometer, the pump passes through this element before undergoing downconversion inside the nonlinear crystal only for one polarization, leading to an effective polarization-dependent loss. To counteract this loss, the pump polarization is tuned so that the pump power for both the clockwise and counterclockwise interferometer paths is equal at the nonlinear crystal. After preparing the pump in the necessary polarization state, the pump then enters a polarizing Sagnac interferometer. This interferometer contains a periodically-poled lithium niobate (PPLN) crystal, inside which the pump undergoes type-0 spontaneous parametric downconversion. The horizontal (vertical) component of the pump traverses the Sagnac loop clockwise (counterclockwise) to produce a pair of horizontally (vertically) polarized photons with wavelengths 810 nm and 1550 nm. The interferometer also contains a broadband halfwave plate, causing the output of the counterclockwise path to be vertically polarized, while the clockwise path's output is horizontally polarized, leading to polarization entanglement. Because the pump is in a superposition of time-bins, these pairs are also entangled in timing mode, with a total state after the Sagnac interferometer of

$$|\psi\rangle = \frac{1}{2}(|HH\rangle + |VV\rangle) \otimes (|t_1 t_1\rangle + |t_2 t_2\rangle). \tag{4.5}$$

In general, there will be phases between the two polarization modes, the two timing modes, and the polarization and timing modes. In our experiments, these phases are set to 0 using liquid crystals elements in the 810-nm measurement system.

4.5.2 Measurement System

After exiting the Sagnac interferometer, the downconversion photons are separated from each other and the pump using dichroic mirrors, before being routed to the measurement system. The measurement system is designed so that both the time and polarization measurements can be carried out using polarization optics. This requires a coupling between the polarization and time modes, achieved using another unbalanced Mach-Zehnder interferometer that contains a final polarizing beamsplitter (see Fig. 4.5 (b)): the PBS couples the timing and polarization modes, allowing the timing mode to be effectively controlled using waveplates. For example, with QWP2 at 0 degrees and HWP2 at 0 degrees, Detector 1 will project onto the first timing mode with the polarization mode set by QWP1 and HWP1 (e.g., with HWP1 and QWP1 at 0 degrees, the measurement state for Detector 1 is $\langle H_t 1|$ or, with HWP1 at 22.5 degrees, $\langle D_t 1|$), while with HWP2

at 22.5 degrees, Detector 1 will project onto an equal superposition of both timing modes with orthogonal polarizations (e.g., with QWP1 and HWP1 at 0 degrees, the measurement state will be $\frac{\langle Ht_1| + \langle Vt_2|}{\sqrt{2}}$). In this way, the relevant amplitude and phase between the timing modes can be controlled with HWP2 and QWP2, while the polarization state is controlled by HWP1 and QWP1.

Because of the probabilistic nature of the first nonpolarizing beamsplitter in this interferometer, three possible measurement bins are generated. The early (late) time-bin arises from photon pairs created by the pump pulse that went through the short (long) arm of the pump interferometer and were subsequently sorted into the short (long) arm of the analysis interferometer; the middle time-bin corresponds to photons that took the short path in one interferometer and the long path in the other. In this work, we post-select on this middle time-bin. Because we have matched the path length differences (long path minus short) in the two interferometers, these "short-long" and "long-short" processes are indistinguishable and, therefore, can interfere. In order to preserve the phase of the time modes, it is necessary to actively stabilize these measurement interferometers to match the pump interferometer. This active stabilization is carried out by sending a portion of the pump laser backwards through both measurement interferometers and measuring its intensity on photodiodes. This feedback signal is then used to adjust the length of the long arm of the measurement interferometers using a piezoelectric element (Thorlabs AE0505D16F with Thorlabs TPZ001 driver) to vary the exact positions of the mirrors in the long arm. The 810-nm photons are detected using four silicon avalanche photodiodes (Excelitas SPCM-AQ4C), while the 1550-nm photons are detected with four silicon nanowire detectors (SNSPDs). The APDs have a measured efficiency of $\sim 45\%$ at 810 nm, while the SNSPDs have an efficiency of $\sim 80\%$. The outputs of these detectors are then sent to fast time-tagging electronics (UQDevices UDQ-Logic-16) with a timing resolution of 156 ps, and the resulting time tags are processed to determine coincidence events between detections in the middle time-bin.

4.6 Results

4.6.1 CHSH Bell Inequalities

For reference, we first measured a CHSH inequality on each of the degrees of freedom individually, where the CHSH Bell parameter can be calculated from the quantum correlation between measurement results made in different bases for the two photons:

$$S_{Bell} = E(a, b) + E(a', b) + E(a, b') - E(a', b'). \tag{4.6}$$
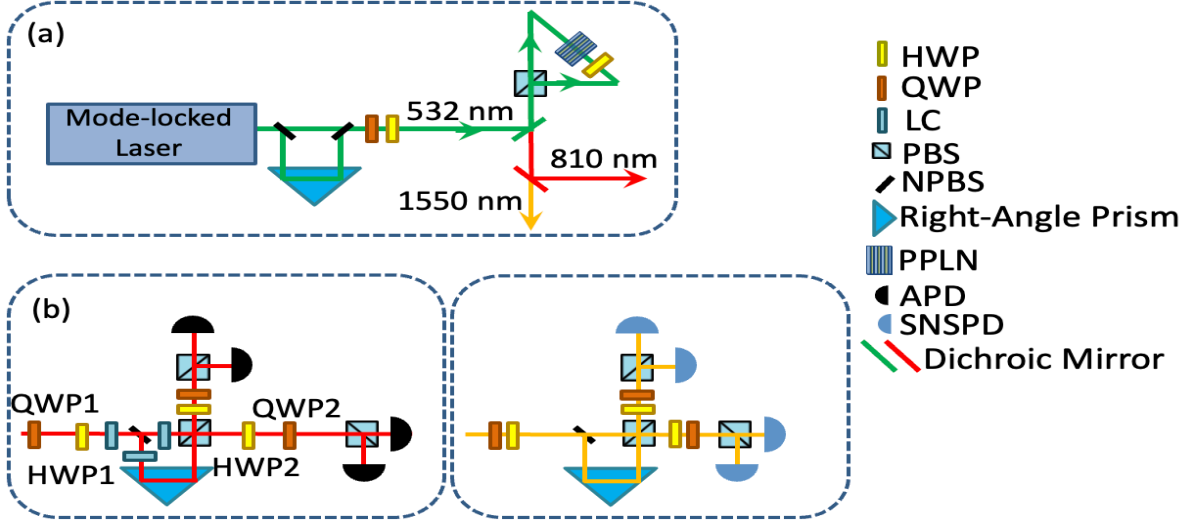
Figure 4.5: Schematic of the hyperentangled photon system. (a) The entanglement source. The pump is prepared in a superposition of timing modes so that, when it passes through the Sagnac polarization entanglement source, the output is entangled in both polarization and time-bin. The photons are then separated by a dichroic mirror for further analysis. (b) The measurement system. The system combines two standard polarization analysis systems with a polarization unbalanced interferometer. This leads to coupling between the polarization and timing modes so that timing measurements can be controlled with polarization optics. This interferometer is actively stabilized to match the pump interferometer, in order to ensure phase stability during measurements. This measurement system uses probabilistic passive sorting, so three timing modes are generated (see text); the results from this system post-select on events in the middle timing mode. This measurement system is duplicated for both photons, with the 810-nm photons being detected by APDs and the 1550-nm photons being detected by SNSPDs. The liquid crystals in the 810-nm measurement system are used to tune the phases in the generated state.

