Characterization of Quantum States of Light

by

Rob Adamson

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Physics
University of Toronto

Copyright © 2009 by Rob Adamson
Abstract

Characterization of Quantum States of Light

Rob Adamson
Doctor of Philosophy
Graduate Department of Physics
University of Toronto
2009

I present a series of experimental and theoretical advances in the field of quantum state estimation. Techniques for measuring the quantum state of light that were originally developed for distinguishable photons fail when the particles are indistinguishable. I develop new methods for handling indistinguishability in quantum state estimation. The technique I present provides the first complete description of states of experimentally indistinguishable photons. It allows me to derive the number of parameters needed to describe an arbitrary state and to quantify distinguishability. I demonstrate its use by applying it to the measurement of the quantum polarization state of two and three-photon systems.

State characterization is optimal when no redundant information is collected about the state of the system. I present the results of the first optimal characterization of the polarization state of a two-photon system. I show an improved estimation power over the previous state of the art. I also show how the optimal measurements lead to a new description of the quantum state in terms of a discrete Wigner function.

It is often desirable to describe the quantum state of a system in terms of properties that are not themselves quantum-mechanical observables. This usually requires a full characterization of the state followed by a calculation of the properties from the parameters characterizing the state. I apply a technique that allows such properties to be determined directly, without a full characterization of the state. This allows one such
property, the purity, to be determined in a single measurement, regardless of the size of
the system, while the conventional method of determining purity requires a number of
measurements that scales exponentially with the system size.
Acknowledgements

The last six years have been the most fulfilling, productive and intellectually exciting period of my life. I feel incredibly fortunate to have been able to spend my time doing something I love and to have worked with and learned from such talented, brilliant and generous people as the ones I have encountered at the University of Toronto. I won’t try to mention all of them, but there are a few who deserve my particular thanks.

Morgan Mitchell was an invaluable mentor to me in my first year of experimental research, when he patiently and thoughtfully answered my questions, taught me a thousand laboratory tricks and showed me by example what it took to make good science happen. Jeff Lundeen, Krister Shalm, Xingxing Xing and all the others on members of the Steinberg lab have been the best colleagues and collaborators one could hope for. Their perpetual curiosity, optimism and resourcefulness helped convince me that in the end things really could be made to work the way they were supposed to, given sufficient persistence. Finally I owe a huge debt of gratitude to Aephraim Steinberg. His way of approaching problems, of thinking through the consequences of solutions and of refusing to accept a complex answer when a simple one was at the heart of the issue are qualities that I will always try to emulate, although perhaps never as well as he manages to do. His leadership has kept our lab at the forefront of scientific research while still allowing each of us to pursue our own ideas and interests. I can only hope that someday I’ll be able to be as good a research supervisor as he has been.

To my parents I owe a lifetime of thanks. From the electronics and chemistry kits under the Christmas tree to my Dad’s patient help in building DC motors in the basement and the trips to the library with my Mom, they have always encouraged my interests in science without ever trying to force me along a certain path. I am especially grateful for their support in my decision to abandon my engineering career and head to grad school and for their unwavering support since then in good times and bad.

I also have to acknowledge the people who let me lead a fulfilling life outside the
MacLennan Physical Labs basement, particularly the Boyballz, the McGoeys, the Smiley Guys and the Buddies. Finally, I owe a huge debt of thanks to Cara who for the past three years has kept me sane, grounded and thoroughly entertained. She showed me that there’s at least one thing I love better than physics.
Preface

I came to the Steinberg group in 2002 after being voluntarily laid off from an engineering job in the telecom sector. In industry I had been fascinated by light as a tool for communications and as a physical phenomenon, but disappointed by the lack of curiosity in fundamental questions about how light worked among my engineering colleagues and at my own ignorance about the basic physics underlying what we were doing. This interest drew me to the University of Toronto where I quickly found what I was looking for. My master’s year was the most intellectually stimulating of my life, and by April I had developed a profound appreciation for the simplicity, ubiquity and elegance of quantum mechanics. Within it lay the unified description of nature that I had never been able to grasp in my undergraduate studies and in particular, a description of how light behaved at a fundamental level.

I was recruited for the Steinberg group by two PhD students, Jeff Lundeen and Kevin Resch, and when I started to understand the research in it I realized it was a perfect fit for me, a fusion of practical, hands-on optics applied to very deep and beautiful questions in quantum mechanics. Also at the lab was Morgan Mitchell, a post-doc who was and remains one of the best scientists I know. He showed me how to fix an argon-ion laser, align an interferometer and cool a Ti:Sapph laser with a motorcycle radiator. Throughout it all he demonstrated to me by example that with resourcefulness, patience, persistence and hard thinking about problems it was possible to produce beautiful, novel and important experiments, even in a place like the basement of McLennan Physical Labs.

With Morgan I started on a project to study quantum mechanical processes without doing full quantum process tomography. With his help I put together a laser diode-based spontaneous parametric downconversion source, detected the coincident radiation from it and set up motorized polarization analyzers for preparation and detection of polarization states. I remember the profound satisfaction and sense of wonder I experienced when I
observed for myself, for the first time, the violation of Bell’s inequalities.

Unfortunately the project got bogged down in both experimental and theoretical difficulties. We were using some inefficient numerical tools to simulate the experiment, and this caused the simulations to take upwards of 12 hours per run. Despite this the simulations showed us that for a particular process we were interested in, measuring the process’s effect on two input states would be enough to determine whether the process possessed a decoherence-free subspace. However we had no analytical understanding of this result and were unsure of the validity of some of the approximations that we used to make the numerical calculations tractable.

At the same time the experiment was having serious difficulties as well. While I was later able to separate out the systematic errors I experienced due to bad dielectric polarizing beamsplitters, inaccurate waveplates and misbehaving rotation stage motors, in 2004 I made the rookie mistake of trying to figure out what was happening by staring at a 256-element superoperator for a complex system instead of testing each component separately. In the end I could get close to what looked like the right results for some processes but wasn’t able to produce anything publishable.

Eventually we decided to publish our partial results in a SPIE conference proceeding and move on to other things. For me the other thing was tomography of indistinguishable photon states. Morgan and Jeff had published their work on making 3-N00N states in Nature that year, but while this work had some very nice interference curves, it hadn’t included a full density matrix measurement of the generated state, and we felt that it would be a worthwhile extension of the work to produce one. When Morgan initially tried to do this he realized that a straightforward application of quantum state tomography would be incapable of characterizing the most significant experimental error, namely the imperfect interference visibility. For a two-photon state it was pretty clear what the ‘missing’ bit of information was, namely the two-photon interference visibility, but it wasn’t at all clear how to incorporate it into a density matrix description or how to
extend it to the three-photon states we were really interested in. Morgan and I started playing around with simulations of the projective measurements we were doing and we soon saw the triplet/singlet structure emerge from the results. Meanwhile, Krister Shalm and I were working on rebuilding the Ti:Sapph and doing some two-photon density matrix characterizations. In a couple of months we had the data and the density matrices, but still had only a rough understanding of what they really meant. Only over two years of writing and revising the manuscript submission did we start to get a full handle on the physics of indistinguishable photon characterization. After that experiment we decided that we wanted to move on to three photons, but none of us had the slightest idea how to extend the successful two-photon characterization techniques to three photons.

I started working on the math and through much trial and error figured out how many elements the three-photon density matrix would have and what measurements we needed to do, but without really understanding what those measurements meant. It was clear to me that the problem was a group-theoretical one, and while I tried to learn enough group theory to work it out, I didn’t quite get there. Aephraim suggested I consult with David Rowe’s student Peter Turner who was good at the math, but new to quantum optics, and between the two of us we managed to muddle our way to the Schur-Weyl duality, the angular momentum structure and the arguments for how tomography should scale with the number of photons.

In the meantime, Krister decided to devote himself full-time to making three-photon states with the Ti:Sapph, while I took up a suggestion of Aephraim’s to use some of Todd Brun’s ideas to measure photon polarization purity in the two-photon setup. This work saw me setting up a Hong-Ou-Mandel interferometer for the first time and after four months of work I again got to experience an enormous sense of satisfaction and wonder when I came into the lab one morning to see the characteristic dip in the overnight scan of coincidence counts.

By the time I was wrapping up the Brun experiment, Krister, along with An-Ning
Zhang and Xingxing Xing, was starting to get some decent three-photon results from the Ti:Sapph setup. Aephraim, meanwhile, had become convinced by some of Polzik’s work that what we really needed to describe these states wasn’t a density matrix, but some kind of approximated spherical Wigner function. Krister and I were a little perplexed by the idea, but Aephraim kept pursuing it and eventually dug up some references to some mid-eighties work by G. S. Agarwal on creating Wigner functions for angular momentum states. As it happened, the data that Krister had been collecting for state tomography was exactly the same as that needed for these Wigner function which turned out to be even better than we’d thought they’d be. Instead of just being an approximated description as in Polzik’s work, they provided an exact description of the states we were making. Over the weekend after Aephraim sent a link to Agarwal’s paper, I made some animations of rotating spherical Wigner functions of N00N states. Playing with the Wigner functions also quickly made me realize that the N00N states were just one of a continuous family of states that we could make. With only minor changes to the setup it would be possible to also generate squeezed states, coherent states, Dicke states and so on. As it turned out, the analytical tools that Peter Turner and I had been developing meshed perfectly with this new description so that we could graphically represent entangled, highly non-classical single-mode polarization states in a mathematically rigorous way. Krister revised his experiment to try to make these other states as well and two years later we had some beautiful direct measurements of Wigner functions to show off at conferences.

While all this was going on I was also collaborating with Ben Fortescue from Hoi-Kwong Lo’s group, on an idea of Hoi-Kwong’s to simulate GHZ entanglement in a cryptography system by employing classical randomness and two-photon entanglement. The idea was elegant and seemed like an interesting thing to try out. It meant adding liquid crystal waveplates to my setup and lowering the accidental coincidence rate by improving our electronics, but these were both good long-term investments of time and gave me a
chance to do some fun engineering work. By spring 2006 I had collected enough data
to generate 32 bits of random key which I proudly presented at CLEO. Ben and I had
realized that there was a gaping loophole in the security of the protocol having to do
with information encoded on unpaired photons and we came up with a neat way of fixing
it by swamping the channel with randomly polarized light. Before we could implement
that idea, though, Aephraim and I decided that we needed to act on a collaboration that
we’d started with Mahsoud Mohseni on a kind of entanglement-assisted process tomog-
raphy. I worked on this throughout late 2006 and early 2007 and presented the results
at conferences in the summer. One of the attendees at one of these conferences quietly
informed me after my talk that I’d been scooped by Guo’s group in a Phys. Rev. A
article published in April of that year.

After a week of self-recrimination and trying to figure out what to do next I decided
to pursue some ideas Aephraim and I had discussed with Andrei Klimov about doing
quantum state tomography using mutually-unbiased bases. This experiment, which I
decided would be my last one before writing up, I managed to finish (without getting
scooped!) by May of 2008, and, after a hectic summer of conferences, thesis-writing, data
analysis and paper-writing I had my committee defense on August 13th.

All the work detailed here has been published in or submitted to peer-reviewed jour-
nals except the scooped work on direct quantum process tomography and the GHZ-
cryptography experiment, none of which form part of this thesis. The GHZ-cryptography
experiment I still hope to complete, probably with some help from other group members.

As I was finishing my corrections to this thesis, I learned that the Wigner function
tomography work has been accepted to Nature. After nearly four years of work on the
theory and experimental production of these states, this is the best conclusion that I
could possibly have hoped for. My thanks go out to Krister, Aephraim, Morgan and
Peter for the hard work, good ideas and cooperation that got us to this point.
List of Publications

Published


R. B. A. Adamson, L. K. Shalm, and A. M. Steinberg, Preparation of pure and mixed polarization qubits and the direct measurement of figures of merit, Physical Review A (Atomic, Molecular, and Optical Physics) 75 (2007), no. 1, 012104


Morgan W. Mitchell, Christopher W. Ellenor, Robert B. A. Adamson, Jeff S. Lundeen, and Aephraim M. Steinberg, Quantum process tomography and the search for decoherence-free subspaces, Proc. SPIE 5436 (2004), 223

Accepted


Submitted

# Contents

## 1 Introduction

1.1 Quantum Information ........................................ 1

1.2 Concepts .................................................... 5

  1.2.1 The density matrix ..................................... 5

  1.2.2 Qubits .................................................. 7

  1.2.3 The Bloch/Poincaré sphere .............................. 9

  1.2.4 Qubit transformations .................................. 10

  1.2.5 Multi-qubit states ..................................... 11

  1.2.6 The Bell/CHSH inequalities ............................ 14

  1.2.7 Quantum measurement .................................. 15

  1.2.8 Two-photon interference ................................ 16

  1.2.9 Bell state filtering .................................... 17

1.3 Summary ..................................................... 18

## 2 Methods ...................................................... 20

2.1 Introduction .................................................. 20

2.2 Quantum state tomography .................................. 20

  2.2.1 Properties of the density matrix ....................... 23

  2.2.2 Measuring the density matrix ......................... 23

  2.2.3 The positivity constraint .............................. 30
2.2.4 Maximum-likelihood estimation ........................................ 31
2.2.5 An example: Two-photon polarization density matrix ............. 34
2.3 Experimental state estimation ........................................... 36
  2.3.1 Spontaneous parametric downconversion ............................ 37
  2.3.2 Polarization-entangled SPDC sources ............................... 41
  2.3.3 Real-world sources ................................................. 42
  2.3.4 Type-I sandwich source ............................................ 43
  2.3.5 Source output ..................................................... 51
  2.3.6 Type-II beam-like emission source ............................... 51
2.4 Photon detection .......................................................... 53
  2.4.1 Coincidence detection ............................................... 55
2.5 Polarization manipulation ............................................... 57
  2.5.1 Polarization projection .............................................. 57
2.6 Polarization rotators ..................................................... 59
  2.6.1 Waveplates ......................................................... 60
  2.6.2 Motorized rotation stages ......................................... 60
  2.6.3 Calibration .......................................................... 62
  2.6.4 Liquid crystal waveplates ......................................... 65
  2.6.5 Liquid crystal waveplate driver ................................... 67
2.7 Polarization analysis ..................................................... 68
2.8 Two-photon interference ............................................... 71
  2.8.1 Alignment .......................................................... 71
  2.8.2 Dips and antidips .................................................. 74
  2.8.3 Crystal walkoff .................................................... 75
  2.8.4 Type-II apparatus .................................................. 78
2.9 Summary ................................................................. 79
3 Hidden differences

3.1 Two-mode systems ........................................... 82
  3.1.1 Characterizing two-mode states of light ............... 83
  3.1.2 Symmetric measurements ................................ 86
  3.1.3 Group theory .......................................... 87

3.2 Hidden information ........................................... 100
  3.2.1 Structure of the accessible density matrix .......... 102
  3.2.2 Implications for distinguishability .................. 104

3.3 Experimental measurements .................................. 106

3.4 The two-photon experimental polarization density matrix ........................................... 108
  3.4.1 Measuring the two-photon experimental density matrix ........................................... 111
  3.4.2 Measuring the three-photon accessible density matrix ........................................... 118

3.5 Summary and conclusions ................................... 122

4 QST with MUBs .................................................. 124

4.1 Introduction ............................................... 124

4.2 Comparing different tomography strategies ................. 128

4.3 Convergence of the estimate towards the true density matrix ........................................... 129

4.4 The choice of measurement bases ............................. 133

4.5 Constructing MUBs for two qubits ........................... 138

4.6 Experiment ................................................ 139
  4.6.1 Separable measurements ................................. 139
  4.6.2 MUBs measurements ..................................... 141
  4.6.3 Analysis ............................................... 144
  4.6.4 Results ............................................... 144

4.7 The discrete Wigner function for two qubits ............... 148
  4.7.1 Finite fields ............................... 149
  4.7.2 Discrete Wigner functions ............................. 151
### 4.7.3 Reconstruction

152

### 4.8 Summary

153

### 5 Measuring figures of merit

5.1 Introduction 155

5.2 Multiqubit measurements 158

5.3 Experimental techniques 162

5.3.1 Singlet state projection 162

5.3.2 Making impure states 165

5.4 Results 168

5.4.1 Density matrices 168

5.5 Discussion 172

### 6 Conclusions

6.1 Major results 175

6.2 Future outlook 176

**Bibliography**

178

**A Electronic circuit designs**

193
Chapter 1

Introduction

Information is physical.
– Rolf Landauer

1.1 Quantum Information

Quantum mechanics is not only a theory of physical systems – of subatomic particles, atoms, molecules, crystalline solids and so on – it is also, and perhaps primarily, a theory of information. Atoms and molecules, when static, store information. When engaged in physical processes they can be thought of as performing a computation. Moreover, it is possible to abstract away the information-theoretic part of quantum mechanics from its embodiment in particular physical systems, to talk of quantum bits and unitary logic gates instead of atoms and Hamiltonians. Some researchers even hope that the postulates of quantum mechanics can be entirely replaced by purely information-theoretic statements about the fundamental limits placed by Nature on our ability to encode, decode and transmit information[45].

This reinvention of quantum mechanics as a theory of information has led to some startling discoveries in the past twenty years. It has revealed that computers that store and manipulate quantum systems can use exponentially fewer computational resources
than ordinary computers[106], even in the presence of errors[36]. It has shown us that information stored in quantum systems, unlike ordinary information, cannot be copied. It can, however, be ‘teleported’ from one system to another, through processes that destroy the information in the original system and transfer it to the other[11]. Furthermore, the peculiar way that quantum states can be perfectly correlated while being perfectly random permits completely secure communications, something that is viewed as impossible in classical information theory[13].

At the same time, our ability to manipulate, control and measure all aspects of quantum systems has improved dramatically. Single atoms can now be reliably trapped, manipulated and interacted[54]. Single photons can be generated at will, either alone or in highly correlated multi-photon states[68]. Single atoms can be made to interact with single photons with exquisite control[83]. Large, many-body systems can be created and interacted coherently with a high degree of tunability in the system parameters[15]. These advances have permitted the development of an experimental branch of quantum information science and to a great many proof-of-concept demonstrations of the main ideas of quantum information theory and foundational quantum mechanics. These landmark experiments include the first unambiguous violation of Bell’s inequalities[7], the first demonstration of secure quantum cryptography[13], the first demonstrations of teleportation[22, 90] and the demonstration of quantum logic gates[92] and simple quantum algorithms[29, 71]. While these experiments have given us concrete examples of quantum information at work, they have also created a need for a fuller understanding of how best to connect the information gleaned from quantum measurements to a full description of the quantum state. It is the main aim of this thesis to contribute to this important area of research.

The interaction between a physical system and the information it contains is mediated via measurement. Measurement has always been a deep concern in quantum mechanics, assuming a leading role in Heisenberg’s uncertainty principle, the postulates of quantum
mechanics and important thought experiments such as the EPR experiment\cite{41}, Bell’s inequalities\cite{10} and the Kochen-Specker theorem\cite{63} which formalize the most irreconcilable differences between the classical and quantum world. In quantum mechanics, measurement is disconnected from the underlying state of the system in a way that it is not in classical mechanics. As Niels Bohr put it,

On the one hand, the definition of a physical system, as ordinarily understood, claims the elimination of all external disturbances. But in that case, according to the quantum postulate, any observation will be impossible, and, above all, the concepts of space and time lose their immediate sense. On the other hand, if in order to make observation possible we permit certain interactions with suitable agencies of measurement, not belonging to the system, an unambiguous definition of the state of the system is naturally no longer possible, and there can be no question of causality in the ordinary sense of the word. The very nature of the quantum theory thus forces us to regard the space-time coordination and the claim of causality, the union of which characterizes the classical theories, as complementary but exclusive features of the description, symbolizing the idealization of observation and definition respectively\cite{19}.

Despite this impossibility of simultaneously knowing everything about a given quantum system, quantum systems can, nevertheless, be fully characterized if enough identical copies are made of them. This characterization is conceptually similar to the way that a large number of systems obeying classical mechanics with unknown individual properties can be characterized by a probability distribution. In fact, it is becoming an increasingly popular viewpoint among foundational researchers that, at its heart, quantum mechanics is a statistical mechanical theory of something\footnote{This is one way of describing the so-called epistemic view of quantum states. This viewpoint holds that quantum states are fundamentally states of knowledge rather than states of reality.}, although what that something is remains
elusive[45, 109].

While this thesis does not provide any answers that will revolutionize foundational quantum mechanics, it does take seriously the notion that developing methods for characterizing quantum states made in the lab can deliver insight into what quantum states really are. To this end we attempt to uncover the underlying symmetries in the description of quantum states. We look at how this description changes when the individual identity of particles becomes distinguishable or indistinguishable to experimental measurements. Finally, we ask what information it is possible to gain from a state directly, without requiring that all the information about a state be available. These investigations into the art and practice of quantum state estimation form the core of this thesis.

The remainder of this chapter will introduce some fundamental concepts that are essential to understanding the original work presented in this thesis. The basic theoretical and experimental methodology employed in state creation and state characterization will be laid out in Chapter 2. After that we will begin examining how quantum state estimation can be extended, improved and understood in a variety of contexts.

While in some quantum systems, estimating the quantum state is straightforward, in others, the effects of interactions, indistinguishability, and unaccounted degrees of freedom can make the question ‘What is the quantum state of my system?’ an extremely thorny one to even phrase in terms of clear experimental observables, let alone answer. Chapter 3 of this thesis examines just such a situation where the indistinguishability of photons makes their quantum state impossible to describe using the standard methods of quantum state estimation which rely on the treatment of each particle as a distinct entity. In contrast, by concentrating on the observables which it is possible to measure, we arrive at a complete, elegant and scalable method of characterizing these states.

Chapter 4 examines how to structure a set of measurements so as to maximize the information extracted with each one, leading to optimal state estimation on a fixed number of copies of the state. This optimality is deeply connected to the geometry of
Chapter 1. Introduction

the Hilbert space of measurements and helps to relate that geometry to operationally relevant parameters. Additionally, this optimal set of measurements is intimately tied to a description of quantum states on a discrete phase-space instead of a Hilbert space. This new description has several useful and intuitive properties that can provide insight into the structure of quantum correlations.

Chapter 5 looks at how some properties of states that are usually thought to require a lengthy and complete characterization of the state can be obtained directly through a judicious choice of measurements. This will be an important technology as quantum systems become larger and the exponential scaling of the complexity of state estimation with the size of the system makes full state characterization technically infeasible.

1.2 Concepts

1.2.1 The density matrix

While one often uses state vectors to describe the quantum state of a system in terms of its wavefunction, in experimental work (and in many other situations) it is usually preferable to describe the state using the more general density matrix formulation. The density matrix or density operator is a linear, Hermitian operator on the Hilbert space of wavefunctions. One way to think of it is as a probability distribution over projectors onto different wavefunctions. That is to say

\[ \rho = \sum_i p_i |\psi_i\rangle \langle \psi_i| \]  

(1.1)

where the \( p_i \) are probabilities, i.e. real numbers on \([0, 1]\) such that \( \sum_i p_i = 1 \). The non-negativity of the \( p_i \) implies that the eigenvalues of \( \rho \) are non-negative and, as a consequence, that for any column \( x \) and row \( y \), \( |\rho_{xy}| = |\rho_{yx}| \leq \sqrt{\rho_{xx}\rho_{yy}} \). The off-diagonal elements for which \( x \neq y \) are called coherences while the diagonal elements are called populations.
While the density matrix can be rotated to any basis by applying unitary operations, we usually choose to express $\rho$ in a preferred basis called the computational basis. The computational basis has the property that all of the basis states are separable, that is to say they can be written as a tensor product of single-particle states. This is not true for all possible density matrices as will be seen in the section on entanglement.

When the magnitudes of all coherences take on their maximum values, i.e. when $|\rho_{yx}| = \sqrt{\rho_{xx}\rho_{yy}}$ for all $x$ and $y$, the state is said to be pure. Only one of the $p_i$ is non-zero, and the density matrix is a projector onto a single state-vector.

If all of the coherences are zero for a density matrix written in the computational basis, the state can be thought of as a classical mixture and its measurement statistics are governed solely by ordinary probability theory applied to the single-particle measurement outcomes.

A maximally-mixed state is a density matrix proportional to the identity operator with all of its coherences zero and all of its populations equal.

The purity is a useful measure of whether the state behaves more like a classical statistical mixture or more like a pure quantum state. The purity is defined as

$$P = \text{Tr} \left[ \rho^2 \right].$$

This measure is invariant under unitary operations.

If $\hat{O}$ is a Hermitian operator on the Hilbert space of states then the expectation value of the operator for a system in the state $\rho$ is given by

$$\langle \hat{O} \rangle = \text{Tr} \left[ \rho \hat{O} \right].$$

The advantage of the density matrix description is that it is capable of describing both the statistics of quantum states and quantum measurement and the classical statistics induced by experimental randomness. This makes it the most appropriate description for experimentally generated states. The density matrix also has the property that, because
it is a Hermitian operator, it is itself an observable. This makes it possible to reconstruct the density matrix from measurements.

Like a probability distribution, a density matrix is a statistical description of a quantum state. Depending on one’s preferred philosophy one can view it as being ‘really’ a description of the frequency of observation of certain measurement outcomes, a state of knowledge about the outcome of such measurements, or a description of reality for a subsystem of a larger, pure, system. These interpretational differences make identical predictions for the outcome of experiments, of course, but one or the other may be more convenient in understanding the density matrix in a particular context.

1.2.2 Qubits

One of the deepest insights of classical computer science is the realization that all information-carrying systems are formally equivalent to a binary system of ones and zeros called bits. This insight allowed the theory of redundancy, data compression and informational entropy to be formulated completely independent of the physical system carrying the information.

Some of the earliest results in quantum information theory have to do with the analogue of data compression[100] for quantum information. It was in developing this theory that Schumacher first coined the term qubit as a contraction of quantum bit to describe the smallest quantity of quantum information[91]. Ever since, the qubit has been used as an abstraction capable of distilling the information-theoretic essence from a given physical quantum system.

A qubit is a quantum two-level system described by a state vector in a two-dimensional Hilbert space. In the quantum information literature it is typical to take the basis states to be $|0\rangle$ and $|1\rangle$. The experiments discussed in this dissertation all involve a particular implementation of the qubit, namely the polarization state of a single photon. Since this is the only type of qubit we will be discussing, we adopt the conventional notation of
using $|H\rangle$ and $|V\rangle$, the horizontal and vertical polarization states as the basis states for our qubit\(^2\). The beauty of the qubit concept, though, is that any information-theoretic development realized in one physical system like polarization is immediately applicable to all other physical qubits like the spin state of a trapped ion, the direction of current in a superconducting loop, the magnetic moment of a hydrogen atom in nuclear magnetic resonance, the excitation state of a neutral atom in a shallow potential lattice or the spin state of an electron trapped on a quantum dot.

It is useful to label particular superpositions of the $|H\rangle$ and $|V\rangle$ basis states which will come up frequently in work with polarization. Following the conventions for Jones vectors\(^{[56]}\) (and as used in \([55]\)), we define

\[
\text{Diagonal } |D\rangle \equiv \frac{1}{\sqrt{2}} (|H\rangle + |V\rangle) \tag{1.4}
\]
\[
\text{Anti-diagonal } |A\rangle \equiv \frac{1}{\sqrt{2}} (|H\rangle - |V\rangle) \tag{1.5}
\]
\[
\text{Left circular } |L\rangle \equiv \frac{1}{\sqrt{2}} (|H\rangle + i|V\rangle) \tag{1.6}
\]
\[
\text{Right circular } |R\rangle \equiv \frac{1}{\sqrt{2}} (|H\rangle - i|V\rangle). \tag{1.7}
\]

Sometimes in the quantum information literature $|L\rangle$ and $|R\rangle$ will have opposite signs to those given here.

Along with $|H\rangle$ and $|V\rangle$, these states are eigenstates of the Pauli operators defined as

\[
\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.8}
\]

$|H\rangle$ and $|V\rangle$ are the +1 and −1 eigenstates of $\sigma_z$. $|D\rangle$ and $|A\rangle$ are the +1 and −1 eigenstates of $\sigma_x$. $|L\rangle$ and $|R\rangle$ are the +1 and −1 eigenstates of $\sigma_y$.\(^{[56]}\)

---

\(^2\)The treatment of polarization as a vector in a two-dimensional Hilbert space dates back to Jones calculus invented by R. C. Jones in 1941\(^{[56]}\).
1.2.3 The Bloch/Poincaré sphere

Since the state of a qubit must be normalized, a convenient way of writing it is

\[ \cos \theta \left| H \right\rangle + e^{i\phi} \sin \theta \left| V \right\rangle \]  

(1.9)

This parameterization of the qubit in terms of angles \( \theta \) and \( \phi \) is suggestive. If we make the mapping

\[ x = \sin 2\theta \cos \phi \]  

(1.10)

\[ y = \sin 2\theta \sin \phi \]  

(1.11)

\[ z = \cos 2\theta \]  

(1.12)

then by varying \( \theta \) on \([0, \pi/2]\) and \( \phi \) on \([0, 2\pi]\) this fully parameterizes a unit sphere called the Bloch sphere. This sphere provides a convenient visual representation of the qubit. Points that are antipodal on the Bloch sphere represent orthogonal states of the qubit. Overlaps between states can be calculated solely from the relative angle between the two corresponding points on the Bloch sphere.

Mixed states can also be represented in this description; they are points on the interior of the Bloch sphere. Let \( \rho \) be a single-qubit density matrix

\[ \rho = \frac{1}{2} \begin{pmatrix} 1 + z & x - iy \\ x + iy & 1 - z \end{pmatrix} \]  

(1.13)

It follows from the positivity constraint on density matrices that \( \det \rho = \frac{1}{4} (1 - |r|^2) \geq 0 \) where \( r \) is the real-space vector \((x, y, z)\). Any point satisfying this inequality, namely points on the surface and interior of the Bloch sphere, represents a valid qubit density matrix.

Before there was the Bloch sphere there was the Poincaré sphere. Invented in 1891 by Henri Poincaré, the Poincaré sphere represents classical polarizations in exactly the same way that the Bloch sphere represents qubits. There is a difference of convention
between the two descriptions. The north pole of the Poincaré sphere represents left-
circular polarization (i.e. the $\sigma_y +1$ eigenstate) whereas the north pole of the Bloch
sphere represents $|H\rangle$ (i.e. the $\sigma_z +1$ eigenstate). The Poincaré sphere has the convenient
feature that all linear polarization states (i.e. those with a real-valued coherence between
$|H\rangle$ and $|V\rangle$) are located in the equatorial plane. In this dissertation we will primarily
make use of the Poincaré sphere description of polarization states since it is more natural
for polarization. To change to the Bloch sphere picture the reader need only rotate his
head 90° to the right.

1.2.4 Qubit transformations

The class of transformations that can be applied to a qubit without changing its purity
form a representation of the group $SU(2)$. They are most easily pictured as being ro-
tations on the Bloch/Poincaré sphere. For this reason (and for brevity), we often speak
of polarization rotations as including the full range of $SU(2)$ transformations, not just
rotations of linear polarizations.

For polarization, any $SU(2)$ transformation can be made by making a series of dif-
ferent phase delays about different axes. A phase delay of $\phi$ about an axis at angle $\theta$
rotates a point on the Poincaré sphere through the angle $\phi$ about an axis on the equa-
torial plane making an angle of $2\theta$ with the H/V axis. One popular method of making
arbitrary polarization transformations uses waveplates. These are thin slices of birefrin-
gent material meant to impart a fixed phase delay $\phi$ at a variable angle $\theta$. Typically,
half waveplates with $\phi = \pi$ and quarter waveplates with $\phi = \pi/2$ are used, but one
occasionally encounters plates with $\phi = 2\pi$ and $\phi = \pi/4$.

A half waveplate can take any linear polarization to a different linear polarization since
a rotation of a ray on the equatorial plane about another ray on the equatorial plane
by an angle $\pi$ will result in a ray on the equatorial plane. Quarter waveplates rotate
by $\pi/2$ and so can take a linear polarization into an elliptical polarization anywhere on
the hemisphere of states whose long axis is polarized in the same direction as the initial linear polarization.

A quarter waveplate and half-waveplate together can take a linear polarization to any point on the Poincaré sphere. A quarter waveplate, half waveplate and quarter waveplate can perform an arbitrary rotation that takes any point on the Poincaré sphere to any other point on the Poincaré sphere.

While this is useful in principle, it can be difficult to enact in practice since the parameters under control, the value of $\theta$ for each waveplate, are tightly coupled, making the system very hard to fine-tune except for some very specific sets of angles. A much more experimentally convenient polarization controller is one that allows $\theta$ to be fixed and a variable $\phi$ to be applied. This is exactly the situation with liquid crystal variable waveplates (LCWPs), which will be discussed in detail in the next chapter.

1.2.5 Multi-qubit states

The Hilbert space of two qubits is spanned by the tensor products of the basis states of the individual qubits, namely $|HH\rangle$, $|HV\rangle$, $|VH\rangle$, $|VV\rangle$. The density matrix has $4 \times 4 = 16$ elements. While unitary transformations can be applied to the two qubits individually there is also a class of measurements called entangling operations that cannot be separated into actions on the individual qubits.

Entanglement

If two separated bodies, about which, individually, we have maximal knowledge, come into a situation in which they influence one another and then again separate themselves, then there regularly arises that which I just called entanglement of our knowledge of the two bodies. At the outset, the joint catalogue of expectations consists of a logical sum of the individual catalogues; during the process the joint catalogue develops necessarily according to the
known law... Our knowledge remains maximal, but at the end, if the bodies have again separated themselves, that knowledge does not again decompose into a logical sum of the knowledge of the individual bodies[103].

–Erwin Schrödinger

Entanglement is viewed by many physicists as the strangest aspect of quantum mechanics. It arises when the density matrix of a system of two or more particles is non-separable. That is to say that the system cannot be described by specifying the properties of the individual particles within it. In a sense, the information about the state is contained not in the individual particles’ properties, but in the correlations between those properties. Entanglement is one of the main properties that makes quantum information different from classical information.

For example, a state like the following two-photon polarization state

\[
|\phi^+\rangle = \frac{1}{\sqrt{2}} (|HH\rangle + |VV\rangle) \tag{1.14}
\]

\[
= \frac{1}{\sqrt{2}} (|DD\rangle + |AA\rangle) \tag{1.15}
\]

\[
= \frac{1}{\sqrt{2}} (|RL\rangle + |LR\rangle) \tag{1.16}
\]

is entangled. Each photon individually has an equal probability of being horizontally polarized or vertically polarized, left or right circularly polarized or diagonally or anti-diagonally polarized. In classical polarization theory a beam of light having this property would be considered unpolarized. The quantum state, though, also has correlations. The two photons in the state will have the same polarization when measured in the horizontal/vertical and diagonal/anti-diagonal bases and will have opposite polarizations in the left and right circular bases. It was shown by John Bell in 1964[10] that the randomness of individual particle measurements coupled with the correlations between them in all multiple bases is inconsistent with classical probability theory and local realism.
Chapter 1. Introduction

$|\phi^+\rangle$ is one of four maximally entangled states called Bell states that form a basis for two-photon polarization states. Reference will often be made to these states in this thesis, so we list them here:

$$|\phi^+\rangle = \frac{1}{\sqrt{2}} (|HH\rangle + |VV\rangle) = \frac{1}{\sqrt{2}} (|DD\rangle + |AA\rangle) = \frac{1}{\sqrt{2}} (|RL\rangle + |LR\rangle) \quad (1.17)$$

$$|\phi^-\rangle = \frac{1}{\sqrt{2}} (|HH\rangle - |VV\rangle) = \frac{1}{\sqrt{2}} (|DA\rangle + |AD\rangle) = \frac{1}{\sqrt{2}} (|RR\rangle - |LL\rangle) \quad (1.18)$$

$$|\psi^+\rangle = \frac{1}{\sqrt{2}} (|HV\rangle + |VH\rangle) = \frac{1}{\sqrt{2}} (|DD\rangle - |AA\rangle) = \frac{1}{\sqrt{2}} (|RR\rangle + |LL\rangle) \quad (1.19)$$

$$|\psi^-\rangle = \frac{1}{\sqrt{2}} (|HV\rangle - |VH\rangle) = \frac{1}{\sqrt{2}} (|DA\rangle - |AD\rangle) = \frac{1}{\sqrt{2}} (|RL\rangle - |LR\rangle) \quad (1.20)$$

$|\psi^-\rangle$ is sometimes called the singlet state and $|\psi^+\rangle$ the triplet state.

