

ENCODING, MANIPULATING AND MEASURING QUANTUM INFORMATION IN OPTICS

Nathan K. Langford

B.A., B.Sc., B.Sc. (Hons. Class I)

A thesis submitted for the degree of
Doctor of Philosophy at The University of Queensland
in February, 2007.



School of Physical Sciences, The University of Queensland

© Nathan K. Langford, 2007.

Typeset in L^AT_EX 2_ε.

Statement of Originality

Except where acknowledged in the text and in the *Statement of Contribution to Jointly-published Work and Contributions by Others*, the research presented in this thesis is, to the best of my knowledge, my own original work and has not been submitted in whole or in part for a degree at this or any other university.

Statement of Contribution to Jointly-published Work and Contributions by Others

The published work that has been described in this thesis has been carried out in many different groups both in Australia and the United States. Appropriate acknowledgements of author contributions have been included at the beginning of each relevant chapter [Part III, Chs 7–10].

Nathan K. Langford

Professor Andrew G. White

“But I don’t want to go among mad people,”
Alice remarked.
“Oh, you can’t help that,” said the Cat; “we’re
all mad here. I’m mad. You’re mad.”
“How do you know I’m mad?” said Alice.
“You must be,” said the Cat, “or you wouldn’t
have come here.”

Alice’s Adventures In Wonderland, Lewis
Carroll

Acknowledgements

My first thanks must go to Andrew White who has persevered with me through the eventful years of both my Honours and PhD projects (only two broken bones ain’t bad). Your enthusiasm and passion for science in all its forms is infectious. Your advice, patience and encouragement, particularly in the difficult times, have been greatly appreciated.

Thanks also to Tim Ralph, who has always been available for a chat when I needed one, even though our research has been mostly following different paths.

My most exciting moments in science have always taken place when working with others and this relationship is never more intense or rewarding than in the lab. A profound thank you to Rohan and Julio for their sense of humour, patience and tolerance through long hours in the dark with only the light of a computer monitor and a little yellow LED as guide.

I feel very privileged to have worked with many talented physicists during my varied research projects. Thanks to all my co-authors and colleagues who have made my scientific journey such an exciting and enjoyable experience. I look forward to continuing our collaborations for many years to come. In particular, for tolerating me bouncing into their offices and asking them silly, random or obscure questions (and for their normally patient responses), I would like to thank Stephen Bartlett, Andrew Doherty, Michael Harvey, Timo Nieminen, Howard Wiseman, and most especially, Geoff Pryde, Jeremy O’Brien and Alexei Gilchrist. Thanks also to Bill Munro and Paul Kwiat who always made time for me when it was needed, and whose advice has made me a better scientist and colleague.

The work in this thesis would certainly not have been possible without a number of people in and around UQ Physics who operate largely behind the scenes. My thanks go to Carl Warner for his expertise with photography, to Ian Mortimer and Sam Zammit for all things computer-related, and to Anna Rogers and Jo Hughes who were always able to get things done—even finding money for a meagre undergrad to go to a conference in Helsinki. Thank you also to the “workshop boys”, Alan, Dave and Evan, for their exceptional work on many a widget and thingummybob. In their hands, my blithe requests for the impossible were made reality and they always managed to fit me in somehow, even when I needed something yesterday.

I must also mention a few people who are the very reasons I am doing physics. A tremendous thank you must go to Halina Rubinsztein-Dunlop and Norm Heckenberg who have watched over my academic career ever since I came to the UQ Physics Department as a work experience student in Grade 11 and who gave me so many early opportunities to do real science. Thanks also to Jeff Willmer and John Mainstone, two exceptional and

inspiring teachers of science who grabbed my curiosity and made it grow.

During my PhD, I was extremely fortunate to have the opportunity to travel twice overseas to work in Paul Kwiat's laboratory in Illinois. Thanks to Andrew and Paul for making that possible and to all my friends in Champagne-Urbana for making a strange place feel like a home. Most particularly, thank you to Julio and Elizabeth who, despite my unintelligible accent, virtually adopted me while I was there and whose hospitality made me feel so welcome.

Throughout my time at university, I have been lucky enough to make and keep many new friends who are spread far and wide around the world. To all my friends, thank you for your patience and support, for your sense of the ridiculous, and for sharing with me in my crusade to save the world through silliness. In particular, thank you to Margaret for movies, chats and general procrastination; to Janet for emergency physio when long hours in the lab became such a pain and (with Hamish) for the permanent welcome into your home; to Raph, co-founder of the Pyjama Productions (long may they continue!), and Marty for musical flights of fancy whenever they were required; and a special thanks to Charles who made a foolish offer to proof-read this thesis and who undertook the Herculean task with good cheer and great thoroughness. Most especially, I wish to thank Cavin ("Calvin") for many hours studying special relativity and electromagnetism, for many lunches at the grassy knoll when we escaped to the "other side" of the uni, for soccer, for endless conversations and for inspiration.

Finally, I could not have reached this point without the continual support and love of my family. To Carl, the big brother who, more than anyone else, can make me feel like a special human being, to Ben for taking me to other worlds and starting me off on the search, to Donna and Suzy just for being my big sisters, and to Zac, Sian and Aaron for your unconditional hugs, smiles and laughter—thank you so very much. And most importantly, to Mum and Dad—your example has given me the courage to do what I love, the strength to survive the difficult times, and the confidence to believe in myself. Thank you for bringing me dinner when I was working late at the lab and for helping me when I was packing at the last minute for a trip. And thank you for your unswerving faith in me—it is finally done!

List of Publications

Publications relevant to the Thesis

- [1] J. T. Barreiro, N. K. Langford, N. A. Peters, and P. G. Kwiat. Generation of hyper-entangled photon pairs. *Physical Review Letters*, 95(26):260501, 2005.
- [2] N. K. Langford, T. J. Weinhold, R. Prevedel, K. J. Resch, A. Gilchrist, J. L. O'Brien, G. J. Pryde, and A. G. White. Demonstration of a simple entangling optical gate and its use in Bell-state analysis. *Physical Review Letters*, 95(21):210504, 2005.
- [3] A. Gilchrist, N. K. Langford, and M. A. Nielsen. Distance measures to compare real and ideal quantum processes. *Physical Review A*, 71(6):062310, 2005.
- [4] N. K. Langford, R. B. Dalton, M. D. Harvey, J. L. O'Brien, G. J. Pryde, A. Gilchrist, S. D. Bartlett, and A. G. White. Measuring entangled qutrits and their use for quantum bit commitment. *Physical Review Letters*, 93(5):053601, 2004.
- [5] T. C. Ralph, N. K. Langford, T. B. Bell, and A. G. White. Linear optical controlled-NOT gate in the coincidence basis. *Physical Review A*, 65(6):062324, 2002.

Additional Publications relevant to the Thesis but not Forming Part of it

- [1] A. G. White, A. Gilchrist, G. J. Pryde, J. L. O'Brien, M. J. Bremner, and N. K. Langford. Measuring two-qubit gates. *Journal of the Optical Society of America B*, 24(2):172, 2007.
- [2] D. T. Pope, D. Berry, and N. K. Langford. The practicality of adaptive phase estimation. *Optics and Spectroscopy*, 99(3):397, 2005.
- [3] D. T. Pope, H. M. Wiseman, and N. K. Langford. Adaptive phase estimation is more accurate than non-adaptive phase estimation for continuous beams of light. *Physical Review A*, 70(4):043812, 2004.
- [4] J. L. O'Brien, G. J. Pryde, A. Gilchrist, D. F. V. James, N. K. Langford, T. C. Ralph, and A. G. White. Quantum process tomography of a controlled-NOT gate. *Physical Review Letters*, 93(8):080502, 2004.

Abstract

Encoding, manipulating and measuring quantum information in optics

Nathan K. Langford
School of Physical Sciences
The University of Queensland

In this thesis, I present experimental and theoretical work which assesses and develops a range of tools that are required for performing quantum information processing, particularly in photonic systems.

I investigate the three degrees of freedom of a single photon—its polarisation, spatial-momentum and time-frequency distributions. For polarisation, I show that using wave plates to implement arbitrary, single-qubit rotations is more complicated than is commonly appreciated. In the spatial-momentum and time-frequency domains, I develop ways to perform tomographic analysis of quantum states, and I report the first demonstrations of these techniques. In the time-frequency domain, the tomography technique utilises entanglement in the photon polarisation as a resource to store and provide access to the time-frequency information.

In my first two experiments, I use spontaneous parametric down-conversion to produce entanglement between pairs of single photons in all three degrees of freedom. I demonstrate the first characterisation of entanglement in spatial modes and the time-frequency domain, the first quantitative measurement of entangled qutrit states, and the highest quality entangled states yet measured in both polarisation and spatial modes. I also report the first realisation of complete hyperentanglement, and a full, black-box tomography of a 36-dimensional two-photon state—the largest system to be characterised in this way to date.

In my final experiment, I model and implement a new architecture for a controlled-Z gate which is much simpler to align than previous implementations. This gate requires only one non-classical interference condition, the visibility of which is the main limitation to its performance. I show that the gate operates effectively as a means of both creating entanglement and discriminating between the four elements in a basis of maximally entangled, Bell-type states. Indeed, its observed performance as a Bell analyser would be sufficient to build a quantum state teleporter which would guarantee that the recipient would be left with a better copy of an unknown input state than any eavesdropper.

In a series of numerical simulations, I investigate some of the practicalities that arise when using tomographic reconstruction techniques, including how to estimate errors, which measurements to make and what is actually the optimal reconstruction. In particular, I show that tomographies perform better when based on the results from over-complete sets of measurements.

Finally, I discuss the important issue of how to compare two processes, particularly when trying to assess the quality of a measured process by comparing it to some expected ideal. Judging possible candidate measures against a set of experimentally and theoretically motivated criteria eliminates all but a small number which have particularly promising characteristics.

To Mum and Dad.