Here $E(a, b)$ can be calculated from the coincidences between Alice and Bob's detector pairs when Alice measures in basis a and Bob measures in basis b:

$$E(a, b) \equiv \frac{N_{11} + N_{22} - N_{12} - N_{21}}{N_{11} + N_{22} + N_{12} + N_{21}}. \tag{4.7}$$

Here N_{12} is the number of events for which Alice's measurement outcome was 1 and Bob's measurement outcome was 2, and similarly for the other N_{ij}. For local realistic states, $|S_{Bell}| \leq 2$. For these measurements, the entanglement source was adjusted to display entanglement in only one of the degrees of freedom at a time. To generate photon pairs that were only entangled in polarization, the long arm of the pump interferometer was blocked so that the nonlinear crystal was being driven by only one time mode. To create pairs entangled only in time-bin, we rotated the pump waveplate so that only the clockwise process in the Sagnac source was activated and the downconversion photons had a definite horizontal polarization. We then carried
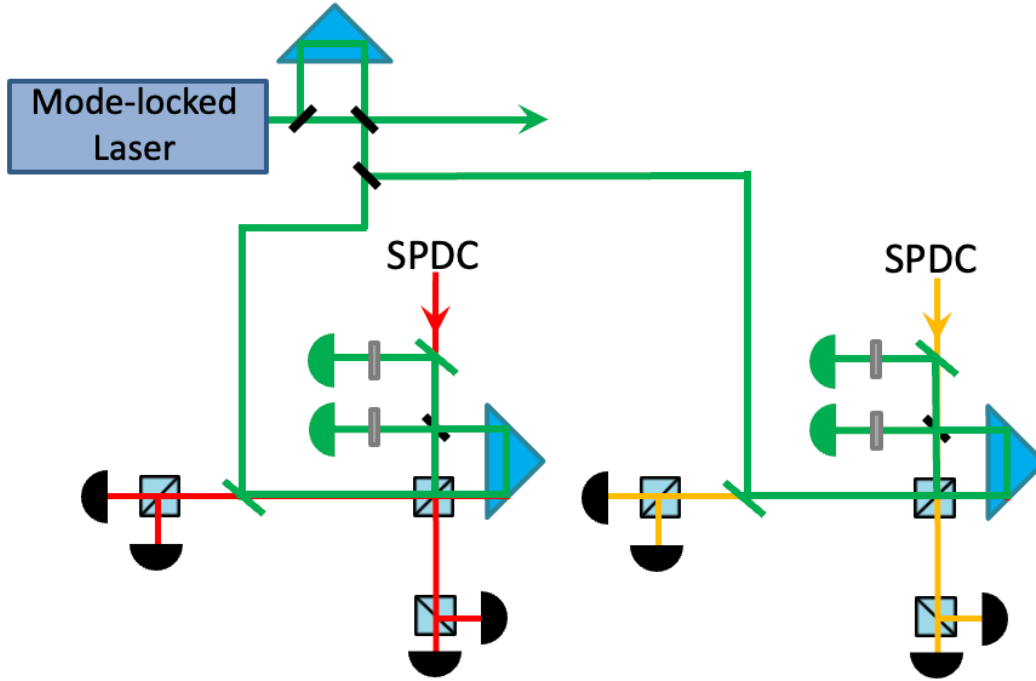
Figure 4.6: Diagram of the active stabilization system. Some of the pump is sent to each of the measurement systems. The pump enters the measurement interferometer through the PBS, counterpropagating relative to the downconversion signal. After exiting the interferometer, the intensity of each port is measured by a silicon photodiode after passing through a polarizer to erase the polarization path information. The relative intensities of these photodiodes are used in a PI feedback system to phase stabilize the measurement interferometer.

out a Bell measurement using the optimal measurement settings for a maximally entangled state ($a = 0°$, $a' = 45°$, $b = 22.5°$, $b' = 67.5°$). For the polarization basis-measurements, these values were set using the halfwave plates before measurement interferometer; for the time-bin basis measurements, they were set by the halfwave plates after the measurement interferometer. We observed a Bell parameter of 2.58±0.02 for the polarization-entangled state and 2.40±0.02 for the time-bin entangled state. From earlier measurements on the quality of the source entanglement, we would expect a polarization Bell parameter of 2.75 and a time-bin Bell parameter of 2.68. The observed values are lower due to imperfect phase tuning and imperfect PBS extinction. The time-bin value is further lowered due to phase averaging caused by path length fluctuations in the interferometers, as well as slight static mismatches between the three unbalanced interferometers, which had to be retuned for this measurement. Both of these Bell parameters indicate that the source was entangled in each degree of freedom separately. Notably, measuring a CHSH inequality of this type cannot be used to infer that the state was entangled in both degrees of freedom simultaneously because

the CHSH inequality can always exhibit a violation even if the source is only entangled in one degree of freedom. This is true even if the measurement settings are chosen to depend on both polarization and time mode, instead of measuring each degree of freedom separately. If an assumption is made that the states Alice and Bob projected on were indeed the ones they intended (or announced they would measure), it is possible to conclude that the state was hyperentangled from just a CHSH Bell measurement. However, an important feature of Bell tests is that they need not rely on such an assumption about what measurement were carried out. The synthesis of a Bell parameter and any conclusions drawn from it rely only on the quantum correlations between Alice and Bob's results, and in this way, the CHSH inequality cannot provide information about a state's dimensionality. Consider the case in which Alice and Bob both try to measure in a hybrid basis between polarization and time-bin that should only yield a violation for hyperentangled states, but problems in their measurement devices make their result insensitive to timing information, so the actual bases used depend only on polarization. In this case, a state only entangled in polarization would lead Alice and Bob to incorrectly conclude that they shared a hyperentangled state, because their assumption about what measurement basis they used was violated. To provide dimensionality information via Bell tests, it is critical that the conclusion not depend on trust of the measurement devices used.

4.6.2 Symmetric Four-Setting Bell Inequalities

Bell inequalities can be characterized by the number of bases measured and the number of outcomes of those measurements. In general, these values can be different for the two parties making measurements, so bipartite inequalities require four parameters to be classified. For example, the CHSH Bell inequality uses two measurement bases and two measurement outcomes, and can be referred to as I_{2222}. For that particular set of parameters, there is only one possible inequality[2]. For bases and outcomes above 2 there can be multiple inequalities for a given set of parameters. We focus on the symmetric case of using four measurement bases and two measurement outcomes for both photons in the pair: I_{4422}. With these parameters, there are at least 27 inequalities [41]. Like the standard Bell parameter, these I_{4422} inequalities are a synthesis of three types of probabilities. The first type of probability is the chance that Alice observes a particular measurement outcome (e.g., 1) when she measures in a particular basis (e.g., a). Similarly, the second type of probability is the chance that Bob observes a particular measurement outcome when he measures in a particular basis. The final type of probability is the chance that Alice and Bob both measure particular outcomes when they measure in particular bases. The inequality then consists of a set of coefficients for these probabilities, such

---

[2]Putting the minus sign on a different term in S_{Bell} (e.g., $\tilde{S}_{Bell} = -E(a, b) + E(a', b) + E(a, b') + E(a', b')$) does not constitute an independent Bell inequality, since it can be transformed to the standard form by relabeling the measurement bases.