From a mathematical point of view, entanglement is an expression of the impossibility of factoring the density matrix on the full Hilbert space of the joint system into a product of density operators on the subsystems.

While the problem of characterizing entanglement for states of more than two particles is complex, quantification of the entanglement of bipartite systems is well-understood. If the state of the whole system is pure, then one may characterize the degree of entanglement by performing a partial trace over one of the particles and measuring the von Neumann entropy of the density matrix for the remaining subsystem[12]. The von Neumann entropy is defined as

$$S(\rho) = -\text{Tr}\rho \log \rho \quad (1.21)$$

It is the natural extension of the Shannon entropy to quantum states.

For a density matrix $\rho = \sum p_i |\psi_i\rangle \langle \psi_i|$ the natural definition of entanglement is the average value of the entanglement of the $|\psi_i\rangle$ weighted by the probabilities $p_i$. Unfortunately, the decomposition of $\rho$ into $|\psi_i\rangle \langle \psi_i|$ is not unique, and the choice of basis $|\psi_i\rangle$ will have an effect on the average value of entanglement. The most sensible definition of entanglement is therefore the minimum value of the average entanglement over all such decompositions. As Wootters and Hill have shown[52], this minimum is a simple
function of the *concurrency*. The concurrence can be obtained by first calculating the matrix \( R = \rho \Sigma \rho^T \Sigma \) [55] where

\[
\Sigma = \begin{pmatrix}
0 & 0 & 0 & -1 \\
0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 \\
-1 & 0 & 0 & 0
\end{pmatrix}
\] (1.22)

The concurrence \( C(\rho) \) can then be obtained from the ordered eigenvalues of \( R \), \( \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 \) as

\[
C(\rho) \equiv \max \left( 0, \sqrt{\lambda_1} - \sqrt{\lambda_2} - \sqrt{\lambda_3} - \sqrt{\lambda_4} \right)
\] (1.23)

\[
(1.24)
\]

The entanglement of \( \rho \) is usually called the *entanglement of formation* and it can be expressed as

\[
E(\rho) = H\left( \frac{1}{2} + \frac{1}{2} \sqrt{1 - C(\rho)^2} \right)
\] (1.25)

where \( H(x) \) is the binary entropy function \( H(x) = - [x \log_2 x + (1 - x) \log_2 (1 - x)] \).

### 1.2.6 The Bell/CHSH inequalities

Following John Bell’s derivation of inequalities violated by quantum mechanics and inconsistent with local realism [10], Clauser, Horne, Shimony, and Holt produced a simpler version of them [30] suitable for testing in an optics experiment [44]. We imagine two photons being distributed to two parties, one of whom measures the counts at two output ports of a polarizing beamsplitter (PBS) at angle \( \alpha \) and the other at the two output ports of a PBS at angle \( \beta \). We define the joint polarization correlation visibility as

\[
V(\alpha, \beta) = \frac{N_{++} + N_{--} - N_{+-} - N_{-+}}{N_{++} + N_{--} + N_{+-} + N_{-+}}
\] (1.26)

where we have labeled the output ports of the PBSs as + and − and \( N_{+-} \), for instance, is the number of times that the first party sees a photon at the + port and the second party sees a photon at the − port.
We construct the CHSH $S$-function as

$$S(\alpha_1, \beta_1, \alpha_2, \beta_2) = V(\alpha_1, \beta_1) - V(\alpha_1, \beta_2) + V(\alpha_2, \beta_1) + V(\alpha_2, \beta_2) \quad (1.27)$$

It can be shown that no local, realistic hidden-variable theory will predict a value for $|S| > 2$ but for $\alpha_1 = 0^\circ$, $\alpha_2 = 45^\circ$, $\beta_1 = 22.5^\circ$, and $\beta_2 = 67.5^\circ$, quantum mechanics predicts a value of $S = 2\sqrt{2}$ when the input state is $|\phi^+\rangle$.

The quantity $S$ and the number of standard deviations by which it exceeds the classical limit of 2 is sometimes used as a measure of the quality of a source of entangled photons\[69\]. Violation of some form of Bell’s inequalities are sufficient to demonstrate entanglement, but not all entangled states are capable of violating Bell’s inequalities. Consequently, the violation of Bell’s inequalities is a more stringent test of entanglement than simply demonstrating a value of concurrence greater than zero.

### 1.2.7 Quantum measurement

Traditionally, measurements in quantum mechanics have been described by projectors. These are orthogonal, idempotent, Hermitian operators. A projective value measure or PVM is a set of projectors that sum to unity,

$$\sum_i \hat{P}_i = \mathbb{I} \quad (1.28)$$

We call the individual projectors in a PVM elements of the PVM. Sometimes we will use the term *basis* interchangeably with PVM since by virtue of their orthogonality the PVM elements, when of rank one, form a basis for the projective Hilbert space.

PVMs are a special case of a larger class of measurements called positive-operator valued measures (POVMs)\[91\] which have been of great importance in theoretical quantum information theory\[45\]. POVMs can be implemented by first interacting the system under study with a larger quantum system and then doing a projective measurement on the larger system. For a variety of reasons, many argue that POVMs represent a more
fundamental measurement primitive than do PVMs[45]. That said, for experimental simplicity, for the experiments described in this thesis we will restrict ourselves to PVMs or, occasionally, to measurement operators that are convex sums over PVMs.

1.2.8 Two-photon interference

When two photons are incident on the input ports of a non-polarizing beamsplitter, an interesting interference effect called Hong-Ou-Mandel(HOM) interference takes place[53]. If we label the input modes of the beamsplitter 1 and 2 and the output modes 3 and 4, the action of the beamsplitter is to map modes 1 and 2 onto modes 3 and 4 as follows:

\begin{align*}
a_1^{\dagger} & \rightarrow \frac{1}{\sqrt{2}} \left( a_3^{\dagger} + a_4^{\dagger} \right) \\
a_2^{\dagger} & \rightarrow \frac{1}{\sqrt{2}} \left( a_3^{\dagger} - a_4^{\dagger} \right)
\end{align*}

If we put one photon in mode 1 and another in mode 2 at the same moment in time, then the beamsplitter will map the two photon state \(a_1^{\dagger}a_2^{\dagger}\) onto

\[
a_1^{\dagger}a_2^{\dagger} \rightarrow \frac{1}{2} \left( a_3^{\dagger} + a_4^{\dagger} \right) \left( a_3^{\dagger} - a_4^{\dagger} \right) = \frac{1}{2} \left( a_3^{\dagger}a_3^{\dagger} - a_4^{\dagger}a_4^{\dagger} \right)
\]

Through destructive interference, the probability amplitude for a photon to be in mode 3 and another in mode 4 has disappeared. This interference effect cannot be thought of classically in terms of interference of the electromagnetic field, but it does have an interpretation in terms of Feynman paths. Since the events where both photons are reflected at the beamsplitter and the events where both photons are transmitted have indistinguishable outcomes, we expect them to interfere. That the interference is destructive is a consequence of the quantum statistical nature of photons as bosons. Were the same experiment to be done with fermions, the particles would leave out opposite ports all the time and never out the same port.
While Hong-Ou-Mandel interference is a simple consequence of linear quantum field theory, it appears very much like an interaction. Two photons meet and seem to ‘stick together’ after meeting. Indeed, it can be shown that this effect can be used to mediate ‘effective interactions’ that can provide the basis for a quantum computing scheme called *linear optics quantum computing* [62].

Two-photon interference is used in this thesis in a closely related way. We use it to make projective measurements onto Bell states [123]. While this method of performing Bell-state measurements works, it has been proven that it is impossible to perform a complete Bell-state PVM deterministically using only linear-optics [80]. For many quantum information applications, though, even a partial Bell-state PVM capable of distinguishing only one or two of the Bell states is still a very useful experimental tool.

### 1.2.9 Bell state filtering

Quantum mechanics allows for a multiparticle system to be measured in an entangled basis. This is called an entangling measurement or a Bell-state measurement (BSM). BSMs were crucial to demonstrating the quantum teleportation protocol [22] and other quantum protocols [26]. They have the interesting property that they determine correlations between single-particle states without collapsing those states individually. This means that with entangling measurements, correlations can be probed in multiple bases at once. The work in Chapter 4 and Chapter 5 relies on entangling measurements to improve how information is extracted from quantum systems.

In the previous section we implicitly assumed that the two input photons had the same polarization. Let’s now assume that the input photons are in some arbitrary quantum polarization state $|\psi\rangle = \left( \alpha a_{H1}^\dagger a_{H2}^\dagger + \beta a_{H1}^\dagger a_{V2}^\dagger + \gamma a_{V1}^\dagger a_{H2}^\dagger + \delta a_{V1}^\dagger a_{V2}^\dagger \right) |\text{vac}\rangle$ where $\alpha$, $\beta$, $\gamma$ and $\delta$ are arbitrary complex constants selected so that the state is normalized.

The two photon interference will map the spatial modes in the same way as before to
produce an output state

\[ |\psi\rangle = \left[ \alpha \left( a_{H3}^\dagger a_{H3}^\dagger - a_{H4}^\dagger a_{H4}^\dagger \right) + \beta \left( a_{H3}^\dagger a_{V3}^\dagger - a_{H4}^\dagger a_{V3}^\dagger + a_{H3}^\dagger a_{V4}^\dagger - a_{H4}^\dagger a_{V4}^\dagger \right) + \right. \]
\[ \gamma \left( a_{V3}^\dagger a_{H3}^\dagger + a_{V4}^\dagger a_{H3}^\dagger - a_{V3}^\dagger a_{H4}^\dagger - a_{V4}^\dagger a_{H4}^\dagger \right) + \delta \left( a_{V3}^\dagger a_{V3}^\dagger - a_{V4}^\dagger a_{V4}^\dagger \right) \right] |\text{vac}\rangle \]
\[ = \left( \alpha \left[ a_{H3}^\dagger a_{H3}^\dagger - a_{H4}^\dagger a_{H4}^\dagger \right] + \left( \beta + \gamma \right) \left[ a_{H3}^\dagger a_{V3}^\dagger - a_{H4}^\dagger a_{V4}^\dagger \right] \right) \]
\[ + \left( \beta - \gamma \right) \left[ a_{H4}^\dagger a_{V3}^\dagger - a_{H3}^\dagger a_{V4}^\dagger \right] + \delta \left[ a_{V3}^\dagger a_{V3}^\dagger - a_{V4}^\dagger a_{V4}^\dagger \right] \right] |\text{vac}\rangle \]

We note that the only term that will give rise to a coincidence detection (i.e. that has photons in both modes 3 and 4) is the term with \((\beta - \gamma)\) as a coefficient. It is easy to show that \((\beta - \gamma) = \langle \psi^- | \psi \rangle\) so that the only component of the state that gives rise to a coincidence count is the one lying along the maximally entangled state \(|\psi^\pm\rangle\). For this reason, two-photon interference followed by coincidence detection can be viewed as a \(|\psi^\pm\rangle\) filter or singlet state filter.

Similarly, among the components of the state that do not give rise to coincidences in modes 3 and 4, only the term with coefficient \((\beta + \gamma)\) contains both a horizontal and a vertical photon. It can be shown that \((\beta + \gamma) = \langle \psi^+ | \psi \rangle\). Thus if one looks at only one output port of the beamsplitter, and detects events when one horizontal and one vertical photon leave from that port, then one will have measured the \(|\psi^+\rangle\) or triplet state projection.

In reality, the interference visibility of two-photon interference is not perfect and so one has to slightly modify this theory to take account of real experimental conditions. These modifications will be discussed when we look at the experimental use of two-photon interference in Chapters 4 and 5.

1.3 Summary

We have looked at a number of concepts that will be important in understanding the work presented in this thesis. In the next chapter we will go deeply into the theory and
practice of quantum state estimation and explain the experimental techniques used to prepare, manipulate and measure quantum states.
Chapter 2

Methods

In theory, theory and practice are the same. In practice, they are not.

2.1 Introduction

This chapter presents the theoretical and experimental methods used throughout this thesis. The first part deals with data analysis techniques for characterizing quantum states of light including the theory of quantum state tomography. The second section discusses laboratory methods for creating, manipulating and measuring those states of light.

2.2 Quantum state tomography

One of the primary tasks for an experimentalist is to correctly determine the state of a system under control in the laboratory. Quantum systems are no exception in this regard, but it is only quite recently that experimentalists have had sufficiently well-behaved, well-isolated systems and the necessary measurement tools to unambiguously determine the quantum state of a system directly from experimental data[55]. The proce-
dure of completely estimating a quantum state from a set of experimental measurements is called quantum state tomography (QST). QST has been performed on such varied single-particle quantum systems as the electronic state of hydrogen\cite{6}, the vibrational state of an ensemble of molecules, the motional state of a trapped ion\cite{74, 96}, the internal angular momentum state of an ensemble of cesium atoms\cite{61} and the state of nuclear spins of a molecule\cite{29}. Continuous variable quantum state tomography has long been used to measure the Wigner function of various states of a single optical mode\cite{108} and has more recently been applied to measuring the Wigner functions of helium atoms in a double-slit experiment\cite{65} and the vibrational state of atoms in an optical lattice\cite{38}. In optics, quantum state tomography can be traced back to the work of G.S. Stokes\cite{114} who developed a minimal set of measurements to describe the polarization of light. The first tomography of an entangled quantum state was performed on the polarization state of two photons generated by spontaneous parametric downconversion\cite{55}. Since then, entangled states have been measured in the same way in trapped ions\cite{50}, superconducting qubits\cite{110}, quantum-dot light sources\cite{113} and other systems. It is no exaggeration to say that quantum state tomography is now considered necessary for verifying claims that a particular quantum state has been generated in the lab. The major aims of this thesis are to extend quantum state tomography to systems of indistinguishable particles as will be done in Chapter 3, and to try to improve upon quantum state tomography as a measurement tool by selecting optimal sets of measurements as in Chapter 4, or by finding better methods of obtaining information about certain aspects of the quantum state as in Chapter 5. The present chapter will introduce the tools that are now routinely used to determine the quantum state of an experimental system and, in particular, describe how the quantum state was measured in the experiments described in subsequent chapters.

In attempting to measure the state of a system under his control the experimentalist is confronted by a fundamental problem – any measurement that he does will disturb the
system in an uncontrolled way, thereby destroying some information about the state. The measurement postulate tells us that any precise measurement of an unknown observable will put the system into a different state, altering, in the process, information about all observables that do not commute with the observable being measured. This makes it impossible to characterize the quantum state of a single copy of a quantum system.

Since determining the quantum state of a single system is impossible, the next best thing from an experimental point of view is to measure the state of a set of identically prepared systems. This can either be done by repeating the same preparation steps multiple times on a single system as is done, for instance, in ion trap quantum systems[50] or in developing a source that produces many systems all in the same state as is usually done in experiments with photons[55]. An additional problem now arises, however. The cautious experimentalist is not certain that the same preparation steps will result in the system being put in the same state unless he is able to measure that the system is in the same state after multiple preparations. Similarly, he needs a way of experimentally verifying that a source of a large number quantum systems successfully puts them all in the same state. Since he cannot measure the state of a single system this seems like a difficult business. Moreover, since the measurements being made are quantum mechanical, one expects to obtain different outcomes for the same measurement from one repetition to the next due to quantum uncertainty even if the systems are identically prepared. How can the randomness due to experimental error be distinguished from the inherent randomness that one expects for quantum measurements? This conundrum can be resolved by describing the system in terms of the density matrix introduced in the last chapter.

Imagine a source that produces some set of states \{|\psi_i\rangle\} with probabilities \{p_i\}. The inclusion of classical probabilities can take into account any random fluctuation in the
Chapter 2. Methods

2.2.1 Properties of the density matrix

We formally restate the properties of the density matrix mentioned in chapter 1. Detailed proofs of these properties can be found in Nielsen and Chuang[91].

- **Hermiticity**: The density matrix is a hermitian operator
- **Positive semi-definiteness**: All of the eigenvalues of the density matrix are positive (they are, in fact, just the probabilities \( \{p_i\} \)). Equivalently \( \langle \phi | \rho | \phi \rangle \geq 0 \) for any state \( |\phi\rangle \).
- **Trace condition**: \( \text{Tr} \rho = 1 \). This follows from the fact that \( \sum p_i = 1 \).
- **Measurement**: For a system in state \( \rho \), the expectation value of a measurement described by Hermitian operator \( \hat{A} \) is \( \text{Tr} \rho \hat{A} \).

2.2.2 Measuring the density matrix

Since the density matrix offers a complete description of the quantum state created by a laboratory process (over many repetitions), the most general state characterization is a
measurement of the density matrix. For a general unknown state this cannot be done in a single shot, but rather requires non-commuting measurements to be done in different bases. The measurement process can be thought of as a set of rotations of the state into different bases, followed by projections onto orthogonal rays in the Hilbert space of states. This procedure involving rotations followed by snapshot measurements bears some resemblance to classical imaging techniques such as Computer Assisted Tomography (CAT) scans where a three-dimensional image is constructed out of many two-dimensional x-ray absorption images taken through different planes.

One assumes in QST that nothing is known about the state to be measured \textit{a priori}. The measurements taken should be able to reproduce the density matrix regardless of the input state.

We assume that the experimentalist has available to him some set of projective measurements (PVMs)\footnote{More generally one could look at Positive Operator Valued Measurements (POVMs), but PVMs are sufficient and far more common in real experiments. The analysis is essentially the same.} denoted \( \{ P^{(i)} \} \). He may perform each of these measurements on an arbitrarily large number of individual systems from the source in order to determine \( \langle P^{(i)} \rangle \) with arbitrary precision. He wishes to determine the density matrix in some particular basis \( \{ |\phi_q \rangle \} \). Since

\[
\rho = \sum_i p_i |\psi_i \rangle \langle \psi_i | = \sum_i p_i \left( \sum_q c_{q,i} |\phi_q \rangle \right) \left( \sum_q c_{q',i}^* \langle \phi_{q'} | \right) = \sum_{i,q,q'} p_i c_{q,i} c_{q',i}^* |\phi_q \rangle \langle \phi_{q'} | ,
\]

(2.2)

this amounts to measuring each of the complex numbers \( \sum_i p_i c_{q,i} c_{q',i}^* \). Writing the density matrix in this way highlights the impossibility of knowing whether variation in measurement outcomes arises due to experimental errors or fundamental quantum mechanical uncertainty. The classical probabilities associated with experimental errors \( p_i \) and the
quantum mechanical amplitudes $c_q$ always occur together as a product so that the contribution due to each cannot be determined for any particular measurement.

**Linear inversion**

The simplest and most intuitive approach to quantum state estimation is to regard the mapping of experimental measurements to a density matrix as a linear inversion problem. Linear inversion is possible because Hermitian operators form a Hilbert space, and so the density matrix will be related to a complete set of measurement expectation values through a linear map (i.e. a matrix) that is a function only of the particular measurement operators that were implemented. The major downside to this approach is that it does not allow one to easily take account of the positivity constraint on the density matrix. Linear inversion is still important because it is the only available analytic inversion tool which makes it important for conceptual understanding of state estimation and as a means of calculating error propagation without resorting to Monte Carlo techniques.

Consider the expectation value of the measurement $P^{(a)}$

\[
\langle P^{(a)} \rangle = \text{Tr} \{ \rho P^{(a)} \} = \sum_m \langle \phi_m | \rho P^{(a)} | \phi_m \rangle = \sum_m \langle \phi_m | \sum_i p_i | \psi_i \rangle \langle \psi_i | P^{(a)} | \phi_m \rangle = \sum_{m,i} p_i \langle \phi_m | \psi_i \rangle \langle \psi_i | P^{(a)} | \phi_m \rangle = \sum_{m,i} p_i c_{m,i} \left( \sum_{q'} c^*_{q',i} \langle \phi_{q'} | P^{(a)} | \phi_m \rangle \right) = \sum_{m,q'} \left( \sum_{i} p_i c_{m,i} c^*_{q',i} \right) \langle \phi_{q'} | P^{(a)} | \phi_m \rangle .
\]

We note that since the basis vectors $| \phi_m \rangle$ and the measurement operator $P^{(a)}$ depend only on the measurement setup, not the measurement outcomes, $\langle \phi_{q'} | P^{(a)} | \phi_m \rangle$ are fixed for a given quantum state tomography setup, independent of the input state. The quantities
\[ \sum_i p_i c_{m,i} c_{q,i}^* \], on the other hand, depend on the density matrix \( \rho \). In fact, as was seen in equation 2.2, they uniquely define it. Once they are extracted from measurement outcomes \( \rho \) can be reconstructed.

If many expectation values \( \langle P^{(a)} \rangle \) are measured then many equations of the form of (2.3) are obtained. These can be arranged in a matrix form as

\[
\begin{bmatrix}
\langle P^{(1)} \rangle \\
\vdots \\
\langle P^{(n)} \rangle 
\end{bmatrix} =
\begin{bmatrix}
\langle \phi_1 | P^{(1)} | \phi_1 \rangle & \cdots & \langle \phi_d | P^{(1)} | \phi_1 \rangle & \cdots & \langle \phi_1 | P^{(1)} | \phi_d \rangle \\
\vdots & & \vdots & & \vdots \\
\langle \phi_1 | P^{(n)} | \phi_1 \rangle & \cdots & \langle \phi_d | P^{(n)} | \phi_1 \rangle & \cdots & \langle \phi_1 | P^{(n)} | \phi_d \rangle
\end{bmatrix}
\begin{bmatrix}
\sum_i p_i c_{1,i} c_{1,i}^* \\
\vdots \\
\sum_i p_i c_{d,i} c_{d,i}^*
\end{bmatrix}
\] (2.8)

More compactly, we can define a matrix

\[
M =
\begin{bmatrix}
\langle \phi_1 | P^{(1)} | \phi_1 \rangle & \cdots & \langle \phi_d | P^{(1)} | \phi_1 \rangle & \cdots & \langle \phi_1 | P^{(1)} | \phi_d \rangle \\
\vdots & & \vdots & & \vdots \\
\langle \phi_1 | P^{(n)} | \phi_1 \rangle & \cdots & \langle \phi_d | P^{(n)} | \phi_1 \rangle & \cdots & \langle \phi_1 | P^{(n)} | \phi_d \rangle
\end{bmatrix}
\] (2.9)

and vectors

\[
\vec{P} =
\begin{bmatrix}
\langle P^{(1)} \rangle \\
\vdots \\
\langle P^{(n)} \rangle
\end{bmatrix},
\vec{\rho} =
\begin{bmatrix}
\sum_i p_i c_{1,i} c_{1,i}^* \\
\vdots \\
\sum_i p_i c_{d,i} c_{d,i}^*
\end{bmatrix}
\] (2.10)

Then equation (2.8) becomes

\[ \vec{P} = M \vec{\rho} \] (2.11)

Clearly if \( M \) has an inverse \( M^{-1} \) then one can write the density matrix elements \( \vec{\rho} \) in terms of the measurements \( \vec{P} \) as

\[ \vec{\rho} = M^{-1} \vec{P} \] (2.12)

In order for \( M \) to have an inverse, it must be a square matrix so that the number of rows (i.e. the number of measurements) is the same as the number of columns (i.e. the number of density matrix elements, \( d^2 \)).
If the number of linearly-independent measurements is less than the number of density matrix elements (i.e. if the rank $M < d^2$), then $M$ is not invertible and the inversion problem is under-determined.

When the number of measurements exceeds the number of density matrix elements, the problem is over-determined. Linear inversion is still possible by using techniques from linear regression theory[16]. Imagine that there are $m$ measurements so that $\vec{P}$ is of length $m$ and $M$ is an $m \times d^2$ dimensional matrix. We can multiply both sides of equation 2.11 by $M^\dagger$ to obtain

$$M^\dagger \vec{P} = M^\dagger M \vec{\rho}$$ (2.13)

Since $M$ is $m \times d^2$, $M^\dagger$ is $d^2 \times m$, so $M^\dagger \vec{P}$ is a $d^2$ dimensional vector and $M^\dagger M$ is a $d^2 \times d^2$ square matrix of rank $d^2$ whose inverse, $(M^\dagger M)^{-1}$ allows us to solve the equation for $\vec{\rho}$

$$(M^\dagger M)^{-1} M^\dagger \vec{P} = \vec{\rho}$$ (2.14)

$\vec{\rho}$ represents the density matrix that minimizes the sum of the variances between its predicted expectation values and the measured data[16]. It is, in other words, the least-squares estimator of the true density matrix.

Unfortunately, because no positivity constraint was imposed, it is possible that the estimated $\rho$ will have negative eigenvalues. This makes it impossible to calculate quantities from $\rho$ that depend on its positivity, including the entanglement-related properties like concurrence and the fidelity with other states. This fact usually makes it necessary to employ a more sophisticated methodology like maximum-likelihood fitting to be discussed in the next section. It should be noted, though, that if the density matrix obtained from linear fitting is positive, then it will also be the maximum-likelihood estimator for the least-squares likelihood function and is usually computationally much easier to obtain.
Chapter 2. Methods

Error analysis

Because the linear inversion is analytic, it can give insight into how errors propagate from experimentally collected counts to the density matrix. We will be making use of the error analysis derived here in Chapter 4.

In photon counting experiments, non-systematic errors arise largely due to the counting statistics or shot noise. For the intensities encountered in the experiments in this thesis, the probability of obtaining more than one photon pair in a single-photon coherence time is vanishingly small\(^2\). As a consequence, the probability of creating a photon pair in an infinitesimal time \(dt\) is essentially constant, giving rise to Poissonian counting statistics over finite times. This can be seen from experimental evidence in Figure 2.1.

Expectation values \(\langle \hat{P}^{(i)} \rangle\) are estimated from counts by dividing the number of counts for a given PVM element by the total number for a complete PVM, or basis, over a given counting period. For a two-photon polarization state we can write

\[
P_{kq} \equiv \frac{\langle \hat{P}^{(i)} \rangle_{\text{est}}}{n_{kq}} = \frac{n_{kq}}{\sum_{i=1}^{4} n_{iq}}, \tag{2.15}
\]

where \(n_{kq}\) is the number of observed counts for the \(k\)th PVM element in the \(q\)th basis. In order to calculate the error \(\delta P^{(i)}_{kq}\) in \(\langle \hat{P}^{(i)} \rangle_{\text{est}}\), we first calculate the dependence of the expectation value estimates on the counts by evaluating the partial derivatives

\[
\frac{\partial P_{kq}}{\partial n_{kq}} = \frac{\sum_{i\neq k} n_{iq}}{\left(\sum_{j=1}^{4} n_{jq}\right)^2} \tag{2.16}
\]

\[
\frac{\partial P_{kq}}{\partial n_{mq}} = \frac{-n_{kq}}{\left(\sum_{j=1}^{4} n_{jq}\right)^2} \quad m \neq k \tag{2.17}
\]

\[
\frac{\partial P_{kq}}{\partial n_{mt}} = 0 \quad t \neq q. \tag{2.18}
\]

\(^2\)Although in high-intensity pulsed SPDC experiments such as those described in [35, 43], bosonic stimulation can make for highly non-Poissonian counting statistics.
The covariance between $P_{kq}$ and $P_{ab}$ is given by

$$\delta P_{kq}\delta P_{ab} = \sum_{x,y} \left( \frac{\partial P_{kq}}{\partial n_x} \right) \left( \frac{\partial P_{ab}}{\partial n_y} \right) \delta n_x \delta n_y$$  \hspace{1cm} (2.19)

$$= \sum_{y} \left( \frac{\partial P_{kq}}{\partial n_y} \right) \left( \frac{\partial P_{ab}}{\partial n_y} \right) \delta n_y^2$$  \hspace{1cm} (2.20)

$$= \sum_{y} \left( \frac{\partial P_{kq}}{\partial n_y} \right) \left( \frac{\partial P_{ab}}{\partial n_y} \right) n_y,$$  \hspace{1cm} (2.21)

where in the second line we have used the fact that the different photon counts are statistically uncorrelated and in the last line we have replaced the variance with the estimated mean since the photon counting statistics are Poissonian.

The errors in the density matrix can then be calculated from this covariance. Care
must be taken to separate the variance in the real and imaginary parts of $\rho$.

$$
(\Delta \Re \rho_{ij})^2 = \sum_{kq,ab} \Re \left[ \frac{\partial \rho_{ij}}{\partial P_{kq}} \right] \Re \left[ \frac{\partial \rho_{ij}}{\partial P_{ab}} \right] \delta P_{kq} \delta P_{ab} 
$$

$$
= \sum_{kq,ab} \Re \left[ \left( (M^\dagger M)^{-1} M^\dagger \right)_{ij,kq} \right] \Re \left[ \left( (M^\dagger M)^{-1} M^\dagger \right)_{ij,ab} \right] \delta P_{kq} \delta P_{ab} \tag{2.23}
$$

$$
(\Delta \Im \rho_{ij})^2 = \sum_{kq,ab} \Im \left[ \frac{\partial \rho_{ij}}{\partial P_{kq}} \right] \Im \left[ \frac{\partial \rho_{ij}}{\partial P_{ab}} \right] \delta P_{kq} \delta P_{ab} 
$$

$$
= \sum_{kq,ab} \Im \left[ \left( (M^\dagger M)^{-1} M^\dagger \right)_{ij,kq} \right] \Im \left[ \left( (M^\dagger M)^{-1} M^\dagger \right)_{ij,ab} \right] \delta P_{kq} \delta P_{ab} \tag{2.24}
$$

A similar analysis may be carried out for systematic errors such as waveplate calibration errors, although for datasets containing less than 10,000 counts per basis, the random errors were found to be dominant.

### 2.2.3 The positivity constraint

Linear inversion will sometimes result in unphysical density matrix estimates for two reasons. First it is impossible to exactly measure an expectation value in a finite number of measurements. While one can obtain an arbitrary precision by measuring for an arbitrarily long time, the error in the estimation of the expectation value will be of order $1/\sqrt{N}$ for $N$ measurements done on uncorrelated copies of the system. Second, a systematic error in the measurement apparatus such as a misalignment of waveplate axes can result in the measurements that could not have been produced by any quantum state, even after an arbitrarily long measurement.

These errors, though small, can result in the linear density matrix reconstruction described in the previous section giving a negative density matrix. This problem is especially pronounced when the true state of the system under study is very pure. In that case all but one of the eigenvalues of the density matrix are already zero or close to it, and a small error in one of the measurements is enough to make some of the estimated eigenvalues negative.
For some applications this is not a significant problem. The magnitude of the negative eigenvalues will be small relative to one if the errors are also small, in which case this negativity might be negligible. There are, however, many calculations that depend on the strict positive-semi-definiteness of the density matrix and that give nonsensical results when a non-positive-semidefinite matrix is used. Consequently it is often useful to turn to a more sophisticated analysis which is guaranteed to generate a positive-semidefinite density matrix from any experimental data.

### 2.2.4 Maximum-likelihood estimation

In statistics the problem of matching the parameters of a physical model to noisy data is well studied and is addressed by a technique known as maximum likelihood estimation[16]. This approach takes the view that given a dataset, an error model and a parameterized model of the system, one can calculate the probability that a given set of parameters in the model generated a particular datum in the data set. A likelihood function can be defined that represents the probability that the entire data set was generated from the model for a given set of parameters. A numerical search can then be used to find the set of parameters that maximize this likelihood function. This technique has been used successfully in quantum state tomography to determine the density matrix that was most likely to have produced a particular set of measurement outcomes[55].

Early work on state tomography relied on an explicit parameterization of a finite-dimensional density matrix based on a Cholesky decomposition[16, 55]. The Cholesky decomposition is akin to a square root operation for matrices. Any positive semi-definite hermitian matrix $A$ can be decomposed as $A = \tau\tau^\dagger$ where $\tau$ is a lower triangular matrix with real-valued diagonals. The trace condition for density matrices can be satisfied by explicitly normalizing to express $\rho$ as $\rho = \tau\tau^\dagger/\text{Tr}\{\tau\tau^\dagger\}$. Parameterizing $\tau$ instead of parameterizing $\rho$ reduces a constrained multi-dimensional search problem to an unconstrained one which can be solved much more efficiently.
This method has a few problems, however. First, the Cholesky decomposition is not unique for positive semi-definite matrices (it is unique for strictly positive matrices), meaning that if one or more eigenvalues of $\rho$ is zero there will generally be multiple maxima of the likelihood function which could lead to convergence problems in finding the global maximum. Moreover, there is no guarantee that all the local maxima will correspond to the same density matrix. In fact, maximization starting from a randomly generated set of parameters will generally result in different final density matrices. In practice the maximization routine is usually seeded with a density matrix obtained by linear inversion and will converge properly for any well-behaved likelihood function, but it remains a problem that in principle this method does not guarantee convergence to the global maximum\[64]\[3]

More recently many groups working on quantum state tomography have begun solving the maximum-likelihood problem using the techniques of convex optimization. This is a class of optimization methods that efficiently reduce and solve optimization problems in simply connected convex spaces. It was pointed out by Kosut\[64] that the positivity and unit-trace constraints on density matrices make maximum-likelihood fitting of a density matrix a convex problem to which convex optimization techniques can be applied. In particular, the Matlab toolkit SeDumi\[115] handles these problems quite well and was used for much of the maximum-likelihood analysis done in this thesis.

The likelihood function

The objective of maximum-likelihood fitting is to find the density matrix that maximizes the likelihood of having generated the measured data. We can write the conditional probability of obtaining data set $D$ given that the system is described by the density

\[3\text{Based on discussions at this summer's Workshop on Quantum State Estimation it appears that Joe Altepeter and Paul Kwiat have developed techniques for guaranteeing global convergence using a Cholesky decomposition method. The comments on the problems with local minima are based on my own experiences in following the prescription of James et al.}[55].\]
matrix $\rho$ as

$$\text{Prob}\{D, \rho\} = \prod_{\alpha, \gamma} (p_{\alpha, \gamma}(\rho)^{n_{\alpha, \gamma}}),$$

(2.26)

where $\alpha$ and $\gamma$ respectively index the outcome and the measurement basis for different waveplate settings. $n_{\alpha, \gamma}$ is the measured number of occurrences of a particular outcome and $p_{\alpha, \gamma}(\rho)$ is the probability of occurrence for that outcome when the system is described by density matrix $\rho$. The product is the probability of obtaining exactly the data set $D$ consisting of the counts $n_{\alpha, \gamma}$ if the state were described by $\rho$. The maximum-likelihood estimator of the true density matrix is the density matrix that maximizes this probability.

In practice, it is common to take the logarithm of both sides of the equation

$$\log \text{Prob}\{D, \rho\} = \sum_{\alpha, \gamma} n_{\alpha, \gamma} \log p_{\alpha, \gamma}(\rho)$$

(2.27)

$$= \sum_{\alpha, \gamma} n_{\alpha, \gamma} \log \text{Tr}\{\rho O_{\alpha, \gamma}\},$$

(2.28)

where the $\{O_{\alpha, \gamma}\}$ are the projective operators measured in the experiment. Since the probabilities are positive numbers between zero and one, this sum will be negative. Because the logarithm is a monotonic function, the density matrix that maximizes the likelihood of equation 2.26 will also maximize the log-likelihood of equation 2.28. Numerical solvers often require a minimization rather than a maximization problem, so one can equivalently try to minimize the negative-log-likelihood which is simply $-1$ times equation 2.28.