Contents

Acknowledgements	v
List of Publications	vii
Abstract	ix
List of Figures	xv
List of Tables	xix
1 Introduction	1
1.1 Overview of the thesis	4
1.2 How to read the thesis	5
1.3 Chapter 1 References	7
I Quantum information	11
2 Quantum information background	13
2.1 Quantum information basics	13
2.1.1 Quantum states	13
2.1.2 Quantum operators	16
2.1.3 Quantum processes	19
2.1.4 The mixed-state fidelity	23
2.1.5 Quantum measurement theory	24
2.1.6 Quantum correlations and entanglement	26
2.2 Quantum information in optics	30
2.2.1 Fields and photons	30
2.2.2 Different pictures of evolution	32
2.2.3 The single photon	36
2.3 Chapter 2 References	37
3 Quantum state and process tomography	39
3.1 Heisenberg's uncertainty principle	40
3.2 Linear quantum state tomography	41
3.2.1 Single qubit tomography	43
3.2.2 Qudit tomography	44
3.3 Maximum likelihood quantum state tomography	44
3.3.1 Parametrising the physical density matrix	45
3.3.2 The penalty function	46

3.3.3	The numerical optimisation	47
3.4	Convex optimisation tomography	47
3.4.1	Convex optimisation problems	47
3.4.2	Tomography as a semidefinite programme	48
3.5	Quantum process tomography	49
3.6	Calculating errors in tomographic reconstructions	52
3.6.1	Fit quality	53
3.6.2	The Monte-Carlo approach	54
3.7	Over-complete measurement sets	59
3.8	Weighting the likelihood function	66
3.8.1	Poissonian errors	66
3.8.2	Variable normalisations	69
3.9	Chapter 3 References	70

II Encoding and measuring information in photons 73

4 Encoding in polarisation 75

4.1	The electromagnetic field	75
4.2	Polarisation and the Poincaré sphere	76
4.3	Birefringence	77
4.4	Wave plates	80
4.4.1	Using wave plates to perform polarisation analysis	82
4.5	Real wave plates in quantum state tomography	83
4.5.1	Real wave plates are not ideal	83
4.6	Arbitrary single-qubit operations using wave plates	84
4.6.1	Unitary operations	86
4.6.2	Berry's or Pancharatnam's phase	87
4.6.3	Using Berry's phase to implement Bloch rotations	88
4.6.4	Optimal method for implementing arbitrary unitaries	90
4.6.5	Simple, practical method for implementing arbitrary unitaries	95
4.6.6	Non-local Berry's phase effects	96
4.7	Chapter 4 References	97

5 Encoding in transverse spatial modes 99

5.1	The paraxial wave equation	100
5.2	The Gaussian mode families	101
5.2.1	Self-similarity and the Gaussian mode generations	104
5.2.2	Example: nondegenerate vortex mode superpositions	105
5.3	Spatially encoded qudits	107
5.3.1	The degenerate qubit	108
5.3.2	The nondegenerate qubits	108
5.3.3	The nondegenerate qutrit	109
5.4	Holograms	109
5.4.1	Different types of holograms	110
5.4.2	Example: the sinusoidal plane-wave hologram	111
5.4.3	Example: engineered plane-wave holograms	112
5.4.4	An intuitive description of off-axis holograms	114
5.5	Creating higher-order Gaussian modes	116

5.5.1	Using plane-wave holograms	116
5.5.2	Optimising mode-conversion efficiency	118
5.5.3	Using Gaussian mode holograms	118
5.6	Spatial mode analysis	122
5.6.1	Detecting higher-order modes	122
5.6.2	Single-mode fibres—the spatial filters	123
5.6.3	Measuring the tomographic spatial modes	125
5.6.4	Spatial mode tomography with coherent states	126
5.6.5	Theory: using plane-wave holograms in degenerate QST	129
5.6.6	Theory: using plane-wave holograms in nondegenerate QST	133
5.7	Chapter 5 References	136
6	The Type-I down-conversion process	139
6.1	Nonlinear optical materials	139
6.2	Spontaneous parametric down-conversion	140
6.3	Ideal, noncollinear Type-I phase matching	142
6.4	Imperfect, collinear Type-I phase matching	144
6.A	Appendix: Proof—down-conversion expansion	146
6.2	Chapter 6 References	149
III	Applications in quantum information	151
7	Measuring entangled qutrits and their use for quantum bit commitment	153
7.1	Introduction	153
7.2	Making computer generated holograms	154
7.2.1	The computer generated images	154
7.2.2	Making the master images	156
7.2.3	Producing the holograms	157
7.3	Experimental design	159
7.4	Results	167
7.5	Quantum bit commitment	171
7.6	Tomography results with unbalanced measurements	175
7.7	Conclusions	176
7.8	Chapter 7 References	178
8	Hyperentangled photons	181
8.1	Multi-degree entanglement	181
8.2	Hyperentanglement in down-conversion	183
8.3	Encoding in the time-frequency domain	184
8.3.1	Entanglement in the time-frequency domain	185
8.3.2	Polarisation-entanglement-assisted Bell violations	187
8.3.3	Time-frequency tomography	188
8.4	Experimental design	191
8.5	Results	195
8.5.1	Individual subspace tomographies	195
8.5.2	Time-frequency tomography	196
8.5.3	CHSH Bell violations	197
8.5.4	Multi-degree quantum state tomography	198

8.6	Conclusion	201
8.7	Chapter 8 References	203
9	Distance measures for comparing quantum processes	207
9.1	Introduction	207
9.2	Distance measures for quantum states	211
9.3	Error measures for quantum processes	214
9.3.1	Error measures based on the process matrix	215
9.3.2	Error measures based on the average case	217
9.3.3	Error measures based on the worst case	219
9.4	Application to quantum computing	221
9.4.1	Function computation	221
9.4.2	Sampling computation	222
9.5	Summary, recommendations, and conclusion	224
9.A	Appendix: Worst case proofs	227
9.A.1	Proof of worst-case stabilisation	227
9.A.2	Proof of convex optimisation property for F_{stab}	228
9.B	Appendix: Application to quantum computing	229
9.B.1	Function computation in the worst case	229
9.B.2	Function computation in the average case	229
9.B.3	Sampling computation in the worst case	230
9.3	Chapter 9 References	231
10	A simple entangling optical gate and its use in Bell-state analysis	235
10.1	Introduction	235
10.1.1	A note on pulsed experiments	236
10.2	The two-photon CZ gate	237
10.2.1	Non-classical interference	237
10.2.2	The original design	238
10.2.3	The all-polarisation approach	240
10.3	Experimental design	244
10.3.1	State preparation—the entangled source	244
10.3.2	State discrimination—the entangling (disentangling) gate	246
10.4	Results	247
10.4.1	Characterising the source	247
10.4.2	Characterising the alternative beam paths	251
10.4.3	The CZ as an entangling gate	251
10.4.4	Process tomography	254
10.4.5	The CZ as a disentangling gate for Bell-state analysis	259
10.5	Conclusions	265
10.6	Chapter 10 References	267
11	Discussion and conclusions	269
11.1	Chapter 11 References	273

List of Figures

2.1	The qubit sphere.	15
2.2	Evolution of “classical” coherent states through a 50:50 beam splitter. . . .	34
2.3	A general beam splitter.	36
3.1	Quantum state and process tomography.	50
3.2	Convergence of Monte-Carlo errors for randomly generated one- and two-qubit states.	56
3.3	Convergence of Monte-Carlo errors for a range of experimental tomographies.	58
3.4	The effect of over-complete measurement sets on the tomography of single-qubit pure states (qubit sphere view).	62
3.5	The effect of over-complete measurement sets on the tomography of single-qubit pure states (flattened view).	63
3.6	The effect of over-complete measurement sets on the tomography of typical two-qubit states.	64
3.7	Weighting the tomographic penalty function.	68
4.1	The Poincaré sphere.	77
4.2	The refractive index ellipse.	78
4.3	Walk-off in a negative uniaxial crystal.	79
4.4	The arbitrary polarisation analyser.	82
4.5	Apparatus for characterising real wave plates.	84
4.6	Expected results for characterising real wave plates.	85
4.7	An arbitrary single-qubit unitary operation.	86
4.8	Pancharatnam’s phase.	87
4.9	Implementing Bloch rotations with waveplates.	89
4.10	The QWP step.	91
4.11	The first QWP step of the optimal unitary rotation protocol.	92
4.12	The phase shift from the optimal unitary rotation protocol.	93
4.13	A simple, practical protocol for implementing an arbitrary unitary rotation.	95
5.1	The spatial mode families for a Gaussian beam.	104
5.2	The spatial superposition of vortex modes with the lowest-order Gaussian mode.	107
5.3	The standard bases of the spatial-mode qubits.	108
5.4	The fringe profiles for sinusoidal, blazed and binary holograms.	113
5.5	An intuitive picture of how off-axis holograms work.	115
5.6	Plane-wave holograms.	117
5.7	Gaussian mode holograms.	118
5.8	Optimising the VM_{+1} Gaussian mode hologram.	120

5.9	VM_{+1} hologram outputs.	121
5.10	Fourier transform recognition for spatial mode analysis.	122
5.11	Demonstrating higher-order mode rejection by a SMF.	124
5.12	Filtering with a single-mode fibre.	125
5.13	The spatial tomography hologram.	125
5.14	The experimental layout for spatial tomography using coherent states.	126
5.15	Prepared and reconstructed spatial modes used for tomography with coherent states.	127
5.16	Reconstructed single-“qubit” density matrices from spatial tomography using coherent states.	128
5.17	Degenerate pseudo-basis measurement probabilities.	131
5.18	Testing the degenerate spatial mode tomography technique.	131
5.19	nondegenerate pseudo-basis measurement probabilities.	135
5.20	Testing the nondegenerate spatial mode tomography technique.	135
6.1	Type-I phase-matching geometry.	142
6.2	Type-I down-conversion cone.	143
6.3	A quantum treatment of collinear, Type-I phase matching.	145
7.1	A schematic overview of the process for making CG holograms.	155
7.2	Sinusoidal holograms: examples.	155
7.3	Typical errors in binary computer-generated holograms.	156
7.4	Two test images for characterising the slide film and the Polaroid slide writer.	156
7.5	Diffraction efficiencies of holograms created with varying exposure.	158
7.6	A conceptual schematic of the experimental apparatus.	160
7.7	Down-conversion with Type-I phase matching.	160
7.8	The basic conceptual design for the down-conversion mode-matching optics.	162
7.9	A detailed schematic of the experimental apparatus.	164
7.10	Modes imaged by the SMFs in each arm of the down-conversion.	165
7.11	Spatial tomography within the degenerate and nondegenerate qubit subspaces.	168
7.12	Spatial tomography within the nondegenerate qutrit subspace.	170
7.13	Bit commitment: Step 1) Logical state preparation.	172
7.14	Bit commitment: Step 2) The commitment.	173
7.15	Bit commitment: The protocol security—Alice’s Control vs Bob’s Knowledge Gain.	174
7.16	Reanalysed nondegenerate qubit subspace: non-optimal measurement settings.	175
7.17	Reanalysed nondegenerate qutrit subspace: non-optimal measurement settings.	177
8.1	The Franson interferometer—a highly unbalanced Mach-Zehnder interferometer.	185
8.2	The Franson time-frequency Bell violation experiment in parametric down-conversion.	186
8.3	The modified Franson experiment.	187
8.4	Conceptual experimental layout for the creation and analysis of hyper-entangled photons.	192
8.5	The Type-I crystal sandwich source of polarisation entanglement.	193