	Basis 1	Basis 2	Basis 3	Basis 4
Alice HWP 1	45°	24°	58°	8°
Alice HWP 2	12°	43°	22°	15°
Bob HWP 1	12°	42°	7°	−33°
Bob HWP 2	20°	49°	22°	15°

Table 4.1: Numerically optimized settings for the I_{4422} Bell inequality for the experimental setup in Fig. 4.5. With these settings, the system can achieve a Bell parameter of 0.46, while a local model is limited to a value of 0. While the optimization was carried out over all HWP and QWP settings, for the optimum, all QWP settings were 0.

that the sum of the probabilities multiplied by their respective coefficients must not exceed a certain limit in a local theory. In our measurement device, we take outcome 1 to be a detection event on Detector 1, and outcome 2 to be a detection event on *any of* the other three detectors (of course we could have assigned any detector to outcome 1, and the other three to outcome 2.)
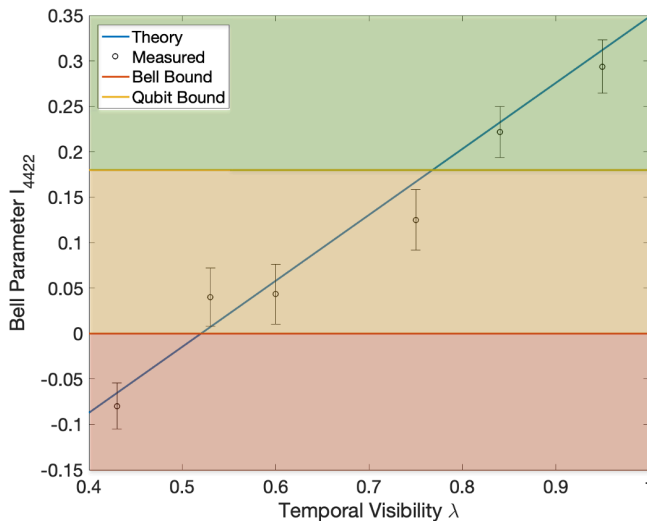
Figure 4.7: Measured I_{4422} values showing the Bell parameter visibility dependence. The visibility of the temporal qubit was changed by unbalancing the measurement interferometer's path length difference relative to the pump. The predicted values were calculated by modifying the input entangled state (see text for details) to match the measured polarization and time-bin visibilities. In the red region, there is no evidence of nonlocality. In the yellow region, the state is nonlocal but could be a pair of entangled qubits. In the green region, the state is nonlocal and must have a dimensionality higher than that of a pair of qubits.

In order to select an inequality to measure, we carried out numerical simulations of the maximum Bell parameter attainable with our system for each of these 27 inequalities, using a maximally entangled ququart state as an input:

$$|\Phi_4\rangle = \frac{1}{2}(|00\rangle + |11\rangle + |22\rangle + |33\rangle), \tag{4.8}$$

where $|0\rangle \equiv |Ht_1\rangle, |1\rangle \equiv |Ht_2\rangle, |2\rangle \equiv |Vt_t\rangle$, and $|3\rangle \equiv |Vt_2\rangle$. In these simulations, we used numerical optimization to find the waveplate settings that produced the maximum violation. Because our measurement system cannot project onto an arbitrary ququart state (e.g., we cannot project onto $|H\rangle \otimes \frac{|t_1\rangle+|t_2\rangle}{\sqrt{2}}$ because the two time-modes always have orthogonal polarizations in our system), it is not possible to obtain a violation for some of the inequalities, and for those with a violation, the maximum violation achievable with our system does not necessarily correspond with the theoretic limit of the inequality. Some of the inequalities exhibited the interesting feature that the maximum violation obtainable in our system using a pair of entangled ququarts exceeded the theoretic bound for a pair of qubits. This feature implies that measuring a sufficiently large Bell parameter using these inequalities could be used as an indication of the dimensionality of the measured entangled state. This means that, in addition to identifying nonlocal states, these inequalities can distinguish the dimensionality of the entangled states in the style of a dimensional witness [42]. We chose to measure the inequality with the largest difference between our system's violation using entangled ququarts and the theoretic bound using entangled qubits:

$$
\begin{aligned}
I_{4422} =&\, 2(Pr(a_1b_1) + Pr(a_1b_2) + Pr(a_1b_3) + Pr(a_2b_1) \\
&+ Pr(a_3b_1) + Pr(a_4b_2) + Pr(a_2b_4)) \\
&- 2(Pr(a_3b_2) + Pr(a_2b_3) + Pr(a_3b_3) + Pr(a_3b_4) \\
&+ Pr(a_4b_3) + Pr(a_1) + Pr(b_1) + Pr(a_2) + Pr(b_2)) \\
&+ Pr(a_4b_1) + Pr(a_1b_4) + Pr(a_2b_2) - Pr(a_4b_4).
\end{aligned}
\tag{4.9}
$$

For local theories, this parameter cannot exceed 0; for entangled qubits, it is bounded by 0.18.

We measured this I_{4422} inequality using our best approximation to the maximally entangled ququart state given in Equation 4.8. With our system, the theoretical maximum value of the Bell parameter was 0.46. With optimal source tuning, we observed a parameter of 0.45±0.03 after measuring 85000 coincidence events (summed over all measurement basis combinations). This value obviously exceeded the inequality bound of 0, indicating that the state was entangled; furthermore, the parameter also exceeded the qubit-entangled bound of 0.18, providing proof that the dimensionality of the state must be larger than that of a qubit-entangled state.

We then investigated the robustness of this Bell parameter to noise by reducing the amount of entangle-

ment in the time-bin degree of freedom. This leads to a state of the form

$$\rho = (\frac{\lambda_{pol}}{2}|\phi_{pol}^+\rangle\langle\phi_{pol}^+| + \frac{1-\lambda_{pol}}{2}(|HH\rangle\langle HH| + |VV\rangle\langle VV|))$$
$$\otimes(\frac{\lambda_{time}}{2}|\phi_{time}^+\rangle\langle\phi_{time}^+| + \frac{1-\lambda_{time}}{2}(|t_1t_1\rangle\langle t_1t_1| + |t_2t_2\rangle\langle t_2t_2|)),$$

(4.10)

where λ_{time} was modified for each measurement and λ_{pol} was held fixed. Decoherence in time-bin qubits can be implemented by shifting the modes relative to each other in time. Such a method only works in combination with a measurement scheme in which the two time-bins are mapped onto one time-bin before detection, which is achieved in our setup through the use of the second two unbalanced Mach-Zehnder interferometers, as well as the post-selection on the middle time-bin pulse.

This functions as an approximation of decoherence in combination with the measurement method; because the state could be made pure again by delaying one pulse relative to the other to achieve total overlap, the effective decoherence is actually reversible.[3] Because these measurements involved adding noise to a maximally entangled state, effort was not made to perfectly tune the polarization state, leading λ_{pol} to be 0.9 for these measurements. As seen in Fig. 4.7, we were able to observe a I_{4422} Bell parameter over the qubit-entangled limit while the temporal visibility was above 0.75. A Bell violation excluding all local realistic models was observed for temporal visibilities of 0.53 and above, in agreement with our theoretical model after accounting for our produced input state. In particular, from the data we extract the slope m of I_{4422} vs. the temporal visibility λ. We find $m = 0.68 \pm 0.04$, in close agreement with the predicted value of $m = 0.72$.