The log-likelihood problem does not assume any particular distribution function on $\rho$ for the probabilities. If one assumes that the outcomes $n_{\alpha, \gamma}$ are large enough to give a good estimate of the relative probabilities of occurrence for the outcomes $p_{\alpha, \gamma}^{\text{emp}} = n_{\alpha, \gamma}/l_\gamma$ where $l_\gamma$ is the total for all the outcomes in a given basis, $l_\gamma = \sum_{\alpha} n_{\alpha, \gamma}$, then one can replace the log-likelihood function with the least-squares likelihood function[64].

The log-likelihood is a log-convex function of $\rho$ and so the maximization problem is a convex one. Unfortunately the numerical solver we use, SeDumi, is as of yet incapable
of optimizing log-convex problems. It can, however, solve the least-squares problem, offering a solution that is quite close to the log-convex minimizing solution. From this starting point we can iteratively search for the log-convex minimizing solution and we will generally achieve fast convergence.

In practice the least-squares fit is usually a very good approximation to the log-convex solution, with density matrix terms differing by less than 1 percent when there are more than 100 counts per basis. Using the least-squares fit as a starting point, the log-convex solution could usually be found to within a part in \(10^4\) within five iterations of our search algorithm.

### 2.2.5 An example: Two-photon polarization density matrix

We can illustrate these approaches to quantum state tomography by reconstructing a two-qubit polarization state from experimental data. We will use the set of 36 measurements of coincidence counts taken in the lab and shown in Table 2.1.

If we apply a linear inversion to this over-complete data set we obtain the following density matrix:

\[
\rho_{\text{lin}} = \begin{pmatrix}
0.4602 & -0.0033 - 0.0463i & -0.0391 - 0.0044i & 0.4257 - 0.0095i \\
-0.0033 + 0.0463i & -0.0119 & 0.0060 - 0.0217i & -0.0149 + 0.0557i \\
-0.0391 + 0.0044i & 0.0060 + 0.0217i & 0.0262 & -0.0346 + 0.0382i \\
0.4257 + 0.0095i & -0.0149 - 0.0557i & -0.0346 - 0.0382i & 0.5255
\end{pmatrix}. \tag{2.29}
\]

The eigenvalues of the matrix are 0.9300, 0.0755, 0.0225 and -0.0279, meaning that the density matrix is not positive semi-definite and is therefore unphysical. In fact, the negativity of the density matrix is obvious from the negative value of the population \(\rho_{22}\).

If we instead apply maximum likelihood estimation with a least-squares likelihood function we obtain

\[
\rho_{\text{ml-ls}} = \begin{pmatrix}
0.4478 & -0.0066 + 0.0460i & -0.0397 + 0.0053i & 0.4318 + 0.0093i \\
-0.0066 - 0.0460i & 0.0073 & 0.0043 + 0.0079i & -0.0090 - 0.0526i \\
-0.0397 - 0.0053i & 0.0043 - 0.0079i & 0.0314 & -0.0339 - 0.0370i \\
0.4318 - 0.0093i & -0.0090 + 0.0526i & -0.0339 + 0.0370i & 0.5135
\end{pmatrix}. \tag{2.30}
\]
Table 2.1: Laboratory data for the closest state to $\frac{1}{\sqrt{2}} (|HH\rangle + |VV\rangle)$ that we can make. The projections column gives the measurements performed on the two photons, the counts column gives the measured number of coincidence counts. The frequencies column gives that number of counts normalized to the total number of counts measured for that basis.
The values of the individual density matrix elements are quite close to those for the linear inversion with the crucial difference that this density matrix is positive semi-definite with eigenvalues of 0.9234, 0.0612, 0.0154 and 0.0000.

If we instead perform correct maximum-likelihood estimation using the log-likelihood function, we obtain

\[
\rho_{\text{ml-ll}} = \begin{pmatrix}
0.4478 & -0.0066 + 0.0460i & -0.0397 + 0.0053i & 0.4318 + 0.0093i \\
-0.0066 - 0.0460i & 0.0073 & 0.0043 + 0.0079i & -0.0090 - 0.0526i \\
-0.0397 - 0.0053i & 0.0043 - 0.0079i & 0.0314 & -0.0339 - 0.0369i \\
0.4318 - 0.0093i & -0.0090 + 0.0526i & -0.0339 + 0.0369i & 0.5135
\end{pmatrix}
\] (2.31)

which is almost identical to the least-squares maximum-likelihood estimate, differing only on in terms of order \(10^{-5}\).

From this density matrix we can calculate all of the figures of merit mentioned in Chapter 1. The concurrence is 0.8585. The purity is 0.8567. The von Neumann entropy is 0.1911. The entanglement of formation is 0.8010.

We can also estimate the errors by calculating the covariance matrix for \(\rho_{\text{lin}}\). The statistical errors on the individual density matrix elements are

\[
\Delta \rho = 10^{-4} \times \begin{pmatrix}
5.372 & 4.232 + 1.300i & 4.427 + 0.781i & 6.619 + 0.510i \\
4.232 + 1.300i & 8.581 & 6.619 + 0.510i & 3.158 + 0.680i \\
4.427 + 0.781i & 6.619 + 0.510i & 6.412 & 3.186 + 0.897i \\
6.619 + 5.10i & 3.158 + 0.680i & 3.186 + 0.897i & 8.298
\end{pmatrix}
\] (2.32)

That these errors are less that the discrepancy between \(\rho_{\text{lin}}\) and \(\rho_{\text{ml-ll}}\) indicates that systematic rather than statistical errors constitute the dominant error source.

### 2.3 Experimental state estimation

Much of my early thesis work was devoted to building a computer-controlled system for doing quantum state estimation. This formed the basic experimental system used throughout this thesis. This section will explain the functioning of the system, describe measurements taken on it and discuss possible improvements.
2.3.1 Spontaneous parametric downconversion

The discovery of Bell’s inequalities[10] in the mid-sixties triggered a search for sources of entangled particles that could be used to test them. Realistic proposals for testing Bell’s inequalities with photons were published in 1969[30] and eventually realized in atomic cascade systems[44, 7]. Compared with modern sources, these sources were limited in their rates by the fact that emission was isotropic and hence weak in any particular measurement direction, and by their narrow resonances and hence limited re-excitation rate. Additionally, these cascade sources were inherently non-degenerate in frequency, limiting the sorts of two-photon interference effects that could be implemented with them.

The discovery of spontaneous parametric downconversion (SPDC)[27] in $\chi^{(2)}$ non-linear materials delivered solutions to both of these problems while providing a much more convenient and efficient source of paired photons. SPDC was initially used [53, 111] for two-photon interference experiments, but a much wider range of experimental possibilities was opened up by the development of polarization-entangled SPDC sources[68, 69]. These sources allowed very large violations of Bell’s inequalities[69], including violations over large distances[122], closing the communication loophole. They allowed entanglement to be studied in much greater depth than it had been up to that point[124]. They also led very quickly to four-photon[94], and later six-photon[128], sources which were used to demonstrate teleportation[94], GHZ-type entanglement[93], cluster state quantum computing[120], quantum logic-gates[92], quantum algorithms[71] and a wide host of other technologies. While there have been exciting developments in alternative sources of photons such as quantum dot sources[102], cavity QED sources[83], and sources making use of $\chi^{(3)}[76]$ or spontaneous Raman scattering[28], $\chi^{(2)}$ SPDC remains the standard tool for studying entanglement, quantum information and linear optical quantum computing systems.

Spontaneous parametric downconversion can be thought of as the time-reversal of sum-frequency generation. Sum-frequency generation is a classical effect that occurs when
the mechanical response of the charge in a dielectric medium is non-linear in the applied field. This results in a polarization field in the material with frequency components that were not present in the applied field. Since the non-linear response of all ordinary materials is extremely weak, this response is very well approximated by the first non-linear term in the Taylor series expansion of the material polarization. In materials that lack inversion symmetry, such as many crystals, this is a term quadratic in the applied field. Such materials are called $\chi(2)$ materials. In centro-symmetric materials like glasses, this is a term cubic in the applied field and the materials are called $\chi(3)$ materials. In $\chi(2)$ materials, sum-frequency generation involves two input fields with frequencies $f_1$ and $f_2$ generating an output field with a frequency of $f_3 = f_1 + f_2$. In the special case that $f_1 = f_2$, sum-frequency generation is called second-harmonic generation.

Time-reversal invariance implies that it ought to be possible to put a frequency $f_3$ into a $\chi(2)$ non-linear material and see two frequencies $f_1$ and $f_2$ come out with $f_1 + f_2 = f_3$. Indeed, Maxwell’s equations predict that if two seed beams with frequencies $f_1$ and $f_2$ are present at the input to the medium then they will both be amplified, an effect known as optical parametric amplification. When quantum theory is applied to this situation, the main result is that light at $f_1$ and $f_2$ is generated even in the absence of a seed field. One interpretation of this is that the seed is provided by quantum vacuum fluctuations of the electromagnetic field. Thus in quantum theory a $\chi(2)$ material illuminated with light at a frequency $f_3$ will generate a polarization field at every pair of frequencies $f_1$ and $f_2$ such that $f_1 + f_2 = f_3$. The two SPDC fields are, for historical reasons, called signal and idler fields.

Because the excitations of the field are quantized, the non-linear interaction must destroy a whole photon from the pump and create whole photons in the signal and idler beams when the interaction takes place. This effect, together with Planck’s Law $E = hf$, means that the relationship $f_i + f_s = f_p$ is simply an expression of conservation of energy
at the single-photon level,

\[ E_s + E_i = E_p. \] (2.33)

So far the discussion of polarization fields has considered their generation, but not their propagation. In order for a polarization field to propagate in the crystal, the polarization fields generated by the pump at different locations in the crystal must interfere constructively in a given direction. Since the polarization created locally by the pump will be in phase with the pump, the polarization field will only propagate if it remains in phase with pump during propagation. This amounts to a requirement that the wave-vectors of the signal and idler sum to the wave-vector of the pump. Interestingly, this puts no restrictions on the individual wave-vectors of the signal and idler, only their sum. The requirement can be expressed as

\[ \vec{k}_p = \vec{k}_s + \vec{k}_i \]

\[ \hbar \vec{k}_p = \hbar \vec{k}_s + \hbar \vec{k}_i \]

\[ \vec{p}_p = \vec{p}_s + \vec{p}_i, \] (2.34)

where we have used de Broglie’s relationship to show that this condition on the wave-vectors amounts to nothing more than momentum conservation for the three photons. Satisfying this condition is known as phase-matching.

Most dielectric materials have normal dispersion, meaning that frequency increases sub-linearly with wave number \( k_0 = |\vec{k}_0| \). This makes equation 2.34 impossible to satisfy since it means that \( k_p \) will always be greater than \( k_s + k_i \). A clever solution to this problem is to employ birefringent materials\cite{24} which have the property that different polarizations have different phase velocities. If the pump is polarized along the direction with a relatively fast phase velocity, then there will generally be some emission directions for which the signal and idler will be phase-matched at some frequencies. Not all \( \chi(2) \) materials are birefringent\cite{4}, nor do all \( \chi(2) \) materials that are birefringent have non-

\footnote{The most notorious case of this is GaAs which has a very large \( \chi(2) \) non-linear susceptibility, but}
zero non-linear susceptibility for the combinations of polarizations that can be phase-matched. There are, however, many materials for which birefringent phase-matching is possible with some of the most popular being beta-phase barium borate ($\beta$-BBO), lithium borate (LBO), potassium titanyl phosphate (KTP), lithium niobate (LiNbO$_3$) and bismuth borate (BiBO).

In most birefringently phase-matched materials there are two ways to select polarizations of the pump, signal and idler to achieve phase-matching. In type-I phase-matching, the signal and idler have the same polarization and both are orthogonal to the pump polarization. In type-II phase-matching the signal and idler have opposite polarization, and the pump is parallel to one of them and orthogonal to the other. Most birefringent materials allow either type of phase-matching to be used, although a few materials allow type-I but not type-II due to crystal symmetry (i.e. no non-linear polarization field develops in the directions required to give the signal and idler polarizations). When both types of phase-matching are allowed, type-I tends to have a slightly larger component of the non-linear susceptibility tensor.

The two types of phase-matching lead to very different SPDC emission patterns due to the behavior of light in birefringent materials. Type-I emission is the more straightforward of the two. Conservation of momentum dictates that if the signal and idler are degenerate in wavelength then they must make the same angle with the pump $k$-vector in order to conserve momentum. The emission pattern is therefore in the shape of a cone with the angle of emission being determined by simultaneously solving equation 2.33 and equation 2.34 along with the birefringent dispersion relations $n_e(\lambda)$ and $n_o(\lambda)$. In Type-II phase-matching in a uniaxial crystal one downconverted photon will be extraordinarily polarized while the other will be ordinarily polarized. Because the two photons see different indices of refraction, even when they are degenerate in frequency, the emissions cones

---

has so far proved impossible to phase-match, despite a considerable amount of effort expended on the problem.
are not coincident on each other. One is centered above and the other below the pump
direction where above and below are in the direction of the optical axis. By adjusting
the angle between the crystal axis and the pump, the cones can be made to overlap[68]
or can be reduced to point-like regions[116].

2.3.2 Polarization-entangled SPDC sources

While early experiments using SPDC sources examined the time-frequency and position-
momentum entanglement[67, 97], the development of polarization-entangled sources rev-
olutionized SPDC’s role in quantum information and quantum computing. The reason
is that polarization provides an easily manipulated, easily measured qubit and SPDC
sources can provide entangled two-qubit states of unparalleled quality[49].

Both kinds of SPDC phasematching can be used to create entangled photons. Type-
II phasematching provides entangled photons without any additional effort so long as
the phase-matching is chosen so as to make the ordinary and extraordinarily polarized
cones intersect at two points. If light is collected from the two intersection points there
is an amplitude for point 1 to contain the extraordinarily polarized photon and point 2
to contain the ordinarily polarized photon and another amplitude for point 1 to contain
the ordinary photon and point 2 to contain the extraordinary photon. The relative phase
between these two amplitudes will depend on the details of the phase matching and the
length of the crystal, but can be easily adjusted by applying a birefringent phase shift to
one of the two photons. The result is an entangled photon state of a very high quality[68].

Type-I SPDC is not naturally polarization-entangled since both the SPDC photons
have the same polarization. A clever idea due to Kwiat[69], though, was to sandwich
two Type-I crystals together with their axes at 90° to each other. A pump polarized
at 45° to the two crystal axes will excite non-linear polarization fields in both crystals
with a definite phase relationship between them. If the emission rates are low then
downconversion will only occur in one of the crystals at a time, but there will be a
definite phase between the amplitudes for downconversion in each crystal. If the output of the two crystals overlaps then at any two points on the emission cone there is an equal amplitude for the photons to have come from either of the two crystals and hence for both to be horizontally polarized or both to be vertically polarized.

This source has the advantageous property that every pair of photons that it produces is entangled and, what is more, it allows the degree of entanglement to be controlled losslessly by changing the polarization of the pump. This makes this source especially useful in systems that want to examine non-maximally entangled states[124] or experiments where the degree of entanglement needs to be controlled. The experiments described in chapters 4 and 5 make use of a type-I entangled source while those described in chapter 3 make use of a type-II unentangled source.

2.3.3 Real-world sources

While SPDC sources are capable of supplying very high-quality entangled states, they are subject to imperfections that can limit entanglement or otherwise produce less-than perfect photons. This section will explain the effects that need to be considered in order to correct source imperfections and present some results on trying to improve our sources. This is an ongoing process, with new techniques for improving sources being a very active area of current research[4, 89]. The sources we use currently are still quite far from optimal, and improving them will be an ongoing project for the next few years.

Since our experimental work made use of two specific sources for two very different applications, we will first discuss each source and its desired properties and then consider the experimental imperfections that will affect these properties. Since 2005 I have been working exclusively on the Type-I source, so most of the discussion will be centered on describing it and less on the Type-II source.
2.3.4 Type-I sandwich source

The type-I source used in chapters 4 and 5 consisted of two 2.5 mm long pieces of $\beta$-BBO glued together by the crystal manufacturer Crystran. The crystals were rotated 90° to each other and the optical axes were tilted 29.2° away from the surface normal. The particular choice of optical axis angle was chosen so that the crystals would be phase-matched for degenerate type-I SPDC at 810 nm at a 3° opening angle with a 405 nm pump at normal incidence. In practice it was found that one of the crystals emitted at 810 nm at normal incidence while the other had to be tilted approximately 0.8° away from normal to achieve phase-matching for the same opening angle. This is likely due to a manufacturing error, but the cause of the discrepancy was never fully investigated.

The source was pumped with a 28 mW output Coherent Vioflame laser containing a gallium-nitride laser diode chip selected to have a 405 ± 0.3 nm center wavelength. The laser spectrum as measured by an Ocean Optics HR2000 spectrometer is shown in figure 2.2. The spectral width is 0.5 nm with a corresponding coherence time of 1.1 ps or a coherence length of 330 $\mu$m.

Before being sent into the crystal, the laser was polarized with a Glan-Thomson calcite polarizer (Thorlabs GTH10M) and then rotated to the desired polarization with a 405 nm true zero-order quartz half-waveplate from FocTek.

SPDC emission from the crystal is very broadband. It was estimated to be approximately 60 nm based on observed reduction in counts when 13 nm interference filters were used to reduce the spectrum. In principle, the spectrum could be measured accurately using the Hong-Ou-Mandel effect [53], but this measurement has not been undertaken.

Longitudinal walkoff

In order for the SPDC from the two crystals to be coherent, the emission from the crystals must be spatially and temporally indistinguishable. Naively one might think that this would involve matching the group delays in the two crystals to better than an...
Figure 2.2: Measured spectrum of 405 nm pump laser. Centre frequency is found to be 404.8 nm with a full-width at half maximum bandwidth of 0.5 nm.

SPDC coherence length. Surprisingly this is not the case – even if emission from one crystal is significantly advanced with respect to the other, the emission amplitudes for the two crystals will be coherent, so long as the delay between them is less than the coherence length of the pump[69]. The reason for this that with a continuous wave (CW) pump the SPDC emission time is completely random, so it makes no sense to talk about early or late emission from the two crystals. Even if emission from the two crystals experiences wildly different group delay, this delay generates no information about which crystal produced the downconversion and so coherence between the two possibilities is conserved. However if the pump has a finite coherence time then emission amplitudes at delays larger than that coherence time will have random relative phases and will not be coherent. In our crystal geometry, a pair of SPDC photons downconverting in the middle
of the first crystal will experience a total optical group delay of 8.272 mm before exiting whereas a pair created in the middle of the second crystal will experience an optical group delay of 8.588 mm. This includes the group delay that the pump experiences before the downconversion happens. Thus there is a difference in the group delays of 318 µm. This can be compared with the 330 µm coherence length of the pump.

Only the extraordinarily polarized component of the pump will create SPDC in each of the crystals. It is therefore possible to precompensate for this decoherence in the SPDC by delaying one component of the pump polarization relative to the other. This can be done simply by inserting a piece of birefringent material with its axis aligned with the SPDC crystal axes. In our experiment a piece of α-phase BBO was used for this purpose. One can calculate the expected concurrence with and without the compensating crystal. With the compensator the expected concurrence is 0.963 and without it it is 0.274. The expected concurrence with the compensator is significantly higher than maximum measured concurrence of 0.89, indicating that there are other sources of decoherence besides longitudinal walkoff.

**Transverse walkoff**

In addition to longitudinal walkoff, the crystals exhibit transverse walkoff. While longitudinal walkoff results from a frequency-dependent phase, transverse walkoff results from a direction-dependent phase. In uniaxial birefringent materials, extraordinary rays experience an index of refraction that depends on the angle the ray makes with the optical axis. This dependence of index on angle results in a dependence of the phase of the extraordinary beam on propagation direction. The linear component of this phase dependence causes the center of the beam to ‘walk off’ in the transverse direction. This effect is easy to see in thick pieces of quartz or calcite. An unpolarized light source viewed through a thick piece of calcite will appear in double because the extraordinarily polarized part of the image will have walked off relative to the ordinarily polarized part. It is also the
principle used in making calcite polarizers and polarizing beamsplitters.

In type-I sandwich SPDC the effect causes the SPDC produced in the first crystal to be displaced relative to the SPDC produced in the second crystal since the SPDC from the first crystal will be extraordinarily polarized in the second crystal. Moreover, the non-local phase between the extraordinary and ordinary polarizations will be spatially-dependent in the far field. This means that for each iris position, the phase $\phi$ in the measured state $\frac{1}{\sqrt{2}} (|HH\rangle + e^{i\phi}|VV\rangle)$ will be different. This effect can be seen in figure 2.3 where the measured non-local phase is plotted against the position of one of the detector irises. When SPDC is collected over a finite collection aperture, this spatially-dependent birefringent phase results in reduced entanglement and reduced purity. The effect can be overcome by reducing the collection aperture, but this also reduces the total number of counts. Figure 2.4 shows a plot of the concurrence of an entangled state versus the collection aperture size and compares it to the concurrence expected from the phase dependence on iris position. The measured concurrence is considerably lower than the expected concurrence, so there is clearly some other decoherence mechanism operating.

Another option that is currently being explored as an upgrade to the setup is to collect into single-mode fiber instead of through an aperture. A single-mode fiber would act as a perfect spatial filter, eliminating any spatial variation in phase. This option will also be lossy, although perhaps less-so than using very small apertures without focusing.

A better solution, proposed by Altepeter, Jeffrey and Kwiat[3] was to use additional crystals after the SPDC crystal to apply a direction-dependent phase shift that exactly cancels the one obtained from the SPDC crystals. These extra crystals would in turn create an additional longitudinal walkoff that would also have to be compensated with yet another set of crystals. We attempted this for our setup using calcite crystals cut with axes at 45°, but the addition of the crystals was not found to increase the concurrence. The reason why this did not work is unclear. One orientation of the crystals was found to reduce the concurrence, but no orientation was found that could increase it. Two
Figure 2.3: Plot of the measured non-local phase $\phi$ in $\frac{1}{\sqrt{2}} (|HH\rangle + e^{i\phi} |VV\rangle)$ against iris position. The linear dependence is expected due to the transverse walkoff experience by downconversion from the first crystal in the second crystal.
Figure 2.4: Plot of the measured concurrence against the diameter of the two collection apertures. The theory is calculated based on the measured phase dependence from figure 2.3. The discrepancy indicates that some other iris diameter-independent decoherence mechanism is also playing a role.
conclusions are possible: either the transverse walkoff is not the limiting factor in the concurrence or the calculated crystal lengths were incorrect. More investigation will be needed to determine which of these possibilities is the right one.

**Alignment of two crystals**

To align the two SPDC crystals the following procedure was used:

1. Align the crystal axes with a polarization reference by putting the crystal between crossed polarizers and minimizing the transmission.

2. Polarize the pump along the axis of one of the crystals so that only one crystal generates SPDC. Align the pump laser to retroreflect off of the crystal surface.

3. Place an iris a few tens of centimeters away from the crystal at a 3 degree angle to the pump

4. Align a visible alignment laser so as to pass through both the irises and the crystal

5. Direct the alignment beam to the detectors

6. Align the lenses on the detectors by looking through the lens and trying to produce a clear image of the detector element

7. Maximize the photon counts due to the alignment beam

8. Turn off the alignment beam. If everything was done carefully there should be some counts from SPDC. Adjust the lenses and mirror so as to maximize these counts. Once the rates at the individual detectors are optimized there should be some coincident detection events and the optimization procedure can be repeated to maximize these.
9. Insert spectral filters before the detectors. Hopefully there are still some coincidence counts even with the filters in place. Adjust the crystal angles in the plane of the pump polarization and the pump direction to maximize the counts.

10. Rotate the pump polarization by $90^\circ$. Tilt the other crystal axis direction so as to maximize the coincidences seen at the detectors. If the pump polarization is now rotated to $45^\circ$, both crystals will emit and the two-photon state will be entangled.

In recent discussions with other groups working on SPDC\cite{9} it has become clear to me that crystal and detector alignment is significantly easier with single-mode fibers. With single-mode fibers, an alignment laser can be sent backwards through the fibers destined for the detectors and the focusing optics adjusted so as to obtain an intersection of the beam waists for the pump and the two alignment lasers. The waist sizes can also be adjusted so as to be the same or so as to have a smaller waist for the SPDC than for the alignment beams which is predicted to be the optimal configuration\cite{37}. The spatial frequency of the standing-wave interference pattern created by the two alignment beams gives an accurate measure of the angle between them, allowing a very precise definition of the geometry, and the overlap of the beams can be made very accurate by having all of them pass through the same pinhole. When the crystal is inserted into this setup, coincidences should be immediately observable. The next SPDC experiments undertaken in the Steinberg group should definitely try to make use of this technique. In order to make this work, a pump laser that is either fiber-coupled or circularized so as to make coupling into fiber more efficient would be a significant asset. The Coherent vioflame lasers have highly elliptical emission and could not be coupled into single-mode fiber with more than 20% efficiency, even when anamorphic prisms were used to try to correct the beam shape.
2.3.5 Source output

The type-I source has been in regular use since it was built in 2004. A typical entangled density matrix is the one reconstructed in equation 2.32, while a typical CHSH correlation measurement[30] is shown in Figure 2.5. The $90.1 \pm 0.5\%$ and $89.7 \pm 0.5\%$ visibility of these fringes is a typical signature of Bell inequality violation[7], and leads to a value of the CHSH[30] $S$ function of $2.61 \pm 0.02$, clearly violating the local realistic maximum value of $2$.

Although our source has a reasonable degree of entanglement suitable for our quantum information experiments, it still falls short of the results obtained by James et al.[55] using $500 \, \mu m$ long crystals and a narrow-band Ar-ion pump with no compensation for either longitudinal or transverse walkoff. While they achieved a concurrence of $0.963 \pm 0.016$ our concurrence was $0.86 \pm 0.02$. This number is comparable to the degree of entanglement seen by at least one other source using long crystals and a unstabilized diode laser pump[99].

While the calculations of this chapter make clear that the current setup is still showing entanglement less than what ought to theoretically be achievable, the results of James et al. make it clear that by eliminating longitudinal and transverse walkoff by having a long pump coherence length and short crystals it is possible to achieve nearly perfectly entangled states. Modifying out setup to collect into single-mode fiber with a narrow-band grating-stabilized pump ought to be able to match or exceed the quality of entanglement that they achieved.

2.3.6 Type-II beam-like emission source

The purpose of the type-II source used in the experiments in Chapter 3 and in several of the other experiments in our group[1, 105, 86] was to produce non polarization-entangled photons that would be coupled into single-mode optical fiber. The photons were produced
Figure 2.5: CHSH measurements. Half-waveplates are placed in front of polarizers at detectors A and B. The different curves are for waveplate A at angles 0°, 22.5°, 45° and 67.5° degrees (effectively creating a polarizer at the angle α in the legend). The x-axis gives the angle of waveplate B. The value the curves at angle 11.25° and 33.75° can be used to calculate the CHSH S-function. The high visibility of 90% for α = ±45° is indicative of a Bell’s inequality violation.
by SPDC of a doubled 50 fs pump pulse from a Ti:Sapph oscillator running at an 82 MHz repetition rate. For reasons to be explained in more detail in Chapter 3, it was desirable that the SPDC photons be temporally indistinguishable from the pulses produced directly from the Ti:Sapph.

In type-II phase-matching, changing the angle between the crystal axis and the pump will change the opening angle of the SPDC cones at a given wavelength. When that angle is made small enough, the cones collapse to spots, a phenomenon first documented by Takeuchi[116]. At the time the pulsed experiments in our lab were being set up it was believed that this sort of emission would prove optimal for coupling into single-mode fiber. Indeed better coupling was obtained for this geometry than for the Type-II collinear geometry, although it was difficult to determine whether coupling had been truly optimized in either situation[66]. Recent theoretical work has provided some better tools for analyzing such questions[78, 117].

2.4 Photon detection

The photodetectors used in all the experiments presented in this thesis were Perkin Elmer Single Photon Counting Modules (SPCMs) based on an actively-quenched silicon avalanche photodiode (APD) operated in ‘Geiger mode’[33]. Avalanche photodiodes are photodiodes in which electron-hole pairs are accelerated by a large applied reverse-bias field which can create additional electron-hole pairs by impact ionization. This allows a signal consisting of many electron-hole pairs to be generated by a single photon. An APD in Geiger mode has a reverse-bias field so large that it exceeds the breakdown field of silicon. In such a device a single electron-hole pair is sufficient to trigger a macroscopic current of around a milliamp which can then be picked up by standard discriminator electronics, triggering a pulse. A photon incident on a Perkin-Elmer detector will produce

\footnote{Some modifications to the laser undertaken in 2005 increased the pump pulse length to 100 fs and reduced the repetition rate to 40 MHz. This laser was then replaced with a Coherent Mira.}
a 25 ns long 4V signal capable of driving a 50Ω transmission line. While the pulse is 25 ns long, the rise time jitter between the arrival of the photon and the beginning of the pulse was measured to be between 300 and 600 ps, depending on the detector.

The avalanche nature of the detectors makes them insensitive to the strength of the incident pulse. The detectors are not number-resolving and so can only distinguish no photons from some photons. Occasionally the detectors will fire even in the absence of an incident photon. Such events, called dark counts, are due to crystal and surface defects in the semiconductor. For the detectors used in these experiments dark counts were between a few dozen and a few hundred counts per second whereas (unpaired) SPDC counts were several tens of thousands per second. For most practical purposes the
detector dark counts could be safely ignored.

The active detector area of the SPCM is a square element 100 $\mu$m on a side. For the fiber-coupled detectors used in the experiments in Chapter 3, a fiber coupler was mounted to the detector and simply butt-coupled to this element. For the free-space experiments in Chapters 4 and 5, the incoming SPDC light was coupled to the detectors through a 5 cm focal-length lens mounted on a three-axis translation stage. A magnified image of the detector element was clearly visible through the lens, a fact that was used in the alignment of the system.

2.4.1 Coincidence detection

The very tight sub-picosecond timing correlations\cite{53} between two photons in an SPDC pair allow the pairs to be very accurately filtered out from background light and detector dark counts by coincidence detection. For this to work, the pulses from the two detectors must be fed to an electronic circuit that fires when two pulses arrive at the same time. Ideally such a circuit should be able to filter out coincident detection to within the detector jitter time of $\sim$500 ps. A coincidence circuit is characterized by a coincidence window, the maximum time difference that can occur between two pulses before they are filtered out. For any coincidence window larger than zero, background light will cause some level of accidental coincidences when two background photons happen to trigger the detectors within a time less than the coincidence window. Assuming the background rates at the two detectors are uncorrelated, the rate of accidental coincidences will be

$$R_C = R_1 R_2 \tau_c$$

(2.35)

where $R_C$, $R_1$ and $R_2$ are the rates of accidental coincidence and independent detections at the two detectors and $\tau_c$ is the coincidence window. The main source of accidental coincidences is between unpaired SPDC events. For instance, in the type-I system, the conditional probability of detecting a photon at one detector given that one was detected
at the other detector ranged from 0.1% to 12%, depending on the iris sizes. This meant that the rate of unpaired SPDC photons was 10 to 1000 times higher than the rate of coincidences.

When I first arrived at the lab, coincidence detection was done by a crude discrete TTL logic gate circuit that simply applied an AND operation to the two pulses. The resulting coincidence window was 50 ns, so that with typical singles rates at the detectors of 30,000 per second the coincidence rate was 45 which represented a significant background for tomography, two-photon interference, cryptography and entanglement measurements, given that the coincidence rate was on the order of 120 coincidences per second.

In 2006 I designed and built a more sophisticated coincidence detection circuit that used a high-pass filter to isolate the rising edge of each signal and fast PECL logic to perform the AND operation between them. It also expanded capacity to measure coincidences to four channels so that light could be collected from both output ports of a polarizing beamsplitter in the analyzer. With this circuit I was able to reduce the coincidence window to a minimum of 1.2 ns, although working with such a short coincidence window reduced the overall coincidence rate. With a 3 ns window rates were unreduced, and the accidental coincidence rate brought down to 3 per second.

The schematic and printed circuit board layout of the circuit can be found in the appendices. A high-pass RC filter differentiates the input signal from each of two detectors and the resulting signal is discriminated by a high-speed comparator. The comparators drive the inputs of a PECL-logic AND gate. When the AND gate triggers, a RS latch, consisting of two PECL NAND gates, is set. An 80 MHz FPGA continuously polls these latches and resets them once they are read. The FPGA runs counters for each of the signals from the latches and the inputs and sends the counter values to a Delcom 802600 microcontroller which passes them on to the computer over USB.
2.5 Polarization manipulation

2.5.1 Polarization projection

A polarizer is an optical element that transmits one polarization component while either absorbing or reflecting the other. Typically polarizers project onto a linear basis, acting as either polarization filters or polarizing beamsplitters (PBSs).

There are four kinds of polarizer commonly used in our lab.

1. **Birefringent crystal polarizing beamsplitters and polarizers:** These polarizers are typically made of highly birefringent materials like calcite and rely on the large transverse walkoff in these materials to separate the two polarizations. Typically two crystals are cut in a prism configuration to enhance the spatial splitting between the two polarizations. In the Glan-Thomson configuration the extraordinary beam is directed into the side of the mount while the ordinary beam is transmitted with a small spatial displacement. In Glan-Taylor configuration the extraordinary beam is reflected out of the prism at an angle while the ordinary beam is transmitted, resulting in a polarizing beamsplitter. Extinction ratios of $1 : 10^5$ or better are claimed for these polarizers, making them an order of magnitude better than other polarizers. They are also extremely broadband, being limited only by the material absorption of calcite and the quality of the anti-reflection coating. This is the only type of polarizer I would recommend using in any future experiments. The only downside is the slight displacement of the transmitted beam, and for application that are sensitive to this, a plate polarizer can be used instead.

2. **Dielectric polarizing beamsplitters:** This is the type of PBS that I initially used. It consists of multiple layers of glass and relies on Fresnel reflection\([20]\) at Brewster’s angle to reflect only the $s$-polarized light. After multiple reflections the reflected light should all be $s$-polarized and the transmitted light all $p$-polarized. I
tried many such polarizers before giving up on them altogether. Not only was the polarization of the reflected light extremely sensitive to incidence angle, detection of the transmitted light did not even project onto a linear polarization state, but rather typically one with a 3 – 4% circular component. This can be seen in plot of figure 2.7 which shows the transmission of light between parallel polarizers as a quarter-waveplate is rotated between them. Figure 2.7(a) shows the transmission when the second polarizer of the parallel pair was a Thorlabs GTM10 Glan-Thompson polarizer while figure 2.7(b) shows the result when it was a Thorlabs PBS3 dielectric polarizing beamsplitter. The troughs of the curves indicate where the light output from the quarter-waveplate is right or left circularly polarized. A linear polarizer will have 50% transmission for either of these, but the dielectric polarizer preferentially transmits the one circular polarization more than the other. This kind of thing wreaks havoc on a polarization analyzer because whereas rotations of linear polarizations can be easily corrected with a half-waveplate, rotations to elliptical polarizations generally require three waveplates to fix. This means that three mutually dependent angles have to be properly set. Finding these angles was found to be an exercise in futility.

3. **Plate polarizers:** Plate polarizers work quite well, but they are lossy elements that absorb one of the two polarizations. Typically they contain a material, either polymer or metal, that conducts in one direction leading to strong absorption for one polarization and weak absorption for the other. The main advantage these have over calcite polarizers is that they do not displace the beam, making them useful for dropping in to an interferometer without affecting alignment. The plate polarizers tested were found to project onto linear bases, but with extinction ratios measurably lower than for the calcite polarizers. Measured extinction ratios varied from $10^{-3}$ for the Edmund optic Polaroid polarizers to $10^{-4}$ for the Polarcore polarizers.
Figure 2.7: Comparison of calcite and dielectric polarizers. The unequal transmission of left and right circular light for the dielectric polarizer is difficult to correct for in a polarization analysis system.