8.6	Detailed experimental layout for the creation and analysis of hyperentangled photons.	194
8.7	Typical two-qubit tomographies within the (a) polarisation and (b) spatial mode subspaces.	195
8.8	Liquid crystal voltage curves.	196
8.9	Time-frequency tomography.	197
8.10	Density matrices for $2 \times 2 \times 3 \times 3$ -dimensional polarisation and spatial mode subspace.	199
8.11	Density matrices for $2 \times 2 \times 2 \times 2$ -dimensional polarisation and spatial mode subspace.	200
10.1	The two-photon CZ gate.	238
10.2	A partially polarising beam splitter (PPBS).	241
10.3	Real-world effects in the PPBS CZ gate.	243
10.4	A detailed schematic of the entangled source apparatus.	244
10.5	Characterising the crystal-sandwich entangled source.	249
10.6	The effect of the laboratory air-conditioner cycle on tomographic counts.	250
10.7	An example entangled state created by our CZ gate.	252
10.8	Quantum process tomography of the cw CZ gate.	255
10.9	The CZ gate operating as a Bell-state analyser.	260
10.10A	A simple, simulated teleportation protocol using our CZ gate.	262
10.11	Example input states for the simple teleportation protocol.	264

List of Tables

3.1	Two-qubit measurement order.	61
4.1	Typical wave plate settings for polarisation analysis.	83
5.1	Engineered phase and amplitude holograms.	114
5.2	Hologram reconstruction efficiencies.	128
5.3	Theoretical fidelities for the pseudo-first-order, degenerate basis states. . .	130
5.4	Theoretical fidelities for the pseudo-first-order, nondegenerate basis states.	133
5.5	Theoretical fidelities for the pseudo-first-order, nondegenerate basis states— <i>balanced</i> measurements.	134
7.1	Typical count rates during the alignment process.	166
7.2	Mapping out positions of hologram singularities.	166
7.3	Entanglement for the two-qubit subspaces extracted from the two-qutrit spatial tomography.	171
8.1	Bell parameter S showing CHSH-Bell inequality violations in every degree of freedom.	198
10.1	Typical count rates during the alignment process.	247
10.2	Typical entangled states created by the PPBS CZ gate.	252
10.3	Consistency results for the PPBS CZ gate—three successive process tomo- graphies.	256
10.4	Ideal CZ gate truth tables.	259
10.5	Key properties of measured input and output states for Bell measurements.	261

Well, this bit which I am writing, called Introduction, is really the *er-h'r'm* of the book, and I have put it in, partly so as not to take you by surprise, and partly because I can't do without it now. There are some very clever writers who say that it is quite easy not to have an *er-h'r'm*, but I don't agree with them. I think it is much easier not to have all the rest of the book.

Er-h'r'm to Now we are Six, A. A. Milne

“And what is the use of a book,” thought Alice,
”without pictures or conversations?”

Alice's Adventures in Wonderland, Lewis Carroll

Chapter 1

Introduction

In 1982, Richard Feynman gave birth to a new field of physics—quantum computation—when he suggested that the best way to simulate a quantum mechanical system would be to use another quantum system—that is, a machine which could genuinely exploit its fundamental quantum features [1]. Then, from 1994, this field began to attract an ever-increasing interest when Peter Shor described an algorithm which used a quantum computer for factoring large semi-prime numbers. Shor's algorithm could far out-perform the best known classical algorithms [2]. Quantum computing has since expanded into the more general field of quantum information—the study of manipulating and using information stored in a quantum system.

Perhaps the most immediate and significant consequence of attempting a physical computation using an explicitly quantum machine is that the machine can exhibit entanglement—correlations which cannot be observed between classical systems. Considered by Erwin Schrödinger to be “not one but rather *the* characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought,” [3] entanglement has sparked great interest ever since the beginnings of quantum mechanics. Albert Einstein felt it was one of the most controversial aspects of the theory he helped to create, and debates about explanations for its effects continue even today. One key development in these debates was Bell's pioneering theory showing that quantum systems could, in principle, exhibit far higher levels of correlation than any system constrained by the typically classical notions of locality and realism [4].

In the field of quantum information, many consider entanglement to be the fundamental resource for quantum computing, directly responsible for its enormous potential capabilities, and entangled states are a vital part of quantum communication protocols such as superdense coding [5] and teleportation [6]. Consequently, the ability to create, manipulate, measure and characterise entanglement has been one of the defining measures

of success for many experimental quantum information groups, whether they have been demonstrating a resource for other applications or the signature of a successfully implemented quantum process. In quantum optics, spontaneous parametric down-conversion has provided a particularly versatile source of photon-based entanglement in many different forms. For example, many of the most successful experimental demonstrations of Bell-type violations of correlation limits were performed using down-conversion sources (see, e.g., [7–14]).

There are also many difficulties that arise when trying to perform physical computations on a quantum machine. In most quantum information protocols, processes must be implemented accurately and without introducing significant mixture into the quantum state. This places stringent and competing demands on experiments, and particularly the experimental systems. Since mixture arises when quantum systems couple to an inaccessible environment, the elements of a quantum computer need to be strongly isolated from the environment. However, in order to do anything interesting, we (the users) must also be able to influence the evolution of the computer by applying operations to it, some of which will have to involve inducing strong interactions between its different components. Clearly, this requires some sort of interaction with the environment (namely, us), but in a very restricted way.

Optical systems are perhaps the most readily accessible systems capable of exhibiting controlled quantum mechanical behaviour. Many other systems, like atoms, ions and superconductors, will only achieve this behaviour at very low temperatures or very high vacuum levels (or both). Photons, however, are generally extremely isolated from both their environment and other photons, and so information stored in optical systems is “safe” (they have long coherence times) without the need for refrigeration or strong vacuums. For many quantum experiments, it is therefore often easiest, quickest and cheapest to perform them initially with photons, making quantum optics a popular test-bed for quantum information tasks in small-scale systems. This “prototyping” feature has allowed experimentalists to explore many fundamental properties of these tasks, even if they may ultimately be translated to another physical system. For example, the field of quantum optics produced practical demonstrations of teleportation [15, 16], dense coding [17], and quantum cryptography [18–22] before similar experiments in other systems. As an aside, using light is also particularly suited to communication tasks, since it provides the fastest possible transmission speeds.

The key difficulty with using optics for quantum information processing is that photons are so strongly isolated that it is extremely difficult to produce the photon-photon interactions vital to universal quantum logic. They can be induced by nonlinear atom-photon interactions, but the nonlinearities required are generally far larger than what is realisable in most experimental systems [23]. In 2001, Knill, Laflamme and Milburn (KLM) proposed a scheme for “efficient” linear optical quantum computing (LOQC) which avoided the problem of weak nonlinearities by using measurements—which are highly nonlinear—to produce *nondeterministic* logic gates [24]. These can then be made near-deterministic, as required for a scalable quantum computer, with teleportation and quantum error correction. It would be brave to say that the KLM scheme provides a *practically feasible* approach to building a quantum computer without some astonishing technological advances, but it did stimulate an active theoretical effort to simplify the original scheme [25–29]. This has led to a series of demonstrations of related quantum logic gates [30–

34], including some of the earliest successful quantum logic experiments in any physical system.

One of the most important outcomes of the KLM proposal was that it made single-photon optics a legitimate candidate for producing a working quantum computer, not just for “test-bed” experiments. Unfortunately, implementing the two key components of teleportation and quantum error correction is extremely difficult, and the simplifications described above have made no significant inroads into this problem. There are, however, promising alternative methods for achieving scalability in optical quantum computing.

One approach replaces the standard circuit model of quantum computing with a completely different paradigm, called *one-way* or *cluster state* quantum computation [35], which has particular advantages for optical quantum computing [36]. In this approach, the nondeterministic logic gates are used “off-line” to produce a complicated, highly-entangled initial state, and the computation is performed with a series of carefully chosen, single-qubit, projective measurements. Since the choice of measurements will generally depend on the preceding results, this also requires the feeding forward of information. In optics, such measurements are generally straightforward, and the feed-forward, though difficult, is achievable. The only effect of the nondeterministic optical gates is to make the initial state preparation probabilistic, but with a success rate far higher than the equivalent KLM scheme [36].

In another new approach, a much higher effective photon-photon nonlinearity is achieved by replacing single ancilla photons with much brighter coherent states [37, 38]. This general principle has been shown theoretically to allow near-deterministic controlled-NOT (CNOT) gates [39], non-destructive Bell-state analysis [38], quantum non-demolition photon-number-resolving detectors [37], single-photon sources [40], and simplified quantum error correction [41]. The main difficulty of this scheme is in successfully implementing the basic theoretical idea—i.e. coupling the bright coherent states and the single photons, and then performing sufficiently accurate homodyne detection on the coherent states. This has not yet been achieved.

From a theoretical perspective, it does not matter what physical system is used to encode and manipulate quantum information. In fact, any realistic quantum computer is likely to use a combination of quantum systems, both optics- and matter-based. However, there are several practical tasks we must be able to perform to produce a working quantum computer, regardless of its eventual form, and the same is true of any quantum information task. We must be able to prepare, measure and characterise quantum states (potentially distributed and entangled), and manipulate them using quantum operations which we must also be able to test and characterise. These techniques are particularly important in experimental quantum information, where experimentalists must be able to gauge their achievements and justify their claims with measured data.

Much of the excitement in quantum information and attractiveness of the field as a whole lies in its highly cross-disciplinary nature. Scientists from diverse fields are all trying to solve the same fundamental problems, using disparate techniques which may sometimes be integrated for unique results. In this thesis I present results in a range of areas within quantum optics, but there are three overarching ideas which form a statement of purpose for the work: first, to explore the characterisation techniques described above which will

be vital to the success of building a quantum computer; second, to investigate and utilise different methods for encoding information in single photons; and finally, because of the importance of entanglement as an information resource, to study a range of ways to generate it in optical systems.

It is also important to make a statement of what this thesis is *not*. In this research, I am not concerned with testing quantum mechanics itself, nor with questioning its fundamental interpretations. Instead, I begin with the assumption that quantum mechanics provides the best explanation of the physical world, and taking a pragmatic approach, I use whichever interpretation is appropriate. I see the study of quantum information as an opportunity to explore fundamental features of quantum mechanics, in my case, using optical quantum information experiments.

1.1 Overview of the thesis

The work which has contributed to this thesis spans a range of contexts within quantum optics and quantum information. Therefore I have tried to bring together and summarise the knowledge and expertise I have accumulated in a unified way, hopefully to help new researchers move between these areas more freely. Chapters 2–6 provide an overview of concepts which will be used in analysing and interpreting the material in the later chapters. They also define the technical quantities I will use and the conventions I will follow throughout. Chapters 7–10 are the main research chapters in the thesis, but large parts of Chs 3–6 also contain the results of my own original work.