4.6.3 Quantum Steering

Next, we verified the ability of one photon in the pair to "steer" the other photon. We generated the maximally entangled state in Equation 4.8 and distributed them so that Alice measured the 810-nm photon and Bob measured the 1550-nm photon. Bob used two measurement bases. For the first basis, Bob's +1 outcome corresponded to projecting onto the state $|Ht_1\rangle$, while the -1 outcome corresponded to projecting onto the state $|Vt_2\rangle$. In the second basis, the +1 outcome corresponded to projecting onto the state $\frac{1}{\sqrt{2}}(|Ht_1\rangle+|Vt2\rangle)$, and the -1 outcome corresponded to projecting onto the state $\frac{1}{\sqrt{2}}(|Ht_1\rangle - Vt2\rangle)$. Because of the symmetric nature of their shared state, Alice's optimal measurement strategy was to measure in the same basis as Bob. For this case using two bases, the steering bound takes its highest value of $\frac{1}{\sqrt{2}}$:

---

[3]Note, however, that this is often true of dephasing-based decoherence: in *principle*, it could be undone if one knew the correct 'anti-phase' fluctuations to apply.

$$S_{Steer} = \frac{1}{2} \sum_{j=1}^{2} \langle A_j \hat{b}_j \rangle \leq \frac{1}{\sqrt{2}}, \tag{4.11}$$

while the largest possible value of S_{Steer}, using a maximally entangled state, is 1. Using these settings, we achieved a steering parameter of 0.94±0.02, indicating Alice's ability to steer Bob's measurement and the entangled nature of the shared state. Accounting for previous measurements of the entangled state's quality, we would expect this system to produce a steering parameter of 0.95, in good agreement with the value measured.

We then investigated the robustness of the steering parameter to non-maximally entangled states. Specifically, we produced states of the form

$$\begin{aligned}
\rho = (\frac{\lambda}{2}|\phi_{pol}^+\rangle\langle\phi_{pol}^+| + \frac{1-\lambda}{2}(|HH\rangle\langle HH| + |VV\rangle\langle VV|))\otimes \\
(\frac{\lambda}{2}|\phi_{time}^+\rangle\langle\phi_{time}^+| + \frac{1-\lambda}{2}(|t_1 t_1\rangle\langle t_1 t_1| + |t_2 t_2\rangle\langle t_2 t_2|)).
\end{aligned} \tag{4.12}$$
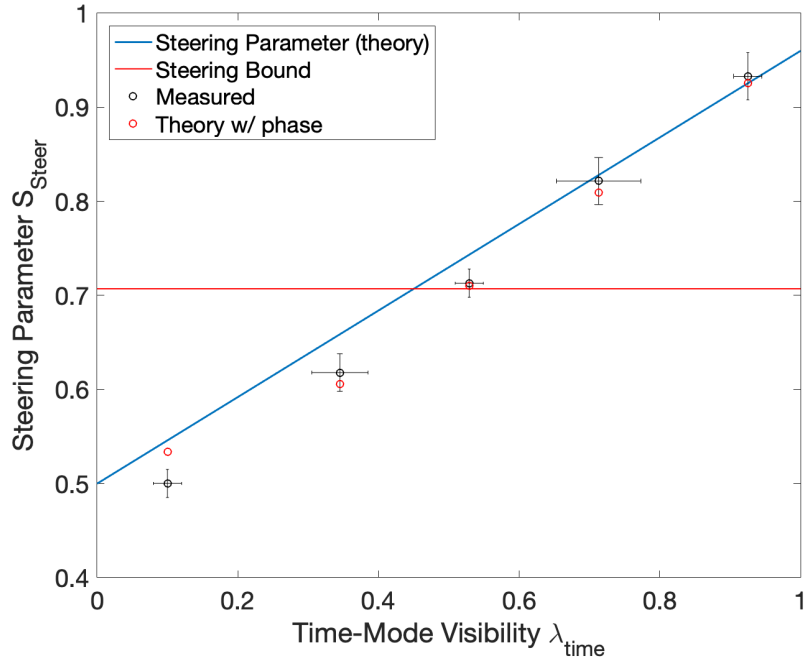
Figure 4.8: Measured steering parameter values. The theory curve assumes a state exactly of the form given in Equation 4.12, while the "theory with phase" prediction includes an additional relative phase between the timing modes, as determined from a subset of the measurements.

This state differs slightly from the one generated in the previous Bell measurements in that the degree of purity in the timing and polarization modes are the same. In order to do this, simultaneous noise had to

be added to both the time-bin and polarization degrees of freedom. In the polarization degree of freedom, the λ_{pol} parameter was tuned by inserting quartz crystals into the 810-nm arm of the source, leading the two polarization components to walk off temporally from each other; alternatively, this can be interpreted as frequency-dependent birefringent phase shift — tracing over the relatively broadband ($\sim$0.4-nm full-width half maximum) of 810-nm photons leaves the photons in a partially mixed state. λ_{time} was set in the same manner as for the Bell test in Section 4.6.2 by unbalancing the interferometers relative to each other. Special care must be taken when using the method in conjunction with polarization decoherence techniques using birefringence because both methods rely on displacing modes relative to each other in time. If polarization and timing modes are coupled (which must be the case in this measurement system), it is possible for these effects to act in opposite directions, leading to reduced effective decoherence in both modes.

Because the quartz elements produce discrete changes in λ_{pol}, the path length difference of the pump interferometer was tuned to produce a value of λ_{time} in the time mode that closely matched that of the polarization. Because the phase of the complete entangled state depends on the relative path length differences of the interferometers, the phase of the entangled state had to be readjusted to zero for each noise value measured. At low values of λ_{time}, this readjustment process becomes more challenging, as the phase-insensitive noise dominates, leading to slightly different values for the entanglement phase at each noise level despite efforts to retune this phase to zero between measurements. In order to accurately predict the steering parameter observed at each level, it was necessary to take into account this phase variation. For each noise level, we first made measurements to determine the entanglement phase ϕ. Specifically, measurements in the second basis are sensitive to this phase. If p_1 is the coincidence probability of Alice measuring -1 and Bob measuring +1, and p_2 is the probability of Alice and Bob both measuring -1, then this phase is given by

$$\phi = \arccos(4(p_2 - p1)). \tag{4.13}$$

The results of these steering measurements can be seen in Fig. 4.8. This shows that the effects of this phase retuning were significant, as a model that assumed this phase was zero did not accurately predict the observed steering parameter. When this phase was included, the model was able to accurately predict the observed steering parameters.

4.7   Conclusion

In this chapter, we have shown a number of ways to detect entanglement in a hyperentangled photonic state. Only the use of Bell tests beyond the CHSH inequality gives an indication of the higher-dimensional

nature of the generated state. This new feature could be more useful with increased study on the theoretic bounds of these inequalities. For example, if the spectrum of maximum Bell parameters as a function of entangled state dimension was known, then a more precise conclusion about the state dimension could be reached. Currently, observing a sufficiently large Bell parameter only provides evidence that the state is more than qubit-entangled, instead of indicating a specific dimension. In order to provide precise dimensionality evidence, it is necessary to establish exact bounds for these Bell parameters as a function of state dimension. Further investigation into these less commonly measured Bell inequalities could thus lead to improved state characterization techniques. Further, we demonstrated the ability to steer hyperentangled states, which could potentially enable a broad range of quantum applications that involve a trusted measurement system.