## 2.6 Polarization rotators

A polarizer makes a measurement along a ray on the Poincaré sphere. In order to obtain a complete set of polarization measurements, a given measurement must be rotated around two non-collinear axes on the Poincaré sphere. The usual way to accomplish this is to use two waveplates, a quarter-waveplate and a half-waveplate and to rotate them to different angles. With good waveplates and accurate positioning, measurements can be made accurate to 0.1% or less.

An alternative approach that we tried with some success was to use two liquid crystal waveplates at fixed angles and to vary the birefringent phase delay induced by the liquid crystal. This had the advantage of being much faster since the system had no moving parts. In general, though, the liquid crystals were found to be of poorer quality than waveplates in terms of extinction through parallel polarizers. Additional effects such as long-term temperature drift and slight non-repeatability of phase control also contributed to errors. Generally, though, these effects amounted to a 1%-2% error and were tolerable for many measurements.
Where liquid-crystal waveplates were most useful was in creating random or mixed polarization states by applying pseudo-randomly generated voltages to them. Since the waveplates could typically be switched at a rate comparable to the incident photon pair flux, a liquid crystal could turn a polarized photon pair train into completely unpolarized one.

In the following sections we will discuss the experimental considerations involved in using these active optical elements for polarization control.

2.6.1 Waveplates

2.6.2 Motorized rotation stages

In 2004 we purchased nine Newport PR50 stepper motor rotation stages with built-in optical encoders and three Newport ESP300 GPIB-enabled three-axis stepper motor drivers. The motors were rated to an accuracy of 0.1° with a precision of 0.01°. The reality was somewhat different as can be seen from the plot in figure 2.8 which shows the measured transmission through parallel polarizers with a quarter-waveplate being rotated between them through 360° according to the motor encoder. Clearly at the end of travel the waveplate is not at the same angle that it was at the beginning of travel as can be seen by comparing it to the expected sine wave. This particular encoder loses 5° degrees in 360°. Of the nine motors we had, four were accurate to within manufacturer specifications while the others lost between 2° and 5° in going through 360°. This problem was difficult to identify, in large part because errors in waveplate accuracy, the polarizing beamsplitters, the polarization alignment of the SPDC and the statistical errors inherent in single-photon counting also contributed to unexpected results. Many months of work could most likely have been saved if the motors had been fully characterized on a simple test rig before being put into the experiment.

Once the problem was understood it was possible to correct for it by only moving
Figure 2.8: Plot showing the effect of the waveplate losing 5° of angle over 360° of travel. Note that the intensity at $-135^\circ$ differs noticeably from that at $225^\circ$ and the best-fit sine curve leads the data for low angles and lags for large angles.

the waveplates to known positions from a known starting position. This was a bandaid solution that caused additional grief whenever any part of the polarization manipulation had to be re-calibrated. In retrospect, it would clearly have been better to have returned the bad motors to Newport to be fixed as soon as the problem was discovered, even at risk of delaying experimental work. The necessity of being able to set a waveplate angle accurately and repeatably to an arbitrary known position in the work in this dissertation cannot be overstated.
2.6.3 Calibration

Initial efforts at calibration of waveplate angles relied on the usual techniques for polarimetry, namely the measurement of the Stokes parameters by setting the waveplates to four known sets of angles and measuring the transmission through crossed polarizers. This was a bad approach because it failed to distinguish between errors caused by incorrect waveplate angles, errors caused by bad polarizers and errors caused by waveplate inaccuracies. Although considerable effort was put into trying to disentangle these different error sources (along with other possible error sources that were dreamed up), it ultimately proved futile.

Significant improvement was achieved by taking a massively over-complete set of measurements by stepping each waveplate through $360^\circ$ a few degrees at a time. This allowed the problems caused by the inaccurate encoder values to be distinguished from other problems and made it clear that the dielectric polarizing beamsplitters were treating left and right circular polarizations differently. This in turn allowed the effects of other optical elements in the system on polarization to be better understood. The most significant problems were caused by a set of prisms that rotated the polarization by a few degrees and dielectric mirrors that caused a relative phase shift of up to $12^\circ$ between the $p$ and $s$ polarizations. Once these effects were recognized it was possible to correct for the rotations with waveplates and for the phase shifts with liquid-crystal waveplates. The alternative approach of correcting phase shifts with thick pieces of quartz tilted away from normal incidence was found to be impractical because it induced polarization-dependent losses due to Fresnel reflections and caused decoherence due to longitudinal walkoff.

In order to calibrate waveplates accurately one needs to either calibrate out or reduce background light levels. Coincidence measurements or the use of an alignment laser with a photodiode allowed essentially background-free measurements to be made. When calibration was done using SPDC singles rates, the background had to be separately measured and subtracted out.
The final calibration method is given below. It assumes that SPDC has already been found and that alignment beams are set up to follow the same optical path as the collected SPDC.

- **Align polarizing beamsplitters with the table and the incident beam**: This is done by sending in a beam parallel with the table and making sure that surface of the PBS retroreflects it and that the beam out the reflected port is also parallel to the table.

- **Align a plate polarizer with the PBS** This is done by measuring transmission through the plate polarizer followed by the PBS. For crossed alignment, minimize the transmitted port light, for parallel alignment the reflected port light. We will assume that the plate polarizer is aligned parallel to the PBS of the remainder of the calibration.

- **Insert a rotation stage with a half-waveplate between the two polarizers and rotate through 360°**: Dirt, scratches and poor centering will generally result in a trace with no symmetry, one lobe of the sine wave being smaller than the other three, for instance. A half-waveplate with a phase delay less than or greater than 180° results in non-zero transmission even when the waveplate is 45° to the polarizer axis. More generally, a waveplate with a phase delay of $\phi$ will result in a transmission through parallel polarizers as a function of waveplate angle $\theta$ given by

$$T(\theta, \phi) = \frac{1}{4} \left( 2 \cos^2 \frac{\phi}{2} - 2 \cos^2 2\theta - \cos \theta \cos 4\theta + 1 \right)$$

(2.36)

Thus if one can trust the rotation stage to accurately set the waveplate angle, one can simply use a one-parameter fit to the transmission to obtain the true waveplate phase delay. This angle can be recorded and modeled in future polarization analysis. Using this technique, the FocTek half-waveplates in the lab have been measured to have phase delays that differ from 180° by up to 0.2°.
• **Insert a rotation stage with a quarter-waveplate between the two polarizers** The quarter-waveplate is important because it tests the linearity of the incident light and final polarizer. Again, one can plot transmission through the apparatus over the 360° rotation of the waveplate. If both the polarizers project onto a linear basis, then the sinusoidal transmission will only have one frequency component, even if the quarter-waveplate is imperfect. If, either the input polarizer or the output polarizer project onto an elliptical basis, then there will be an additional frequency component present in the transmission with twice the period of the main component (i.e. with 180° symmetry rather than 90° symmetry). If the polarizers are both linear then only a single frequency component will be present, and the resulting curve can be fit to equation 2.36. The FocTek quarter-waveplates in the lab were measured to have phase delays differing from 90° by up to 3°. Again, by keeping track of the measured phase delay, tomography procedures can be adjusted to account for it.

• **Move the first polarizer and waveplate back through the setup one optical element at a time, noting any phase shifts in the transmission curves** This will detect polarization rotations due to mirrors, beamsplitters, prisms and other elements. Pure rotations can be corrected with half-waveplates until the phase difference in the transmission curves is eliminated. To measure pure phase shifts, both the parallel polarizers should be rotated by 45° and the quarter-waveplate rotated between the optical element being tested and the final polarizer. Assuming a true quarter-waveplate, if the optical element induces a phase shift φ and the quarter-waveplate rotation angle is θ, the transmission will be

\[
T(\theta, \phi) = \frac{1}{2} + \frac{1}{4} \cos \phi - \frac{1}{8} \cos (\phi - 4\theta) - \frac{1}{8} \cos (\phi + 4\theta) - \frac{1}{4} \sin (\phi - 2\theta) - \frac{1}{4} \sin (\phi + 2\theta)
\]  

(2.37)

The phase shift φ can be found by fitting to this curve.
• When all the optical elements have been corrected for, adjust the angle of the crystal to minimize SPDC transmission through the detector PBS: This should get the SPDC polarization aligned with the polarizer axes. Since the SPDC is polarized orthogonal to the crystal axis, it is very linear and will align well with the polarizer axis. A problem may occur if the two crystals aren’t at exactly 90°. For the two crystals currently in the lab this misalignment appears to be 0.3°, which means that only one of the two crystals can be aligned perfectly with the polarizer axis. The misalignment of the other crystal is negligible for most purposes.

2.6.4 Liquid crystal waveplates

For some experiments it was necessary to have the ability to switch quickly between two polarizations. Ideally this would have been done with an electro-optic modulator (EOM), but due to the high price of these devices we opted to use liquid crystal waveplates (LCWPs) instead. LCWPs are much slower than EOMs, having a response on the order of milliseconds rather than nanoseconds, but are easy to drive and can be purchased for under $1000. They are a useful addition to the polarization control toolbox because, unlike waveplates, they have a variable phase delay. This makes them ideal for correcting phase delays that accumulate in the system as well as providing reasonably fast switching between different polarization states. A downside is that the delay they impart varies non-linearly with the applied voltage, making calibration and interpolation more involved. In addition, when placed between parallel calcite polarizers at 45° the LCWPs we tested could not achieve the same isolation as half-waveplates, showing a minimum transmission of 0.6%. Additionally, the LCWPs displayed some longterm drift in phase-shift at a given applied voltage. This is likely due to the 0.4%/° C temperature dependence of the crystals[84] and the lack of temperature control in the lab.

We have purchased five LCWPs from BolderVision Optik and four LCWPs from
Meadowlark. Devices from both companies had similar performance in terms of retardation, but the mechanical construction of the Meadowlark devices was much better. Most notably, the BolderVision LCWPs induce a beam deflection of 2°, presumably due to poor alignment or manufacture of the optical flats. The Meadowlark devices do not do so. Practically this meant that inserting a BolderVision device in the input arm of an interferometer generally destroyed the interference whereas doing the same with a Meadowlark device did not.

Liquid crystal waveplates are based on nematic liquid crystals. These are anisotropic polymers for which it is energetically favourable for each molecule to align with its neighbours. Liquid crystal material can be sandwiched between two optical flats structured in such a way as to induce alignment of the crystals along a particular direction. When a voltage is applied across the two flats, the liquid crystal axis tilts to align with the field and the birefringent phase shift induced on light by the crystal is reduced. Even at very large applied voltages there remains a residual phase shift which can be compensated by attaching a fixed retarder to the LCWP. In this way commercial LCWPs can achieve a variable retardance that ranges, depending on the unit, from a little less than 270° to nearly 360° at our operating wavelength of 810 nm. The liquid crystals can be damaged by electromigration resulting from the application of a DC electric field. For this reason the LCWPs needed to be driven with an AC field. An AC field defines an alignment direction for the crystals, but without inducing electromigration.

Figure 2.9 shows some typical coincidence data when an LCWP with its axis at 45° is used to rotate the pump polarization while one polarization of SPDC is collected. The plot shows the pronouncedly non-linear response of the waveplates phase delay with applied voltage. The LWCPs also show some small amount of hysteresis which was corrected for by always making transition from low voltages to high voltages.
Figure 2.9: Plot showing the coincidence rate of horizontally polarized SPDC photons while a voltage was applied to a LCWP with its axis at 45° to the pump polarization. The x-axis gives the peak-to-peak voltage of a 2KHz square wave applied across the liquid crystal.

### 2.6.5 Liquid crystal waveplate driver

Following the recommendations of the engineer at BolderVision, I designed a driver for the LCWPs that produced a 2KHz AC square wave with a variable peak-to-peak voltage of between 0 and 30V. The voltage was referenced to a 10-bit MAX5201 digital-to-analog converter with a temperature-controlled voltage-reference. The reference voltage was modulated with a MAX4614 analog switch and increased to the desired level with an OPA445 high-voltage op-amp. The whole unit was controlled over USB through a Delcom 802600 microcontroller. The unit could control the voltage applied to four
different LCWPs. Three units were built, and although one of them has two output ports that never worked, the three units have been in continual operation for two years without any issues. The schematic and PCB layout for the driver can be found in the appendices.

The response time of the LCWPs is faster when going from a low to a high voltage where the applied field drives the motion of the crystal molecules than when going from high to low voltage when the crystal must relax thermally into its equilibrium state. Figure 2.10 shows the measured transmission through crossed polarizers with a BolderVision waveplate between them at 45° to the polarizer axes. The 20% to 80% rise time in response to an applied voltage was 800 µs whereas for the removal of the applied voltage it was 13 ms.

2.7 Polarization analysis

We have now described all the elements that go into a polarization analysis system. The standard approach to measuring the polarization of light is to put a quarter and half-waveplate in front of a polarizer or polarizing beamsplitter. Table 2.2 shows the usual waveplate settings followed by the quantum projector implemented by detection of a photon at the transmission path of the beamsplitter: The single-qubit density matrix[91] or, equivalently, the Wolf coherence matrix[81] can be constructed directly from these values. In fact, because these measurements form a set of mutually unbiased bases (see chapter 4), the density matrix can be written directly in terms of their expectation values.
Figure 2.10: Rise and fall times for LCWP waveplate transmission when a step voltage is applied and removed. The fall time is longer because the liquid crystal must relax back to the equilibrium state.

\[ \rho = \langle H | \rho | H \rangle \cdot |H\rangle \langle H| + \langle V | \rho | V \rangle \cdot |V\rangle \langle V| + \langle D | \rho | D \rangle \cdot |D\rangle \langle D| + \langle A | \rho | A \rangle \cdot |A\rangle \langle A| + \langle R | \rho | R \rangle \cdot |R\rangle \langle R| + \langle L | \rho | L \rangle \cdot |L\rangle \langle L| - I_2 \]

where \( I_2 \) is the two-by-two identity matrix.

It is easy to extend this polarimetry scheme to two-photon polarization states following the method of James et al.\[55\]. The measured quantity becomes the rate of coincidence counts between different polarizations for the two photons. This can be
measured as the rate of coincident detection of photons passing through two independently set single-photon polarization analyzers. The two-qubit density matrix has fifteen independent elements plus an overall normalization, so it is sufficient to measure 16 linearly-independent coincidence rates obtained by setting the two analyzers to sixteen specific pairwise combinations of the settings in Table 2.2. Later it was pointed out by Altepeter et al. [3] that better results could be obtained if all 36 pairwise combinations of measurements in Table 2.2 were used. Later tomography work in the group took advantage of this insight.

In the type-II apparatus, waveplates were used in the same manner, but the interpretation of the results was more complicated. This will be explained in detail in Chapter 3.

Some tomography work was also done using the liquid crystal waveplates. Two LCWPs were placed in front of a polarizing beamsplitter with their axes at 22.5° and 45° in the order LCWP at 22.5°, LCWP at 45°, polarizer. The phase delays at the two LCWPs were set to the values in table 2.3.

<table>
<thead>
<tr>
<th>QWP</th>
<th>HWP</th>
<th>Projector</th>
</tr>
</thead>
<tbody>
<tr>
<td>0°</td>
<td>0°</td>
<td>$</td>
</tr>
<tr>
<td>0°</td>
<td>45°</td>
<td>$</td>
</tr>
<tr>
<td>45°</td>
<td>22.5°</td>
<td>$</td>
</tr>
<tr>
<td>45°</td>
<td>−22.5°</td>
<td>$</td>
</tr>
<tr>
<td>0°</td>
<td>22.5°</td>
<td>$</td>
</tr>
<tr>
<td>0°</td>
<td>−22.5°</td>
<td>$</td>
</tr>
</tbody>
</table>

Table 2.2: Settings for polarization analysis with waveplates
2.8 Two-photon interference

Since Hong, Ou and Mandel first demonstrated it in 1987, two-photon interference or Hong-Ou-Mandel (HOM) interference has become a vital experimental tool in all branches of quantum optics. It has been used for applications as diverse as measuring tunneling times\[112\], compensating dispersion in white-light interferometry \[111\], teleportation\[22\], implementing quantum logic gates\[92\], and making GHZ states\[93\].

To the cynic it can often seem that all of experimental linear quantum optics consists of two curves, the sine-wave and the HOM dip. In this section we will discuss the alignment of the Hong-Ou-Mandel interferometer, the effects of experimental imperfections and some unpublished results obtained over many hours spent trying to make the HOM behave as it was supposed to.

### 2.8.1 Alignment

The Hong-Ou-Mandel effect is obtained when two time and frequency-correlated beams of photons are incident on a beamsplitter in such a way that the outputs match perfectly in direction, position and time. Setting up such an interferometer in free space requires precision alignment of beams with low photon fluxes, making it an experimentally non-
trivial task. A major advantage fiber-coupled sources over the sources described in this chapter is that the well-defined spatial modes of a fiber-system make aligning Hong-Ou-Mandel interferometers much easier.

**Iris scanning**

The trick to aligning the HOM is to accurately find and overlap the two beams that are reflected and transmitted from the two input ports of the beamsplitter. For this, two irises are required in one output arm, one as close to the beamsplitter as possible, and the other as close to the detector as possible. If the beams overlap at both the irises then they are said to be mode-matched. The beamsplitter is mounted on a translation stage and on a kinematic mount with two tilt axes. The two beams can be overlapped at the beamsplitter by translating the stage to the point where they cross and at the detector by changing the angle on the kinematic mount.

The SPDC at the two input ports can be labeled using polarization. If one input port sends in H light and the other V light then when the irises are scanned across the beam the H detector will give the profile of one input port and the V detector of the other. The position and angle of the beamsplitter can be adjusted iteratively to achieve overlap of 50 $\mu$m or better between the two beams which is the requirement to see two-photon interference in the current setup.

Overlap in the vertical direction is trickier. While the iris near the beamsplitter can still be used to judge positional overlap, there is no independent degree of freedom that can be used to adjust it. The only way is to change the angle of a mirror well before the beamsplitter and then adjust the beamsplitter angle to compensate. Even once this is done, difficulties can arise since tilting the beamsplitter vertically just makes each side look at different positions on the SPDC cone without changing the SPDC rate. In practice it was found that the vertical alignment was best achieved by carefully aligning an alignment beam to be parallel to the table and to the pump beam and using this as a
reference for setting the height of one of the detector irises. The other iris could then be scanned vertically to maximize coincidences. By carefully matching the alignment beams at the two detector irises, HOM interference will usually be seen if the other degrees of freedom are sufficiently well-matched.

In order to interfere, the beams must also overlap longitudinally to within less than 30 $\mu$m. This was achieved by balancing the path lengths of alignment beams before and after the crystal using white light interferometry. While the coherence length of the alignment laser is too long for the presence of interference to be a useful guide to longitudinal overlap, if the laser is turned down below threshold, it becomes essentially no different than an LED and the coherence length decreases to 100 $\mu$m or so. The two alignment beam paths can be made equal before the crystal by imaging the crystal plane onto a webcam with a lens and observing the standing-wave interference. As the laser is turned down below threshold the standing-wave interference will only persist if the pre-crystal path-lengths are equal to within 100 $\mu$m. Similarly, a detector placed at the output of the beamsplitter will record Mach-Zehnder interference whenever the interferometer is disturbed by, say, tapping a mirror. This interference will disappear when the laser is turned below threshold unless the path lengths from the crystal to the beamsplitter are equal to within 100$\mu$m.

With position, direction and time all overlapped one can scan any one of these degrees of freedom to observe a characteristic dip in the coincidence rate when photons of the same polarization are sent into the beamsplitter. In our experiments we scanned the longitudinal degree of freedom by translating a retroreflecting prism on a New Focus picomotor-powered translation stage.

While not fool-proof, this alignment procedure succeeded most of the time in generating HOM interference. Figure 2.11 shows the result of a typical HOM scan. Once the HOM dip is located, the same degrees of freedom can be fine-tuned to maximize the visibility. The highest visibility achieved was 94% as compared to the maximum achievable
Figure 2.11: Measured rate of coincidences as the delay in one input arm of the HOM interferometer is scanned

visibility of 96% given our beamsplitter splitting ratio. The most likely explanation for this 2% discrepancy is that in going through different optics the two input beams acquire different spatial profiles and do not interfere perfectly. It will be interesting to see if the discrepancy disappears when the system is upgraded to single-mode fiber.

2.8.2 Dips and antidips

Often, after coarse-aligning for Hong-Ou-Mandel interference, a scan of delay in one of the input arms resulted in a peak rather than a dip, even when the two photons had the same polarization. An example of this is shown in figure 2.12(a). This can be explained by considering the multi-mode nature of the collection system. A misalignment
of the beamsplitter in the vertical direction will couple SPDC at a different angle for the transmission-transmission path than for the reflection-reflection path. This slight angle mismatch of 1 mrad between the two coincidence amplitudes results in additional phase being acquired for the reflection-reflection path. The additional phase alters the usual $\pi$ phase difference between transmission-transmission and reflection-reflection events to some other value that depends on the angle of the beamsplitter. Figure 2.12(b) plots the visibility of the HOM dip as a function of iris position. It trends towards a sinusoidal oscillation. This is the same phenomenon found in a misaligned Mach-Zehnder interferometer where a small angular error away from optimal alignment gives rise to a standing-wave interference pattern consisting of sinusoidal fringes. Here, though, the interference is only visible in the coincidence rate between the two detectors and not in the singles rate.

Unfortunately, after discovering this new interference effect, we learned that it had been published by another group a few months earlier[59].

### 2.8.3 Crystal walkoff

Photons produced in the two SPDC crystals exhibited a HOM dip minimum for slightly different relative path length differences. The observed 20 $\mu$m dip displacement can be explained by considering that the two photons generated in the early crystal are extraordinarily polarized in the second crystal. Since in traveling through this second crystal they make different angles with the optical axis, the two photons will see different indices of refraction and hence different group velocities. Thus while two SPDC photons from the second crystal can be expected to emerge at the same time, and hence have an HOM dip at zero path length difference, two SPDC photons from the first crystal will be offset, and will exhibit a corresponding shift in the path-length difference for maximum dip visibility. The calculated relative group delay difference for the two photons from the first crystal due to the angle-dependent refractive index in the second crystal was 17 $\mu$m,
(a) A Hong-Ou-Mandel antidip with a visibility of 55%

(b) Visibility of the Hong-Ou-Mandel antidip as a function of vertical iris position. The sinusoidal variation is analogous to the standing-wave interference seen in a misaligned Mach-Zehnder interferometer.
in reasonable agreement with experiment. Once it was understood, this mismatch in dip locations was easily corrected with a 1.5 mm thick piece of quartz chosen to compensate the delay between the two photons from the first crystal.

Another effect was that the location of the beamsplitter for which the dip visibility was maximized differed by approximately 150 $\mu$m for the two crystals. Figure 2.8.3 shows the result of scanning the beamsplitter location on dip visibility, while 2.8.3 shows the explanation. Since the SPDC from the first crystal will walk off horizontally in the second crystal, we expect an effective displacement between the two sources. When parallel rays emanating from these sources are traced out it becomes apparent that they meet at horizontally displaced locations which is exactly what we observe. This provides an interesting method of measuring the walkoff in the crystal, but presents an impediment to using two-photon interference as a polarization filter for states that are superpositions of the output of both crystals. As a compromise, the beamsplitter was usually located to the point halfway between the two visibility maxima. At that point the Hong-Ou-Mandel visibility was equal to 86% for both the crystals.
Figure 2.13: Diagram explaining the dependence of HOM visibility on beamsplitter location. The horizontal photons walk off in the crystal that generates vertical photons. This leads to a displacement in the location of the minimum dip visibility as the beamsplitter is scanned horizontally.

2.8.4 Type-II apparatus

Two-photon interference also occurs when two photons of different polarizations are put into the same spatio-temporal mode. For example, if a horizontal and a vertical photon are put into the same spatio-temporal mode, the state is described by

$$|\psi\rangle = a_H^\dagger a_V^\dagger |\text{vac}\rangle.$$  

(2.38)

When that same state is measured in the diagonal basis it is convenient to rewrite it as

$$|\psi\rangle = \frac{1}{2} \left( a_D^\dagger + a_A^\dagger \right) \left( a_D^\dagger - a_A^\dagger \right) |\text{vac}\rangle$$  

(2.39)

$$= \frac{1}{2} \left( a_D^\dagger a_D^\dagger - a_D^\dagger a_A^\dagger + a_A^\dagger a_D^\dagger - a_A^\dagger a_A^\dagger \right) |\text{vac}\rangle$$  

(2.40)

$$= \frac{1}{2} \left( a_D^\dagger a_D^\dagger - a_A^\dagger a_A^\dagger \right) |\text{vac}\rangle.$$  

(2.41)

The interference that occurs between the two $a_A^\dagger a_D^\dagger$ terms is exactly analogous to the interference that occurs between the reflected and transmitted amplitudes in the Hong-Ou-Mandel effect. It is even possible to generate a dip by delaying one of the photons.
relative to the other. If the photons are temporally distinguishable then the interference does not occur, and so the rate of coincidences between $A$ and $D$ is determined by classical probability theory, whereas when the two photons are spatio-temporally indistinguishable the two-photon interference eliminates coincidences between $A$ and $D$. The dip visibility on the type-II apparatus was $98 \pm 1\%$. Most likely it was higher than on the type-I apparatus because of an effectively perfect 50/50 splitting ratio for the 'beam-splitter' (really just a polarizer in the $45^\circ$ basis) and better indistinguishability due to spatial filtering by the single-mode fiber. Further discussion of this phenomenon is left to Chapter 3 where its relationship to distinguishability and symmetry of the photon state is discussed in detail.

2.9 Summary

We have presented the main experimental methods used in measuring the quantum state of light using tomography and in performing experimental work discussed in this thesis. These methods allow two-photon SPDC states to be generated and collected, allow their polarizations to be manipulated and allow the observation of two-photon interference. They will be used throughout the remainder of this thesis.
Bose's theory of radiation and my analogous theory of ideal gases have been re-proved by Mr. Ehrenfest and other colleagues because in these theories the quanta or molecules are not treated as structure statistically independent of one another, without this circumstance being especially pointed out in our papers. This is entirely correct. If one treats the quanta as being statistically independent of one another in their localization, then one obtains the Wien radiation law; if one treats the gas molecules analogously, then one obtains the classical equation of state for ideal gases, even if one otherwise proceeds exactly as Bose and I have[42].

—Albert Einstein

Quantum statistics, a subject which has played a central role in the development of quantum mechanics, is only rarely mentioned in quantum information. It is usually not necessary to know whether a qubit is a boson or a fermion in order to understand a quantum circuit. On the contrary, one of the central achievements of quantum information science has been to abstract away the particular physical systems in which quantum circuits are implemented and so to select model systems where the fermionic or bosonic character of the particles involved plays no role. The circuit model of quantum computing wherein each qubit sits on a separate rail and interacts with other qubits via gates leaves no room for the exchange effects of quantum statistics.
While this abstraction has been useful in teasing out the information-theoretic aspects of the quantum theory, when it comes to real-world implementations of quantum ideas, the particles that carry quantum information have a quantum statistical nature that often cannot be ignored. The quantum-statistical nature of real-world particles is important in many systems ranging from superconductors to neutron stars, so it should come as no surprise that it also arises in linear optics quantum information[62]. In linear optics quantum computing the ‘interactions’ between photons are regulated by the quantum statistical Hong-Ou-Mandel effect. While post-selection is used to select those experimental trials where only one photon occupies each output mode of the beamsplitter, those occasions where the post-selection ‘fails’ result in multiple photons occupying a single spatio-temporal mode. This means that quantum statistics lies just below the surface of most post-selected linear quantum optics experiments where photons carry the qubits, including the demonstration of quantum teleportation[22], the non-deterministic CNOT[92], the generation of GHZ states[93] and the one-way quantum computer[120].

For this reason, it is important that techniques developed for quantum state estimation and quantum state tomography be made compatible with the quantum-statistical nature of states. This, as it turns out, is a non-trivial task, and one to which the remainder of this chapter will be devoted. Along the way we will establish an operational distinction between fundamental indistinguishability and indistinguishability caused merely by an inability to measure differences. We will introduce an accessible density matrix for an experimental system that contains all the information about a system that can be ascertained given a limited number accessible of degrees of freedom for a particle. Using this tool, we will be able to determine how distinguishing information in the state will affect all future measurements. Finally, we will put this theory into practice by applying it to two and three-photon systems.
3.1 Two-mode systems

Before we can examine the effects of distinguishability and quantum statistics in systems of few photons, we need an adequate description of photons as indistinguishable particles. The quantum theory of optics treats each optical mode (i.e. each linearly independent solution to Maxwell’s equations) as an independent linear harmonic oscillator. The quantum excitations of these oscillators are called photons. The orthogonal polarization modes $|H\rangle$ and $|V\rangle$ represent one important set of modes; other modes include, for instance, the Hermite-Gauss modes of Gaussian cavities and the different arms of a Mach-Zehnder interferometer.

We consider a system with a fixed number of photons $N$. These photons can be distributed between two modes in any way so that the quantum state will be a superposition of states of the form $|m, N - m\rangle$ where $m$ is the number of photons in the first mode and $N - m$ is the number of photons in the second mode.

It was first pointed out by Schwinger\[101\] that two uncoupled simple harmonic oscillator modes such as two optical polarization modes, are formally equivalent to a spin system. If one of these modes is the horizontal polarization mode and the other the vertical mode, we can define the usual raising and lowering operators with the commutation relations $[a_H, a_H^\dagger] = 1$ and $[a_V, a_V^\dagger] = 1$, but with $[a_H, a_V^\dagger] = 0$ since the two modes do not interact. We can also define operators that couple the modes together $J_+ = \hbar a_H^\dagger a_V$ and $J_- = \hbar a_V^\dagger a_H$. Rather than commuting like $a$ and $a^\dagger$, these two operators satisfy $[J_+, J_-] = 2\hbar J_z$ where $J_z = \frac{\hbar}{2} \left( a_H^\dagger a_H - a_V^\dagger a_V \right)$. It can also be shown that $[J_z, J_\pm] = \pm \hbar J_\pm$. This set of commutation relations is identical to the ones for angular momentum, and so the entire quantum mechanical machinery devoted to analyzing angular momentum can also be applied to these states.

For the specific case of photon polarization, the angular momentum operators have historically been replaced with Stokes operators\[107\] which give the total number of photons ($S_0$) and the degree of polarization along the $D - A \ (S_1)$, $R - L \ (S_2)$ and $H - V$
(S₃) axes of the Poincaré sphere

\[S_0 = a_H^\dagger a_H + a_V^\dagger a_V\]  \hspace{1cm} (3.1)

\[S_1 = \frac{1}{2} \left( a_H^\dagger a_V + a_V^\dagger a_H \right)\]  \hspace{1cm} (3.2)

\[S_2 = \frac{1}{2i} \left( a_H^\dagger a_V - a_V^\dagger a_H \right)\]  \hspace{1cm} (3.3)

\[S_3 = a_H^\dagger a_H - a_V^\dagger a_V.\]  \hspace{1cm} (3.4)

### 3.1.1 Characterizing two-mode states of light

We now have a formalism capable of describing multi-photon polarization states as angular momentum states. It seems that it should be easy to adapt the methods of quantum state tomography to this description and to thereby measure the density matrix on the angular momentum Hilbert space. A problem arises, though, when one considers what measurements are needed to perform the tomography. Take, for example, the simple case of a system of two photons. Via the Schwinger formalism, the polarization can easily be seen to form a spin-1 system. If we were to do a projective measurement onto the diagonal elements of this system we would project the state onto

\[a_H^\dagger a_H, a_H^\dagger a_V = a_V^\dagger a_H \text{ and } a_V^\dagger a_V\]

or, rewriting these second-quantized states as first-quantized Dirac kets through the use of Clebsch-Gordan coefficients,

\[|j = 1, m = 1\rangle = |HH\rangle\]  \hspace{1cm} (3.5)

\[|j = 1, m = 0\rangle = \frac{1}{\sqrt{2}} (|HV\rangle + |VH\rangle)\]  \hspace{1cm} (3.6)

\[|j = 1, m = -1\rangle = |VV\rangle.\]  \hspace{1cm} (3.7)

The problem lies in the \(|j = 1, m = 0\rangle\) term which is a maximally-entangled state of the two single-photon polarizations. It is a little strange to think of single-photon polarizations when the photons are fundamentally indistinguishable, but in this case we can think of a simple operational meaning for single-photon polarization. We can deterministically split the \(H\) and the \(V\) photons at a polarizing beamsplitter into two different spatial
modes 1 and 2, and then the particles in these two modes can be safely thought of as
qubits and the polarization correlations between them will be exactly those predicted for
the maximally-entangled Bell state $|\psi^+\rangle$.

If we reverse this reasoning then it seems that simply measuring one of the photons to
be $H$ and the other to be $V$ should be enough to project onto the state $|j = 1, m = 1\rangle$, but
this seems too good to be true. Typically projections onto maximally-entangled states
are difficult and involve multi-photon interference and post-selection. Clearly no such
interference or post-selection is needed if all we are doing is counting $H$ and $V$ photons.
Can such a measurement really be viewed as a projection onto a maximally-entangled
state?

Let us assume that our measurement detects only polarization, but is completely in-
sensitive to the other properties that may define different modes for the two photons.
Now imagine that one of the photons is blue and the other is red. Then, rather than
having a single state $a_H^\dagger a_V^\dagger$, one has two different states $a_{H,b}^\dagger a_{V,r}^\dagger$ and $a_{H,r}^\dagger a_{V,b}^\dagger$. Surpris-
ingly, this difference in the colour of the two photons can lead to different results in the
polarization measurements.

Say, for example that we measure the rate at which we will detect both photons to
be diagonally polarized. We can rewrite $a_H^\dagger a_V^\dagger$ as

$$a_H^\dagger a_V^\dagger |\text{vac}\rangle = \left[ \frac{1}{\sqrt{2}} (a_D^\dagger + a_A^\dagger) \right] \left[ \frac{1}{\sqrt{2}} (a_D^\dagger - a_A^\dagger) \right] |\text{vac}\rangle = \frac{1}{2} (a_D^\dagger a_D^\dagger + a_D^\dagger a_A^\dagger - a_A^\dagger a_D^\dagger - a_A^\dagger a_A^\dagger) |\text{vac}\rangle
= \frac{1}{\sqrt{2}} (|2_D, 0_A\rangle - |0_D, 2_A\rangle). \quad (3.8)$$

The probability of the state containing two diagonal photons can immediately be seen
to be 50%. On the other hand if the two photons are different colours so that the state
is $a^\dagger_{H,b}a^\dagger_{V,r}$ then we have

$$a^\dagger_{H,b}a^\dagger_{V,r}|\text{vac}\rangle = \left[ \frac{1}{\sqrt{2}} \left( a^\dagger_{D,b} + a^\dagger_{A,b} \right) \right] \left[ \frac{1}{\sqrt{2}} \left( a^\dagger_{D,r} - a^\dagger_{A,r} \right) \right] |\text{vac}\rangle$$

$$= \frac{1}{2} \left( a^\dagger_{D,b}a^\dagger_{D,r} + a^\dagger_{A,b}a^\dagger_{A,r} - a^\dagger_{A,b}a^\dagger_{D,r} - a^\dagger_{A,b}a^\dagger_{A,b} \right) |\text{vac}\rangle$$

$$= \frac{1}{\sqrt{2}} \left( |1_{D,b},1_{D,r},0_{A,b},0_{A,r}\rangle + |1_{D,b},0_{D,r},0_{A,b},1_{A,r}\rangle \right) \quad (3.9)$$

$$- |0_{D,b},1_{D,r},1_{A,b},0_{A,r}\rangle - |0_{D,b},0_{D,r},1_{A,b},1_{A,r}\rangle \right). \quad (3.10)$$

In this case the probability of obtaining two diagonal photons is 25%. Even though our measurement apparatus is insensitive to colour, the mere existence of distinguishing colour information can affect the outcome of polarization measurements.