In Chapters 2 and 3, I summarise the quantum information concepts which form the basis of the research in this thesis. I briefly introduce the tools required to analyse optical quantum information experiments [Sec. 2.2], and in some detail, I discuss the very important techniques of quantum state and process tomography [Ch. 3]. In particular, I report the results of my recent theoretical investigations into the properties of tomographic measurement and reconstruction techniques [Secs 3.6, 3.7 and 3.8].

In Chapters 4 and 5, and later in Sec. 8.3, I discuss in depth the theory and experimental techniques which are useful when encoding information in the three degrees of freedom of a single photon: polarisation [Ch. 4], spatial-momentum [Ch. 5], and time-frequency [Sec. 8.3]. In relation to polarisation, I investigate the important quantum information task of implementing arbitrary single-qubit unitary rotations with birefringent wave plates [Sec. 4.6]. In the section on transverse spatial modes, I present the results of new work exploring the problems associated with using thin holograms to create and analyse Gaussian spatial modes [Sec. 5.5], and how they affect spatial mode tomography [Sec. 5.6]. Finally, in Section 8.3.1, I describe a new technique for performing tomography in the time-frequency domain which exploits the simultaneous polarisation entanglement of hyperentangled photons (an extension of a technique introduced in Ref. [42]).

In Chapter 6, I provide a brief summary of the nonlinear process of spontaneous parametric down-conversion and type-I phase matching.

In the work described in Chapter 7, we performed the first tomographic characterisation

of photonic qutrits entangled in the spatial degree of freedom. We also showed how to use the entangled state in a purification quantum bit commitment protocol. Finally, using the more recent results of Secs 5.5 and 5.6, I have shown how to improve the tomography of nondegenerate spatial quantum states.

In Chapter 8, I discuss experiments where we demonstrated the first production and characterisation of completely hyperentangled photons using nonlinear down-conversion. We used Bell inequality violations to verify the presence of the hyperentanglement and full quantum state tomography to provide a more quantitative analysis.

In Chapter 9, I report the results of a theoretical research project in which we studied different ways to compare and characterise quantum processes, such as the logic gates which must be implemented when carrying out quantum information tasks.

In Chapter 10, I describe and model a new architecture for a two-photon, polarisation-based controlled-Z (CZ) gate. I characterised it experimentally using quantum process tomography, and combined it with a source of polarisation-entangled photon pairs to demonstrate completely discriminating Bell-state analysis.

Finally, I review the main results of the thesis in Chapter 11, and discuss future directions for experimental and theoretical investigations which have arisen out of this work.

1.2 How to read the thesis

*But the principal failing occurred in the sailing,
And the Bellman, perplexed and distressed,
Said he had hoped, at least, when the wind blue due East
That the ship would not travel due West!*

The Hunting of the Snark, an Agony in Eight Fits,
Lewis Carroll

Because of the diversity of topics covered, there are many different paths through this thesis which should be reasonably self-contained for those readers with particular interest areas. This section is meant as a guide to those who wish to read the thesis more selectively.

For readers who are new to the field of quantum information or who simply wish to gain a rough handle on the language of the field (the local dialect, one might say), then Ch. 2 would be useful, particularly the first section.

Those who are interested in classical holography should read Sec. 5 [with reference to the notational conventions introduced in Secs 2.1 and 2.2]. If the interest also extends to quantum imaging, coupling photons into fibres or encoding information in spatial modes, then Chs 7 and 8 are also pertinent.

For readers who wish to focus on quantum information encoded in higher-dimensional

systems, these are used in the experiments in Chs 7 and 8, which build on the background provided in Sec. 2.1 and Chs 4 and 5.

For readers involved with polarisation-based optical quantum computing, then Ch. 2, Ch. 4, Ch. 6 and Ch. 10 are the most relevant sections, along with selected sections of Ch. 9 for some useful theoretical quantities.

Those who particularly wish to explore the techniques of tomographic reconstruction and the related experimental issues should read Ch. 3, Sec. 5.6, Sec. 8.3.1 and the results sections of Chs 7, 8 and 10.

Finally, for the reader with a more theoretical bent, various sections of particular interest may include Secs 2.2, 4.6, 6.4, 7.5, and 8.3.1, Ch. 9, and Sec. 10.2.

1.3 Chapter 1 References

- [1] R. P. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21:467, 1982.
- [2] P. W. Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In *Proceedings of the 35th Annual Symposium on the Foundations of Computer Science*, page 124, 1994.
- [3] E. Schrödinger. *Proceedings of the Cambridge Philosophical Society*, 31:555, 1935.
- [4] J. S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195, 1964. As reprinted in *Speakable and Unspeakable in Quantum Mechanics: Collected papers on quantum philosophy*, Cambridge University Press (1987).
- [5] C. H. Bennett and S. J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Physical Review Letters*, 69:2881, 1992.
- [6] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 70:1895, 1993.
- [7] Z. Y. Ou and L. Mandel. Violation of Bell's inequality and classical probability in a two-photon correlation experiment. *Physical Review Letters*, 61:50, 1988.
- [8] Y. H. Shih and C. O. Alley. New type of Einstein-Podolsky-Rosen-Bohm experiment using pairs of light quanta produced by optical parametric down conversion. *Physical Review Letters*, 61:2921, 1988.
- [9] J. G. Rarity and P. R. Tapster. Experimental violation of Bell's inequality based on phase and momentum. *Physical Review Letters*, 64:2495, 1990.
- [10] P. G. Kwiat, A. M. Steinberg, and R. Y. Chiao. High-visibility interference in a Bell-inequality experiment for energy and time. *Physical Review A*, 47:R2472, 1993.
- [11] P. R. Tapster, J. G. Rarity, and P. C. M. Owens. Violation of Bell's inequality over 4 km of optical fiber. *Physical Review Letters*, 73:1923, 1994.
- [12] A. Vaziri, G. Weihs, and A. Zeilinger. Experimental two-photon, three-dimensional entanglement for quantum communication. *Physical Review Letters*, 89(24):240401, 2002.
- [13] R. T. Thew, A. Acín, H. Zbinden, and N. Gisin. Experimental realization of entangled qutrits for quantum communication. *Quantum Information and Computation*, 4(2):93, 2004.
- [14] R. T. Thew, A. Acin, H. Zbinden, and N. Gisin. Bell-type test of energy-time entangled qutrits. *Physical Review Letters*, 93:010503, 2004.
- [15] D. Bouwmeester, J.-W. Pan, K. Mattle, M. Eibl, H. Weinfurter, and A. Zeilinger. Experimental quantum teleportation. *Nature*, 390:575, 1997.

- [16] D. Boschi, S. Branca, F. De Martini, L. Hardy, and S. Popescu. Experimental realization of teleporting an unknown pure quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 80:1121, 1998.
- [17] K. Mattle, H. Weinfurter, P. G. Kwiat, and A. Zeilinger. Dense coding in experimental quantum communication. *Physical Review Letters*, 76:4656, 1996.
- [18] C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. In B. Bederson and H. Walther, editors, *Proceedings of the IEEE International Conference on Computers, Systems & Signal Processing, Bangalore, India*, page 175, New York, 1984. IEEE.
- [19] A. K. Ekert. Quantum cryptography based on Bell's theorem. *Physical Review Letters*, 67:661, 1991.
- [20] A. K. Ekert, J. G. Rarity, P. R. Tapster, and G. M. Palma. Practical quantum cryptography based on two-photon interferometry. *Physical Review Letters*, 69:1293, 1992.
- [21] A. Muller, J. Breguet, and N. Gisin. Experimental demonstration of quantum cryptography using polarized photons in optical fibre over more than 1 km. *Europhysics Letters*, 23:383, 1993.
- [22] W. T. Buttler, R. J. Hughes, P. G. Kwiat, S. K. Lamoreaux, G. G. Luther, G. L. Morgan, J. E. Nordholt, C. G. Peterson, and C. M. Simmons. Practical free-space quantum key distribution over 1 km. *Physical Review Letters*, 81:3283, 1998.
- [23] Q. A. Turchette, C. J. Hood, W. Lange, H. Mabuchi, and H. J. Kimble. Measurement of conditional phase shifts for quantum logic. *Physical Review Letters*, 75:4710, 1995.
- [24] E. Knill, R. Laflamme, and G. J. Milburn. A scheme for efficient quantum computation with linear optics. *Nature*, 409:46, 2001.
- [25] T. C. Ralph, A. G. White, W. J. Munro, and G. J. Milburn. Simple scheme for efficient linear optics quantum gates. *Physical Review A*, 65:012314, 2001.
- [26] E. Knill. Quantum gates using linear optics and postselection. *Physical Review A*, 66:052306, 2002.
- [27] T. B. Pittman, B. C. Jacobs, and J. D. Franson. Demonstration of nondeterministic quantum logic operations using linear optical elements. *Physical Review Letters*, 88(25):257902, 2002.
- [28] T. C. Ralph, N. K. Langford, T. B. Bell, and A. G. White. Linear optical controlled-NOT gate in the coincidence basis. *Physical Review A*, 65(6):062324, 2002.
- [29] H. F. Hofmann and S. Takeuchi. Quantum phase gate for photonic qubits using only beam splitters and postselection. *Physical Review A*, 66:024308, 2002.
- [30] T. B. Pittman, M. J. Fitch, B. C. Jacobs, and J. D. Franson. Experimental controlled-NOT logic gate for single photons in the coincidence basis. *Physical Review A*, 68:032316, 2003.

-
- [31] J. L. O'Brien, G. J. Pryde, A. G. White, T. C. Ralph, and D. Branning. Demonstration of an all-optical quantum controlled-NOT gate. *Nature*, 426:264, 2003.
- [32] J. L. O'Brien, G. J. Pryde, A. Gilchrist, D. F. V. James, N. K. Langford, T. C. Ralph, and A. G. White. Quantum process tomography of a controlled-NOT gate. *Physical Review Letters*, 93(8):080502, 2004.
- [33] S. Gasparoni, J.-W. Pan, P. Walther, T. Rudolph, and A. Zeilinger. Realization of a photonic controlled-NOT gate sufficient for quantum computation. *Physical Review Letters*, 93:020504, 2004.
- [34] Z. Zhao, A.-N. Zhang, Y.-A. Chen, H. Zhang, J.-F. Du, T. Yang, and J.-W. Pan. Experimental demonstration of a nondestructive controlled-NOT quantum gate for two independent photon qubits. *Physical Review Letters*, 94:030501, 2005.
- [35] R. Raussendorf and H. J. Briegel. A one-way quantum computer. *Physical Review Letters*, 86:5188, 2001.
- [36] M. A. Nielsen. Optical quantum computation using cluster states. *Physical Review Letters*, 93:040503, 2004.
- [37] W. J. Munro, K. Nemoto, R. G. Beausoleil, and T. P. Spiller. High-efficiency quantum-nondemolition single-photon-number-resolving detector. *Physical Review A*, 71:033819, 2005.
- [38] S. D. Barrett, P. Kok, K. Nemoto, R. G. Beausoleil, W. J. Munro, and T. P. Spiller. Symmetry analyzer for nondestructive Bell-state detection using weak nonlinearities. *Physical Review A*, 71:060302(R), 2005.
- [39] K. Nemoto and W. J. Munro. Nearly deterministic linear optical controlled-NOT gate. *Physical Review Letters*, 93:250502, 2004.
- [40] W. J. Munro, K. Nemoto, and T. P. Spiller. Weak nonlinearities: a new route to optical quantum computation. *New Journal of Physics*, 7:137, 2005.
- [41] F. Yamaguchi, K. Nemoto, and W. J. Munro. Quantum error correction via robust probe modes. *Physical Review A*, 73:060302(R), 2006.
- [42] D. V. Strekalov, T. B. Pittman, A. V. Sergienko, Y. H. Shih, and P. G. Kwiat. Postselection-free energy-time entanglement. *Physical Review A*, 54:R1, 1996.