Appendix A

Stopping Condition Data

Here we present the data from Fig. 3.7 (a) binned by state dimension and count rate. These results show that the required stopping parameter does not significantly depend on the count rate or the state dimension.
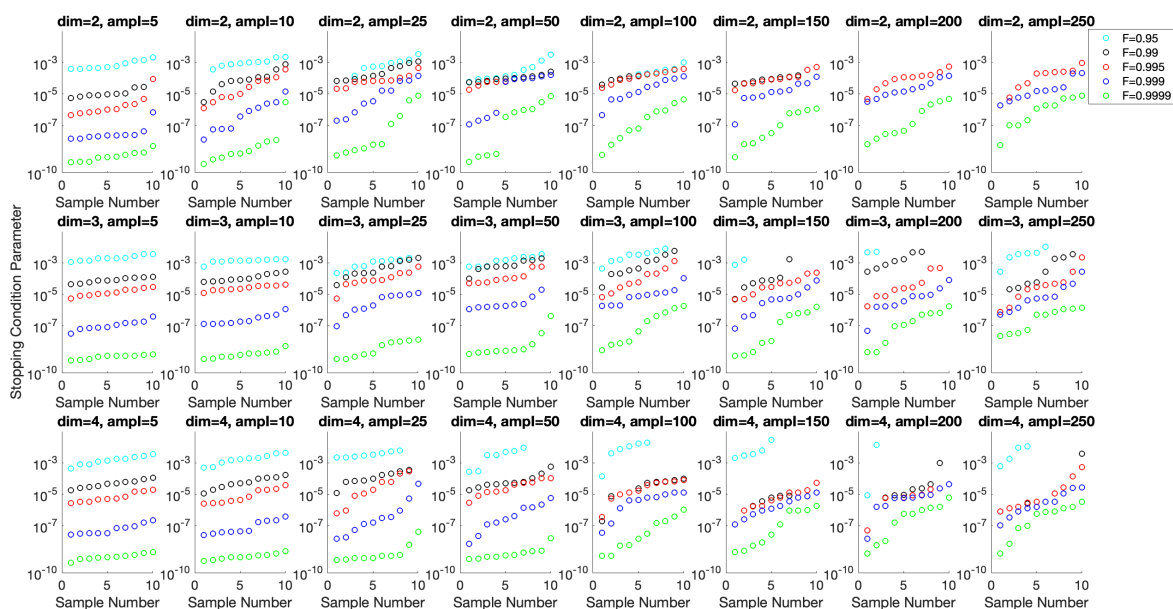
Figure A.1: Binned stopping condition data. This data is the same as the result presented in Fig. 3.7 (a), but has been binned by dimension and count rate.

Appendix B

Bayesian mean estimation tomography code

This appendix contains the BME algorithm described in Chapter 3, as well as relevant subfunctions that facilitate the algorithm.

B.1   bayesianTomography

This is the main BME algorithm.

```
1  function [p,randomStates,randLikelihood,currentStability,SCi,samplesToSC,meanT
       ,covarianceMat] = bayesianTomography(counts, settings,guessState)
2  %inputs: counts, m x 1 array of measured count values.
3  %settings: m x d x d matrix of projectors.
4  %guessState: d x d density matrix for initialization
5
6  %outputs:
7  %p: Bayesian mean estimate.
8  %randomStates: sampled states
9  %randLikelihood: log likelihoods of the sampled states
10 %stabilityHistory: history of the stability stopping stoppingCondition
11 %stoppingstoppingConditionTestLocations: sample number when the stability was
12 %measured
13 %samplesToSC: sample number at which the stopping stoppingCondition was met
14 %meanT: mean value of the posterior
15 %covarianceMat: covariance of the posterior
16
17 %algorithm parameters
18 preallocationSize=100000; %preallocate memory for speed
```

```matlab
19  changeThreshold=10^−9; %stopping condition threshold
20  numberOfPriorSamples=5000; %number of samples to sample from the inital prior
21  updateFrequency=1000; %controls how often the estimate of the posterior is
        updated, and how often stability is checked
22
23  %initialize and preallocate memory
24  d=length(guessState);
25  p=guessState;
26  randomStates=zeros(d,d,preallocationSize);
27  randLikelihood=zeros(preallocationSize,1);
28  randomT=zeros(preallocationSize,d*d);
29  meanT=zeros(d*d,1);
30  currMeanU=zeros(d,d);
31  stabilityHistory=zeros(floor(preallocationSize/updateFrequency),1);
32  stoppingConditionTestLocations=zeros(size(stabilityHistory));
33  sectionLikelihood=0;
34  cumLikelihood=0;
35  i=1;
36  stillUsingPrior=1;
37  samplesToSC=0;
38  iterationCounter=0;
39
40  %estimate the state insensity for the guess state
41  norm=sum(counts/(length(counts)*d));
42  [norm,baseLike]=fminsearch(@(x) BMELikelihood(counts,p,settings,x),norm);
43  stoppingCondition=1;
44
45  while stoppingCondition
46
47      %expand memory in chunks if preallocation size is exceeded
48      if i>length(randLikelihood);
49          randomStates(:,:,round(i*1.2))=zeros(d,d);
```

68

```matlab
50            randLikelihood(round(i*1.2))=0;
51            randomT(round(i*1.2),:)=zeros(1,d*d);
52        end
53
54        %once we have exceeded the prior samples, try to switch to sampling the
55        %posterior estimate
56        if i>numberOfPriorSamples && stillUsingPrior
57
58            %calculate the covariance matrix
59            normalizedLikelihood=normalizeExponentLikelihood(randLikelihood(1:i-1)
                );
60            contributingParams=normalizedLikelihood>0;
61            nLMax=max(normalizedLikelihood);
62            nLNext=max(normalizedLikelihood(normalizedLikelihood<nLMax));
63            covarianceMat=weightedcov(randomT(contributingParams,:),
                normalizedLikelihood(contributingParams));
64            for g=1:d*d
65                meanT(g)=sum(randomT(1:i-1,g).*normalizedLikelihood);
66            end
67
68            %switch to adaptive distributions only if the covariance matrix is
69            %valid
70            [~,e]=cholcov(covarianceMat);
71            if e~=0 || (nLNext/nLMax)<.01
72                stillUsingPrior=1;
73                numberOfPriorSamples=numberOfPriorSamples+updateFrequency;
74            else
75                stillUsingPrior=0;
76            end
77        end
78
```

```matlab
79        %periodically update the posterior estimate if we aren't using the prior
              anymore
80        if mod(i,updateFrequency)==0 && ~stillUsingPrior
81
82            normalizedLikelihood=normalizeExponentLikelihood(randLikelihood(1:i-1)
                  );
83            contributingParams=normalizedLikelihood>0;
84            covarianceMatNew=adaptFactor*weightedcov(randomT(contributingParams,:)
                  ,normalizedLikelihood(contributingParams));
85          nLMax=max(normalizedLikelihood);
86          nLNext=max(normalizedLikelihood(normalizedLikelihood<nLMax));
87
88            [~,e]=cholcov(covarianceMatNew);
89          %only update the posterior estimate if the covariance matrix is
90          %valid
91            if e==0 && (nLNext/nLMax)>.001
92                covarianceMat=covarianceMatNew;
93                for g=1:d*d
94                    meanT(g)=sum(randomT(1:i-1,g).*normalizedLikelihood);
95                end
96            end
97        end
98
99      %when not using the initial prior, sample from the posterior estimate
100       if ~stillUsingPrior
101          %sample the posterior estimate, and compute its likelihood
102          randT=matricizeTVector(mvnrnd(meanT,covarianceMat));
103          randP=randT*randT';
104          randP=randP./trace(randP);
105          randomStates(:,:,i)=randP;
106          randomT(i,:)=vectorizeTMatrix(randT);
107          rLtemp=BMELikelihood(counts,randP,settings,norm);
```