It should be clear from this example that counting $H$ and $V$ photons does not project the state onto $|j = 1, m = 0\rangle$. Instead it projects the state onto the subspace of states having one $H$ and one $V$ photon, or equivalently onto the space of states with $m = 0$. Although for indistinguishable photons the only polarization state with $m = 0$ is the $|j = 1, m = 0\rangle$ state, if the photons are distinguishable the state $|j = 0, m = 0\rangle$ is also allowed. Insofar as the goal of tomography is to provide the best estimate of the quantum state, it cannot be assumed from the outset that the photons are indistinguishable. Rather tomography should measure all the information available about the state and try to determine from that whether the photons are indistinguishable or not.

If this seems a trivial insight, consider that the first quantum state tomography scheme[18, 17] applied to states of two ‘indistinguishable’ photons only characterized the $j = 1$ subspace and assumed that the measurement of an $H$ and a $V$ photon constituted a projection onto the state $|j = 1, m = 0\rangle$. A careful examination of the data in [17] shows rather poor agreement between the density matrices and the measurements which may possibly be due to this incorrect assumption.

This makes state estimation for indistinguishable particles fundamentally different from state estimation for distinguishable particles. With distinguishable particles one can safely ignore those degrees of freedom that don’t encode information. For example
when qubits are encoded in polarization one can safely ignore wavelength and spatial mode degrees of freedom of the photon and measure the polarization density matrix. If the polarization state is correlated to these other degrees of freedom, the result is simply a reduction in the purity of the state.

For indistinguishable particles that may have hidden distinguishing information, the unmeasured degrees of freedom can not only affect purity, but can even change the effective size of the Hilbert space! As we have seen, for two particles, instead of only having to consider the $j = 1$ space, distinguishing information can expand the state into the $j = 0$ space as well. Luckily, with a few group-theoretic tricks, the effect of distinguishing information can be understood within the context of the density matrix description of the quantum state. Through the next several sections we will develop the group theoretical tools needed to understand distinguishable and indistinguishable quantum particles and use them to develop a new framework for quantum state tomography that can address cases of particles with hidden distinguishability.

### 3.1.2 Symmetric measurements

Let us assume that we are given one end of an optical fiber. At the other end is a black-box source emitting pulses containing $N$ photons. We have equipment capable of separating the polarization modes and counting how many photons are in each mode. We also have waveplates that allow us to perform arbitrary unitary rotations of the polarization modes prior to measurement. What can we learn about the state of the photons from our measurements? Or, slightly rephrased, what range of density matrix descriptions of the state will be consistent with all the possible measurements taken with this system?

To answer this question we first need to consider what range of states are allowed by the laws of physics. The most important restriction on the allowed states is given by the bosonic nature of photons which implies that if all of the properties of two of the photons
in the state are exchanged then the state vector must be unchanged. Determining which
states satisfy this requirement will necessitate a departure into group theory, the relevant
aspects of which will be developed in the next few sections.

3.1.3 Group theory

The symmetric group

The possible permutations of $N$ particles form a mathematical group called the symmetric
group denoted by $S_N$. As a reminder, the statement that the permutations form a group
means that they are closed under composition, that all elements have an inverse, that
there is an identity element and that composition is associative. This can easily be seen
by considering the possible permutations of three numbers, $(1, 2, 3)$. Permuting the first
two numbers gives $(2, 1, 3)$, the second two $(1, 3, 2)$, the first with the second, the second
with the third and the third with the first $(3, 1, 2)$ and so on. In fact the permutations
can be written down in just this way, by labeling how each permutation in $S_N$ affects the
order of $N$ particles. For example, the six permutations in $S_3$ can be written as:

$$
\begin{pmatrix}
1 & 2 & 3 \\
1 & 2 & 3
\end{pmatrix},
\begin{pmatrix}
1 & 2 & 3 \\
2 & 1 & 3
\end{pmatrix},
\begin{pmatrix}
1 & 2 & 3 \\
3 & 1 & 2
\end{pmatrix},
\begin{pmatrix}
1 & 2 & 3 \\
1 & 3 & 2
\end{pmatrix},
\begin{pmatrix}
1 & 2 & 3 \\
2 & 1 & 3
\end{pmatrix},
\begin{pmatrix}
1 & 2 & 3 \\
3 & 2 & 1
\end{pmatrix}.
$$

(3.11)

(3.12)

(3.13)

One may obtain a more compact expression of these permutations by resolving them into
‘cycles’. A cycle $(r_1, r_2, \ldots, r_n)$ is a permutation that replaces each element $r_k$ by the
element $r_{k+1}$ that follows it except for the last element $r_n$ which is replaced by $r_1$. The
six permutations in $S_3$ listed above can be resolved into cycles as

$$\{(1)(2)(3), (12)(3), (1)(23), (13)(2), (123), (321)\}. \quad (3.14)$$

We say that two permutations containing the same number of cycles and whose cycles are of the same length belong to the same class. This is to say that they can be transformed into each other by a ‘symmetry transformation’ that maps each permutation element to another permutation element of the same class.

There are three permutation classes in $S_3$, one containing $(1)(2)(3)$, one containing $(12)(3), (1)(23), (13)(2)$ and one containing $(123), (321)$.

**Representations**

A representation of a group is a mapping between the elements of the group and a set of matrices that have the same relationships under multiplication as the group elements do under composition. It can be shown that any group can be represented by a set of unitary matrices and in particular, the symmetric group $S_N$ can always be represented by a set of symmetric matrices (which gives the group its name). To give an example for
S₃, we can make the following correspondences between permutations and matrices:

\[(1)(2)(3) \rightarrow \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \] (3.15)

\[(12)(3) \rightarrow \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \] (3.16)

\[(1)(23) \rightarrow \begin{pmatrix} -\frac{1}{2} & \frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3} & \frac{1}{2} \end{pmatrix} \] (3.17)

\[(13)(2) \rightarrow \begin{pmatrix} -\frac{1}{2} & -\frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3} & \frac{1}{2} \end{pmatrix} \] (3.18)

\[(123) \rightarrow \begin{pmatrix} -\frac{1}{2} & \frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3} & -\frac{1}{2} \end{pmatrix} \] (3.19)

\[(321) \rightarrow \begin{pmatrix} -\frac{1}{2} & -\frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3} & -\frac{1}{2} \end{pmatrix} \] (3.20)

The reader can check that the matrices have the same multiplicative relations as the permutations to which they correspond.

The representation of S₃ given above is by no means unique. One could easily construct another set of matrices having the same multiplicative relations by, for example, taking each matrix A and constructing a three-by-three matrix

\[
\begin{pmatrix}
A & 0 \\
0 & 1
\end{pmatrix}
\] (3.21)

Clearly such a transformation adds extra dimensions to the matrices without changing the multiplicative relationships between the matrices. Such representations are called reducible, while representations that cannot be put into the form of 3.21 through a unitary transformation on the representation elements are called irreducible representations or irreps for short.
Even more simply, one could assign each permutation to the number 1. Since $1 \times 1 = 1$, all the multiplicative relations between the permutations are trivially satisfied. This is called the trivial irrep. While this representation is irreducible it is called unfaithful whereas a representation like table 3.15 that assigns a unique matrix to each group element is called faithful.

It can be shown that the number of distinct irreducible representations of a group is equal to the number of classes in the group, that is to say the sets of group elements that can be exchanged without changing the multiplicative relationships in the group. Again, for the symmetric group the number of classes is simply the number of different possible lengths and numbers of cycles.

The dimensionality of an irrep refers to the dimensionality of its matrices. Thus the irrep of table 3.15 has dimension two while the trivial irrep has dimension one.

For any symmetric group $S_N$ there will always be two one-dimensional irreps, the trivial representation and another irrep called the alternating irrep that can be obtained by assigning to each group element either the number 1 or -1 consistent with the multiplicative relationships. The name alternating comes from the fact that one can properly assign 1 to all elements obtained by an even number of permutations of two particles and -1 to those obtained through an odd number of permutations. All the other irreps of $S_N$ are multidimensional.

We mention these two irreps because they play a special role in quantum statistics. It is a fact of nature that fundamental particles are identical. That is to say that no physical measurement can tell apart two multiparticle quantum states in which the two of the particles are exchanged along with all the properties of those particles (spatial, temporal and polarization properties for photons, say). In group theory language, all physical observables of the system must be invariant under the action of any permutation operator. This means that the quantum states of the system must be invariant up to a global phase. At most, a permutation of all the degrees of freedom of two particles will multiply a state
by a unit modulus constant. Thus the states must transform according to one of the two
one-dimensional irreducible representations of the symmetric group. We call particles
transforming under the trivial representation bosons and particles transforming under
the alternating representation fermions. When the quantum state of a system transforms
according to a particular irrep of a group we say that it carries that irrep.

This is not to say that these two representations are the only useful ones in quantum
mechanics. Often we are concerned with the permutation of only some degrees of free-
dom of particles in which case fundamental particles no longer need to be regarded as
identical because the remaining degrees of freedom can still distinguish them. A classic
example of this is the generalized Hong-Ou-Mandel effect in which the spatial modes
of some number of photons are exchanged at a beamsplitter that has no effect on the
polarization properties of the photons. The polarization state at the output of such
a beamsplitter is not limited to the trivial and alternating group but can carry other
irreducible representations as well.

Counting the irreps

The total number of distinct irreps of $S_N$ can be obtained by simple counting arguments.
The number of irreps corresponds to the number of classes which is to say the number
of different ways that the numbers from 1 to $N$ can be divided into cycles. This is called
the number of partitions of $N$ i.e. the number of different ways that we can select a set
of integers $\{\lambda_i\}_k$ such that $\lambda_1 + \lambda_2 + \cdots + \lambda_k = N$.

A convenient way of labeling the partitions is to use Young diagrams, a series of boxes
arranged according to the rule that the number of boxes in each row must be equal to or
less than the number in the previous row. Let’s consider the three partitions of 3 listed
below:

\[
\begin{align*}
&1 + 1 + 1 \\
&2 + 1 \\
&3
\end{align*}
\]

(3.22)

To each of these Young diagrams there corresponds a distinct partition of the numbers from 1 to \( N \) and hence a distinct irreducible representation of \( S_N \).

Young diagrams can be used to obtain the dimensionality of each of the irreps of \( S_N \) through the construction of Young tableaux. A standard Young tableau is the arrangement of the number from 1 to \( N \) in a Young diagram so that the numbers in any row or column of the table are non-decreasing.

We state without proof the following theorem

The number of dimensions of an irrep of \( S_N \) represented by a particular Young diagram is equal to the number of distinct standard Young tableaux that it can accommodate. Figure 3.23 shows the five standard Young tableaux for the Young diagram

\[
\begin{align*}
1 & 2 & 3 & 1 & 2 & 4 & 1 & 3 & 5 & 1 & 2 & 5 & 1 & 3 & 4 \\
4 & 5 & 6 & 3 & 5 & 6 & 2 & 4 & 6 & 3 & 4 & 6 & 2 & 5 & 6
\end{align*}
\]

(3.23)

In the case of Young diagrams with two rows the number of distinct standard Young tableaux can be shown from simple combinatorics to be \( \binom{n}{k} - \binom{n}{k-1} \), where \( n \) is the total number of blocks in the tableau and \( k \) is the number of blocks in the second row. \( \binom{n}{k} \) is the number of combinations of \( k \) elements selected from among \( n \) elements. \( \binom{n}{k} = \frac{n!}{k!(n-k)!} \)
row. For the example given in figure (3.23) we have

$$\binom{6}{3} - \binom{6}{3-1} = \frac{6!}{3!3!} - \frac{6!}{2!4!}$$

$$= 20 - 15$$

$$= 5.$$  

(3.24)

(3.25)

(3.26)

The irreps corresponding to two-row Young diagrams are especially important in considering the permutations of two-level systems such as photon polarizations.

While one can also use Young tableaux to help construct a linearly independent basis spanning the irreps of \( S_N \) it is easier to make use of a very powerful duality between the irreps of \( S_N \) and those of \( SU(2)^\otimes N \) described in the next section.\(^2\)

**The Schur-Weyl duality**

Now that we have developed the theory of permutations as applied to quantum states we are ready to use it to describe a system of identical photons. We will consider a system of photons each of which carries with it a \( m \)-dimensional Hilbert space that completely describes its properties (polarization, spatial mode and frequency mode). The limitation to a finite-dimensional Hilbert space may seem strange since properties like spatial mode and frequency are often considered as continuous variables, but since our interest is in tomography we will have to discretize the Hilbert space at some point in order to have a hope of measuring the state of the system. Moreover, our main concern in the experiments that will be presented is with polarization which, as a two level system, is already finite-dimensional.

The Hilbert space of a system of \( N \) photons will be the tensor product of the \( N \) spaces describing the individual photons \( \mathcal{H}_m^\otimes N \). To make clear the action of permutations we would like to find a set of basis vectors for this space that carry the irreps of \( S_N \).

\(^2\)In fact the duality holds between any \( \text{GL}(n) \), but we will only make use of it for \( SU(2) \).
Luckily the Schur-Weyl duality[51] in group theory shows that this is always possible. Their theorem can be stated as

\[ \mathcal{H}_m^\otimes n = \bigoplus_{\lambda} Q_{\lambda}^m \otimes P_{\lambda}. \]  

This decomposes the Hilbert space \( \mathcal{H}_m^\otimes n \) into a direct sum of tensor products between spaces \( Q_{\lambda}^m \) that carry an irreducible representation of \( U(m) \) and \( P_{\lambda} \) that carries a particular irreducible representation of \( S_N \). The theorem states that this decomposition is multiplicity-free so that each irrep of \( U(m) \) and \( S_N \) appears exactly once in the sum and so can be labeled with the same index \( \lambda \).

In the case where \( m = 2 \) we have a particularly instructive example, both because it nicely describes photon polarization and because the irreps of \( U(2) = U(1) \otimes SU(2) \) are easy to work with because of their relationship to the irreps of the group of spatial rotations generated by angular momentum operators\(^3\)

This being the case, we can label the irreducible representations of \( SU(2) \) with a total angular momentum \( j \) which is either an integer or a half-integer. The irrep labeled \( j \) will be \( 2j + 1 \) dimensional. This is the same \( j \) that we discussed in the context of the Schwinger formalism. Its basis states are labeled with an index \( m \) that ranges from \(-j\) to \( j\). The action of \( SU(2) \) on these states will generally map a state \( |j, m\rangle \) onto a linear combination of states \( \sum_{m=-j}^{j} c_m |j, m\rangle \) all with the same \( j \).

The \( N \) copies of \( SU(2) \) present in the Hilbert space \( \mathcal{H}_2^\otimes N \) of \( N \) photon polarizations allow irreps with \( j \) between 0 and \( N/2 \) when \( N \) is even and between 1/2 and \( N/2 \) when \( N \) is odd. The Schur-Weyl theorem in this case is

\[ \mathcal{H}_2^\otimes N = \bigoplus_{j=0 \text{ or } 1/2}^{N/2} Q_j \otimes P_j. \]  

A set of basis states taking advantage of this decomposition can be created using the

---

\(^3\)While we already derived this from the Schwinger formalism, a group theorist would say that this is because \( SU(2) \) is the universal covering group of \( O(3) \), the group of spatial rotations in 3-dimensional space. The two groups share a common Lie algebra and any irrep of \( O(3) \) is also an irrep of \( SU(2) \).
Chapter 3. Hidden differences

machinery of Clebsch-Gordan for building states carrying a specific irrep $j$ out of tensor products states.

One familiar example of this is the singlet-triplet decomposition of the space of two two-level states, $\mathcal{H}_2 \otimes \mathcal{H}_2 = \mathcal{H}_{j=1} \oplus \mathcal{H}_{j=0}$. The Schur-Weyl theorem express the fact that this decomposition, constructed to yield states with well defined total angular momentum $j$, also carry irreps of $S_2$.

Using polarization as our two level system so that $m = 1/2$ is labeled with $H$ and $m = -1/2$ with $V$ we can write out this decomposition explicitly as

\[
| j = 1, m = 1 \rangle = |HH\rangle \\
| j = 1, m = 0 \rangle = \frac{1}{\sqrt{2}} (|HV\rangle + |VH\rangle) \\
| j = 1, m = -1 \rangle = |VV\rangle \\
| j = 0, m = 0 \rangle = \frac{1}{\sqrt{2}} (|HV\rangle - |VH\rangle).
\]

Notice that the $j = 1$ states all carry the trivial representation of $S_2$, since if the two particle labels are permuted the states are unchanged. The $j = 0$ state carries the other irrep of $S_2$, the antisymmetric irrep, and will be multiplied by one if the identity permutation is applied and by -1 if the polarizations are permuted.

If polarization is the only degree of freedom describing the photons then the requirement that the whole state be invariant under permutations would only allow the $j = 1$ states. The $j = 0$ state does not have the requisite symmetry to be a valid boson state (although it would be a valid state for fermions whereas the $j = 1$ states would not).

In the second-quantized formalism this fact is built into the raising and lowering operator commutation (and anti-commutation) relations. Usually this is thought of as a convenient property of the formalism, but when non-trivial permutation symmetries are possible due to hidden distinguishability it can lead to confusion\textsuperscript{4} and to incorrect re-

---

\textsuperscript{4}Two different referees of our paper on this topic objected that all of our results would be trivial if we would only have used second-quantized notation!
sults such as the Bogdanov tomography scheme’s equating of a measurement of $m$ with a projection onto $|j = N/2, m\rangle$.

With the Schur-Weyl theorem establishing that we can span the Hilbert space of $N$ particles with states that carry irreps of both $U(m)$ and $S_N$ we now wish to take the decomposition one step further. Among the $m$ levels describing the complete state of each particle we would like to separate the $d$ levels that are visible to our experimental measurements from the $D$ levels that are hidden from them. Clearly $D \times d = m$, and for the moment we will also require that $D \geq d$ which will make the math easier and, in any event is certainly true in the experiments we will be considering\footnote{In the case where $d \geq D$ we can always formally add more hidden levels until $d \geq D$ without affecting the derivation.}. In this case we can apply the Schur-Weyl theorem separately to the visible and hidden degrees of freedom.

\begin{equation}
\mathcal{H}^{\otimes N}_m = (\mathcal{H}^{\otimes d}_{\text{vis}}) \otimes (\mathcal{H}^{\otimes D}_{\text{hid}}) = \left( \bigoplus_{\lambda} Q^d_{\lambda} \otimes P_{\lambda} \right) \otimes \left( \bigoplus_{\lambda'} Q^D_{\lambda'} \otimes P_{\lambda'} \right),
\end{equation}

A general state will be a superposition of states in these spaces, so that

\begin{equation}
|\psi\rangle = \sum_{\lambda} c_{\lambda} |\lambda, q^d_{\lambda}, p_{\lambda}\rangle_{\text{vis}} |\lambda, q^D_{\lambda}, p_{\lambda}\rangle_{\text{hid}}.
\end{equation}

If we are dealing with bosons, however, this is too general because we have not yet applied the restriction that bosonic states must be invariant under the action of all permutations operators, which is to say that they must carry the trivial irrep of $S_N$.

One obvious way of achieving this is to restrict the kets to those that carry the trivial irrep of $S_N$ in both the visible and hidden degrees of freedom. For those states a permutation will transform the visible part trivially and the hidden part trivially, so clearly the whole state will also transform trivially.

Can states which are invariant under all permutations also be constructed from states that carry non-trivial irreps of $S_N$? Consider a state that carries the antisymmetric irrep
of $S_N$ in both the visible and hidden degrees of freedom. Permutation operators on the visible and hidden parts of the state act either by leaving the state unchanged or by multiplying the state by -1. If the same permutation operator is applied to both the hidden and visible degrees of freedom then either the whole state is multiplied by 1 or by $(-1)(-1) = 1$ so such a state is also invariant under all permutations.

What about multi-dimensional irreps of $S_N$? Let’s say we have a state that carries a mixed-symmetry irrep of $S_N$ like $\begin{array}{c} \boxed{\phantom{\text{HHV}}} \\ \boxed{\phantom{\text{HVH}}} \end{array}$ for three photons. The $S_N$ irrep is two-dimensional and can be represented by the matrices in table 3.15.

We label the two states that transform according to these measurements $|a\rangle_{\text{vis}}$ and $|b\rangle_{\text{vis}}$. If the visible degree of freedom under consideration were polarization then these states could be

$$|a\rangle_{\text{vis}} = \frac{1}{\sqrt{6}} |HHV\rangle + \frac{1}{\sqrt{6}} |HVH\rangle - \sqrt{\frac{2}{3}} |VHH\rangle$$

(3.36)

$$|b\rangle_{\text{vis}} = \frac{1}{\sqrt{2}} (|HHV\rangle - |HVH\rangle).$$

(3.37)

The reader can verify that applying the permutations of $S_3$ to the ordering of the polarizations does indeed affect these states in the same way as the matrix representation in table 3.15.

As long as the dimensionality of the space of hidden degrees of freedom is not less than that of the visible degree of freedom ($D \geq d$) there will also be states of the hidden degree of freedom carrying the $\begin{array}{c} \boxed{\phantom{\text{HHV}}} \\ \boxed{\phantom{\text{HVH}}} \end{array}$ irrep of $S_N$. We may therefore create two states $|a\rangle_{\text{hid}}$ and $|b\rangle_{\text{hid}}$ that transform in the same way as $|a\rangle_{\text{vis}}$ and $|b\rangle_{\text{vis}}$ under the action of the matrices in table 3.15.

Now if we construct the following state

$$\frac{1}{\sqrt{2}} (|a\rangle_{\text{vis}} |a\rangle_{\text{hid}} + |b\rangle_{\text{vis}} |b\rangle_{\text{hid}}).$$

(3.38)

then it will be invariant under the action of all the permutations in table 3.15 (as long as the same permutation is performed on the visible and hidden parts of the state).
That this is so follows from the fact that all of the matrices of table 3.15 are orthogonal matrices.

Consider a particular matrix

$$
M = \begin{pmatrix}
    q & r \\
    s & t
\end{pmatrix},
$$

(3.39)

that takes \(|a\rangle\) to \(q|a\rangle + s|b\rangle\) and \(|b\rangle\) to \(r|a\rangle + t|b\rangle\). If the same matrix acts on both parts of the state \(|a\rangle|a\rangle + |b\rangle|b\rangle\) it will map it to

$$
q^2|a\rangle|a\rangle + qs|a\rangle|b\rangle + sq|a\rangle|b\rangle + s^2|a\rangle|b\rangle \\
+ r^2|a\rangle|a\rangle + rt|a\rangle|b\rangle + tr|a\rangle|b\rangle + t^2|a\rangle|b\rangle
$$

(3.40)

$$
= (q^2 + r^2)|a\rangle|a\rangle + (qs + rt)|a\rangle|b\rangle + (sq + tr)|a\rangle|b\rangle + (s^2 + t^2)|a\rangle|b\rangle.
$$

(3.41)

For the mapping to leave the state unchanged we require that

$$q^2 + r^2 = 1
$$

(3.43)

$$qs + rt = 0
$$

(3.44)

$$s^2 + t^2 = 1.
$$

(3.45)

However, this is precisely the requirement that the columns of \(M\) be orthonormal to each other. Since all the matrices in table 3.15 are orthogonal, the state \(\frac{1}{\sqrt{2}} (|a\rangle_{\text{vis}}|a\rangle_{\text{hid}} + |b\rangle_{\text{vis}}|b\rangle_{\text{hid}})\)

is invariant under the action of all permutations in the \(\n\) irrep.

More generally, whenever we construct an equally-weighted superposition of kets that transform in the same way for the visible and hidden parts over an irrep of \(S_N\), we will always obtain a state that is invariant under all permutations. This follows directly from the fact that \(S_N\) can always be represented using orthogonal matrices and is a straightforward generalization of the two-dimensional case demonstrated above. It also follows that such an equally weighted sum is the only linear combination of basis vectors of a given irrep of \(S_N\) that will be invariant under all the permutations.
Using the notation introduced previously, any state of the form

\[ \frac{1}{\sqrt{\dim P_\lambda}} \sum_{p_\lambda} |\lambda, q^d_\lambda, p_\lambda \rangle |\lambda, q^D_\lambda, p_\lambda \rangle \]  

(3.46)

will be invariant under all permutations and hence will constitute a valid state for bosons.

Finally, it should be clear that any state carrying a different irrep in the visible and hidden degrees of freedom cannot be invariant under all permutations since the two parts of the state will transform in completely different way under permutations.

The most general state of \( N \) bosons can therefore be written as

\[ |\Psi^N_{d,D} \rangle = \sum_{\lambda} \sum_{q^d_\lambda, q^D_\lambda} A^{\lambda}_{q^d_\lambda, q^D_\lambda} \sum_{p_\lambda} |\lambda, q^d_\lambda, p_\lambda \rangle |\lambda, q^D_\lambda, p_\lambda \rangle , \]  

(3.47)

where the \( \lambda \) labels the terms in the Schur-Weyl decomposition for the visible and hidden states respectively, \( q^d_\lambda \) runs over all the states in the \( \lambda \)th irrep of \( U(d) \) in the visible degree of freedom and \( Q^d_\lambda \) does the same for the \( \lambda \)th irrep of \( U(D) \) for the hidden degree of freedom. Since the only way to construct bosonic terms is from states carrying the same irrep of \( S_N \) in both the hidden and visible degree of freedom, we can use the same index \( \lambda \) to label them both (rather than a separate indices \( \lambda \) and \( \lambda' \) for the visible and hidden parts). The final sum runs over the basis states for the \( \lambda \)th \( S_N \) irrep. Note that the visible and hidden part of the terms of this sum are represented by the same basis vector of the \( S_N \) irrep.

The complex amplitudes \( A^{\lambda}_{q^d_\lambda, q^D_\lambda} \) are assumed to carry the normalization so as to satisfy the requirement that

\[ \sum_{\lambda} \dim P_\lambda \sum_{q^d_\lambda, q^D_\lambda} |A^{\lambda}_{q^d_\lambda, q^D_\lambda}|^2 = 1. \]

(3.48)

We’ll take polarization as the visible degree of freedom and make use of the total angular momentum to label the SU(2) irreps. We can therefore replace the label \( \lambda \) with \( j \). We’ll assume that we have three photons so that \( j \) runs from 1/2 to 3/2. We can also replace the index \( \kappa \) with \( m \) running from \(-j\) to \( j \). Finally we can label the basis
states of $S_3$ with an index label $s$. The basis states for the hidden degrees of freedom will, from the Schur-Weyl duality, also carry an irrep of $S_3$, and there can be arbitrarily many orthogonal states carrying a given $S_3$ irrep. We can denote these many orthogonal states with the labels $\lambda_j$ and $q^D_\lambda$. Using these conventions, a properly bosonic state can be written as

\[
\sum_{j=1/2}^{3/2} \sum_{\lambda, m=-j}^{j} \sum_{q^D_\lambda} C_{m,j;\lambda,j} \sum_{p_\lambda} |j, m, s\rangle \langle \lambda_j, q^D_\lambda, s| .
\]

(3.49)

Note that the sum only consists of tensor products of states carrying the same $S_3$ irrep in both the visible and hidden degrees of freedom.

### 3.2 Hidden information

In the last section we showed how to write down a state that is invariant under all permutations and hence is an acceptable description for a system of bosons. In this section we will show what information is accessible in the visible degrees of freedom once the hidden degrees of freedom have been traced out.

In an experiment that measures the visible degree of freedom only, the measurements we do can tell us about the state of the hidden degrees of freedom only insofar as they correlate to the visible ones. We project the state onto every possible state of the hidden degrees of freedom and sum these projections to obtain a density matrix in the visible degrees of freedom only which we will call the *accessible density matrix* $\rho_{d,\text{acc}}^N$.

\[
\rho_{d,\text{acc}}^N = \text{Tr}_{\text{hid}} \left[ |\Psi_{d,D}^N \rangle \langle \Psi_{d,D}^N | \right]
\]

(3.50)

\[
= \sum_{\lambda} \sum_{q^D_\lambda, q'^D_\lambda} B^\lambda_{q^D_\lambda, q'^D_\lambda} \sum_{p_\lambda} |\lambda, q^d_\lambda, p_\lambda\rangle \langle \lambda, q'^d_\lambda, p_\lambda| ,
\]

(3.51)

where $B^\lambda_{q^d_\lambda, q'^d_\lambda} = \sum_{q^{D}_\lambda} A^\lambda_{q^{d}_\lambda, q^{D}_\lambda} A^{\lambda^*_D}_{q^{d}_\lambda, q^{D}_\lambda}$. It should be clear that since we started from the most general possible boson state, the set of numbers $B^\lambda_{q^d_\lambda, q'^d_\lambda}$ provides an informationally-complete description of the state
of the visible degrees of freedom. Any property of the visible degree of freedom, and the outcome of any measurement on them has to be a function of the $B^\lambda_{q^d \lambda', q^d \lambda}$ alone.

How many numbers $B^\lambda_{q^d \lambda', q^d \lambda}$ are required to describe the state? It is clear from the structure of the formula (3.50) that each irrep of $S_N$ will contribute one unique $B^\lambda_{q^d \lambda', q^d \lambda}$ for each unique pair $q^d \lambda, q^d \lambda'$. Thus if the $\lambda$th irrep of SU($d$) is $k$-dimensional it will contribute $k^2$ different numbers $B^\lambda_{q^d \lambda', q^d \lambda}$ regardless of dimensionality of the $\lambda$th irrep of $S_N$.

We can make this even more precise by applying the Weyl character formula for SU($N$)[51]. This formula gives the dimension of the $\lambda$th irrep of SU($N$) as

$$\dim(\lambda) = \prod_{1 \leq i < j \leq d} \frac{\lambda_i - \lambda_j + j - i}{j - i},$$

(3.52)

where $\lambda$ is a partition of $N$ i.e. an ordered set of numbers $\lambda_n, \lambda_{n-1}, \ldots, \lambda_m$ such that $\lambda_i$ and $\sum_i \lambda_i = N$ and $\lambda_i \geq \lambda_{i-1}$. Because of the one-to-one correspondence between partitions and Young diagrams, this sum can be thought of as running over the different Young diagrams with $N$ boxes.

The SU($N$) irrep associated with the trivial irrep of $S_N$ corresponds to the partition of $N$ with only one non-zero $\lambda_i = N$ (i.e. a partition $(N, 0, 0, \ldots, 0$). The dimension of this irrep can be written down explicitly as

$$\prod_{j=2}^{d} \frac{N + j - 1}{j - 1} = \binom{N + d - 1}{N}.$$  

(3.53)

When the particles in a bosonic state are fundamentally indistinguishable, this is the full dimension of the Hilbert space since there are no other degrees of freedom that can distinguish the particles. In all other cases where there are distinguishing, but hidden, degrees of freedom, expression 3.52 must be summed over all $\lambda$ to arrive at total number of $B^\lambda_{q^d \lambda', q^d \lambda}$ in $\rho_{acc}$. Thus the number of free parameters in $\rho_{acc}$ is

$$\sum_{\lambda_i, \lambda_j} \prod_{1 \leq i < j \leq d} \frac{\lambda_i - \lambda_j + j - i}{j - i} = \binom{N + d^2 - 1}{N},$$

(3.54)

as can be proved using the Cauchy formula for the general linear group$^6$.

$^6$This was shown by my co-author Peter Turner based on discussions with Trevor Welsh.
For the case of polarization where $d = 2$, this expression simplifies to

$$
\sum_{j=0}^{N/2} (2j + 1)^2 = \binom{N + 3}{3},
$$

so that, for example, three polarizations are completely described by 20 real numbers.

### 3.2.1 Structure of the accessible density matrix

Expression 3.50 implies a particular structure for the accessible density matrix. If we consider for the moment only SU(2), the structure of the density matrix will be block-diagonal in angular-momentum space, with one or more $2j + 1$ by $2j + 1$ blocks for each allowed value of $j$. For a three-photon polarization state it would look like this:

$$
\rho_{\text{acc}} = \begin{bmatrix}
\begin{pmatrix}
* & * & * & * \\
* & * & * & * \\
* & * & * & * \\
* & * & * & *
\end{pmatrix}_{j=3/2}
\begin{pmatrix}
* & * \\
* & *
\end{pmatrix}_{j=1/2}
\end{bmatrix}.
$$

The trace in equation 3.50 eliminates any coherences between sectors with different values of $j$. This isn’t to say that such information does not exist, only that it will be undetectable by measurements done on polarization and, by the same token, will not affect any polarization operations one might like to perform. The reason for the elimination of these coherences is that states with different values of $j$ necessarily carry different representations of the symmetric group by the Schur-Weyl duality. Saying that we are ignorant of the particle ordering (which is what we do when we take the trace
in equation 3.50) applies all possible permutations to the particles. Coherences between states carrying two different irreps of $S_N$ will necessarily transform differently under these permutation and, it can be shown, will average to zero under all the permutations.

The trace also has an interesting effect on orthogonal vectors of multi-dimensional irreps of $S_N$, for instance the spanning vectors of the two $j = 1/2$ spaces for three polarizations. For these states the trace in equation 3.50 reduces the equal superposition over the orthogonal permutation spaces into an equal convex sum over those spaces. This means that two orthogonal states with the same $j$ but carrying different permutation irreps will be indistinguishable to polarization measurements. The polarization measurements can only access the average value of the density matrix in these two orthogonal spaces. For three photons, the 16 elements of the $j = 3/2$ space and the four elements consisting of averages over the $j = 1/2$ spaces exhaust the 20 free parameters calculated from equation 3.55.

A final interesting property of this accessible density matrix is that the number of elements in it scales polynomially in the dimension of the Hilbert space. This means that particles that are distinguishable in principle, but not in practice, form an intermediate case between indistinguishable particles which are required to occupy states that transform trivially under permutations and for which the Hilbert space grows linearly in the number of particles and distinguishable particles or qubits for which the Hilbert space grows exponentially. The possibility of having particles distinguishable by hidden information allows the state to contain more information than would be possible if there were no distinguishability, but does not give rise to the exponential scaling of Hilbert space size that is considered essential to quantum computation and other quantum information processing tasks.
3.2.2 Implications for distinguishability

In most experiments on photons there are two degrees of freedom of importance. One degree of freedom is the one in which the quantum state is described and measurements are performed, and the other is a distinguishing degree of freedom used to label the two photons. The prototypical quantum optics experiment is the Bell’s inequality violation[7] where two photons are identified by their different propagation directions and measurements are performed on their polarization. What is really being measured here is a joint correlation function between spatial mode and polarization. This is entirely consistent with the classical statistical view that the photons at the two locations are really ‘different’ particles and that we are measuring the polarization ‘of each photon’.

This all changes when photons travel in a single spatio-temporal mode, usually defined by a single-mode optical fiber. In some cases another degree of freedom like wavelength can still be used as a labeling degree of freedom[88]. In most cases, though, the goal is to get the photons into the same single mode and any distinguishing information between the photons is there by accident[86, 43]. It is this sort of experiment that this work addresses. In such experiments one would like a way of answering the question “Is there, in principle, some measurement on the degrees of freedom that I can’t measure whose outcome would distinguish one particle from the other?” . The discussion of the previous section helps both to make this question operationally precise and to provide an answer.

We first note that the visible degrees of freedom can at most provide us with the experimental density matrix elements. At best, therefore, we can answer the question, “Are the density matrix elements $B^{\lambda, q}_q, d^{q'}_q$ that I can obtain from measurements on the degrees of freedom that I can measure consistent with there being no measurement, I could do, even in principle, on the other degrees of freedom that could distinguish the particles?” This is the best one can hope for under the assumption that we are limited as to which degrees of freedom we can access.

If the photons are indistinguishable in the hidden degrees of freedom then of the $D$
possible states that each photon could be in, all $N$ of the photons must be in the same state. The hidden part of the state could be written $|\phi\rangle^{\otimes N}$ where $|\phi\rangle$ is a valid single-particle state in the hidden degrees of freedom. Such a state naturally carries the trivial irrep of $S_N$ since if all the particles are in the same state, any permutation of them will leave the state unchanged. From the discussion in the previous section it follows that the state of the visible degrees of freedom must also carry the trivial irrep in the visible degrees of freedom.