Part I

Quantum information

“The time has come,” the Walrus said,
 “To speak of many things:
 Of shoes—and ships—and sealing-wax—
 Of cabbages—and kings,—
 And why the sea is boiling hot—
 And whether pigs have wings...”

Alice’s Adventures in Wonderland, Lewis Carroll

“When you wake up in the morning, Pooh,” said Piglet at last, “what’s the first thing you say to yourself?”
 “What’s for breakfast?” said Pooh. “What do *you* say, Piglet?”
 “I say, I wonder what’s going to happen exciting *to-day*?” said Piglet.
 Pooh nodded thoughtfully.
 “It’s the same thing,” he said.

Winnie-the-Pooh, A. A. Milne

Chapter 2

Quantum information background

2.1 Quantum information basics

2.1.1 Quantum states

The simplest unit of quantum information is the *quantum bit*, or *qubit*. A direct quantum analogue of the classical bit, the qubit is a binary quantum system with two logical basis states, $|\mathbf{0}\rangle$ and $|\mathbf{1}\rangle$, which can also occupy any superposition of these two basis states, i.e. $|\psi\rangle = \alpha_0|\mathbf{0}\rangle + \alpha_1|\mathbf{1}\rangle$ where α_j are complex numbers that satisfy $|\alpha_0|^2 + |\alpha_1|^2 = 1$. Of course, this only describes a qubit which is in a pure state. In order to include the possibility of mixed states, in reality impossible to avoid, the state must be written in terms of a density matrix, ρ . For a pure state,

$$\rho = |\psi\rangle\langle\psi| = |\alpha_0|^2|\mathbf{0}\rangle\langle\mathbf{0}| + \alpha_0\alpha_1^*|\mathbf{0}\rangle\langle\mathbf{1}| + \alpha_0^*\alpha_1|\mathbf{1}\rangle\langle\mathbf{0}| + |\alpha_1|^2|\mathbf{1}\rangle\langle\mathbf{1}|, \quad (2.1)$$

$$= \begin{bmatrix} |\alpha_0|^2 & \alpha_0\alpha_1^* \\ \alpha_0^*\alpha_1 & |\alpha_1|^2 \end{bmatrix}. \quad (2.2)$$

The diagonal elements, $p_j = |\alpha_j|^2$, represent the populations of the basis states, or probability that a measurement will find the system in that state. The off-diagonal elements, $c_{jk} = \alpha_j^*\alpha_k$, contain information about the coherence between the two basis states, that is, about the purity or mixedness of the superposition. For the state to be physical,

the density matrix must satisfy several conditions. It must be Hermitian ($\rho^\dagger = \rho$), non-negative (real, non-negative eigenvalues— $\lambda_j \geq 0$), and its trace ($\sum |\alpha_j|^2$) must equal one (normalisation).

Just as with classical computers, we can store more information by combining many qubits together to form larger systems. If \mathcal{H}_1 is the Hilbert space containing all possible states of the first qubit, then the state space for the combined quantum system¹ is $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3 \otimes \dots$. These qubits can also be entangled and can therefore display correlations not observed in classical systems.

In principle, there is no reason why we should be restricted to systems with only two levels. Another way of storing more information is to use individual quantum systems with larger state spaces, e.g. a *qudit* is a system with d basis states. Although a quantum computer built with qudits in principle offers no formal advantage over one which uses qubits ([1], p. 203), there still may be many practical advantages to building a qudit computer. For example, if the number of units in the computer is limited, as may perhaps be dictated by space, more complicated calculations could be performed with qudits than by a computer with the same number of qubits. Moreover, physical systems which naturally occupy a two-dimensional state space are relatively rare, and often a significant amount of effort is necessary to artificially restrict the system to two levels. In such cases, it may make more sense to utilise the extra complexity rather than to fight it. There are also situations in the broader field of quantum information, such as some quantum communication protocols, where qudits have the potential to offer possibilities which are not accessible to qubit-based systems [e.g. see Sec. 7.5].

As with two-level systems, the state space accessible to qudits includes all possible superpositions of the computational basis states, $|\psi\rangle = \sum_{j=0}^{d-1} \alpha_j |\mathbf{j}\rangle$. Mixed states must again be described in terms of density matrices. For example, for a *qutrit* (3 levels),

$$\rho = |\psi\rangle\langle\psi| = \begin{bmatrix} p_0 & c_{01}^* & c_{02}^* \\ c_{01} & p_1 & c_{12}^* \\ c_{02} & c_{12} & p_2 \end{bmatrix}, \quad (2.3)$$

where the diagonal and off-diagonal elements again describe the populations (p_j) and coherences (c_{jk}) respectively. In this form, the normalisation constraint becomes $\sum_{j=0}^{d-1} p_j = 1$.

Because of the greater complexity of mixed states over pure states, more basis elements are necessary to span the space accessible to the quantum system. From Eq. (2.3) above, the number of variables required to describe a d -dimensional density matrix is d^2 —one for each of the population elements and two for each of the coherences (cf. $2d-1$ variables to describe a pure state²). This is also the number of matrices in the mixed state basis.

There are many different ways of constructing a basis for matrices, and these bases will vary in usefulness depending on the situation. Perhaps the simplest is just the basis, $e_{jk} = |\mathbf{j}\rangle\langle\mathbf{k}|$ —I will call this the *elementary basis* since each e_{jk} describes a single matrix element. However, while they are trivial to list, not all these elements represent physical states, which is often inconvenient in a basis for density matrices. One convenient way

¹ \otimes is the symbol for the standard Kronecker tensor product.

²In both of these cases, one variable corresponds to the normalisation of the state.

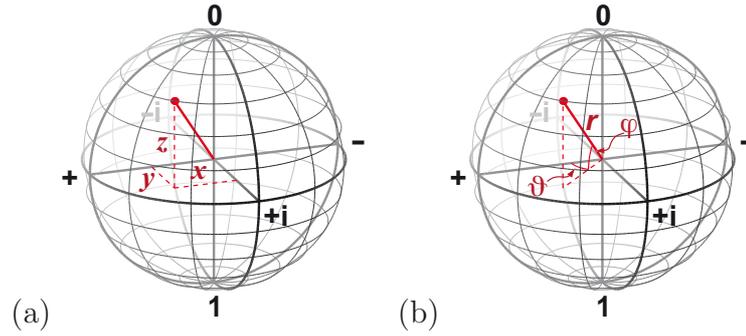


Figure 2.1: The qubit sphere in: (a) Cartesian coordinates; (b) polar coordinates. The Bloch form of the sphere (shown here) is generally convenient when using the logical representation of the qubit (i.e. with the computational basis states).

of constructing a basis from physical states is to use the computational basis states, $|\mathbf{j}\rangle$ (and $|\mathbf{k}\rangle$), and the real and imaginary, equal two-state superpositions, $1/\sqrt{2}(|\mathbf{j}\rangle + |\mathbf{k}\rangle)$ and $1/\sqrt{2}(|\mathbf{j}\rangle + i|\mathbf{k}\rangle)$. In matrix form, the basis elements for qutrits are:

$$\begin{aligned}
 \Omega_0 &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, & \Omega_3 &= \frac{1}{2} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, & \Omega_6 &= \frac{1}{2} \begin{bmatrix} 1 & -i & 0 \\ i & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \\
 \Omega_1 &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, & \Omega_4 &= \frac{1}{2} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}, & \Omega_7 &= \frac{1}{2} \begin{bmatrix} 1 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 1 \end{bmatrix}, & (2.4) \\
 \Omega_2 &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, & \Omega_5 &= \frac{1}{2} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}, & \Omega_8 &= \frac{1}{2} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & -i \\ 0 & i & 1 \end{bmatrix}.
 \end{aligned}$$

The advantage of this physical basis is that the elements correspond to what is normally most convenient in the lab.

For qubits, the physical basis states described above correspond to the pure states which are in some sense the 3 “standard” qubit bases: $\{|\mathbf{0}\rangle, |\mathbf{1}\rangle\}$, $\{|\pm\rangle \equiv |\mathbf{0}\rangle \pm |\mathbf{1}\rangle\}$, $\{|\pm i\rangle \equiv |\mathbf{0}\rangle \pm i|\mathbf{1}\rangle\}$ (ignoring normalisation constants). These three measurement bases define the axes of the *qubit sphere*³—a very useful representation of the state of a single qubit as a point lying on or inside a unit sphere in a three-dimensional phase space, $|\psi\rangle = \psi(x, y, z)$ [Fig. 2.1]. In each dimension, the position of the state represents the population imbalance

³The qubit sphere is more commonly known as the *Bloch sphere* in the context of spin- $\frac{1}{2}$ quantum systems, and the *Poincaré sphere* when discussing the polarisation of light. The Bloch terminology is perhaps more common in quantum information, but I have used a non-specific name to emphasise that this representation can be applied to any qubit system. There is only a slight cosmetic difference between the two forms of the sphere, which results from the conventions used to define the lines of “latitude” and “longitude” (or alternatively, the position of the “poles”). Here, in the logical representation, it is convenient to use the Bloch form of the sphere. Later, when appropriate, I will introduce the Poincaré form (see Ch. 4).

in that measurement basis.

$$\begin{aligned} x &= \langle +|\rho|+ \rangle - \langle -|\rho|- \rangle = \text{Tr} \{ \sigma_x \rho \} \\ y &= \langle +i|\rho|+i \rangle - \langle -i|\rho|-i \rangle = \text{Tr} \{ \sigma_y \rho \} \\ z &= \langle \mathbf{0}|\rho|\mathbf{0} \rangle - \langle \mathbf{1}|\rho|\mathbf{1} \rangle = \text{Tr} \{ \sigma_z \rho \} \end{aligned} \quad (2.5)$$

Here, the σ_j are the standard Pauli spin operators,

$$\sigma_x \equiv X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_y \equiv Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_z \equiv Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \quad (2.6)$$

Each of these coordinates varies from -1 to $+1$, and is further constrained by normalisation to satisfy $r^2 \equiv x^2 + y^2 + z^2 \leq 1$. Pure states lie on the surface of the sphere ($r^2 = 1$) and mixed states lie inside the sphere with the completely mixed state at the origin. So r^2 is a parameter which gives an indication of the purity of a single-qubit quantum state.