70

```matlab
108
109         %if the range of likelihoods becomes too large, rescale
110           if rLtemp<baseLike
111               currentBestGuess=randP;
112               oldBase=baseLike;
113               baseLike=rLtemp;
114               [norm,~]=fminsearch(@(x) BMELikelihood(counts,currentBestGuess,
                      settings,x),norm);
115               randLikelihood(1:i-1)=randLikelihood(1:i-1)+oldBase-baseLike;
116               currMeanU=zeros(d,d);
117               for j=1:i-1
118                   currMeanU=currMeanU+randomStates(:,:,j)*exp(-randLikelihood(j)
                          );
119               end
120           end
121
122           randLikelihood(i)=rLtemp-baseLike;
123       else
124           %initially, sample from the ginibre distribution and compute its
125           %likelihood
126           randP=sampleGinibreDensityMatrix(d);
127           randomT(i,:)=vectorizeTMatrix(tMatrix(randP));
128           randomStates(:,:,i)=randP;
129           rLtemp=BMELikelihood(counts,randP,settings,norm);
130
131           if rLtemp<baseLike
132               currentBestGuess=randP;
133               oldBase=baseLike;
134               baseLike=rLtemp;
135               [norm,~]=fminsearch(@(x) BMELikelihood(counts,currentBestGuess,
                      settings,x),norm);
136               randLikelihood(1:i-1)=randLikelihood(1:i-1)+oldBase-baseLike;
```

```matlab
137                  currMeanU=zeros(d,d);
138                  for j=1:i-1
139                      currMeanU=currMeanU+randomStates(:,:,j)*exp(-randLikelihood(j)
                             );
140                  end
141              end
142              randLikelihood(i)=rLtemp-baseLike;
143          end
144
145          %update the mean with the newly sampled state
146          currMeanU=currMeanU+randomStates(:,:,i)*exp(-randLikelihood(i));
147
148          %track the total likelihood and the section likelihood
149          cumLikelihood=cumLikelihood+exp(-randLikelihood(i));
150          sectionLikelihood=sectionLikelihood+exp(-randLikelihood(i));
151
152          %periodically check if the stopping condition is met
153          if mod(i,updateFrequency)==0
154              %normalize the mean and calculate what the mean was a few samples
155              %ago, and avoid dividing by zero if we having sampled a likely
156              %state yet
157              tr=trace(currMeanU);
158              if tr~=0
159                  currMean=currMeanU/tr;
160              else
161                  currMean=zeros(d);
162              end
163              %check if this is the first time we've calculated the posterior
164              %estimate
165              if i~=updateFrequency
166                  %if we haven't found enough high-likelihood states to stop using
167                  %the prior, we know we aren't done sampling yet
```

```matlab
168                 if  stillUsingPrior  ||  tr==0
169                     iterationCounter=iterationCounter+1;
170                     stabilityHistory(iterationCounter)=Inf;
171                     stoppingConditionTestLocations(iterationCounter)=i;
172
173                     %otherwise, calculate the stability and compare it to our
174                     %threshold. If our state is stable, we can stop sampling
                            states
175                     %and return the mean
176                 else
177                     iterationCounter=iterationCounter+1;
178
179                     %calculate the normalization factor based on the fraction
180                     %of total likelihood sampled in this iteration
181                     normFactor=floor(i/updateFrequency)*sectionLikelihood/
                            cumLikelihood;
182                     infidelity=(1-real(fidelity(currMean,prevMean)));
183
184                     %ensure rounding never leads the infidelity to be negative
185                     if  infidelity<0
186                         infidelity=eps;
187                     end
188                     stabilityHistory(iterationCounter)=infidelity/normFactor;
189                     stoppingConditionTestLocations(iterationCounter)=i;
190                 end
191                 %if we've iterated at least once, and the state is stable, we
192                 %can exit
193                 if  (iterationCounter>0) && (stabilityHistory(iterationCounter)<
                        changeThreshold)
194                     p=currMean;
195                     samplesToSC=i;
196                     stoppingCondition=0;
```

```matlab
197              end
198          end
199          %if we haven't reached stability, set the state to compare to next
200          %iteration, and reset the section likelihood
201          prevMean=currMean;
202          sectionLikelihood=0;
203       end
204       i=i+1;
205    end
206    p=p/trace(p);
207    end
```

B.2   BMELikelihood

This function calculates the log of the likelihood of a density matrix.

```matlab
1  function [ esum ] = BMELikelihood( counts,p,settings,Norm)
2  %inputs
3  %counts: n x 1 array of observed counts
4  %p: d x d density matrix whose likelihood is being evaluated
5  %settings: d x d x n matrix of measurement operators that correspond to the
6  %measured counts
7  %norm: intensity of the measured state
8
9  %outputs
10 %esum: log likeihood of p
11
12 unNorm=p*abs(Norm);
13 esum=0;
14 exponent=zeros(size(settings,3),1);
15 for i=1:size(settings,3)
16     %calculate the expected number of counts for each setting
17     expected=trace(unNorm*settings(:,:,i));
```

```
18      expected=real(expected);
19
20      %avoid dividing by zero for pure states
21      if expected==0
22          expected=1;
23      end
24      exponent(i)=(expected-counts(i))^2/(2*expected);
25
26      esum=esum+exponent(i);
27  end
28  end
```

B.3 SampleGinibreDensityMatrix

This function samples a density matrix from the Ginibre distribution.

```
1  function [ p ] = sampleGinibreDensityMatrix( dim )
2  A=normrnd(0,1,dim,dim)+1i*normrnd(0,1,dim,dim);
3  p=(A*A');
4  p=p/trace(p);
5
6  end
```

B.4 normalizeExponentLikelihood

This function converts a list of log likelihoods to normalized likelihoods.

```
1  function [ fractionalLikelihood ] = normalizeExponentLikelihood( likelihoods )
2  %input: n x 1 array of log likelihoods
3
4  %output: n x 1 array of likelihoods, normalized to sum to 1
5
6  nFactor=min(likelihoods);
7  likelihoods=likelihoods-nFactor;
```

```matlab
 8   for  i=1:length(likelihoods)
 9       likelihoods(i)=exp(-likelihoods(i));
10   end
11   norm=sum(likelihoods);
12   fractionalLikelihood=likelihoods./norm;
13
14
15   end
```

Appendix C

I_{4422} Bell Inequalities

This Appendix contains the coefficients and bounds for the I_{4422} Bell inequalities investigated in Section 4.6.2. These tables present coefficients for the different probabilities that constitute a Bell inequality. The first row indicates the coefficients for Alice's local measurements. For example, the first entry in the first row indicates the coefficient in the inequality for the probability that Alice measures outcome 1 when she uses measurement setting 1. Similarly, the first row represents these coefficients for Bob's local measurements. The remaining matrix of values indicates the coefficients for the coincidence outcomes. For example, the first entry in this matrix indicates the coefficient for the probability that, when Alice and Bob both measure setting 1, they both obtain outcome 1. Expanding these coefficients and probabilities, we can obtain an equation for the Bell parameter. For example,

$$
\begin{aligned}
I_{4422}^1 =& Pr(a_1b_3) + Pr(a_1b_4) + Pr(a_2b_2) + Pr(a_2b_4) + Pr(a_3b_1) + \\
& Pr(a_3b_4) + Pr(a_4b_1) + Pr(a_4b_2) + Pr(a_4b_3) - \\
& Pr(a_1) - Pr(a_2) - Pr(a_3) - Pr(b_1) - Pr(b_2) - \\
& Pr(b_3) - Pr(a_2b_3) - Pr(a_3b_2) - Pr(a_3b_3) - Pr(a_4b_4).
\end{aligned}
\tag{C.1}
$$

I_{4422}^{18} is the inequality that was measured, with results presented in Section 4.6.2.