If a series of measurements is done that determines all the $B_{q, q'}^{\lambda}$ and it is found that the only non-zero elements are those with $\lambda = 1$, i.e. those carrying the trivial irrep of $S_N$ in the visible degree of freedom then we can say that the measurements are consistent with the $N$ photons being indistinguishable.

This does not necessarily mean that there is no measurement that could be done on the hidden degrees of freedom that would distinguish the photons. It only means that if there were no such measurement we would obtain the same results on the visible degrees of freedom. The conclusions of the previous section is that this is the most that we can say about the distinguishability of the photons based on our measurements.

By the same token, when we measure one of the $B_{q, q'}^{\lambda}$ with $\lambda \neq 1$ so that there is a component of the visible degree of freedom that carries a non-trivial irrep of $S_N$ then we know with certainty that there is some measurement on the hidden degrees of freedom that could distinguish the particles. This has to be the case since we can infer from the measurement that the hidden degree of freedom also has a component carrying a non-trivial irrep of $S_N$ which immediately implies that the photons cannot all be in the same single-particle state.

Moreover, from the magnitude of the non-trivial $S_N$ component in the visible degree of freedom we obtain a lower bound on the degree to which the particles might be distinguishable in the hidden degrees of freedom. This is really quite remarkable - our knowledge that photons are bosons allows us to infer the presence of information in
degrees of freedom that we can’t measure.

3.3 Experimental measurements

With the theory of hidden distinguishability and the accessible density matrix fully developed we will, in this section, present our reconstructions of $\rho_{\text{acc}}$ for two and three photon polarizations systems. The measurements for two photons were completed by Krister Shalm and me in the summer of 2004. The three photon data was taken by Krister Shalm in spring 2007, following the procedure developed in this chapter. I was only involved in the data analysis.

The measurement system used in both experiments was the same. Conceptually it involved applying to the state an SU(2) operation with a quarter waveplate and half-waveplate followed by a polarizing beamsplitter oriented so as to split horizontal and vertical photons into separate spatial modes and a number resolving detector on each of these modes.

The number resolving detector was, in reality, a series of non-polarizing beamsplitters with a single-photon counting APDs on the output ports of each beamsplitter. Figure 3.1 shows a conceptual tomographic apparatus and the specific beamsplitter networks used for the two and three-photon characterization.

It may not be immediately apparent that such a system is capable of measuring all the density matrix elements $B_{q_j q'_j}^{j}$. However a simple recursive argument shows that this is indeed the case.

With $N$ photons in the state there are $N + 1$ different ways that these can split between the $H$ and $V$ ports of the polarizing beamsplitter so that $N_H$ photons go to the $H$ port and $N_V$ photons go to the $V$ port. Clearly $N = N_H + N_V$, and the angular momentum projection quantum number $m = N_H/2 - N_V/2$. When the photons leaving each port are counted, the measurement implemented will be a convex sum of projectors.
Figure 3.1: Measurement apparatuses for measuring the elements of the accessible density matrix. Figure (a) shows an ideal apparatus that will work for any number of photons. Figures (b) and (c) show the apparatus actually used for characterizing two-photon and three-photon experiments.

onto all states having that number of horizontal and vertical photons. For example, $P_N = |HH \cdots H \rangle \langle HH \cdots H|$, $P_{N-1} = (|VH \cdots H \rangle \langle VH \cdots H| + |HV \cdots H \rangle \langle HV \cdots H| + \ldots )$ and so on. Equivalently, this sum could run over all the states carrying different $j$ irreps of SU(2) and different irrep of $S_N$ but having the same value of $m$.

We begin by looking at the projector $P_N$. It has support only on the $j = N/2$ irrep and is a ‘pure’ projector onto the state $j = N/2, m = N/2$. By changing the angles of the waveplates one can ‘orbit’ this measurement in the $j = N/2$ space thereby obtaining all the density matrix element $B_{q(N)}^{(N)}(q_{(N)}^{(N)})$ of the symmetric states.

It follows that the $j = N/2$ part of the accessible density matrix can be completely characterized by only measuring $P_N$ rotated under various SU(2) transformations.

$P_{N-1}$ is a convex sum of projectors onto all states with $m = N/2 - 1$. Since the $j = N/2, m = N/2 - 1$ projection has already been measured, it can be subtracted off $P_{N-1}$, leaving a projector with support only in the $j = N/2 - 1$ subspace. By changing the waveplate angles one can use this modified operator to completely characterize the $j = N/2 - 1$ space. One can then subtract the $j = N/2$ and $j = N/2 - 1$ terms from the $m = N/2 - 2$ operator, and so on. In this way all the terms in the accessible density matrix can be measured. The number of operators obtained in this way will be exactly
\[ \sum_j (2j + 1)^2 \], or, by equation 3.55, the total number of operators that can be measured on the \( N \) photon polarizations.

In practice, the method we followed was to calculate from equation 3.55 the number of linearly independent density matrix elements. A set of waveplate angles were then chosen and the projection operators for the measurement apparatus determined as a function of the waveplate settings. These projection operators were treated as vectors in a vector space, and the dimensionality of that space was calculated by taking the rank of a matrix with the vectorized projectors as columns. When the dimensionality of the space spanned by the vectors matched the number obtained from equation 3.55 we knew that we could invert the density matrix uniquely from the measurements. The maximum-likelihood techniques discussed in Chapter 2 were applied to obtain the density matrix from measurement of these projectors.

### 3.4 The two-photon experimental polarization density matrix

We began our investigations into state estimation of indistinguishable particle states with the two-photon polarization state. Although at the time of the experiment we had not developed a full understanding of permutation symmetries of visible and hidden degrees of freedom, the two-photon case is simple enough that it can be understood without needing this formalism. In group theory language this simplicity comes about because the permutation symmetries of two particles can be completely understood in terms of the two one-dimensional irreps of \( S_N \), the trivial irrep and the alternating irrep. Consequently there are no values of angular momentum with multiplicity and the elements of the accessible density matrix all represent coherences or populations of states with no averaging over multiple spaces with the same value of \( j \). The two-photon case also has the appealing property that the effects of distinguishing information are
all concentrated in a single accessible density matrix element, namely the singlet state projection.

The two-photon case is an important one because of the many proposals for using two-photon single-spatial mode states as a qutrit for quantum information applications [18, 70, 72]. At the time this experiment was done, Kulik’s group had already developed a quantum state tomography technique [18, 17] for estimating these states, but, as we have already mentioned, it was flawed in that it implicitly assumed indistinguishable particles from the outset rather than trying to obtain all available polarization information and determining distinguishability from that.

The Schur-Weyl duality applied to two polarizations decomposes the Hilbert space into a $j = 1$ space carrying the symmetric permutation irrep and a $j = 0$ space carrying the alternating permutation irrep. In order for the whole state to be symmetric under permutations (which it must be for bosons), the only available states for the hidden degrees of freedom are those carrying the same permutation symmetry as the visible degree of freedom.

A general pure two-photon state can therefore be written as

$$\sum_i c_i \left| \phi^{j=1}_i \right\rangle_{\text{vis}} \left\langle \chi^S_i \right\rangle_{\text{hid}} + \sum_k c_k \left| \phi^{j=0}_k \right\rangle_{\text{vis}} \left\langle \chi^A_k \right\rangle_{\text{hid}}.$$  \hspace{1cm} (3.57)

Here $\left| \chi^S_i \right\rangle_{\text{hid}}$ represents states on the hidden degrees of freedom carrying the trivial symmetry irrep and $\left| \chi^A_k \right\rangle_{\text{hid}}$ represents states on the hidden degrees of freedom carrying the alternating irrep. Similarly $\left| \phi^{j=1}_i \right\rangle_{\text{vis}}$ represents visible states carrying the trivial irrep and $\left| \phi^{j=0}_i \right\rangle_{\text{vis}}$ visible states carrying the alternating irrep. The important point here is that in order to be bosonic the state must be a sum over terms carrying the same irrep of $S_2$ in the hidden and visible spaces.

When we trace over the hidden degrees of freedom we are left with

$$\rho_{\text{acc}} = \sum_{m_1, m_2 = -1}^1 c_{m_1} c_{m_2}^* \left| \phi^{j=1}_{m_1} \right\rangle \left\langle \phi^{j=1}_{m_2} \right| + |c_0|^2 \left| \phi^{j=0}_{m_0} \right\rangle \left\langle \phi^{j=0}_{m_0} \right|.$$  \hspace{1cm} (3.58)
This accessible density matrix takes the form

\[
\rho_{\text{acc}} = \begin{pmatrix}
\rho_{HH,HH} & \rho_{HH,\psi} & \rho_{HH,VV} \\
\rho_{\psi+,HH} & \rho_{\psi+,\psi} & \rho_{\psi+,VV} \\
\rho_{VV,HH} & \rho_{VV,\psi} & \rho_{VV,VV} \\
0 & 0 & (\rho_{\psi-,\psi-})
\end{pmatrix},
\] (3.59)

where \(|\psi^+\rangle = |j = 1, m = 0\rangle = \frac{1}{\sqrt{2}} (|HV\rangle + |VH\rangle)\) and \(|\psi^-\rangle = |j = 0, m = 0\rangle = \frac{1}{\sqrt{2}} (|HV\rangle - |VH\rangle)\).

In the paper [2] we made this argument in a different way by directly applying the properties of the permutation operators in \(S_2\) without using the Schur-Weyl duality (of which we were ignorant until later). The same result as above can be obtained by writing a completely general two-photon state as

\[
|\psi\rangle = \sum_i c_i |\phi_i\rangle_{\text{acc}} |\chi_i\rangle_{\text{hid}},
\] (3.60)

where \(|\phi_i\rangle_{\text{acc}}, |\chi_i\rangle_{\text{hid}}\) are eigenstates of exchange operators \(X_{\text{acc}}\) and \(X_{\text{hid}}\) for the visible and hidden degrees of freedom, respectively, with the same eigenvalue \(\pm 1\). The requirement that the whole state be bosonic so that \(X_{\text{acc}} \otimes X_{\text{hid}} |\psi\rangle = |\psi\rangle\) guarantees that each term can be written with the visible and hidden parts of the state either both symmetric or both anti-symmetric. A completely general state of two photons is a mixture of states such as \(|\psi\rangle\), described by a density matrix \(\rho = \sum_j w_j |\psi_j\rangle \langle \psi_j|\).

The accessible density matrix in equation 3.59 can be thought of as simply a rotated version of the ordinary two-qubit density matrix, but with ‘missing’ coherences between states of different \(j\). The absence of these elements can be understood as being directly due to \textit{inaccessible information} about particle ordering that is present for qubits, but absent in a system where the photons are experimentally indistinguishable. The information contained in these coherences corresponds very directly to distinguishing information that could in principle be used to tell the two photons apart. For example, consider the missing coherence in (3.59) between the singlet state \(|\psi^-\rangle = 1/\sqrt{2} (|HV\rangle - |VH\rangle)\) and
triplet state $|\psi^+\rangle = 1/\sqrt{2} (|HV\rangle + |VH\rangle)$. Note that $|HV\rangle = 1/\sqrt{2} (|\psi^+\rangle + |\psi^-\rangle)$ and $|VH\rangle = 1/\sqrt{2} (|\psi^+\rangle - |\psi^-\rangle)$. The ability to distinguish $|HV\rangle$ from $|VH\rangle$, the property we usually think as distinguishability, amounts to knowing the relative phase between $|\psi^+\rangle$ and $|\psi^-\rangle$. In a complete polarization density matrix for two distinguishable particles this phase would be contained in the coherence between $|\psi^+\rangle$ and $|\psi^-\rangle$. Its absence is an expression of the inability of the photons to be distinguished from the information contained in the accessible density matrix.

The inaccessibility of the other two missing coherences in (3.59) can be explained in the same way. If we knew the value of the coherences between $|HH\rangle$ and $|\psi^-\rangle$ and between $|VV\rangle$ and $|\psi^-\rangle$ then, taken together with the populations of $|HH\rangle$ and $|VV\rangle$ this would tell us the phase between $|\phi^+\rangle = \frac{1}{\sqrt{2}} (|RL\rangle + |LR\rangle)$ and $|\psi^-\rangle = \frac{1}{\sqrt{2}} (|RL\rangle - |LR\rangle)$ and between $|\phi^-\rangle = \frac{1}{\sqrt{2}} (|DA\rangle + |AD\rangle)$ and $|\psi^-\rangle = \frac{1}{\sqrt{2}} (|DA\rangle - |AD\rangle)$. This is just the information need to tell $|DA\rangle$ from $|AD\rangle$ and $|RL\rangle$ from $|LR\rangle$, i.e. the distinguishing information between the two photons in the circular and diagonal bases.

This interpretation of the ‘missing’ information in the accessible density matrix can also be made for more than two photons, although it becomes more complex. For example, the information needed to distinguish $|HHV\rangle$, $|VHV\rangle$ and $|VHH\rangle$ is clearly contained in the coherences between the three $m = 1/2$ states in the accessible density matrix, although since there are three such coherences there is no longer a single number that one can point to as uniquely containing that distinguishing information.

3.4.1 Measuring the two-photon experimental density matrix

We measured two-photon accessible polarization density matrices for two-photon states coming from spontaneous parametric downconversion of a 405-nm, 50 fs pulse in $\beta$-BBO cut for type-II phasematching in the collapsed cone geometry[116]. This system allowed for thorough experimentation with our new characterization technique for a few reasons. First, the two SPDC photons as they emerged from the crystal were quite
Table 3.1: The projective measurements used to determine the two-photon accessible density matrix. The detectors can detect either a coincidence between two photons in the H mode thereby implementing the projector $P_{HH} \equiv |HH\rangle \langle HH|$ or a coincidence between the H and V modes thereby implementing $P_{HV} \equiv |HV\rangle \langle HV| + |VH\rangle \langle VH|$. A quarter- and half-waveplate at angles $q$ and $h$ respectively, placed before the detection apparatus, effectively rotate the detection operators to $U \otimes (U^\dagger)^{\otimes 2}$ where $P$ is either $P_{HH}$ or $P_{HV}$ and $U \equiv \exp \left[i\pi \left(\sigma_z \cos 2h - \sigma_x \sin 2h\right)\right] \exp \left[\frac{i\pi}{2} \left(\sigma_z \cos 2q - \sigma_x \sin 2q\right)\right]$, where $\sigma_x$, $\sigma_y$ and $\sigma_z$ are the Pauli matrices.

<table>
<thead>
<tr>
<th>$h$</th>
<th>$q$</th>
<th>$P$</th>
<th>$h$</th>
<th>$q$</th>
<th>$P$</th>
<th>$h$</th>
<th>$q$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0°</td>
<td>0°</td>
<td>$P_{HH}$</td>
<td>22.5°</td>
<td>0°</td>
<td>$P_{HH}$</td>
<td>45°</td>
<td>0°</td>
<td>$P_{HH}$</td>
</tr>
<tr>
<td>22.5°</td>
<td>45°</td>
<td>$P_{HV}$</td>
<td>11.25°</td>
<td>0°</td>
<td>$P_{HH}$</td>
<td>0°</td>
<td>22.5°</td>
<td>$P_{HV}$</td>
</tr>
<tr>
<td>45°</td>
<td>22.5°</td>
<td>$P_{HH}$</td>
<td>22.5°</td>
<td>0°</td>
<td>$P_{HV}$</td>
<td>22.5°</td>
<td>22.5°</td>
<td>$P_{HH}$</td>
</tr>
<tr>
<td>0°</td>
<td>0°</td>
<td>$P_{HV}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

indistinguishable and distinguishability between them could be added by delaying one with respect to the other by inserting birefringent crystals. This allowed us to explore the full range of distinguishability from indistinguishable to fully distinguishable. Second, the relatively high detection rates for two photons made it easy to acquire data within a time shorter than the time it took the alignment of the system to drift. Detection rates for photons were of order 100 per second and a tomographically complete dataset could be collected in a few minutes. This compares to rates of around 0.1/s for the three-photon data shown in the next section.

The waveplate angles used to generate different projective measurements for two photons are listed in Table 3.1. Note that ten measurements are required to measure the ten different elements of the accessible density matrix.

The complete experimental setup including both the preparation of the state and the tomographic measurement apparatus is shown in Figure 3.2.
Figure 3.2: Experimental implementation of state preparation and tomography protocols showing polarizing beamsplitters (PBS), beamsplitters (BS), non-linear $\beta$-Barium Borate (BBO) crystals, a second harmonic generation crystal (SHG), quarter waveplates (QWP), half-waveplates (HWP), polarization maintaining fiber (PMF), single-mode fiber (SMF) and single photon counting modules (SPCM). A spontaneous parametric downconversion (SPDC) crystal produces pairs of H and V photons. The separation between the H and V photons is controlled with movable quartz wedges and very long delays can be introduced by inserting a thick piece of BBO into the beam. Single-mode fiber and a 10nm interference filter make the photons essentially indistinguishable in the spatio-temporal modes. Tomography is performed with a set of waveplates and a polarizing beamsplitter. This system can implement all the measurements in Table 3.1.
50 fs pulses centered at 810 nm were generated in a Ti:Sapph oscillator with a 90 MHz repetition rate. The average power of the pulse train was 400 mW. These pulses were upconverted at a 0.5 mm BBO crystal to create 405-nm centered pulses which were sent to a second 0.5 mm BBO crystal to be downconverted. The downconversion was phase-matched for collapsed cone ‘beamlike’ emission so that the horizontal and vertical photons left the crystal in individual beams separated by a 6° angle. These two photons were then recombined at a polarizing beamsplitter so as to travel in a single spatial mode. The path lengths of the H and V photons were balanced to within much less than a coherence length (30 µm as determined by the filter bandwidth) so that the two photons were in wavepackets centered at the same point in time.

To ensure minimal hidden distinguishability, filters were used to select a narrow range of frequencies and a single spatial mode. The photons were passed through a 10 nm full-width at half-maximum spectral filter which significantly reduced the spectral distinguishability.

A single-mode optical fiber selected a single spatial mode for the photons which guaranteed that any remaining hidden distinguishing information would be in the temporal degree of freedom. This temporal distinguishability could be manipulated by changing the group delay between the H and the V modes. When this delay was well in excess of the coherence time of the photons they were distinguishable, since a measurement of arrival time would then give complete information about ‘which’ photon had which polarization. When the delay was non-zero but less than the coherence time, a measurement of the arrival time could only give partial information about the polarization.

Measurements were taken by setting the waveplates in figure 3.4 to the ten values in table 3.1. This resulted in ten linearly independent projections from which the accessible density matrix element could be extracted. Maximum-likelihood fitting was achieved by
Figure 3.3: The density matrices measured with quantum state tomography. The imaginary parts of (a) through (f), which are not shown, had all elements less than 0.05. White elements are inaccessible to measurement. (a) indistinguishable H and V photons (b) distinguishable H and V photons (c) partially distinguishable H and V photons (d) indistinguishable photons transformed to the 2-NOON state (e) distinguishable photons with the same transformation applied (f) the same state as it would be characterized by the technique of [17].
Table 3.2: The accessible density matrices of various two-photon polarization states measured with quantum state tomography. (a) indistinguishable H and V photons (b) distinguishable H and V photons (c) partially distinguishable H and V photons (d) indistinguishable photons transformed to the 2-NOON state (e) distinguishable photons with the same transformation applied (f) the same state as it would be characterized by the technique of [17].
fitting to a lower-triangular matrix $T$ of the form

$$
T = \begin{pmatrix}
    t_1 & 0 & 0 & 0 \\
    t_2 + it_3 & t_4 & 0 & 0 \\
    t_5 + it_6 & t_7 + it_8 & t_9 & 0 \\
    0 & 0 & 0 & t_{10}
\end{pmatrix},
$$

(3.61)

so that $\rho_{\text{acc}} = T^\dagger T / \text{Tr}[T^\dagger T]$. A least-squares likelihood function was minimized over all values of $t_i$ using Mathematica’s FindMinimum function.

Six measured density matrices are shown in Figure 3.3 and Table 3.2. The state was initially prepared to contain one $H$ photon and one $V$ photon with zero path length difference. The zero-path length difference point was found by the observation of two-photon interference in the diagonal basis. When tomography was carried out on this state, the density matrix in Figure 3.3(a) was obtained. 95% of the population was found to be in $|\psi^+\rangle$ and less than 2% in the $|\psi^-\rangle$ state. On the other hand when a delay much larger than the photon coherence length was introduced between the two polarizations, the density matrix in Figure 3.3(b) was obtained. The population was split roughly evenly between the $|\psi^+\rangle$ and $|\psi^-\rangle$ states.

The difference between the two density matrices becomes further apparent when the state is rotated using a quarter waveplate with its axis at 45° to the horizontal. When the two polarizations are indistinguishable this results in the state $\frac{1}{2} (a^\dagger_H a^\dagger_H + a^\dagger_V a^\dagger_V)$, or equivalently the 2-N00N state[21] or the $|\phi^+\rangle$ Bell state. Our measured density matrix for this state is shown in figure 3.3(d). This state has zero probability of producing anti-correlated photons in the $H/V$ basis. On the other hand when the distinguishable photons of figure 3.3(b) are sent through the same quarter waveplate, the resulting state still has a $j = 0, m = 0$ component consisting of half of the population as shown in figure 3.3(e). This ensures that anti-correlated counts will be observed in every basis half of the time. The two states really are different, and any good state estimation scheme must measure that difference.
In fact, if we try to use the state characterization scheme proposed in [17] to estimate the state in 3.3(e) we obtain completely different results. We can use our reconstructed density matrix to predict the outcome of measurements used in reference [17]. The result is the $3 \times 3$ density matrix shown in Table 3.2(f). Their reconstruction procedure lumps the two $m = 0$ populations together into the $j = 1, m = 0$ state and incorrectly measures the coherence between the $m = 1$ and $m = -1$ states. This density matrix makes completely incorrect predictions about correlation measurements in all bases except the H/V basis.

Figure 3.3(c) shows what happens when the distinguishing delay is non-zero, but less than a coherence length. While both the $j = 0, m = 0$ and $j = 1, m = 0$ populations are non-zero, the $j = 1, m = 0$ term is larger. In fact, the $j = 0, m = 0$ term can be thought of as a direct measurement of the distinguishability to the extent that the distinguishability affects polarization measurements. Two photons both polarized along $H$ can be distinguishable in some other degree of freedom, but this distinguishability will have no effect on polarization measurements. On the other hand making two photons distinguishable when one is $H$ and the other is $V$ will always affect polarization measurements because the distinguishing degree of freedom is correlated to polarization. It is the degree of correlation between polarization and all the other degrees of freedom of the photons that is measured in the $j = 0, m = 0$ density matrix element.

In summary, ours is the only two-photon polarization state tomography scheme that is able to properly account for distinguishability so as to correctly predict the outcomes of all measurements. It deals equally well with indistinguishable, distinguishable and partially distinguishable photon states.

### 3.4.2 Measuring the three-photon accessible density matrix

The three-photon case proceeds in a similar way to the two-photon case, but three-photon polarization states offer a much richer state space than two photon states. Part
of this richness manifests itself in the more complicated characterization needed. Since $S_3$ has multi-dimensional irreps, the density matrix contains not one, but two $j = 1/2$ sectors that characterize the distinguishability. There are four different numbers in this submatrix. The total population of the two $j = 1/2$ submatrices plays a role similar to the $j = 0, m = 0$ component for two photons, while the distribution of populations and the coherences within the $j = 1/2$ submatrices also contains additional information about the polarization.

Even if we forget about the complexities of $S_3$ and consider only the case of indistinguishable photons we obtain a richer set of states. For two polarizations the N00N is equivalent to the Dicke state with $m = 0$, but for three photons these are distinct states. In fact, a range of coherent, squeezed, and Dicke states are available with three photons. Some of these states have features that make them attractive for quantum metrology applications[40].

We undertook state estimation for a wide range of three-photon polarization states[105] using the techniques developed in this chapter. To create the states we employed the apparatus pictured in figure 3.4. A two-photon polarization state was created in SPDC using the same setup as in the previous section, but then a third photon was added to it from an attenuated laser pulse (marked LO for local oscillator in the figure). This addition of the third photon is achieved by a process known as mode-mashing which involves simply bringing photons together on a beamsplitter and post-selecting on events where all the photons leave in one port. The process is non-deterministic, but when it succeeds there is a clear experimental signal of its success. It can be shown that if, after mashing, the photons are spatio-temporally indistinguishable, then any three-photon polarization state can be created simply by polarizing three photons and mode-mashing them together. The polarization of the state was controlled unitarily with waveplates and also non-unitarily with a variable partial polarizer (VPP). The ability to perform non-unitary transformations meant that the relative angle between the polarizations could be
In the real experiment true single-photon sources were not used, resulting in some probability that the post-selection would incorrectly signal success if, for example, three photons from the laser pulse exited the same output port or if two SPDC pairs did. By adjusting the relative power of SPDC and the attenuated laser pulse these ‘background’ counts could be minimized, resulting in the post-selection correctly selecting for the desired state around 60% of the time. More details of the experiment can be found in [105]. Here we will discuss the characterization of the states.

To characterize the three-photon polarization state, the state was directed to a polarizing beamsplitter to separate $H$ and $V$ photons. The $V$ output was further sent to two cascaded beamsplitters. Assuming perfect detector, the presence of three photons at the $V$ port would trigger a coincidence between detectors B, C and D 7.3% of the time. Similarly, when one photon left in the $H$ port and two in the $V$ port, it was signaled by coincident firing of two of detectors B, C and D and detector A. If the detector were perfectly efficient this would happen 45% of the time. Relative detector efficiencies could also be calibrated out so that the rate of three-fold coincidences between different detectors could be used to measure the average number of times three photons were present at the $V$ port of the PBS, thereby projecting onto $m = -3/2$, and the number of times that two photons were present at the $V$ port and one at the $H$ port, thereby projecting onto $m = -1/2$. 

Figure 3.4: Apparatus for measuring the three-photon accessible density matrix changed.
A quarter and half-waveplate placed before the polarizing beamsplitter allowed these two measurements to be rotated into an informationally complete set of projection operators capable of spanning all the elements of the accessible density matrix. The outcomes of these projective measurements were then fed into the maximum-likelihood convex problem solver SeDumi[115]. The accessible density matrix was constrained to take the block-diagonal form of equation 3.56 with the two \( j = 1/2 \) subspaces constrained to be equal.

Data from a typical experimental tomography run is shown in Table 3.4.2.

After the detector efficiencies were taken into account, this data was fit to the accessible density matrix below

\[
\begin{pmatrix}
0.2448 & 0.0057 + 0.0008i & -0.0323i & 0.0325 - 0.1437i \\
0.0057 - 0.0008i & 0.1577 & 0.0819 - 0.0085i & -0.0653 - 0.0286i \\
0.0323i & 0.081908 + 0.0085i & 0.1763 & 0.0609 - 0.0060i \\
0.0325 + 0.1437i & -0.0653 + 0.0286i & 0.0069 + 0.0060i & 0.2639 \\
0.0494 & 0.0220 + 0.0297i \\
0.0220 - 0.0297i & 0.0293 \\
0.0494 & 0.0220 + 0.0297i \\
0.0220 - 0.0297i & 0.0293
\end{pmatrix}
\]
The non-ideality of the state preparation is apparent from the population in the two \( j = 1/2 \) subspaces which amounted to 16% of the total population of the state. Unlike in the two-photon experiment where the population of the \( j = 0 \) could be essentially eliminated by temporally overlapping the photons, in the three-photon experiment the particles could not be made indistinguishable. This is because the two photons generated via SPDC are fundamentally distinguishable from the photon taken from the laser due to temporal broadening from the SHG and SPDC processes and because of residual spectral entanglement due to phase-matching in the SPDC. All of these effects can be corrected with stronger spectral filtering, and experiments to test the dependence of the \( j = 1/2 \) populations on spectral filtering are ongoing.

Within the \( j = 3/2 \) space, the state looks like a N00N state, with strong coherences between the \( |m = 3/2 \rangle \) and \( |m = -3/2 \rangle \) terms. The extra diagonal population can be largely accounted for as being due to the unwanted background coincidence sources. When these were subtracted from the measured density matrix the result is quite a bit closer to a N00N state[105].

### 3.5 Summary and conclusions

We have developed a complete theory of quantum state tomography for indistinguishable particles. This theory elegantly handles the reduction in the size of the Hilbert space when particles are fundamentally indistinguishable and provides a lower-bound measure of just how distinguishable they are. It is the most complete state characterization possible for experimentally indistinguishable particles. The theory was applied to two and three-photon states, and their accessible density matrices were presented.

It is hoped that the techniques outlined in this chapter will become the standard characterization method for this class of state, which has applications in quantum metrology, lithography and orienteering, just as ordinary quantum state tomography[55] has become
standard for distinguishable particles.
Chapter 4

Quantum state tomography with mutually-unbiased bases

One should perform his deeds for the benefit of mankind with an unbiased approach because bias gives birth to evil, which creates thousands of obstacles in our path.
–The Rig Veda

4.1 Introduction

Since quantum state tomography was first introduced as an experimental tool, relatively little research has been focused on improving it. In fact, as quantum state tomography techniques have spread to other physical systems of entangled particles such as trapped ions[50], superconducting qubits[110], and quantum dots[113], the prescription for doing quantum state tomography given by James, Kwiat, Munro and White[55] has been implemented almost exactly as it was originally laid down.

There is no particular reason why this should be the case. The projectors measured in [55] were picked for experimental convenience, but in principle, any set of measurements could have been used, as long as they form a spanning set for the Hilbert space of
density matrices. The arbitrariness of the set used by James et al. is apparent upon inspection of Table 4.1. Sixteen projections were taken, the minimal number required for completeness, and all them were onto tensor products of eigenstates of the $\sigma_x$, $\sigma_y$ and $\sigma_z$ Pauli operators in the two qubits. Because there are 36 such tensor products, but only 16 projectors in the set, twenty of the combinations are excluded, and since there is no natural way to select the set of 16 included and 20 non-included measurements, the selected set appears unbalanced. The first four measurements are projectors onto eigenstates of $\sigma_z \otimes \sigma_z$, but this is the only basis where all four outcomes are included. $\sigma_y \otimes \sigma_z$, $\sigma_x \otimes \sigma_z$, $\sigma_z \otimes \sigma_z$ and $\sigma_y \otimes \sigma_z$ have two projectors each in the set and $\sigma_x \otimes \sigma_x$, $\sigma_x \otimes \sigma_y$, $\sigma_y \otimes \sigma_x$ and $\sigma_y \otimes \sigma_y$ have one projector each. This choice of projectors results, as one might expect, in a much better estimate of the polarization in the $Z$ direction than in the $X$ and $Y$ directions. Computer simulations show that with 10,000 states available for measurement, the James strategy results in a 53% higher standard deviation for $\sigma_y \otimes \sigma_y$ outcomes than for $\sigma_z \otimes \sigma_z$ outcomes for an input state like $|DD\rangle$ that gives the same average value for the measurements in the two bases.

There is no reason that the measurements should be biased in this way. From an informational point of view, the James measurement strategy is inherently wasteful because it measures only some elements of PVMs it includes, whereas one could collect all the outcomes of all the PVM elements for the same number of input photons. From this perspective, measuring all 36 tensor products of single-qubit projectors is no more costly than measuring only sixteen, but offers more information which should both eliminate the bias in favour of the $\sigma_z \otimes \sigma_z$ basis and give a better overall estimate of the state.

There need be no concern about having too many measurements as there are well-established ways of adapting linear fitting and maximum likelihood fitting to over-complete sets of measurements. If we are interested in understanding what limitations quantum mechanics imposes on our ability to estimate the density matrix from a finite number of copies, we need to consider sets of measurements composed of complete PVMs
Table 4.1: The tomographic projectors used by James et al[55]. Sixteen combinations of Pauli operators were chosen out of the 36 possible combinations, resulting in better estimation of $\sigma_z \otimes \sigma_z$ outcomes than, for instance, $\sigma_y \otimes \sigma_y$. 

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>Mode 1</th>
<th>Mode 2</th>
<th>$h_1$</th>
<th>$q_1$</th>
<th>$h_2$</th>
<th>$q_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$</td>
<td>H\rangle$</td>
<td>$</td>
<td>H\rangle$</td>
<td>45°</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>$</td>
<td>H\rangle$</td>
<td>$</td>
<td>V\rangle$</td>
<td>45°</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>$</td>
<td>V\rangle$</td>
<td>$</td>
<td>V\rangle$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>$</td>
<td>V\rangle$</td>
<td>$</td>
<td>H\rangle$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>$</td>
<td>R\rangle$</td>
<td>$</td>
<td>H\rangle$</td>
<td>22.5°</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>$</td>
<td>R\rangle$</td>
<td>$</td>
<td>V\rangle$</td>
<td>22.5°</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>$</td>
<td>D\rangle$</td>
<td>$</td>
<td>V\rangle$</td>
<td>22.5°</td>
<td>45°</td>
</tr>
<tr>
<td>8</td>
<td>$</td>
<td>D\rangle$</td>
<td>$</td>
<td>H\rangle$</td>
<td>22.5°</td>
<td>45°</td>
</tr>
<tr>
<td>9</td>
<td>$</td>
<td>D\rangle$</td>
<td>$</td>
<td>R\rangle$</td>
<td>22.5°</td>
<td>45°</td>
</tr>
<tr>
<td>10</td>
<td>$</td>
<td>D\rangle$</td>
<td>$</td>
<td>D\rangle$</td>
<td>22.5°</td>
<td>45°</td>
</tr>
<tr>
<td>11</td>
<td>$</td>
<td>R\rangle$</td>
<td>$</td>
<td>D\rangle$</td>
<td>22.5°</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>$</td>
<td>H\rangle$</td>
<td>$</td>
<td>D\rangle$</td>
<td>45°</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>$</td>
<td>V\rangle$</td>
<td>$</td>
<td>D\rangle$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>$</td>
<td>V\rangle$</td>
<td>$</td>
<td>L\rangle$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>$</td>
<td>H\rangle$</td>
<td>$</td>
<td>L\rangle$</td>
<td>45°</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>$</td>
<td>R\rangle$</td>
<td>$</td>
<td>L\rangle$</td>
<td>22.5°</td>
<td>0</td>
</tr>
</tbody>
</table>
that do not waste information in obvious ways.

In recent years a few groups have started doing two-photon polarization quantum state tomography using these 36 different projectors[3], usually by setting up four detectors, one at each output port of two polarizing beamsplitters. This has resulted in the elimination of the bias in favour of a particular single-qubit basis and better overall estimation of the quantum state.

One can still ask, though, whether this is the optimal way of doing quantum state tomography, or whether there exists some other set of measurements that might in some way be better. To be precise, we can pose the question, ‘Given \( N \) copies of a quantum state from a source, what set of measurements will, on average, allow the best estimate of the density matrix of the source?’ This question is important not just for technical reasons, but seems to be related to the deeper epistemological problem how we obtain information about quantum states, and hence to the ontological problem of what quantum states are[75]. In fact, as will be seen in section 4.7, the optimal set of measurements for quantum state tomography is directly tied to a compelling description of quantum states in terms of the discrete Wigner function which has several conceptual advantages over the density matrix as a description of quantum states.

This chapter will present the first experiment to undertake optimal quantum state tomography within the framework of PVM-based measurements\(^1\). As we will show, the key to selecting measurements is to try to minimize measurement overlap so as to also minimize the amount of redundant information being collected. This requirement results in the use of mutually-unbiased bases, a special class of projective measurements.

---

\(^1\)While for practical reasons PVMs on individual copies of a quantum state are the most popular class of measurements implemented in experiments, they are not the only measurements that could potentially be used in state estimation. Theorists have long been interested in other approaches to optimal information extraction, and have shown that certain classes of POVMs[98] can do better than PVMs, and, perhaps more surprisingly, that joint measurements performed over multiple copies of a quantum system can extract more information than measurements on the individual copies[82]. While theoretically interesting, a real-world implementation of either approach would not be possible with present technology, except in very specialized circumstances[79, 77].
particularly well-suited to quantum state estimation.