Another more general measure of purity can be defined directly in terms of the density matrix. Using the physicality constraints, it is straightforward to show that the density matrix must also satisfy $\text{Tr} \{ \rho^2 \} \leq 1$, and is equal to one only for pure states. The quantity $P = \text{Tr} \{ \rho^2 \}$ is called the *purity* of a quantum state. For a d -dimensional density matrix, the purity varies between $1/d$ and 1 , approaching 1 as the state becomes closer to a pure state. In the qubit case, the purity $P = \text{Tr} \{ \rho^2 \} = \frac{1}{2}r^2 + \frac{1}{2}$. This is easily shown by writing the density matrix in terms of the qubit sphere (rectangular) coordinates, i.e.,

$$\rho = \frac{1}{2}I + \underline{x} \cdot \underline{\sigma} = \frac{1}{2} \begin{bmatrix} 1+z & x-iy \\ x+iy & 1-z \end{bmatrix}, \quad (2.7)$$

where $\underline{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$. The density matrix can also be written explicitly in standard polar coordinates [Fig. 2.1(b)]:

$$\rho = \frac{1}{2}(1-r^2)I + |\psi'(r, \vartheta, \varphi)\rangle \langle \psi'(r, \vartheta, \varphi)|, \quad (2.8)$$

where $|\psi'(r, \vartheta, \varphi)\rangle = r \cos \frac{1}{2}(\frac{\pi}{2} - \varphi)|\mathbf{0}\rangle + e^{i\vartheta} r \sin \frac{1}{2}(\frac{\pi}{2} - \varphi)|\mathbf{1}\rangle$ is an unnormalised pure state.

2.1.2 Quantum operators

Operators arise in several different circumstances in quantum mechanics. I have already mentioned one example—formally, the density matrix ρ is just the matrix form of a quantum operator, the *density operator* $\hat{\rho}$, defined by $\rho = (\rho_{jk})$ where $\rho_{jk} = \langle \mathbf{j} | \hat{\rho} | \mathbf{k} \rangle$. However, in essence, the two terms are completely equivalent and I will use them interchangeably. In another example, an *observable*, which is any measurable physical property of a quantum state, can be represented by a Hermitian operator ($\hat{A}^\dagger = \hat{A}$). The final interpretation of an operator is something which actively “operates” to change a quantum state. For pure input states, $|\psi\rangle$, the operator acts to produce an output state $O(\psi) = \hat{O}|\psi\rangle$. Similarly, for mixed input states, ρ , the operator gives an output state $O(\rho) = \hat{O}\rho\hat{O}^\dagger$.

In formal notation, operators are distinguished from other variables with a “hat”, eg. \hat{A} . For simplicity, however, I will generally omit this as a distinction that will be clear from context.

Two important families of operators are the *unitary operators* and *projectors*.

$$\text{Unitary Operators, } \hat{U} : \hat{U}^\dagger \hat{U} = \hat{U} \hat{U}^\dagger = I \quad (2.9)$$

$$\text{Projectors, } \hat{P} : \hat{P} = \hat{P}^\dagger \ \& \ \hat{P}^2 = \hat{P} \quad (2.10)$$

Unitary operators

For an arbitrary quantum system, unitary operators have the following convenient (and interrelated) properties: (i) using the cyclic property of the trace, it is simple to show that they are trace-preserving, i.e.,

$$\text{Tr} \{U(\rho)\} = \text{Tr} \left\{ \hat{U} \rho \hat{U}^\dagger \right\} = \text{Tr} \left\{ \hat{U}^\dagger \hat{U} \rho \right\} = \text{Tr} \{ \rho \}; \quad (2.11)$$

(ii) a similar argument shows that the purity of the state is also conserved, $\text{Tr} \{U(\rho)^2\} = \text{Tr} \{\rho^2\}$, so the unitary operators do not change the mixture of a quantum state; and (iii) any unitary operation satisfies the formal definition of a rotation, because it preserves the inner product (or overlap) between vectors, i.e. $(U(\psi_1), U(\psi_2)) = \langle \psi_1 | U^\dagger U | \psi_2 \rangle = (\psi_1, \psi_2)$.

For a single qubit, a unitary operation performs a geometrical rotation of the qubit sphere.⁴ It can be represented by three parameters describing the axis and amount of rotation, and a global phase which is normally ignored. In standard 3-dimensional physical space, any arbitrary rotation can be decomposed into three successive rotations around two non-parallel axes, called *Euler rotations*. This is also true for single-qubit unitary operations ([1], p. 175).

$$\hat{U} = e^{i\alpha} R_{\hat{n}}(\theta) = e^{i\alpha} R_{\hat{p}}(\beta) R_{\hat{q}}(\gamma) R_{\hat{r}}(\delta), \quad (2.12)$$

where $R_{\hat{n}}(\theta) = \exp(-i\frac{\theta}{2}\hat{n} \cdot \underline{\sigma})$ describes a rotation of the qubit sphere by an angle θ around the axis defined by the unit vector \hat{n} . This means we can implement an arbitrary unitary operation provided we have the ability to perform controllable rotations around at least two non-parallel axes of the qubit sphere. Some special examples are the rotations around the \hat{x} , \hat{y} and \hat{z} axes of the Bloch sphere.

$$\begin{aligned} X_\theta &= e^{-i\frac{\theta}{2}X} = \cos \frac{\theta}{2} I - i \sin \frac{\theta}{2} X = \begin{bmatrix} \cos \frac{\theta}{2} & -i \sin \frac{\theta}{2} \\ -i \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{bmatrix} \\ Y_\theta &= e^{-i\frac{\theta}{2}Y} = \cos \frac{\theta}{2} I - i \sin \frac{\theta}{2} Y = \begin{bmatrix} \cos \frac{\theta}{2} & -\sin \frac{\theta}{2} \\ \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{bmatrix} \\ Z_\theta &= e^{-i\frac{\theta}{2}Z} = \cos \frac{\theta}{2} I - i \sin \frac{\theta}{2} Z = \begin{bmatrix} e^{-i\frac{\theta}{2}} & 0 \\ 0 & e^{i\frac{\theta}{2}} \end{bmatrix} \end{aligned} \quad (2.13)$$

(N.b. These equations can be verified using the fact that $\sigma_j^2 = I$.) I will occasionally refer to these special examples as *Bloch rotations*, and will discuss them in more detail in Ch. 4, particularly in relation to how they can be implemented in optics.

⁴This is not necessarily obvious from the formal definition of a rotation operation. As I have already mentioned, the purity of a single-qubit state is directly related to the distance of the state from the centre of the qubit sphere. Therefore, the condition that unitaries preserve purity immediately restricts them to either rotations or reflections of the qubit sphere. However, unitary operators always have two eigenstates (considering only pure states), whereas reflections have an infinite number of eigenstates (all states lying in the reflection plane). Therefore, single-qubit unitary operations must be a rotation of the qubit sphere. Reflections are not meaningful physical operations.

Projector operators

The action of the projector operators is to mathematically “project” a quantum state onto a smaller subspace of basis states, and therefore they behave physically like a state filter. For example, the projector $P_j = |\mathbf{j}\rangle\langle\mathbf{j}|$ only passes the component of the input state which is parallel to $|\mathbf{j}\rangle$.

Operator bases

I have already mentioned two ways of constructing a basis for operators [the elementary basis and the physical basis in Eq. (2.4)]. For single qubits, the Pauli operators $\{\sigma_0 = I, \sigma_1 = X, \sigma_2 = Y, \sigma_3 = Z\}$ also form a basis with several “nice” properties.⁵ The operators are:

1. traceless except for the identity — $\text{Tr}\{\sigma_0\} = 2; \text{Tr}\{\sigma_j\} = 0$;
2. Hermitian — $\sigma_j^\dagger = \sigma_j$;
3. unitary — $\sigma_j^\dagger\sigma_j = \sigma_j^2 = I$;
4. and orthogonal⁶ — $\text{Tr}\{\sigma_j^\dagger\sigma_k\} = 2\delta_{jk}$.

Conveniently, the Pauli operators are also related straightforwardly to the standard basis states:

$$|\mathbf{0}, \mathbf{1}\rangle = \frac{1}{2}(\sigma_0 \pm \sigma_3), \quad |\pm\rangle = \frac{1}{2}(\sigma_0 \pm \sigma_1), \quad |\pm i\rangle = \frac{1}{2}(\sigma_0 \pm \sigma_2). \quad (2.14)$$

Can these desirable characteristics be retained when generalising to higher dimensions? As it turns out, doing this is not as simple as one might expect, but I note here some alternatives which will be useful later.

The first method generalises the relationship between the Pauli operators and the standard qubit bases. Starting from the physical basis introduced earlier for d -dimensional density matrices, a complete, orthogonal operator basis can be constructed using the computational basis states, $|\mathbf{j}\rangle\langle\mathbf{j}|$, and traceless versions of the two-state superposition matrices, $\frac{1}{2}(|\mathbf{j}\rangle\langle\mathbf{k}| + |\mathbf{k}\rangle\langle\mathbf{j}|)$ and $\frac{1}{2}(-i|\mathbf{j}\rangle\langle\mathbf{k}| + i|\mathbf{k}\rangle\langle\mathbf{j}|)$. These operators are Hermitian and only the “computational basis” elements have non-zero trace. Their main disadvantage is that they are not unitary. For qutrits, the matrices are:

$$\begin{aligned} \Lambda_0 &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, & \Lambda_3 &= \frac{1}{2} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, & \Lambda_6 &= \frac{1}{2} \begin{bmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \\ \Lambda_1 &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, & \Lambda_4 &= \frac{1}{2} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, & \Lambda_7 &= \frac{1}{2} \begin{bmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{bmatrix}, & (2.15) \\ \Lambda_2 &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, & \Lambda_5 &= \frac{1}{2} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, & \Lambda_8 &= \frac{1}{2} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{bmatrix}. \end{aligned}$$

⁵Further discussion of “nice” operator bases can be found in Refs [2, 3].

⁶The inner product for operators—called the *Hilbert-Schmidt norm*—is $(A, B) = \text{Tr}\{A^\dagger B\}$.