$I_{4422}^1 \equiv$

	-1	-1	-1	0
-1	0	0	1	1
-1	0	1	-1	1
-1	1	-1	-1	1
0	1	1	1	-1

≤ 0

$I_{4422}^2 \equiv$

	-3	-1	0	0
-2	2	1	2	0
-1	1	1	-1	1
0	2	-2	-1	0
0	1	1	-1	-1

≤ 0

$A_5 \equiv$

	-1	-1	-1	0
-1	0	1	1	1
-1	1	1	1	-1
-1	1	1	-1	0
0	1	-1	0	0

≤ 0

$A_6 \equiv$

	-1	-1	0	0
-1	1	1	0	1
-1	1	0	1	-1
0	0	1	-1	-1
0	1	-1	-1	-1

≤ 0

$AS_1 \equiv$

	-2	-1	0	0
-2	1	1	1	1
-1	1	1	1	-1
0	1	1	-2	0
0	1	-1	0	0

≤ 0

$AS_2 \equiv$

	-3	-1	-1	0
-3	1	1	2	2
-1	1	2	1	-2
-1	2	1	-2	1
0	2	-2	1	-1

≤ 0

$AII_1 \equiv$

	-1	-1	-1	0
-1	-1	1	1	1
-1	1	0	2	-1
-1	1	2	-1	-1
0	1	-1	-1	0

≤ 0

$AII_2 \equiv$

	-3	-1	0	-1
-1	2	1	1	-1
-1	1	2	-1	1
0	1	-1	-1	1
0	1	-1	0	0

≤ 0

$I_{4422}^3 \equiv$

	-2	-1	-1	0
-1	1	1	1	1
0	0	1	0	-1
0	1	-1	1	-1
0	1	0	-1	0

≤ 0

$I_{4422}^4 \equiv$

	-1	-1	0	0
-1	1	1	1	-1
-1	1	1	-1	1
0	1	-1	-1	-1
0	-1	1	-1	-1

≤ 0

$I_{4422}^5 \equiv$

	-2	-1	0	0
-1	1	0	1	0
-1	1	1	-1	1
0	1	-1	0	0
0	1	1	-1	-1

≤ 0

$I_{4422}^6 \equiv$

	-2	-1	-1	0
-1	1	-1	1	1
-1	1	1	-1	1
0	1	-1	1	-1
0	1	1	-1	-1

≤ 0

$I_{4422}^7 \equiv$

	-1	0	0	0
-1	2	-1	-1	1
0	-1	-1	0	1
0	0	1	-1	0
0	1	0	1	-1

≤ 1

$I_{4422}^8 \equiv$

	-2	-1	-1	0
-2	1	1	2	1
-1	1	2	-2	0
-1	2	-2	-1	1
0	1	0	1	-2

≤ 0

$I_{4422}^9 \equiv$

	-2	-1	-1	0
-2	1	1	2	1
-1	1	2	-2	0
-1	2	-2	-2	1
0	1	0	1	-1

≤ 0

$I_{4422}^{10} \equiv$

	-2	-1	-1	0
-2	1	1	1	2
-1	1	1	2	-2
-1	1	2	-2	-1
0	2	-2	-1	-1

≤ 0

$I_{4422}^{11} \equiv$

	-2	-1	-1	0
-2	1	1	1	2
-1	1	0	2	-1
-1	1	2	-1	-1
0	2	-1	-1	-1

≤ 0

$I_{4422}^{12} \equiv$

	-2	-1	-1	0
-2	1	1	1	2
-1	1	-1	1	0
-1	1	1	2	-2
0	2	0	-2	-1

≤ 0

$I_{4422}^{13} \equiv$

	-2	-1	-1	0
-2	0	1	1	1
-1	1	-2	1	1
-1	1	1	-1	1
0	1	1	1	-1

≤ 0

$I_{4422}^{14} \equiv$

	-2	-1	0	0
-2	2	2	0	1
-1	2	-1	1	-1
0	0	1	-1	-1
0	1	-1	-1	0

≤ 0

$I_{4422}^{15} \equiv$

	-2	-1	0	0
-2	2	1	1	1
-1	1	-1	-1	1
0	1	-1	0	-1
0	1	1	-1	-1

≤ 0

$I_{4422}^{16} \equiv$

	-2	-1	0	0
-2	2	0	1	1
-1	0	1	-1	1
0	1	-1	-1	0
0	1	1	0	-1

≤ 0

$I_{4422}^{17} \equiv$

	-2	-1	-1	-1
-2	-1	1	2	2
-1	1	-1	-1	2
-1	2	-1	2	-1
-1	2	2	-1	0

≤ 0

$I_{4422}^{18} \equiv$

	-2	-2	0	0
-2	2	2	2	-1
-2	2	1	-2	2
0	2	-2	-2	-2
0	-1	2	-2	-1

≤ 0

$I_{4422}^{19} \equiv$

	-3	-2	0	0
-3	2	2	1	2
-2	2	-1	2	-2
0	1	2	-1	-1
0	2	-2	-1	0

≤ 0

$I_{4422}^{20} \equiv$

	-2	-2	-2	0
-2	-1	1	1	2
-2	1	-1	1	2
-2	1	1	-2	2
0	2	2	2	-2

≤ 0

Figure C.1: List of I_{4422} Bell inequalities. Tables from [41] (copyright by American Physical Society).

References

[1] Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge UP, 2010).

[2] Sharma, A. N. *In search of photonic bound entanglement: using hyperentanglement to study mixed entangled states.* Ph.D. thesis, University of Illinois at Urbana-Champaign (2016).

[3] Schrödinger, E. Probability relations between separated systems. *Mathematical Proceedings of the Cambridge Philosophical Society* **32**, 446–452 (1936).

[4] Boyd, R. W. *Nonlinear optics* (Academic press, 2003).

[5] Hong, C. K. & Mandel, L. Theory of parametric frequency down conversion of light. *Phys. Rev. A* **31**, 2409–2418 (1985). URL `https://link.aps.org/doi/10.1103/PhysRevA.31.2409`.

[6] Tanzilli, S. *et al.* Highly efficient photon-pair source using periodically poled lithium niobate waveguide .

[7] Kwiat, P. G., Waks, E., White, A. G., Appelbaum, I. & Eberhard, P. H. Ultrabright source of polarization-entangled photons. *Phys. Rev. A* **60**, R773–R776 (1999). URL `https://link.aps.org/doi/10.1103/PhysRevA.60.R773`.

[8] Evans, P. G., Bennink, R. S., Grice, W. P., Humble, T. S. & Schaake, J. Bright source of spectrally uncorrelated polarization-entangled photons with nearly single-mode emission. *Phys. Rev. Lett.* **105**, 253601 (2010). URL `https://link.aps.org/doi/10.1103/PhysRevLett.105.253601`.