### 4.2 Comparing different tomography strategies

In order to gauge which quantum state estimation strategy is best, we need a metric to determine how close an estimate of the quantum state is to the true state of the system. The ‘true state’ of the system is itself a philosophically troubling notion, since in principle it would take an infinitely large dataset to determine what the true state is. For our purposes we may consider it to simply be the asymptotic estimated state after a sufficiently large number of measurements have been made. Typically in our experimental systems, systematic errors, particularly errors in setting the waveplates accurately, limit the convergence of the density matrix elements at the 1% level. This gives us an experimentally convenient definition of the true density matrix as the best estimate we can obtain before systematic effects become dominant. For numbers of copies of the state small enough that systematic effects are overwhelmed by uncertainty due to quantum randomness, we can probe the fundamental limits of the estimation strategy by comparing a given estimate with the asymptotic estimate.

For a given number of copies of the state we are interested not so much in the estimated density matrix as in the variance of the estimated density matrix over several repetitions of the characterization. For each value of number of copies $N_{\text{tot}}$ we can develop a histogram of estimates and distances of the estimates from the asymptotic estimate.

In order to make this comparison there are several different distance measures available, most notably the fidelity[57], the quantum Chernoff-bound[8] and the Hilbert-Schmidt distance or trace distance[91]. Of these, a convincing argument can be made that the Chernoff bound is the best physically-motivated bound since it represents the limit on one’s ability to distinguish two quantum states from a finite number of copies[34]. Suppose an experimentalist wishes to distinguish whether a quantum state is $\rho$ or $\sigma$, but
is only given \( N \) copies of the state to work with. It can be shown that her probability of incorrectly identifying the state goes asymptotically as 
\[ P_e \approx e^{N \ln \lambda_{cb}(\rho, \sigma)}, \]
where \( \lambda_{cb} \) is the quantum Chernoff bound
\begin{equation}
\lambda_{cb}(\rho, \sigma) = \min_{0 \leq s \leq 1} \text{Tr} \left\{ \rho^s \sigma^{1-s} \right\}.
\end{equation}

When either \( \sigma \) or \( \rho \) is pure, the quantum Chernoff bound is equal to the square root of the fidelity \( F(\sigma, \rho) \). For all states, the square root of the fidelity provides an upper bound for the Chernoff bound \( \sqrt{F(\sigma, \rho)} \geq \lambda_{cb}(\sigma, \rho) \). In a previous study comparing different tomography techniques\[34\] a comparison was made between using the quantum Chernoff bound and using the fidelity and the two were found to give the same qualitative results. In the same work, the Hilbert-Schmidt distance was found to not differentiate between different choices of measurement bases that gave very different results in terms of the Chernoff bound and the fidelity. For this reason the Hilbert-Schmidt distance was not considered a good choice for comparing different tomography configurations, while the Chernoff bound and fidelity were considered equivalent.

In the present work we use the fidelity rather than the Chernoff bound since the fidelity can be represented as an analytic function of the density matrices rather than as a minimization. This makes it both tractable to calculate and allows us to derive analytic results with it. It is also a more familiar measure for experimentalists and is the distance measure most often used in experiments\[91\].

### 4.3 Convergence of the estimate towards the true density matrix

Consider an estimate of the density matrix \( \rho_{\text{est}} \) taken from measurements performed on the true state \( \sigma \). The fidelity of the estimate with the true state is defined as\[91\]
\[
F = \left( \text{Tr} \sqrt{\sigma \rho_{\text{est}} \sqrt{\sigma}} \right)^2.
\]
If the estimated density matrix is reasonably close to $\sigma$, then we can write $\rho_{\text{est}} = \sigma + \delta \rho$ where $\delta \rho$ is a traceless error term whose elements will be small compared to those of $\sigma$. Small here means of order $1/\sqrt{N}$ if $N$ individual measurement outcomes were averaged to create the expectation values necessary for the state estimation. We can quantify the deviation from the true density matrix by the infidelity,

$$I = 1 - F.$$  \hfill (4.3)

In the perturbative limit we can rewrite the fidelity in terms of $\delta \rho$ as

$$F = \left( \text{Tr} \sqrt{\sigma^2 + \sigma \delta \rho \sqrt{\sigma}} \right)^2.$$  \hfill (4.4)

$\sigma$ and $\sqrt{\sigma}$ commute so that

$$F = \left( \text{Tr} \sqrt{\sigma^2 + 2\sigma \delta \rho \sqrt{\sigma}} \right)^2.$$  \hfill (4.5)

In general this is a difficult expression to deal with, but it becomes both simple and illuminating in two special cases. If the true state is the pure state $|\psi\rangle$ then $\sigma = |\psi\rangle \langle \psi|$. It follows that $\sqrt{\sigma} = \sigma$ and

$$\sqrt{\sigma} \delta \rho \sqrt{\sigma} = \sigma \delta \rho \sigma$$  \hfill (4.6)

$$= |\psi\rangle \langle \psi| \delta \rho |\psi\rangle \langle \psi|$$  \hfill (4.7)

$$= \sigma \text{Tr} [\sigma \delta \rho]$$  \hfill (4.8)

And since $\sigma = \sigma^2$,

$$F = \left( \text{Tr} \sqrt{\sigma^2 + \sigma^2 \text{Tr} \sigma \delta \rho} \right)^2$$  \hfill (4.9)

$$= \left( \text{Tr} \sigma \sqrt{1 + \text{Tr} \sigma \delta \rho} \right)^2$$  \hfill (4.10)

$$= 1 + \text{Tr} \sigma \delta \rho.$$  \hfill (4.11)

When $\sigma$ is pure, $\text{Tr} \sigma \delta \rho$ will always be non-positive since to the extent that a measurement in the basis of $\sigma$ is in error, it must mean a reduction in the element of $\rho_{\text{est}}$ corresponding to $|\psi\rangle$ along with an increase in those elements corresponding to states orthogonal to $|\psi\rangle$. [130]
The important thing to note here is that the reduction in the fidelity comes in linearly with the density matrix elements. Since the uncertainty in $\rho$ scales as $1/\sqrt{N}$, we expect the infidelity to also scale as $1/\sqrt{N}$. Indeed this is what we observe. In figure 4.1(a), showing the fidelity between the estimated and true density matrices for an experimentally collected tomography data, the slope on a log-log plot is $N_{\text{tot}}^{-0.42\pm0.06}$, reasonably close to the expected $N^{-1/2}$.

The other simple and illustrative case occurs when $\sigma$ is the maximally-mixed state, $\sigma = \frac{1}{D} I$ where $D$ is the dimensionality of the Hilbert space. When this is the case, we have

$$F = \left( \text{Tr} \sqrt{\sigma^2 + \sqrt{\sigma} \delta \rho \sqrt{\sigma}} \right)^2 \quad (4.12)$$

$$= \left( \text{Tr} \sqrt{\frac{1}{D^2} I + \frac{1}{\sqrt{D}} \frac{1}{\sqrt{D}} \delta \rho \sqrt{D} I} \right)^2 \quad (4.13)$$

$$= \left( \frac{1}{D} \text{Tr} \sqrt{I + D \delta \rho} \right)^2. \quad (4.14)$$

We can now Taylor expand the square root using the expansion $\sqrt{1 + x} = 1 + \frac{1}{2} x - \frac{1}{4} x^2 + \ldots$

$$F = \left( 1 \frac{1}{D} \text{Tr} I + \frac{D}{2} \text{Tr} \delta \rho - \frac{D}{4} \text{Tr} \delta \rho^2 \right)^2 \quad (4.15)$$

$$= \left( 1 - \frac{D}{4} \text{Tr} \delta \rho^2 \right)^2 \quad (4.16)$$

$$= 1 - \frac{D}{2} \text{Tr} \delta \rho^2 + \frac{D}{16} (\text{Tr} \delta \rho^2)^2 \quad (4.17)$$

$$\approx 1 - \frac{D}{2} \text{Tr} \delta \rho^2, \quad (4.18)$$

where we have used the fact that $\delta \rho$ is traceless in the second line. We can recognize $\text{Tr} \delta \rho^2$ as the degree of polarization of $\rho_{\text{est}}$. When the state is completely unpolarized, any polarization results in infidelity with the true state. Unlike in the pure state case, the infidelity now scales quadratically with $\delta \rho$. This is also observed in experimental measurements as shown in figure 4.3 where the infidelity can be seen to be proportional to $N_{\text{tot}}^{-1.03\pm0.06}$. 


(a) Dependence of the infidelity on $N$ for the pure quantum state $|HH\rangle$. The scaling goes roughly as $1/\sqrt{N}$.

(b) Dependence of the infidelity on $N$ for the maximally-mixed state $\frac{1}{4} I_4$. The scaling goes as $1/N$. 
4.4 The choice of measurement bases

These general scaling rules are a universal feature of quantum state tomography, but the
details of how $\delta \rho_{\text{est}}$ depends on the number of measurements that have been made rely
crucially how those measurements relate to one another. In order to distinguish different
states, projections must be made in different ‘directions’ in the Hilbert space. In fact, it
makes intuitive sense that the most powerful measurement strategy will be the one that
makes the direction of different measurements as different as possible. This intuition is
indeed correct, but before it can be applied we need to understand what is meant by
‘direction’ and to develop a mathematical framework for deciding what ‘as different as
possible’ means.

In the Hilbert space of projectors, the natural measure of overlap is the Hilbert-
Schmidt distance defined as

$$\text{Tr} \left[ \hat{P}_1 \hat{P}_2 \right].$$

(4.19)

For a set of projectors to be maximally distant from each other, the RMS value of
this overlap should be made as small as possible. Figure 4.1 plots this overlap for the
36-measurements formed by projecting onto eigenstates of the two-qubit Pauli operators.
We shall call the tomography scheme that makes measurements in these bases standard
separable quantum state tomography (SSQST). It is clear from the graph that these
overlaps are not all equal. In particular, there are pairs of bases that share a common
single-qubit Pauli operator, for example, $\sigma_z \otimes \sigma_z$ and $\sigma_x \otimes \sigma_z$. For these pairs of bases,
the overlaps are either 0 or 0.5 depending on whether a given pair of projectors share
the same eigenstate of $\sigma_z$ for the second qubit. Other pairs of bases such as $\sigma_z \otimes \sigma_z$ and
$\sigma_x \otimes \sigma_x$ have no eigenstates in common and therefore all of the overlaps are 0.25.

Unfortunately there is no way to construct a complete basis where all the overlaps are
equal from tensor products of single-qubit eigenstates. One can interpret the inequality of
the overlaps as being due to the failure of the measurement scheme to reflect the symmetry
Figure 4.1: Plot of the Hilbert-Schmidt overlap of the projectors in standard separable state tomography.
of the underlying Hilbert space. Consider, for example, a maximally-entangled state. Such states can be thought of as having all their information contained in correlations and none in the single-photon polarizations. In order to characterize such a state, a scheme like SSQST must make single-qubit measurements and examine the correlation data to determine the state. In the process of collecting these correlations, the single-qubit polarizations must be measured multiple times. A more efficient scheme could measure each single-qubit polarization once, and then use entangling measurements to determine the correlations that cannot be determined from the single-qubit measurements. Once one allows the possibility of using entangling measurements it becomes possible to create an optimal measurement scheme where the overlap between any pair of PVM elements drawn from different bases is equal. Mathematically, if \( \delta \) and \( \gamma \) label PVM elements while \( \alpha \) and \( \beta \) label PVMs, such bases will have the property

\[
\text{Tr} \left[ \hat{P}_{\alpha,\delta} \hat{P}_{\beta,\gamma} \right] = \delta_{\alpha,\beta} \delta_{\delta,\gamma} 1/D.
\] (4.20)

These bases were first introduced in the context of quantum state estimation by Wootters and Fields\[127\] who argued that this condition 4.20 is precisely the one that needs to be satisfied for optimal state estimation. They called these sorts of PVMs \textit{mutually-unbiased bases} or MUBs. They were able to show that the maximum number of MUBs in a Hilbert space of dimension \( D \) is \( D + 1 \), and that a set of MUBs exists whenever \( D \) is the power of a prime number.\(^2\)

Wootters and Fields calculated the Shannon entropy reduction per measurement in a general complete QST scheme and argued that measurements in MUBs would maximize this quantity. Unfortunately, their geometric argument is hard to apply to over-complete sets of bases such as the bases of SSQST. Here we present an alternate argument based on the density matrix error analysis discussion of Chapter 2.

We will consider the advantage that MUBs have for estimating the maximally-mixed

\(^2\)When \( D \) is not a power of a prime, it is generally believed that MUBs do not exist, although proving this remains an open problem. There is strong numerical evidence supporting this belief for \( D = 6 \)[118].
state from experimental data. Since this is the average of all states on the Hilbert space it seems at least plausible that the scheme that estimates the maximally-mixed state best will, on average, estimate all states better. Clearly if one were interested in estimating a particular range of states, then one could construct a tomographic scheme tailored to those states that could do better.

Recall from Chapter 2 that the variance in the density matrix can be expressed as

$$\langle \Delta \rho \rangle = \sum_{kq,ab} \frac{\partial \rho_{ij}}{\partial P_{kq}} \frac{\partial \rho_{ij}}{\partial P_{ab}} \delta P_{kq} \delta P_{ab}.$$  \hspace{1cm} (4.21)$$

We can simplify the problem by neglecting the error in the normalization so that $P_{kq} = n_{kq}/N$ where $N$ is the number of copies per basis. For the maximally-mixed state, $P_{kq} = 1/D$ in any basis, so $\langle P_{kq} \rangle = 1/4$ and $\delta P_{kq} = \sqrt{1/4N}$. By neglecting the error in the normalization we make all the $\delta P_{kq}$ statistically independent so that

$$\delta P_{kq} \delta P_{ab} = \delta_{kq} \delta_{ab} \frac{1}{4N}.$$  \hspace{1cm} (4.22)$$

Now $N$ refers to the number of copies of the state used per basis, but in comparing tomography schemes we need to consider the total number of copies available to be measured $N_{tot}$. We will assume that the available copies are split equally between the different bases so that $N = N_{tot}/B$ where $B$ is the number of bases in a given scheme. For two qubits $B = D + 1 = 5$ for MUBs tomography and $B = 9$ for standard separable tomography.

Applying this fact and equation 4.22 to equation 4.21, we obtain

$$\langle \Delta \rho \rangle = \frac{B}{4N} \sum_{kq} \left( (M^\dagger M)^{-1} M^\dagger \right)_{ij,kq}^2.$$  \hspace{1cm} (4.23)$$

And given the expression for the error in the fidelity for the maximally-mixed state derived in equation 4.18, we can express the infidelity as

$$I = 1 - F = \frac{1}{4} \text{Tr} [(\Delta \rho)^2]$$

$$= \frac{B}{16N_{tot}} \sum_{i=1}^4 \sum_{kq} \left( (M^\dagger M)^{-1} M^\dagger \right)_{ij,kq}^2.$$  \hspace{1cm} (4.25)$$
This result allows us to calculate the ratio of the infidelity achieved via the two tomography methods for the same value of \( N_{\text{tot}} \).

\[
\frac{I_{\text{SSQST}}}{I_{\text{MUBs}}} = \frac{B_{\text{SSQST}}}{B_{\text{MUBs}}} \frac{\sum_{i=1}^{4} \sum_{kq} \left| \left( (M_{\text{SSQST}}^\dagger M_{\text{SSQST}})^{-1} M_{\text{SSQST}}^\dagger \right)_{ii,kq} \right|^2}{\sum_{i=1}^{4} \sum_{kq} \left| \left( (M_{\text{MUBs}}^\dagger M_{\text{MUBs}})^{-1} M_{\text{MUBs}}^\dagger \right)_{ii,kq} \right|^2}.
\]

(4.26)

We expect that the elements of \((M^\dagger M)^{-1} M^\dagger\) will be large whenever two rows of \( M \) are close to being degenerate. This fits with our intuition that a good tomography scheme will distribute its projectors widely in the Hilbert space. For reasonably well-balanced tomography schemes like SSQST and MUBs we expect that over-complete sets will tend to have smaller values for the sum than merely complete sets because larger sets of measurements will take a smaller contribution from each measurement to arrive at a density matrix element. The sum of the squares of a large number of small contributions will generally be larger than the sum of squares of a smaller number of large contributions.

By this argument we expect the sum to be somewhat smaller for SSQST than for MUBs, but not enough to make up for the leading factor \( \frac{B_{\text{SSQST}}}{B_{\text{MUBs}}} = 9/5 \). When we calculate the ratio of the sums for the two tomography scheme we find that it has a value of 0.7691 which is indeed greater than 5/9. The total advantage expected for MUBs is 0.7691 \times 9/5 = 1.38 in reasonable agreement with measurements, as we shall see.

For states other than the maximally-mixed state, the advantage of MUBs over SSQST will be state-dependent. As a general rule, though, we expect MUBs to display the most significant advantage for entangled states and the least advantage for separable states since SSQST is biased towards better estimation of single-qubit polarizations at the expense of correlations. The magnitude of the advantage will be determined by a relation very similar to equation 4.26, involving a sum of squares of linear map elements, and so as long as the map is non-singular (which it obviously isn’t for either MUBs or SSQST), this ratio will be on the order of unity.
Pauli operators & Eigenstates/Mutually-unbiased bases
\[\sigma_z \otimes I, I \otimes \sigma_z, \sigma_z \otimes \sigma_z\] & \[|HH\rangle, |HV\rangle, |VH\rangle, |VV\rangle\]
\[\sigma_x \otimes I, I \otimes \sigma_y, \sigma_x \otimes \sigma_y\] & \[|DR\rangle, |DL\rangle, |AR\rangle, |AL\rangle\]
\[\sigma_y \otimes I, I \otimes \sigma_x, \sigma_y \otimes \sigma_x\] & \[|RD\rangle, |RA\rangle, |LD\rangle, |LA\rangle\]
\[\sigma_y \otimes \sigma_y, \sigma_z \otimes \sigma_x, \sigma_x \otimes \sigma_z\] & \[\frac{1}{\sqrt{2}} (|RL\rangle + i|LR\rangle), \frac{1}{\sqrt{2}} (|RL\rangle - i|LR\rangle),\]
\[\quad \frac{1}{\sqrt{2}} (|RR\rangle + i|LL\rangle), \frac{1}{\sqrt{2}} (|RR\rangle - i|LL\rangle)\]
\[\sigma_x \otimes \sigma_x, \sigma_y \otimes \sigma_z, \sigma_z \otimes \sigma_y\] & \[\frac{1}{\sqrt{2}} (|RV\rangle + i|LH\rangle), \frac{1}{\sqrt{2}} (|RV\rangle - i|LH\rangle),\]
\[\quad \frac{1}{\sqrt{2}} (|RH\rangle + i|LV\rangle), \frac{1}{\sqrt{2}} (|RH\rangle - i|LV\rangle)\]

Table 4.2: The construction of mutually-unbiased bases according to the method of [73].

The states making up mutually-unbiased bases for two qubits are the eigenstates of sets of three mutually commuting Pauli operators.

### 4.5 Constructing MUBs for two qubits

When \(D\) is a power of two, it can be shown that the eigenstates of mutually commuting sets of Pauli operators (i.e. tensor products of the Pauli matrices) form MUBs[73]. Since mutually-commuting sets of Pauli operators are easy to construct, this offers a simple and elegant construction algorithm for MUBs.

Consider the pairwise tensor products of Pauli operators, including the identity, \(\sigma_\mu \otimes \sigma_\nu\) where \(\sigma_\mu = (\sigma_x, \sigma_y, \sigma_z, I)\) for \(\mu = (1, 2, 3, 4)\). If we exclude the term \(I \otimes I\), we can divide the remaining operators into 5 sets of 3 operators sharing a common set of eigenvalues.

One possible such division is shown in table 4.2.

We note that three of the bases are separable and the other two are maximally-entangled. This represents the optimal partitioning of the measurements on the Hilbert space. The three bases are used to probe the single-qubit polarizations, while the remaining two bases directly probe the correlations missed by this choice of single-qubit measurement basis. From an experimental point of view this division of the measure-
ments into separable and maximally-entangled projections is fortuitous since it means that MUBs tomography can be implemented relatively easily in a linear optics system.

4.6 Experiment

To experimentally demonstrate the superiority of MUBs for quantum state estimation tasks, many individual quantum state estimation experiments must be repeated and statistically analyzed to see which strategy, on average, gives the highest fidelity with the true or asymptotic state. To that end, we developed an automated quantum state tomography system capable of performing the required measurements for both MUBs tomography and SSQST.

The main requirement in such a system is that it be capable of performing projections onto separable states and also onto maximally-entangled states. For two qubits, all maximally-entangled states are related to each other by separable unitary transformations, and all separable states are similarly related to each other\(^3\). Thus if a system is capable of performing projections onto any maximally-entangled state and onto any separable state and of performing arbitrary unitary transformation on the individual qubits, then all the necessary measurements required for MUBs tomography and standard separable tomography can be achieved.

4.6.1 Separable measurements

All of the necessary separable measurements can be achieved by performing ordinary polarization analysis on the two qubits separately and keeping track of correlations between measurements\(^5\). As was seen in Chapter 2, this involves simply a quarter-waveplate, a half-waveplate and a polarizer for each of the beam paths, followed by coincidence

\(^3\)This follows from the Schmidt decomposition of the two qubit state. Using this decomposition it can easily be shown\(^9\) that the only parameter of the state not entirely determined by single-qubit operations is the Schmidt number or, equivalently, the degree of entanglement.
detection to determine when the single-photon detectors for the two qubits fire at the same time.

While it is possible to collect all four PVM elements for a separable measurement at the same time, the unequal collection efficiencies of the detectors makes it difficult to normalize properly. In practice it was found to be easier to use a single pair of detectors and 36 different waveplate settings to record all 36 projectors necessary for standard separable tomography. The twelve separable MUBs projectors were collected using the same setup. This approach tarnishes some of the luster of MUBs tomography since the amount of information being extracted from each photon is not optimal as it would be if the number of measurements were simply the number of different bases. From a practical point of view, though, there is very little difference since measurement outcomes are normalized not to the number of photon pairs produced by the source, but to the number detected. This normalization process sweeps collection inefficiencies under the rug while maintaining the ability to measure information-theoretic inefficiencies in the amount of information extracted about the state from each detected photon pair. This is all that is needed to validate the information-theoretic advantages of MUBs.

The measurement apparatus used to perform the separable tomography is shown in figure 4.2(a). The source of photons is the type-I source described in Chapter 2 and is capable of producing states with a variable degree of entanglement that can be controlled by the pump polarization. Liquid crystal waveplates (LCWPs) can be used to randomize the polarization and produce mixed polarization states. Standard polarization analyzers are followed by single-photon counting modules that fire upon detection of a single photon, while coincidence electronics distinguish SPDC pairs from background light.

The measurement was set by controlling the angle of half and quarter-waveplates mounted in Newport PR50 motorized rotation stages and controlled via GPIB.

Coincidences were recorded during 3000 0.2-second intervals for each of the 36 mea-
measurements. Each interval contained roughly 30 coincidence events per basis. Because changing waveplate angles took several seconds, waveplate angles were only changed after 3000 counting intervals had expired for each of the measurements. The entire cycle of 3000 measurements for each of the 36 settings took approximately 36 hours, a duration that was limited by the slow communication speed of the USB interface between the coincidence counting module and the computer. This measurement scheme was not optimal as long-term power drifts due to day-night temperature variations and other environmental effects would unequally affect the different measurement bases. Although one would not expect such drift to affect polarization or waveplate retardance, it could affect laser power, mode-hopping, liquid-crystal phase shift and detector dark count rates. Were this a problem, it would be observable in the detection rates at the single-photon counting modules. These rates, however, were stable to within shot-noise fluctuations, providing reassurance that such drift was not a statistically significant problem.

4.6.2 MUBs measurements

The maximally-entangled measurements can be obtained by applying two-photon interference as described in Chapter 1. Recall that when a two-photon polarization state is incident on a 50-50 beamsplitter, the outcome is determined by interference. If the polarization state is permutation symmetric, the interference will be constructive for probability amplitudes for the photons leaving the beamsplitter in the same port, and
the photons will never leave from opposite ports. If the polarization state is permutation anti-symmetric, the interference will be constructive for the two photons leaving in opposite ports and the photons will always do so.

If this interference effect is followed by post-selection on one photon being horizontally polarized and one being vertically polarized, then events in which the photons leave in opposite ports can be caused only by a component of the state along the anti-symmetric state with one horizontal photon and one vertical photon, the singlet state \( \frac{1}{\sqrt{2}} (|HV\rangle - |VH\rangle) \). Similarly, events in which the photons leave from the same port can be uniquely identified with the symmetric component with one vertical and one horizontal photon, the triplet state \( \frac{1}{\sqrt{2}} (|HV\rangle + |VH\rangle) \). Thus detection of such events constitute projections onto these states, and measurement of the frequency of such outcomes as a fraction of all outcomes gives a measure of the expectation value of the projectors onto these states.

In a real experiment two major considerations must be taken into account. First, the visibility of the two-photon interference is typically significantly less than 100%. In our experiment it was typically 93%, limited by a number of factors including the unequal transmission and reflection coefficients of the beamsplitter and imperfect spatial overlap at the beamsplitter. Second, for the singlet state projection, if the reflection and transmission coefficients are unequal, and coincidences are measured by projecting one output port onto \( V \) and the other onto \( H \), then the detection probability depends not only on the \( \frac{1}{\sqrt{2}} (|HV\rangle - |VH\rangle) \) projection, but on the \( |HV\rangle \) and \( |VH\rangle \) projections individually, since one of these will correspond to two transmission events and the other to two reflection events. The triplet state projection does not suffer from this problem since triplet state projections require the \( H \) and \( V \) photons to leave the same port of the beamsplitter and hence always involve one reflection event and one transmission event. As it happened, the beamsplitter used in the experiment had a splitting ratio of T57-R43, a significant departure from 50-50. Surprisingly, we could not find a vendor who would
guarantee a non-polarizing beamsplitter to have a 50-50 splitting ratio to better than ±5%. This being the case, it was found more convenient to use the triplet projection and ignore the singlet state projection. The limited visibility was accounted for by reducing the size of the coherence in the projection operator which was written as

\[
\begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & RT & vRT & 0 \\
0 & vRT & RT & 0 \\
0 & 0 & 0 & 0
\end{pmatrix},
\]

(4.27)

where \( v \) is the two-photon interference visibility, and \( R \) and \( T \) are the reflection and transmission coefficients. Numerical simulations showed that the 93% visibility affected the results by less than the statistical error in the data. As with the separable measurements, 3000 0.2-second measurement intervals were taken for each set of waveplate angles needed to rotate the operator in 4.27 to the entangled projectors listed in Table 4.2.

Since \( R, T < 0.5 \), operator 4.27 will generally have a trace less than 0.5 and so fails to produce an outcome at least 50% of the time, even when the input is \( |\psi^+\rangle \). As the splitting ratio becomes less equal, the efficiency of the projection becomes even worse. From an experimental implementation viewpoint this is not a problem as it only necessitates counting for a longer interval to obtain the same number of coincidences. From a conceptual point of view, though, a tomography scheme that fails to work half of the time is problematic.

Recently, some clever schemes for using hyper-entangled states and multiple photon degrees of freedom has allowed for a complete Bell-state measurement to be implemented in linear optics[104]. A MUBs tomography scheme that made use of this Bell-state measurement technique would not suffer the same failure rate as the scheme presented here. It has also been suggested[60] that MUBs tomography would be of great use in trapped ion systems, where strong inter-particle interactions make deterministic Bell-state measurements relatively easy operations to perform. The present experiment may
be considered a proof-of-principle demonstration of the sort of advantage that one might obtain in systems capable of making efficient Bell-state measurements.

4.6.3 Analysis

The 3000 measurements taken in each basis were randomly combined together to form datasets with different total numbers of counts. To each dataset was then applied the maximum-likelihood fitting method to determine the density matrix most likely to have produced the dataset. Each such density matrix was compared to the density matrix most likely to have produced the entire dataset. This was repeated 30 times for each total number of counts and the fidelities averaged over these 30 repetitions. The data were then plotted on a log-log plot of infidelity versus number of copies of the state.

The errors in the regime of interest are dominated by statistical errors. These were calculated by taking the standard deviation of the infidelity over the set of 30 repetitions for each total number of copies of the state.

4.6.4 Results

Figure 4.2 through 4.4 show the results of this analysis. The experiment was repeated for three different quantum states, namely the separable states $|HV\rangle$, the maximally-mixed state and the state

$$
\rho = \begin{pmatrix}
0.5 & 0 & 0 & 0.43 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0.43 & 0 & 0 & 0.5
\end{pmatrix},
$$

(4.28)

which was the closest approximation we could make to a maximally-entangled state given the limitations of our system\textsuperscript{4}.

\textsuperscript{4}See Chapter 2 for an explanation of these limitations
Chapter 4. QST with MUBs

Total number of copies of the state

Figure 4.2: Measured infidelity for $|HV\rangle$
Figure 4.3: Measured infidelity for the maximally-mixed state
Figure 4.4: Measured infidelity for the partially mixed entangled state close to $\frac{1}{\sqrt{2}} (|HH\rangle + |VV\rangle)$
The results are shown in figures 4.2 through 4.4. The MUBs tomography produced a lower infidelity than separable tomography for the maximally-mixed state and the entangled state, while a similar level of infidelity was observed for the separable state $|HV\rangle$. The constant shift on the log-log plot corresponds to a fixed ratio of infidelity between the two estimation methods. For the entangled state this ratio was $1.84 \pm 0.06$. For $|HV\rangle$ it was $1.09 \pm 0.04$. For the maximally-mixed state it was $1.49 \pm 0.05$, in reasonable agreement analytical results of section 4.4.

These results are consistent with Monte Carlo simulations and with the arguments of the preceding sections. Because standard separable state tomography produces better estimates of single-qubit polarizations than correlations, we expect that MUBs tomography will produce a larger improvement for entangled states than for separable states. Indeed, this is what we observe as can be seen by comparing the widely separated lines in figure 4.4 to the essentially overlapping lines of figure 4.2. While MUBs tomography indeed shows a significant advantages for entangled states, it is perhaps more notable that it is no worse than SSQST even for the states for which SSQST performs best. This is evidence that using MUBs can result in significant improvements over the full range of states.

### 4.7 The discrete Wigner function for two qubits

Another appealing feature of MUBs is their close relationship to the discrete Wigner function[75, 47]. While we will not delve too deeply into this rich subject here, we will briefly show how a discrete Wigner function can be obtained from the data collected in MUBs tomography. The analysis follows the framework laid out in reference [47].
4.7.1 Finite fields

The theory of MUBs and discrete Wigner functions is closely related to the theory of finite fields in number theory. We can define a four-element field $\mathbb{F}_4$ consisting of the symbols $\{0, 1, \omega, \bar{\omega}\}$ along with addition and multiplication operations defined by tables 4.3 and 4.4. With the operations of multiplication and addition clearly defined we can define lines on the discrete phase space consisting of a grid of points labeled by an (x,y) pair of elements in $\mathbb{F}_4$. Lines are equations of the form $y = mx + b$ where $m$ and $b$ are also elements of $\mathbb{F}_4$. Of particular importance will be families of parallel lines in the phase space, that is to say lines with the same slope $m$. Since the phase space has 16 discrete points, each such family can contain 4 parallel lines. Each family will have a particular line called a ray that intersects the point (0,0). Each of the $4^2 - 1$ points not at the origin defines a ray, but each ray contains three points other than the origin, so the number of distinct rays is $(4^2 - 1) / 3 = 5$. Consequently, the number of families of parallel lines is also 5. It is left as an exercise for the reader to check that the lines denoted by the blue-circled points in figure 4.5 form lines in this phase space and that each horizontal box contains a family of parallel lines.

Table 4.3: Addition table for $\mathbb{F}_4$. Reproduced from reference [47].

<table>
<thead>
<tr>
<th>+</th>
<th>0</th>
<th>1</th>
<th>$\omega$</th>
<th>$\bar{\omega}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>$\omega$</td>
<td>$\bar{\omega}$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>$\bar{\omega}$</td>
<td>$\omega$</td>
</tr>
<tr>
<td>$\omega$</td>
<td>$\omega$</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$\bar{\omega}$</td>
<td>$\bar{\omega}$</td>
<td>$\omega$</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\times$</th>
<th>0</th>
<th>1</th>
<th>$\omega$</th>
<th>$\bar{\omega}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>$\omega$</td>
<td>$\bar{\omega}$</td>
</tr>
<tr>
<td>$\omega$</td>
<td>$\omega$</td>
<td>0</td>
<td>$\bar{\omega}$</td>
<td>1</td>
</tr>
<tr>
<td>$\bar{\omega}$</td>
<td>$\bar{\omega}$</td>
<td>0</td>
<td>1</td>
<td>$\omega$</td>
</tr>
</tbody>
</table>

Table 4.4: Multiplication table for $\mathbb{F}_4$. Reproduced from reference [47].
Figure 4.5: The five sets of four parallel lines in the $\mathbb{F}_4 \times \mathbb{F}_4$ phase space. Reproduced from reference [47].
4.7.2 Discrete Wigner functions

For each family of parallel lines in figure 4.5 we will associate a MUB from table 4.2, and to each particular line $\lambda$ we assign a projector $Q(\lambda)$, in the same order as the lines and projectors appear in the figure 4.5 and table 4.2. Next, for each point $(x, y)$ in the phase space we associate a \textit{point operator} defined as the sum of all the projectors associated with lines that contain that point

$$A(x,y) = \sum_{\lambda \ni (x,y)} Q(\lambda) - I_4. \quad (4.29)$$

Then we can define the discrete Wigner function of the state to be

$$W(x,y) = \frac{1}{N} \text{Tr} (\rho A(x,y)), \quad (4.30)$$

so that

$$\rho = \sum_{(x,y)} W(x,y) A(x,y). \quad (4.31)$$

Defined in this way, the Wigner function has many useful properties. First, it is complete as equation 4.31 attests. Any density matrix $\rho$ has a corresponding Wigner function. Second, if the Wigner function is summed along a series of parallel lines in the phase space, one obtains a marginal probability distribution, namely the set of expectation values for measurements done in the corresponding MUB. In this way the discrete Wigner function resembles the ordinary continuous Wigner function\cite{125} on $(x,p)$ phase space where a projection of the function along any axis generates a strictly positive marginal probability distribution. Like the continuous Wigner function, the discrete Wigner function is real, but may be negative at some points. Despite this, summing the discrete Wigner function elements along any line in the discrete phase space results in a positive probability.
Table 4.5: Table of measured expectation values from MUBs tomography on an entangled state

<table>
<thead>
<tr>
<th>Projectors</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>HH\rangle,</td>
</tr>
<tr>
<td>$</td>
<td>DR\rangle,</td>
</tr>
<tr>
<td>$\frac{1}{\sqrt{2}} (</td>
<td>RH\rangle + i</td>
</tr>
<tr>
<td>$\frac{1}{\sqrt{2}} (</td>
<td>RH\rangle + i</td>
</tr>
</tbody>
</table>

4.7.3 Reconstruction

We will reconstruct the discrete Wigner function for an entangled state produced in our type-I setup. From experimental data taken during MUBs tomography we can construct the sets of frequencies shown in table 4.5 for each of the 5 MUBs. Each point in the Wigner function is the sum of the probabilities for the projectors associated with the lines going through the point minus 1. As a result we can directly calculate the Wigner function from the data in table 4.5. The resulting Wigner function is:

$$W(x, y) = \bar{\omega} 0.1913 -0.0803 -0.0001 0.1079$$
$$\omega 0.2160 -0.0790 0.0135 0.1239$$

$$W(x, y) = 1 0.0579 0.1286 0.0030 0.1193$$
$$0 0.0602 0.0420 -0.0106 0.1064$$

It will be immediately apparent to the reader that summing along the columns and rows generates the expectation values of the first two rows of table 4.5. The other sets of parallel lines are not so evident, but the reader can verify that summing along them will generate the outcomes of rows three through five.
To my knowledge this is the first time that anyone has reconstructed a discrete Wigner function of a two-qubit state from experimental data. It represents an intriguing way of describing the quantum state of a two-qubit system. While the density matrix contains direct probability information about one basis along its diagonal, the Wigner function contains probability functions for five different bases simultaneously in the geometry of the associated lines in finite phase space. Just as the density matrix can be unitarily transformed to put another basis along the diagonal, the Wigner function can be rotated unitarily so as to change the five bases whose expectation values are directly described, so long as the five bases are mutually-unbiased. This makes it an appealingly compact representation of quantum information. As a bonus, the Wigner function is real, rather than complex-valued like the density matrix.