Alternatively, an elegant example of an orthogonal basis of unitary operators in d dimensions is [4]

$$U_{kl} = X_d^k Z_d^l, \quad \dots \begin{cases} X_d |\mathbf{j}\rangle \equiv |\mathbf{j} \oplus \mathbf{1}\rangle \\ Z_d |\mathbf{j}\rangle \equiv e^{2\pi i \mathbf{j}/d} |\mathbf{j}\rangle \end{cases} \quad (2.16)$$

where \oplus represents addition modulo d . These operators satisfy the orthogonality condition $\text{Tr} \left\{ U_{kl}^\dagger U_{mn} \right\} = \delta_{km} \delta_{ln} d$, but they are not Hermitian. X_d and Z_d are called the *generators* of the basis, and they satisfy the further identity, $(X_d)^d = (Z_d)^d = I$.

Example: Calculating the partial trace the elegant way

An example of the usefulness of operator basis representations arises when trying to calculate the partial trace of a general (possibly mixed) quantum state⁷. Consider a bipartite system with orthonormal operator bases $\{a_j\}$ and $\{b_j\}$. The full state can be expanded in terms of these basis elements,

$$\rho_{ab} = \sum_{jk} \gamma_{jk} (a_j \otimes b_k), \quad (2.17)$$

where $\gamma_{jk} = \text{Tr} \{ [a_j \otimes b_k]^\dagger \rho_{ab} \}$. A similar technique can be used to calculate the reduced density matrix, ρ_a ,

$$\rho_a = \sum_j \alpha_j a_j. \quad (2.18)$$

The coefficients which define ρ_a can be calculated very easily from the original density matrix:

$$\alpha_j = \text{Tr}_a \left\{ a_j^\dagger \rho_a \right\}, \quad (2.19)$$

$$= \text{Tr}_a \left\{ a_j^\dagger \text{Tr}_b \{ \rho_{ab} \} \right\}, \quad (2.20)$$

$$= \text{Tr} \left\{ [a_j \otimes I_b]^\dagger \rho_{ab} \right\}. \quad (2.21)$$

2.1.3 Quantum processes

Not all processes in quantum mechanics can be described just by the action of an operator. The completely general process is in fact a much broader class, encompassing anything that can take any physical input state and produce a physical output state. Though they have many names⁸, I will refer to any such object as a *quantum process*, generally preferring this physically motivated term, or occasionally by the more mathematical alternative, *quantum operation*.

⁷Private communication from Alexei Gilchrist.

⁸There is a confusing morass of terminology in use throughout the literature. Processes are referred to variously as *quantum operations*, *superoperators* and *completely positive maps*, whereas the measurement technique for characterising them is called *quantum process tomography*. A completely positive map is the mathematical name for any object which always maps a positive operator to another positive operator, even when the map is only defined on a subspace of a larger quantum system (for more detail, see Ref. [1], p. 368).

Essentially, quantum processes are related to operators in the same way that mixed states are related to pure states, since they are able to describe stochastic processes that can introduce mixture into a system. There are three common and equivalent ways of describing these operations mathematically which I will now summarise.

The operator-sum representation

In the *operator-sum representation*, a quantum operation, \mathcal{E} , is represented by a set of operators, E_j , which together define the way it acts on an input state, ρ . The output state, $\mathcal{E}(\rho)$, can be calculated via

$$\mathcal{E}(\rho) = \sum_j E_j \rho E_j^\dagger, \quad (2.22)$$

where the *operation elements* must satisfy the inequality⁹ $\sum_j E_j^\dagger E_j \leq I$. Thus, a quantum process is represented by not one, but several operators, each acting in the usual way, and then combined in an incoherent sum, which is needed to explain processes which increase mixture.

The physical interpretation of this representation ([1], p. 362) is that a quantum operation with operation elements $\{E_j\}$ randomly applies the operation $E'_j \rho E_j'^\dagger$ (where $E'_j = E_j / \sqrt{\text{Tr}\{E_j \rho E_j^\dagger\}}$) to the input state, ρ , with a probability $p(j) = \text{Tr}\{E_j \rho E_j^\dagger\}$. In this sense, a process which can be described by a single non-zero operation element is “pure”.

The operator inequality is an important constraint, because it describes the information preserving properties of the process. As outlined above, a legitimate quantum process must produce a legitimate output state. It can be seen directly from the form of the operator-sum representation that the output state, $\mathcal{E}(\rho)$, will be Hermitian and positive. The operator inequality arises out of the remaining physicality constraint on the output state density matrix relating to its trace, i.e.,

$$\text{Tr}\{\mathcal{E}(\rho)\} = \text{Tr}\left\{\sum_j E_j \rho E_j^\dagger\right\}, \quad (2.23)$$

$$= \sum_j \text{Tr}\left\{E_j \rho E_j^\dagger\right\}, \quad (2.24)$$

$$= \sum_j \text{Tr}\left\{E_j^\dagger E_j \rho\right\}, \quad (2.25)$$

$$= \text{Tr}\left\{\left[\sum_j E_j^\dagger E_j\right] \rho\right\}. \quad (2.26)$$

Therefore, if $\text{Tr}\{\mathcal{E}(\rho)\} = \text{Tr}\{\rho\}$, then this implies that $\sum_j E_j^\dagger E_j = I$, and the operation is said to be *trace-preserving*. However, the constraint inequality also allows for the existence

⁹The operator inequality $A \leq B$ means that $B - A$ is a positive matrix. In particular, if $A \leq I$, then A is a Hermitian matrix with eigenvalues less than or equal to 1.

of *trace-decreasing* operations ($\sum_j E_j^\dagger E_j \leq I$). Physically, this corresponds to processes which discard some part of the quantum system, e.g. a polariser.

The main weakness of the operator-sum representation is that the operation elements describing a given process are not unique, analogous to the fact that there are many different ways of writing a mixed state as an ensemble of pure states. In fact, there is a unitary freedom in the choice of the operation elements. Two operations, \mathcal{E} and \mathcal{F} , are equivalent when the corresponding operators are related via $E_j = \sum_k u_{jk} F_k$, where $\tilde{u} = (u_{jk})$ is a unitary matrix.

The process matrix representation

One way to resolve the ambiguity of the unitary freedom in the operation elements [1, 5] is to describe the process using a fixed basis, $\{A_j\}$, which for convenience, can be orthonormal, i.e. $\text{Tr}\{A_j^\dagger A_k\} = \delta_{jk}$. Using this basis to expand the operation elements as $E_j = \sum_m a_{jm} A_m$, the output state can be rewritten, i.e.,

$$\mathcal{E}(\rho) = \sum_j \sum_{mn} a_{jm} a_{jn}^* A_m \rho A_n^\dagger, \quad (2.27)$$

$$= \sum_{mn} (\chi_{\mathcal{E}})_{mn} A_m \rho A_n^\dagger, \quad (2.28)$$

where $(\chi_{\mathcal{E}})_{mn} = \sum_j a_{jm} a_{jn}^*$ are the elements of the *process matrix*. This matrix completely describes the effect of the quantum process, but unlike the operator-sum representation, the χ matrix depends only on the choice of operator basis and not on the choice¹⁰ of E_j .

Consider the process matrix representations of a single quantum operation described in two different orthonormal operator bases,

$$\mathcal{E}(\rho) = \sum_{mn} (\chi_{\mathcal{E}}^A)_{mn} A_m \rho A_n^\dagger, \quad (2.29)$$

$$= \sum_{mn} (\chi_{\mathcal{E}}^B)_{mn} B_m \rho B_n^\dagger, \quad (2.30)$$

where the bases are related via $A_m = \sum_i \beta_{im} B_i$, which makes the coefficients $\beta_{im} = \text{Tr}\{B_i^\dagger A_m\}$. The different matrices are then related via the simple equation,

$$\chi_{\mathcal{E}}^B = \beta \chi_{\mathcal{E}}^A \beta^\dagger, \quad (2.31)$$

where $\beta = (\beta_{im})$ is a unitary matrix¹¹, which is very similar to the equation used to convert density matrices from one basis to another.

¹⁰Let E_j and F_j be two equivalent decompositions of the quantum process, \mathcal{E} , related by $E_j = \sum_k u_{jk} F_k$, and let $\chi_{\mathcal{E}}$ and $\chi_{\mathcal{F}}$ be the corresponding process matrices expanded in the operator basis, A_m . Then, since $E_j = \sum_m e_{jm} A_m$ and $E_j = \sum_k u_{jk} F_k = \sum_{km} u_{jk} f_{km} A_m$, it follows that $e_{jm} = \sum_k u_{jk} f_{km}$. Thus, because \tilde{u} is unitary, direct calculation shows that $\chi_{\mathcal{E}} = \chi_{\mathcal{F}}$.

¹¹Since $\{A_j\}$ and $\{B_j\}$ are orthonormal bases, $\delta_{jk} = \text{Tr}\{A_j^\dagger A_k\} = \sum_i \beta_{ik} \text{Tr}\{A_j^\dagger B_i\} = \sum_i \beta_{ik} \text{Tr}\{B_i^\dagger A_j\}^* = \sum_i \beta_{ik} \beta_{ij}^* = \sum_i \beta_{ji}^\dagger \beta_{ik} = (\beta^\dagger \beta)_{jk}$.

The trace-limiting inequality for the operation elements also translates into a generalised normalisation constraint for the process matrix. When re-expressed, it becomes $\sum_{mn} (\chi_{\mathcal{E}})_{nm} A_m^\dagger A_n \leq I$. It is important to note the transpose on $\chi_{\mathcal{E}}$.

An important property of the process matrix is its structural similarity to the density matrix for states. It follows directly from the definition that χ is Hermitian, and since there must also be an operator basis for which χ is diagonal, then χ must also be positive¹². Finally, the generalised normalisation constraint shows that $\text{Tr} \{ \chi_{\mathcal{E}} \} \leq d$, where d is the dimension of the Hilbert space. Although this has the same form as the normalisation requirement for quantum states, it also incorporates other more complex constraints, so the process matrices are actually a more restricted set than the density matrices.

A physical interpretation of the process matrix can be obtained by carefully examining its operational definition in Eq. (2.28). The diagonal elements of the process matrix, $(\chi_{\mathcal{E}})_{jj}$, determine the probabilities that the process will perform simple basis operations of the form $A_j \rho A_j^\dagger$. The off-diagonal elements then describe the coherence between these basis operations, i.e. the amount of coherence that is maintained in the quantum state which is subject to the different possible operations. So a process matrix that looks like a mixed quantum state will introduce mixture into the output state, while one that looks like a pure state can always be written as a single diagonal element in an appropriate operator basis. This notion can be formalised by defining the *process purity*, $P_{\text{pro}}(\mathcal{E}) = \text{Tr} \{ \rho_{\mathcal{E}}^2 \}$, which I discuss in detail in Ch. 10.