[9] Shi, B.-S. & Tomita, A. Generation of a pulsed polarization entangled photon pair using a sagnac interferometer. *Phys. Rev. A* **69**, 013803 (2004). URL `https://link.aps.org/doi/10.1103/PhysRevA.69.013803`.

[10] Kaminow, I. P. *An Introduction to Electrooptic Devices* (academic press, 2013).

[11] Bennet, C. & Brassard, G. "Quantum cryptography: Public key distribution and coin tossing". In *IEEE Int. Conf. on Comp., Sys. and Sig. Proc.*

[12] Wootters, W. K. & Zurek, W. H. A single quantum cannot be cloned. *Nature* **299**, 802–803 (1982).

[13] Huang, A. *et al.* Laser-damage attack against optical attenuators in quantum key distribution. *Phys. Rev. Applied* **13**, 034017 (2020). URL `https://link.aps.org/doi/10.1103/PhysRevApplied.13.034017`.

[14] Huang, A. *et al.* Laser-seeding attack in quantum key distribution. *Phys. Rev. Applied* **12**, 064043 (2019). URL `https://link.aps.org/doi/10.1103/PhysRevApplied.12.064043`.

[15] Chaiwongkhot, P. *et al.* Eavesdropper's ability to attack a free-space quantum-key-distribution receiver in atmospheric turbulence. *Phys. Rev. A* **99**, 062315 (2019). URL `https://link.aps.org/doi/10.1103/PhysRevA.99.062315`.

[16] Acín, A. *et al.* Device-independent security of quantum cryptography against collective attacks. *Phys. Rev. Lett.* **98**, 230501 (2007). URL https://link.aps.org/doi/10.1103/PhysRevLett.98.230501.

[17] Lo, H.-K., Curty, M. & Qi, B. Measurement-device-independent quantum key distribution. *Phys. Rev. Lett.* **108**, 130503 (2012). URL https://link.aps.org/doi/10.1103/PhysRevLett.108.130503.

[18] Sangouard, N., Simon, C., de Riedmatten, H. & Gisin, N. Quantum repeaters based on atomic ensembles and linear optics. *Reviews of Modern Physics* **83**, 33–34 (2011).

[19] Żukowski, M., Zeilinger, A., Horne, M. A. & Ekert, A. K. "event-ready-detectors" bell experiment via entanglement swapping. *Phys. Rev. Lett.* **71**, 4287–4290 (1993). URL https://link.aps.org/doi/10.1103/PhysRevLett.71.4287.

[20] Duan, L.-M., Lukin, M. D., Cirac, J. I. & Zoller, P. Long-distance quantum communication with atomic ensembles and linear optics. *Nature* **414**, 413–418 (2001). URL https://doi.org/10.1038/35106500.

[21] Broadbent, A., Fitzsimons, J. & Kashefi, E. Universal blind quantum computation. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science*, 517–526 (2009).

[22] Spring, J. B. *et al.* Boson sampling on a photonic chip. *Science* **339**, 798–801 (2013). URL https://science.sciencemag.org/content/339/6121/798. https://science.sciencemag.org/content/339/6121/798.full.pdf.

[23] Muralidharan, S. *et al.* Optimal architectures for long distance quantum communication. *Scientific Reports* **6**, 20463 (2016). URL https://doi.org/10.1038/srep20463.

[24] Jeffrey, E. R., Altepeter, J. B., Colci, M. & Kwiat, P. G. Optical implementation of quantum orienteering. *Phys. Rev. Lett.* **96**, 150503 (2006). URL https://link.aps.org/doi/10.1103/PhysRevLett.96.150503.

[25] Lundeen, J. S., Sutherland, B., Patel, A., Stewart, C. & Bamber, C. Direct measurement of the quantum wavefunction. *Nature* **474**, 188–191 (2011). URL https://doi.org/10.1038/nature10120.

[26] James, D. F. V., Kwiat, P. G., Munro, W. J. & White, A. G. Measurement of qubits. *Phys. Rev. A* **64**, 052312 (2001). URL https://link.aps.org/doi/10.1103/PhysRevA.64.052312.

[27] Altepeter, J. B. *Testing the Limits of Nonlocality*. Ph.D. thesis, University of Illinois at Urbana-Champaign (2000).

[28] Ferrie, C. & Blume-Kohout, R. Maximum likelihood quantum state tomography is inadmissible (2018). arXiv:1808.01072.

[29] Granade, C., Combes, J. & Cory, D. G. Practical Bayesian tomography. *New Journal of Physics* **18**, 033024 (2016). URL https://iopscience.iop.org/article/10.1088/1367-2630/18/3/033024.

[30] Maziero, J. Random sampling of quantum states: a survey of methods. *Brazilian Journal of Physics* **45**, 575–583 (2015). URL https://doi.org/10.1007/s13538-015-0367-2.

[31] Chapman, J. C., Graham, T. M., Zeitler, C. K., Bernstein, H. J. & Kwiat, P. G. Time-bin and polarization superdense teleportation for space applications. *Phys. Rev. Applied* **14**, 014044 (2020). URL https://link.aps.org/doi/10.1103/PhysRevApplied.14.014044.

[32] Einstein, A., Podolsky, B. & Rosen, N. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **47**, 777–780 (1935). URL https://link.aps.org/doi/10.1103/PhysRev.47.777.

[33] Bell, J. S. On the Einstein Podolsky Rosen paradox. *Physics Physique Fizika* **1**, 195–200 (1964). URL https://link.aps.org/doi/10.1103/PhysicsPhysiqueFizika.1.195.

[34] Clauser, J. F., Horne, M. A., Shimony, A. & Holt, R. A. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.* **23**, 880–884 (1969). URL `https://link.aps.org/doi/10.1103/PhysRevLett.23.880`.

[35] Shalm, L. K. *et al.* Strong loophole-free test of local realism. *Phys. Rev. Lett.* **115**, 250402 (2015). URL `https://link.aps.org/doi/10.1103/PhysRevLett.115.250402`.

[36] Giustina, M. *et al.* Significant-loophole-free test of bell's theorem with entangled photons. *Phys. Rev. Lett.* **115**, 250401 (2015). URL `https://link.aps.org/doi/10.1103/PhysRevLett.115.250401`.

[37] Hensen, B. *et al.* Loophole-free bell inequality violation using electron spins separated by 1.3 kilometres. *Nature* **526**, 682–686 (2015). URL `https://doi.org/10.1038/nature15759`.

[38] Lewenstein, M., Kraus, B., Cirac, J. I. & Horodecki, P. Optimization of entanglement witnesses. *Phys. Rev. A* **62**, 052310 (2000). URL `https://link.aps.org/doi/10.1103/PhysRevA.62.052310`.

[39] Wiseman, H. M. The two bell's theorems of john bell. *Journal of Physics A: Mathematical and Theoretical* **47**, 424001 (2014). URL `https://doi.org/10.1088/1751-8113/47/42/424001`.

[40] Orieux, A. *et al.* Experimental detection of steerability in bell local states with two measurement settings. *Journal of Optics* **20**, 044006 (2018). URL `https://doi.org/10.1088/2040-8986/aab031`.

[41] Brunner, N. & Gisin, N. Partial list of bipartite bell inequalities with four binary settings. *Physics Letters A* **372** (2007).

[42] Brunner, N., Navascués, M. & Vértesi, T. Dimension witnesses and quantum state discrimination. *Phys. Rev. Lett.* **110**, 150501 (2013). URL `https://link.aps.org/doi/10.1103/PhysRevLett.110.150501`.