Apart from its practical usefulness, the discrete Wigner function creates a remarkable link between the theory of quantum state estimation and abstract number theory. The very compactness of the description highlights the limitations on the strength of quantum correlations, since a set of changes in the Wigner function to obtain a particular set of outcomes in one basis will inevitably affect the probability distributions in other bases. It cannot be coincidental that this set of tradeoffs is perfectly and succinctly encapsulated in the geometry of finite fields. Understanding this link remains an area of active research.

In short, the discrete Wigner function represents a novel and illuminating description of the quantum state which may some day play an important role in the description of experimental quantum systems and may point to a deeper underlying geometry of quantum measurements.

4.8 Summary

We have, for the first time, demonstrated an optimal PVM quantum state tomography protocol. Optimality was achieved by making use of mutually-unbiased bases, a set of
PVMs that have minimal RMS overlap and hence require the least amount of redundant information to be collected to characterize the quantum state. In our experimental implementation of the protocol we observed a significant improvement over previous methods, especially for entangled states.

MUBs tomography also allows very simple construction of discrete Wigner functions, an alternative to the density matrix as a way of describing quantum states. We obtained the discrete Wigner function of a state directly from measurements and showed that it represents an extremely compact and intuitive description that may help us to understand the geometric structure underpinning quantum correlations.
Chapter 5

Measuring density matrix figures of merit without measuring the density matrix

How little is purity known in the world. How little we value it. What little care we take to preserve it; what little zeal we have in asking for it, since we cannot have it of ourselves.

–St. John Vianney

5.1 Introduction

While quantum state tomography has proven to be a very useful tool in experimentally characterizing quantum systems, it has a very serious limitation in that its complexity grows exponentially in the number of degrees of freedom in the system. In a qubit system of \( N \) qubits, a minimum of \( 4^N \) measurements need to be done to in order to measure the density matrix. For ten qubits this is just over a million measurements. Similarly the difficulty of reconstructing the density matrix from the measurements grows exponen-
tially in the number of qubits, resulting in the need for greater classical computational resources. This exponential complexity scaling in the measurement and calculation of density matrices was most pronounced in the tomographic characterizations of eight qubit states of trapped ions[50]. There, twenty-four hours of measurement were required to characterize the state, followed by several weeks of maximum-likelihood fitting on a computer cluster. The same experimental apparatus was capable of making states of up to 12-qubits, but the difficulty in performing quantum state tomography made the group decide to limit the size of the state to eight qubits.

This poor scaling of quantum state tomography with system size does not bode well for its potential as a tool in characterizing quantum computers, which, in order to be useful for tasks like factoring, will require at least thousands and possibly millions of qubits. Nor is it a particularly surprising characteristic of quantum states. To characterize a classical probability distributions we also need to make a number of measurements that scales with the number of possible states. The number of possible states in turn scales exponentially in the number of bits. Despite this, very complex classical systems are made and characterized, not by trying to map out the complete probability distribution of all possible states of the system, but by defining figures of merit by which the system can be evaluated. For example, an engineered system might be characterized by its failure rate, or a physical system by thermodynamic variables like temperature and entropy. In order to extend our current success in quantum state engineering to devices that can hope to be useful, a similar approach must be taken with quantum states.

Indeed, there are some figures of merit that are already commonplace in the quantum information world. Some of them, like the tangle, concurrence and negativity measures arose because of the desire to develop locally invariant descriptions for entanglement. Others, notably the purity and the von Neumann entropy, try to quantify the width of the density matrix viewed as a probability distribution over states.

These figures of merit used to describe states have typically been measured by first
doing quantum state tomography to determine the state and then calculating the figure of merit from the density matrix. This approach is necessary because generally the figures of merit of interest are non-linear functionals\(^1\) of the density matrix. This means that they cannot be expressed as an expectation value of an operator acting on the space of quantum states because any such operator \(A\) will necessarily be linear in the density matrix: \(\langle A \rangle = \text{Tr} [\rho A]\). Since they cannot be measured directly, the only recourse has been to completely characterize the state in terms of the \(4^N\) density matrix elements and then calculate the figure of merit from these. Clearly this approach will suffer from the same scaling problem as quantum state tomography itself, a resource requirement that scales exponentially in the size of the state.

In this chapter we present the implementation of a solution to this problem proposed by Todd Brun\(^2\) that will work whenever the figure of merit is polynomial in the density matrix. This is an important class of figures of merit which includes the purity and other unitary invariants such as Kempe’s invariant and the 3-tangle\(^{32, 58}\), as well as the Q-measure of Meyer and Wallach\(^{85}\). Other quantities that are analytic in the density matrix such as the von Neumann entropy can be approximated by a truncated Taylor series, while the concurrence\(^{126}\), entanglement of formation\(^{12}\), entanglement of distillation\(^{14}\), and the negativity of the partial transpose\(^{91}\) can be approximately restated as simple function of such truncated Taylor series.

In particular we will apply Brun’s technique to measure the purity of a one-qubit density matrix as a ‘single-shot’ measurement, without measuring the density matrix itself. In so doing, we will investigate different techniques for making impure polarization states and find that they are not all the same and do not all result in the same measured purity using this technique, even when they are described by the same density matrix. This difference arises because of the effect of the ‘environment’ whose entanglement with

\(^1\)A functional is a function defined on the space of operators. Given an operator a functional returns a number.
the system gives rise to the impurity.

Related experimental work on directly measuring the purity has been done by Du et al[39] in NMR systems and by Bovino et al[23] in a four-photon system. Work on measuring another figure of merit, the concurrence[126], was published[119] while this experiment was in progress.

5.2 Multiqubit measurements

Brun’s method relies on the insight that \( m \)-th-order polynomial functionals of the density matrix can be thought of as linear functionals of a density matrix consisting of \( m \) or more copies of the state.

If the state of a quantum system is described by a density matrix \( \rho \),

\[
\rho = \sum_{i,j=0}^{d-1} \rho_{ij} |i\rangle \langle j|,
\]

then Brun defines a polynomial functional of degree \( m \) as a function

\[
f(\rho) = \sum_{i_1,j_1,...,i_m,j_m} c_{i_1j_1...i_mj_m} \rho_{i_1j_1} \rho_{i_2j_2} \cdots \rho_{i_mj_m},
\]

where the \( c_{i_1j_1...i_mj_m} \) are arbitrary complex constants.

Now suppose that we have available \( m \) copies of a system in state \( \rho \). The density matrix for the joint state of the system will simply be the tensor product of \( \rho \) with itself \( m \) times.

\[
\rho^{\otimes m} = \sum_{i_1,j_1,...,i_m,j_m} \rho_{i_1j_1} \cdots \rho_{i_mj_m} |i_1\rangle \langle j_1| \otimes \cdots \otimes |i_m\rangle \langle j_m|.
\]

Now each term in \( f(\rho) \) will correspond to a term of 5.3. If we write an operator

\[
\hat{A}_{i_1j_1...i_mj_m} |i_1\rangle \langle j_1| \otimes \cdots \otimes |i_m\rangle \langle j_m|,
\]

then if we calculate the expectation value of \( \hat{A} \) for the state \( \rho^{\otimes m} \) we obtain:

\[
\text{Tr}\left\{ \hat{A}_{i_1j_1...i_mj_m} \rho^{\otimes m} \right\} = \rho_{i_1j_1} \cdots \rho_{i_mj_m},
\]
which is one of the terms making up $f(\rho)$. Since $f(\rho)$ is linear in such terms, we can always construct an operator $\hat{B}$ as a linear combination of operators $\hat{A}$ such that $f(\rho)$ is the expectation value of $\hat{B}^2$.

We can apply this technique to a particular figure of merit of interest, the purity, defined as

$$P = \text{Tr} \{ \rho^2 \}. \quad (5.6)$$

This figure of merit provides a measure of the width of the distribution $\rho$ over the pure quantum states. It is unitarily invariant and has proven a useful measure of entanglement[91], classicality[46] and even effective temperature[5]. In a $D$-dimensional Hilbert space it is equal to one for a pure state and to $1/D$ for a maximally mixed state. States that are neither pure, nor maximally mixed have intermediate values of the purity.

Treating the purity as a function of the density matrix elements, if we have a single-qubit density matrix

$$\rho = \begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix}, \quad (5.7)$$

then the purity is

$$P = \rho_{00}^2 + \rho_{01}\rho_{10} + \rho_{10}\rho_{01} + \rho_{11}^2. \quad (5.8)$$

Following Brun’s procedure we can define a corresponding two-qubit operator ($m = 2$ here) $\hat{A}$ whose expectation value will be the purity:

$$\hat{A} = |0\rangle \langle 0| \otimes |0\rangle \langle 0| + |0\rangle \langle 0| \otimes |1\rangle \langle 1| + |1\rangle \langle 0| \otimes |0\rangle \langle 1| + |1\rangle \langle 0| \otimes |1\rangle \langle 1|. \quad (5.9)$$

On the face of it this operator looks like a rather complicated one to measure. We can simplify it by using the identity operator $I_4$

$$I_4 = |0\rangle \langle 0| \otimes |0\rangle \langle 0| + |0\rangle \langle 0| \otimes |1\rangle \langle 1| + |0\rangle \langle 0| \otimes |1\rangle \langle 0| + |1\rangle \langle 1| \otimes |1\rangle \langle 1|. \quad (5.10)$$

---

2You may be wondering about terms in $f(\rho)$ of order less than $m$. These can be accounted for by rewriting them as sums of terms of order $m$ and using the property that $\sum_i \rho_{ii} = 1$ to eliminate the unwanted terms.
so that

\[
\hat{A} = \mathbb{I}_4 - |01\rangle \langle 01| - |10\rangle \langle 10| + |01\rangle \langle 10| + |10\rangle \langle 01|
\]

\[
= \mathbb{I}_4 - 2|\psi^-\rangle \langle \psi^-|,
\]

where we have introduced the two-qubit singlet-state \(|\psi^-\rangle = \frac{1}{\sqrt{2}} (|01\rangle - |10\rangle)\). Writing out the expectation value of \(\hat{A}\) we get

\[
P = \text{Tr} \left\{ (\mathbb{I}_4 - 2|\psi^-\rangle \langle \psi^-|) \rho^{\otimes 2} \right\}
\]

\[
= 1 - 2 \langle \psi^-| \rho^{\otimes 2} |\psi^-\rangle.
\]

Thus a single two-qubit expectation value measurement, namely the expectation value of the singlet-state projector, is sufficient to measure the purity. In contrast it would have taken four linearly independent expectation values to perform a full characterization of the single qubit state. This four-fold reduction is perhaps not especially impressive, especially as the single-particle measurements required for tomography will generally be easier to obtain that the two-particle joint measurement. The real advantage of Brun’s method, though, is that it completely changes the scaling. While for tomography the number of measurements scales exponentially in the system size, for Brun’s method the complexity is set by the degree of the polynomial in the density matrix functional. A million-qubit quantum computer which would be inconceivable to characterize using tomography could, in principle have its purity measured by Brun’s technique by running two copies of the computer through a calculation and then having a joint measurement applied to the two copies at the end. This is a truly scalable approach to state characterization, and one that will become increasingly important as quantum machines grow beyond a few qubits.

That the purity in the single-qubit case is so closely related to the singlet-state projection is not so surprising. As has been discussed in previous chapters, the singlet state measurement is an anti-symmetry measurement in every basis. It assumes the form
\[ \frac{1}{\sqrt{2}} \left( |\psi\tilde{\psi}\rangle - |\tilde{\psi}\psi\rangle \right) \] in all bases, where \( |\tilde{\psi}\rangle \) is a single-qubit state orthogonal to the single-qubit state \( |\psi\rangle \). This means that a state of the form \( |\psi\rangle |\psi\rangle \) will have a singlet projection of zero. Thus when \( \rho \) represents a pure state we recover from equation 5.12 the correct value of 1 for the purity.

When \( \rho \) is the maximally mixed state, \( \frac{1}{2}I_2 \), the purity is 1/2. In this case \( \rho^{\otimes 2} = \frac{1}{4}I_4 \). Since the singlet-state is one of the four Bell states and the maximally mixed state is rotationally invariant (meaning it looks the same in the Bell basis as in any other basis), its singlet state projection is 1/4 and equation 5.12 gives the correct value of 1/2 for the purity.

One way to think of this is to imagine the density matrix as literally describing a box of qubits prepared in one of two orthogonal states with equal probability. Two qubits at a time are selected from the box and the singlet state projection of the joint state is measured. There is a 50\% chance that the two selected photons will be in the same state, in which case the singlet state projection is zero. There is also a 50\% chance that the two photons will be in orthogonal states in which case the singlet state projection is 1/2. On average, then, the singlet state projection is \( 0.5 \times 0 + 0.5 \times 1/2 = 1/4 \). A similar reasoning can be applied to states that are mixed but not maximally mixed, by imagining a box filled with states orthogonal in the basis in which the density matrix \( \rho \) is diagonal. If one state has probability \( p \) and the other probability \( 1 - p \) then the probability of selecting two of the same state is \( p^2 + (1 - p)^2 \) while the probability of selecting two orthogonal states is \( 2p(1 - p) \). The former case results in zero singlet-state projection and the latter a projection of 0.5. It follows from formula 5.12 that the purity is \( P = 1 - 2p(1 - p) = p^2 + (1 - p)^2 \) which is what you would get by squaring and taking the trace of a diagonal density matrix with entries \( p \) and \( 1 - p \) along the diagonals.

The problem of measuring the purity without measuring the density matrix therefore reduces to the problem of measuring the singlet state projection, which, as we’ve seen in previous chapters, can be implemented using two-photon interference[53].
Figure 5.1: Experimental setup for (a) the direct purity measurement and (b) quantum state tomography. Labels designate a 50/50 beamsplitter (BS), a non-linear $\beta$-Barium Borate (BBO) crystal, half-waveplates (HWP), liquid-crystal variable waveplates (LCWP), single photon counting modules (SPCM) and a polarizing beamsplitter (PBS). A type-I spontaneous parametric downconversion (SPDC) crystal produces pairs of H-polarized photons. In (a), the same state $\rho$ is prepared in both arms and the beamsplitter acts as a singlet state filter. In (b), a state is prepared in one of the arms and polarimetry is used to measure the density matrix of the state using a quarter-waveplate, a half-waveplate and a polarizing beamsplitter.

5.3 Experimental techniques

5.3.1 Singlet state projection

As in the mutually-unbiased basis tomography experiment, we perform joint measurements by using two-photon interference. Since in order to implement the purity measurement we only need to be able to do a singlet state projection, there is no need to rotate the measurement to other bases or even to use a polarizer (in fact, using a polarizer would break the unitary invariance of the system and we’d no longer be measuring purity!). We simply set up the beamsplitter as shown in Figure 5.1(a), prepare the two input photons to have the same density matrix, and measure the rate of coincidence detections.

Since the visibility of the two-photon interference is not 100%, we need a way to relate
the purity not just to the singlet state projection, but to the measured projection with the limited visibility. We will make the useful assumption that the factors that limit the visibility such as imperfect mode-matching, imperfect splitting ratio and so on, are not themselves polarization-sensitive. Under this assumption the measured projector can be written as a singlet state projector plus a term proportional to identity that carries no polarization information but adds a background to the measurement. If the maximum visibility possible given these limiting factors were 90%, then the projector

$$\hat{P}_{\text{actual}} = 0.10\mathbb{I}_4 + 0.80|\psi^-\rangle \langle \psi^-|$$

satisfies the necessary conditions that it goes to 0.90 when the state is pure singlet, to 0.10 when the state is symmetric and to 0.50 when the state contains equal symmetric and antisymmetric components such as for $|HV\rangle$. Physically we can interpret the first term as representative of the fact that with imperfect visibility all states have some likelihood of producing a coincidence.

Using this formula, the purity can be written as

$$P = 1 - 2 \frac{\langle \hat{P}_{\text{actual}} \rangle - 0.1}{0.80}. \quad (5.14)$$

This adaptation of the problem to the real world of imperfect visibilities introduces a new issue. It is now possible that for a finite number of detector clicks, the value of the measured purity will be greater than 1 due to statistical noise in the estimation of the expectation value. This is not as bad as it seems. The whole analysis has assumed that we can measure expectation values whereas in reality we can only ever estimate them. While it is true that if the two-photon interference visibility were perfect then it would be impossible to observe a purity greater than one, this is more of a fluke than a qualitative difference. It arises because a pure state has a singlet state projection of zero and the variance in a Poisson distribution with a mean of zero is zero. Even with perfect singlet state projections it is still possible to get unphysical purity measurements
at the lower end of the purity scale. The minimum purity for a qubit is 1/2 which will occur when the state is maximally mixed, but then the singlet state projection is 1/4 and so there will be statistical noise on it, so it is certainly possible for formula 5.12 to spit out unphysical values of the purity if the singlet state bracket refers to estimates of the expectation value rather than the true expectation value.

One cannot get unphysical values of the purity from a density matrix, but in that case physicality has been added in as an assumption by requiring that the density matrix be positive semi-definite. One could do the same with these direct approaches to measuring purity by applying maximum-likelihood fitting. Such an approach would rephrase the problem as what value of purity on the interval [1/2, 1] was most likely to have given the observed measurement. This was not done in the experiment, but in hindsight that probably would have been a good idea.

Figure 5.1b shows how we measure the single-photon polarization density matrix for one of the photons. It is a textbook polarization analyzer, with a QWP, HWP and polarizer in front of the detector. Projections onto \{\ket{H}, \ket{V}, \ket{D}, \ket{A}, \ket{R}, \ket{L}\} were measured and the single-qubit density matrix was found by maximum likelihood fitting. In the experiment we measured the single-qubit expectation values from the the singles rate at the detector, not the coincidence rate. We measured the total intensity of light reaching the detector, then turned off the pump laser and measured the background due to residual room lights and detector dark counts. The difference between these two detection rates was used in the tomographic reconstruction. A better approach would have been to use a scissor jack to remove the beamsplitter from the system and then rely on coincidence detection with an analyzer on the side to be characterized and no polarizer on the other side. This approach would have been background-free, and much more elegant. Unfortunately it didn’t occur to me at the time. Luckily, the density matrix is made up of expectation values, not of higher moments in the statistical distribution of counts, so as long as the background rates were stable (which they were over an hour
or so), the density matrices produced by measuring singles were indistinguishable from those taken using coincidences.

### 5.3.2 Making impure states

Unique among candidate quantum information systems, photon polarization states, by their nature, are angelically pure. As Andrew White is fond of saying, the fact that a residual polarization can be measured in the cosmic microwave background indicates that photon coherence times are on the order of $10^{17}$ s. In non-photonic systems impurity arises due to interactions with the ‘environment’, the other quantum systems to which the system under study is coupled. For photons, propagation in free space induces no decoherence. Propagation through linear optical materials induce unitary, rather than non-unitary evolution. In an experiment like this, where we want to study photon impurity, what options do we have to obtain it?

A quantum-information approach is to use quantum randomness to make an impure state. One could, for example, create a Bell-state of two photons and then send one of them into an absorbing medium. The photons start out in a maximally entangled state, and the absorption of one photon creates a maximally entangled state of the other photon and the electron spins in the absorbing medium. These quickly couple to everything else in the medium and we have a situation that looks very much like environment-induced decoherence in that we couldn’t possibly retrieve the correlation information if we wanted to, so we can treat the first photon as being impure. This approach has been used in a few experiments[23], but it requires having at least as many photon pairs as you need impure photons at the end. For the Brun scheme we would need two photon pairs, and we were not set up to create them at the time of the experiment.

Another technique, used, for instance, by Wei et al.[121] is to use the photons’ other degrees of freedom as the ‘environment’ and then simply not detect those other degrees of freedom. So, for example, if the photon is sent through a birefringent medium whose
birefringent delay is greater than its coherence time, then after propagating through the medium, the extraordinary and ordinary polarization components of the photon will have essentially random relative phases. If an equal superposition of extraordinary and ordinary polarizations are sent in (say $|D\rangle$ with $H$ and $V$ as the ordinary and extraordinary crystal axis directions), what emerges will be an entangled state of delay and polarization \( \frac{1}{\sqrt{2}} (|H, E\rangle + |V, L\rangle) \) where $E$ and $L$ are early and late time bins with a separation larger than the coherence time. If the photon detectors are insensitive to the time delay - and they will be, at least in usual SPDC experiments, then they effectively trace over the time information and we are left with a mixture of $|H\rangle$ and $|V\rangle$. One can also think about this decoherence process in the frequency domain where the relative group delay between $|H\rangle$ and $|V\rangle$ can be thought of as a phase shift that depends linearly on frequency. If the detector can collect a wide range of frequencies then it collects photons with a wide range of birefringent phase shifts for the different frequencies in the light. This will appear the same as if the photons had been given random phase shifts.

A third way of generating impurity is to directly apply random unitaries to the photons. For example, one could split the photon amplitudes for $H$ and $V$ at a PBS and put a heat source in one arm to cause random changes in the index of refraction of air before recombining them at a second PBS[87]. Or one could, as we did in this experiment, hook a liquid crystal waveplate up to a pseudo-random number generator and produce well-defined phase shifts that depend on the random number generated. On the face of it, these sources of impurity seem somewhat contrived as compared to the first two, but they are not fundamentally any different. If impurity is generated by the loss of a single photon from a Bell pair, then the information about its polarization is still contained somewhere. If it gets absorbed then it transfers its polarization state to the absorbing material and, in principle, one could find the electron whose spin was flipped and use the Bell state correlation to determine the polarization of the first photon. Sim-
ilarly, a sufficiently fast or spectrally sensitive detector could, in principle, resolve the frequency-dependent phase shift induced by a birefringent group delay. As far as Nature is concerned in all three cases impurity is created because we choose not to look at information that is, in principle, available to be measured.

Of these three possible methods of creating impure photon states we tried applying Brun’s method to the latter two. For the birefringent group-delay-induced impurity we used AR-coated pieces of quartz of varying thickness. The light was prepared in $|D\rangle$ and the quartz was aligned to have its optical axis vertical or horizontal. After the quartz, the light was depolarized to a varying degree that depended on quartz thickness. The thickest piece was 25 mm long and was enough to induce a group delay of 237 $\mu$m between the two polarizations which is far more than the coherence length of 49 $\mu$m determined by the 10 nm bandwidth of the interference filters used. It was therefore possible to produced maximally mixed states with these quartz pieces.

For the impurity created by the liquid-crystal waveplates, two different methods were used. In the first method the waveplates were calibrated so that by applying known voltages, known phase shifts of 0, $\pi$, and $\pi/2$ radians could be obtained between the $H$ and $V$ polarizations. A software pseudo-random number generator selected between these possibilities and the phase shift was applied. In the second method, the function of phase shift versus voltage was mapped out for phase shifts between 0 and $\pi$ and used to create a probability distribution that produced uniformly distributed phase shifts on this interval. A pseudo-random number generator selected voltages from this interval at random.
Chapter 5. Measuring figures of merit

Figure 5.2: Experimental density matrices for various single-photon polarization states.
(a) pure horizontal $|H\rangle$ (b) pure diagonal $|+\rangle$ (c) equal mixture of $|+\rangle$ and $|-\rangle$ (d) equal mixture of $|+\rangle$, $|-\rangle$ and $|R\rangle = (|H\rangle - i|V\rangle)/\sqrt{2}$ (e) A mixture of states of the form $(|H\rangle + e^{i\phi}|V\rangle)/\sqrt{2}$ with $\phi$ distributed equally over $[0, \pi]

5.4 Results

5.4.1 Density matrices

The density matrices for several polarization states measured using the apparatus in figure 5.1(b) are shown in figure 5.2. The pure states $|H\rangle$ and $|D\rangle$ have measured purities of $0.99 \pm 0.01$, and the various mixed states, which were all generated using the LCWP and pseudo-random number generator, have purities that match to within $0.02$ the expectations based on the distribution of states that make them up. For the random state selection the uncertainty had a contribution from the statistics of the state selection process as well as the photon counting statistics. The states examined were $|H\rangle$, $|D\rangle$ obtained by rotating $|H\rangle$ with a half-waveplate, an equal mixture of $|D\rangle$ and $|A\rangle$ obtained by randomly applying either 0 or $\pi$ birefringent phase shifts to $|D\rangle$, a partially mixed state obtained by applying a phase shift of 0, $\pi/2$ or $\pi$ to $|D\rangle$, and finally, the state obtained when random phase shifts selected uniformly over the interval 0 to $\pi$ were applied to $|D\rangle$. 
Table 5.1: The purities measured for five states using the direct joint purity measurement and a full characterization followed by a calculation of $\text{Tr}\{\rho^2\}$. The stated errors arise from counting statistics and the statistics associated with the random state selection.

The purities measured indirectly using quantum state tomography, are compared to the direct purity measurement in table 5.1. The theoretical purities can be calculated by summing density matrices over the probability distribution of having different phases. The calculation is left as an exercise for the reader.

It can be seen that the purities as measured by the two methods are the same to within error, indicating that Brun’s method does indeed work on these states. What is perhaps more surprising are the cases where it does not work. The rest of this section will be devoted to examining those cases.

The results in table 5.1 are for impure states generated with liquid crystal waveplates. What happens when the impurity is instead generated by applying a birefringent group delay so as to entangle the photons’ time degrees of freedom with their polarization? First, if we are going to create impurity in this way then we have to decide whether to apply a shift in the same direction to the two photons (say delaying $H$ relative to
V for both), or in opposite directions. Both actions will give the same single-qubit density matrix for the individual photons, but could potentially affect the two-photon interference, perhaps in such a way that it can no longer be properly interpreted as the purity. In fact, doing it either way affects the two-photon interference profoundly and makes it impossible to interpret as a polarization singlet-state measurement! If the two crystals are aligned in the same direction, then the $H$ amplitude is delayed for both the photons. The $V$ components of the state always arrive at the beamsplitter first and the $H$ components arrive last. Crucially, the $H$ components arrive together and the $V$ components arrive together. The $H$ components will interfere perfectly and the $V$ components will interfere perfectly and one will record a maximal drop in the coincidence detection rate at the detectors, exactly as if the single-photon states were pure! Things get even worse if the crystal axes have opposite alignment. In that case, the $H$ photon will be delayed on one side and the $V$ photon will be delayed on the other. The undelayed $H$ photon will always arrive at the same time as the undelayed $V$ photon and the delayed $H$ photon will arrive at the same time as the delayed $V$ photon. In either case the two photons are completely distinguishable in polarization and no two-photon interference occurs. The singlet state projection in this case is 0.5 (since the singlet state is 50% $|HV\rangle$ and 50% $|VH\rangle$). By formula 5.12 this would constitute a purity of 0 which is impossible.

To get a sense of what is going on we can put a delay in one of the arms and plot the coincidence rate as a function of that delay. In practice we need to do this anyway to figure out where the two-photon interference is. Figure 5.3(a) shows what happens when the delays go in the same direction. As one photon wavepacket slides past the other in time there will be a moment when the early $H$ packet overlaps with the late $V$ packet, but this does not cause interference because these two states are orthogonal. As the packets keep sliding past each other, a point is reached when the two $H$ packets overlap and the two $V$ packets overlap and the maximal 90% visibility is achieved. Finally, on the way
Figure 5.3: (a) The two-photon interference dip when the crystals introduce the same delay at the two beamsplitter inputs (b) The two-photon interference dip when the two crystals introduce opposite delays at the two beamsplitter inputs.

out, the late $V$ packet will overlap the early $H$ packet, but again there is no interference. Contrast that with the situation when the two delays are in opposite directions. Then as the arm is scanned a point is reached when the early $H$ amplitude overlaps with the late $H$ amplitude from the other photon. These two amplitudes interfere perfectly, but each photon is only $H$ 50% of the time. The joint probability for both photons to be $H$ is $1/4$, and the other $3/4$ of the time there is no interference. A similar thing happens when the late $V$ amplitude overlaps the early $V$ amplitude. Figure 5.3(b) shows this situation which results in two distinct dips, each with a 25% visibility.

So why does two-photon interference no longer measure the impurity properly when the impurity is created by entanglement between different photon degrees of freedom? The answer is that two-photon interference is not just a polarization interference effect, but rather an interference effect that depends on the other photon degrees of freedom, in this case the timing information. While the detectors are not sensitive to the timing information, the interference is, and this means that the strength of the interference can no longer be used as a polarization-singlet state filter. Interestingly, if we were to somehow randomize the action of the quartz crystals by sometimes advancing the $H$
polarization and sometimes the $V$, then we’d end up with an equal probability of having the two sides have the same delay and opposite delays. The two-photon interference visibility would be the average of the visibilities for the two relative orientations, namely 0 and 1 and so we would recover the correct result of a 50% visibility or a singlet state projection of 0.25 with a corresponding purity of $1/2$. The point is that this randomness must be stored in a degree of freedom of some system other than the photon in order for the two-photon interference to act properly as a singlet-state filter.

One can also think of this effect as being due to correlations in the birefringent phases of the two photons. Returning to the model of the box full of photons, we assumed at first that the two photons were selected at random from the box. Applying the same group delay to both inputs of the beamsplitter is the same as applying the same linearly dependent phase-shift with frequency with the same slope to the two photons. If somehow our picking of supposedly mixed photons from the box always resulted in photons with the same phase-frequency profile, then really we’d be in the case of a pure state and would expect perfect two-photon interference. Similarly if whenever we picked two photons they always had oppositely sloped phase-frequency profiles we would expect them to never interfere. When using random LCWP phase shifts we can create a similar kind of situation when the phase shift applied to the LCWP is the same for both photons, even though this phase shift is varied randomly from photon pair to photon pair. When this is done, the single-qubit density matrix is measured to be maximally mixed, but the interference is consistent with a perfectly pure state. This can be seen in the two-photon interference dip and measured density matrix in figure 5.4.

### 5.5 Discussion

We have seen that in the ‘ordinary’ case of impurity caused by an external source of randomness, the Brun approach to direct purity measurement produces the same results
Figure 5.4: Hong-Ou-Mandel interference dip taken while random phases were applied to the two LCWPs in figure 5.1a, but in a correlated way so that the phase on either side was the same at any given moment. While the density matrices measured for either photon show a maximally mixed state, this high-visibility dip would usually be indicative of a completely pure state.

as the old indirect approach of calculating purity from the density matrix. The amazing thing about this technique is that the number of measurements required to obtain the purity does not increase as a function of the size of the Hilbert space. No matter how big the system being measured, a single joint measurement on two copies of that system is sufficient to determine the purity. We could equally well imagine a box full of million-qubit quantum computers that had reached the end of some calculation. We could obtain their purity by pulling them out two at a time and performing a joint measurement on their million-qubit state. This will undoubtedly be much easier than taking the $4^{1,000,000}$
individual measurements that would be required to determine the density matrix and calculate the purity from that.

More generally, this approach to finding ways to directly obtain figures of merit on quantum systems will become an increasingly important characterization method as quantum systems become more complex. The work presented in this chapter is an important proof-of-principle in this regard.

This work also shows that states with the same benign-looking single-qubit mixed-state density matrix can exhibit very different behaviour under certain measurements. The writing down of a reduced density matrix describing a system puts up an artificial border between that system and the rest of the universe. To talk about the purity of that reduced density matrix is to assume that the randomness that makes it impure is sufficiently well-separated from the system that no future measurements will have access to it. The example presented in this chapter shows that a common means of creating impurity is fundamentally incompatible with the most common means of making two-photon joint measurements and interactions, and that a new method of creating impurity needs to be employed, one that safely stores the randomness in a completely different system. This effect, which is certainly an important one for photon polarization, may well be important in other systems where the environment is inefficient at ferrying away entropy.
Chapter 6

Conclusions

Do not seek for information of which you cannot make use.
– Anna C. Brackett

6.1 Major results

We have presented three major sets of results that have advanced the field of quantum state characterization. We have extended the techniques of quantum state tomography to systems of indistinguishable particles. In the process we have been able to quantify the amount of information available to be measured in a given system of experimentally indistinguishable particles and used it to put a bound on the degree of distinguishability in the system. We have demonstrated the technique by characterizing states of two and three photons in the laboratory, and it is to be hoped that this technique will be adopted in future research into states of indistinguishable particles.

We have examined how to improve and optimize ordinary two-qubit quantum state tomography by trying to match the tomography procedures to the natural symmetries of the Hilbert space of states. The protocol we introduce minimizes redundant information collected and is therefore able to obtain a better estimate of the quantum state from a given number of copies than any other method implemented to date. We have demon-
strated this advantage by analyzing three different quantum states and found that our method offers a real-world advantage, even given our experimental limitations. Additionally, the information collected in this new tomography procedure can be arranged to form a very natural description of the quantum state in terms of a discrete Wigner function. This function is real-valued and has a connection to classical probability distributions that is absent for the density matrix. As a result, it offers an appealing description of the quantum state that may play a significant role in quantum information research.

Finally, we have demonstrated that it is occasionally possible to do without quantum state tomography when one is interested in measuring specific property of the system, even when that property is not an observable on the Hilbert space of states. By measuring the purity of a single-photon polarization state directly, without first measuring the density matrix, we have demonstrated a technique that will prove useful when the exponential dependence of the number of density matrix elements on the number of particles in the system becomes too onerous to be practically measurable. In the process we highlight the fundamental differences of different methods of preparing mixed polarization states and point out the necessity of removing entropy from the system rather than hiding it away in another mode. The source of randomness that leads to impurity is an important consideration to keep in mind when designing linear optical quantum information systems and may play a role in other systems where particles storing information have many degrees of freedom of which only one is used.

6.2 Future outlook

With dozens of proposals for physical systems as candidates for a quantum computing architecture, the importance of demonstrating high quality entanglement and adequate state control has become essential to advancing research into any particular one. Tomography is now regarded as the gold standard for proving that a particular quantum state
has been made or a given level of entanglement achieved. As a result, interest in quantum state estimation technologies is blossoming, as can be seen by the proliferation of papers and conferences devoted to the topic. The importance of properly understanding quantum state tomography was made clear when it was pointed out that a Nature paper[113] describing a quantum dot photon source had incorrectly inferred that the source was entangled based on a misreading of tomographic data[48]. As tomography methods become part of the standard toolkit for quantum information research, the techniques of quantum state tomography, including those presented in this thesis, will become day-to-day operations in laboratories everywhere.

At the same time, quantum state estimation is becoming an important branch of research in its own right. By examining closely how information is extracted from quantum mechanical systems we can begin to probe the strange and delicate interplay between certainty and randomness that underlies quantum information. In this way, little by little, we are beginning to understand what quantum states really are, learning the limits and uses of quantum information and understanding, as Rolf Landauer first emphasized, that information is a physical quantity, as much tied to the physical world as gravity and angular momentum. We are discovering that there is a deep connection between what Nature will allow us to learn about a system through measurement and the fundamental laws of physics encoded in the postulates of quantum mechanics[31]. This, arguably more than any other discovery of the past fifty years, will be what drives our quest for deeper understanding of the physical world in the twenty-first century.
Bibliography


Appendix A

Electronic circuit designs
Figure A.3: PECL-logic coincidence detection circuit PCB layout
Figure A.4: LCWP driver circuit, page 1