The Jamiolkowski representation

This striking similarity between the form of process and state matrices is the key to the third main way of describing quantum processes, called the *Jamiolkowski isomorphism* [6, 7]. Closely related to the process matrix representation, though more abstract, it maps a quantum operation, \mathcal{E} , onto a quantum state, $\rho_{\mathcal{E}}$, through the relationship,

$$\rho_{\mathcal{E}} \equiv [\mathcal{I} \otimes \mathcal{E}] (|\Phi\rangle\langle\Phi|), \quad (2.32)$$

where $|\Phi\rangle = \frac{1}{\sqrt{d}} \sum_j |j\rangle_a |j\rangle_b$ is a quantum state which is maximally entangled between two copies of the (d -dimensional) system, and $\{|j\rangle\}$ is an orthonormal basis set. This representation is also a complete description of the quantum process, and therefore invertible. In particular, the output state can be directly calculated via the matrix equation,

$$\mathcal{E}(\rho) = \text{Tr}_a \{ (\rho^T \otimes I) \rho_{\mathcal{E}} \}. \quad (2.33)$$

The generalised normalisation constraint for processes can also be expressed in terms of the Jamiolkowski matrix via the inequality,

$$\text{Tr}_b \{ \rho_{\mathcal{E}} \} \leq I/d, \quad (2.34)$$

which can be verified by direct calculation using the definition of \mathcal{E} and normalisation constraints for the operator-sum representation. Note again that not all density matrices represent physical processes—i.e. the set of allowed quantum processes on a qudit Hilbert space is more restricted than the set of allowed two-qudit quantum states.

¹²If χ is diagonal in the orthonormal basis D_j , then $\lambda_j = (\chi_{\mathcal{E}})_{jj} = \sum_k |d_{kj}|^2 \geq 0$.

Since $\chi_{\mathcal{E}}$ and $\rho_{\mathcal{E}}$ both have similar properties (i.e. the features of a density matrix) and contain a complete description of the quantum process \mathcal{E} , it is probably not surprising to find that they are very closely related. In fact, by choosing the appropriate operator basis sets $\{A_j\} = \{|m\rangle\langle n|\}$ [the elementary basis, see Sec. 2.1.1], it can be shown¹³ that $\chi_{\mathcal{E}} = d\rho_{\mathcal{E}}$. Therefore, I will refer to both as the process matrix, and treat them interchangeably. This is an important connection, because the most convenient representation can be used to solve a problem, but the same result will be true for both forms of the process matrix. In practice, $\rho_{\mathcal{E}}$ is often easier to work with mathematically, while the definition of $\chi_{\mathcal{E}}$ gives its elements an obvious physical significance.

To illustrate the value of this connection, I will use it to prove that $\chi_{\mathcal{E}\otimes\mathcal{F}} = \chi_{\mathcal{E}} \otimes \chi_{\mathcal{F}}$. This is a very useful property of process matrices and allows complex quantum tasks to be analysis more simply by splitting them up into their components [see Ch. 9]. Consider two processes, \mathcal{E} and \mathcal{F} , acting on $d_{\mathcal{E}}$ - and $d_{\mathcal{F}}$ -dimensional quantum systems. Using the Jamiolkowski form of the process matrix,

$$\rho_{\mathcal{E}} \otimes \rho_{\mathcal{F}} = [\mathcal{I}_{\mathcal{E}} \otimes \mathcal{E}] (|\Phi^{(d_{\mathcal{E}})}\rangle\langle\Phi^{(d_{\mathcal{E}})}|) \otimes [\mathcal{I}_{\mathcal{F}} \otimes \mathcal{F}] (|\Phi^{(d_{\mathcal{F}})}\rangle\langle\Phi^{(d_{\mathcal{F}})}|), \quad (2.35)$$

$$= [\mathcal{I}_{\mathcal{E}} \otimes \mathcal{E} \otimes \mathcal{I}_{\mathcal{F}} \otimes \mathcal{F}] (|\Phi^{(d_{\mathcal{E}})}, \Phi^{(d_{\mathcal{F}})}\rangle\langle\Phi^{(d_{\mathcal{E}})}, \Phi^{(d_{\mathcal{F}})}|), \quad (2.36)$$

$$= \frac{1}{d_{\mathcal{E}}d_{\mathcal{F}}} \sum_{j,k=0}^{d_{\mathcal{E}}-1} \sum_{l,m=0}^{d_{\mathcal{F}}-1} [\mathcal{I}_{\mathcal{E}} \otimes \mathcal{E} \otimes \mathcal{I}_{\mathcal{F}} \otimes \mathcal{F}] (|j, j, l, l\rangle\langle k, k, m, m|), \quad (2.37)$$

$$= \frac{1}{d_{\mathcal{E}}d_{\mathcal{F}}} \sum_{j,k=0}^{d_{\mathcal{E}}-1} \sum_{l,m=0}^{d_{\mathcal{F}}-1} [\mathcal{I}_{\mathcal{E}} \otimes \mathcal{I}_{\mathcal{F}} \otimes \mathcal{E} \otimes \mathcal{F}] (|j, l, j, l\rangle\langle k, m, k, m|), \quad (2.38)$$

$$= [\mathcal{I}_{\mathcal{E}\otimes\mathcal{F}} \otimes \mathcal{E} \otimes \mathcal{F}] (|\Phi^{(d_{\mathcal{E}}d_{\mathcal{F}})}\rangle\langle\Phi^{(d_{\mathcal{E}}d_{\mathcal{F}})}|), \quad (2.39)$$

$$= \rho_{\mathcal{E}\otimes\mathcal{F}}. \quad (2.40)$$

The mathematical equivalence of quantum processes and states has profound implications for dealing with processes in quantum information. An extensive set of tools has been developed for use with quantum states, and because of the isomorphic relationship, these same tools can generally be mapped directly across to problems involving processes. Often, the main challenge is then trying to interpret these results. I will return to these ideas for a detailed discussion in Ch. 9.

2.1.4 The mixed-state fidelity

In quantum information, it is often necessary to determine how close two quantum states are to one another. There are many different ways to approach this problem, each with its own particular merits and faults [see Sec. 9.2 for more detail]. Possibly the most

¹³We first pointed out this fact in Ref. [8]. It can be proved via a direct calculation outlined as follows. Consider a process, $\mathcal{E}(\rho)$, and its corresponding process matrix, $\chi_{\mathcal{E}}$ —defined via $\mathcal{E}(\rho) = \sum_{jk} (\chi_{\mathcal{E}})_{jk} c_j \rho c_k^\dagger$ —and Jamiolkowski form, $\rho_{\mathcal{E}} = [\mathcal{I} \otimes \mathcal{E}] (|\Phi\rangle\langle\Phi|)$. Define the process basis, $\{c_j\}$, to be the elementary basis (i.e. $c_j = e_{l_j, m_j} = |l_j\rangle\langle m_j|$) given in the order $\{|\mathbf{0}\rangle\langle\mathbf{0}|, \dots, |\mathbf{d}-1\rangle\langle\mathbf{0}|, |\mathbf{0}\rangle\langle\mathbf{1}|, \dots, |\mathbf{d}-1\rangle\langle\mathbf{d}-1|\}$, and note that $[\mathcal{I} \otimes c_j] |\Phi\rangle = \frac{1}{\sqrt{d}} |m_j, l_j\rangle$. Then, using the standard order for a tensor product state basis, $\{|\mathbf{0}, \mathbf{0}\rangle, \dots, |\mathbf{0}, \mathbf{d}-1\rangle, |\mathbf{1}, \mathbf{0}\rangle, \dots, |\mathbf{d}-1, \mathbf{d}-1\rangle\}$, and substituting directly into the Jamiolkowski form, it is easy to show that $(\rho_{\mathcal{E}})_{jk} = \frac{1}{d} (\chi_{\mathcal{E}})_{jk}$, as required.

prolific quantitative measure for comparing states is derived from the overlap between two quantum states¹⁴, the exact analogy of the scalar product for geometric vectors. However, even this is complicated by the fact that at least two different definitions are used regularly throughout the literature.

I will define the *fidelity* between two pure quantum states to be $F(\psi_1, \psi_2) = |\langle \psi_1 | \psi_2 \rangle|^2$. Defined in this way, the fidelity provides the answer to the following question: “given a state $|\psi_2\rangle$, what is the probability of measuring the system in the state $|\psi_1\rangle$?” There does not seem to be any such compelling reason to use the alternative definition which is just the square root of this quantity.

The main complication with the fidelity arises when considering mixed states. The above definition can be rewritten $F(\psi_1, \psi_2) = \text{Tr} \{ |\psi_1\rangle\langle\psi_1| |\psi_2\rangle\langle\psi_2| \}$. However, if replaced by¹⁵ $F(\rho_1, \rho_2) = \text{Tr} \{ \rho_1 \rho_2 \}$, then the result is less than one for two identical mixed states, since $\text{Tr} \{ \rho_1^2 \} = 1$ only when ρ_1 is pure. This is clearly inadequate, since for any reasonable way of comparing two quantum states, by definition the result should be 1 when the two states are the same.

While not at all obvious at first glance, it turns out that the appropriate generalisation of the fidelity for mixed states is $F(\rho_1, \rho_2) = \text{Tr} \{ \sqrt{\sqrt{\rho_1} \rho_2 \sqrt{\rho_1}} \}^2$ [9]. This has the important properties that $F(\rho_1, \rho_2) = F(\rho_2, \rho_1)$, $F(\rho_1, \rho_1) = 1$ and, using the fact that $(|\psi\rangle\langle\psi|)^2 = |\psi\rangle\langle\psi|$, it is easy to see that it reduces to the original definition for two pure states.

Despite these features, however, it is not clear that the fidelity has a good physical interpretation when both states are mixed. The best current interpretation is only valid when at least one state is pure. In this case, $F(\psi_1, \rho_2) = \langle \psi_1 | \rho_2 | \psi_1 \rangle = \text{Tr} \{ |\psi_1\rangle\langle\psi_1| \rho_2 \}$, which is once again in the original form. It then answers the question: “given a mixed state ρ_2 , what is the probability of measuring the state $|\psi_1\rangle$?”

2.1.5 Quantum measurement theory

They walked on, thinking of This and That, and by-and-by they came to an enchanted place on the very top of the Forest called Galleons Lap, which is sixty-something trees in a circle; and Christopher Robin knew that it was enchanted because nobody had ever been able to count whether it was sixty-three or sixty-four, not even when he tied a piece of string round each tree after he had counted it.

The House At Pooh Corner, A. A. Milne

The majority of this thesis is devoted to investigating ways of encoding and manipulating quantum information in physical systems; making measurements is a vital part of this work. At the simplest level, the end of any information processing task will involve a

¹⁴The overlap or *inner product* between two quantum states is $\langle \psi_1 | \psi_2 \rangle = \sum_j \alpha_{1j}^* \alpha_{2j}$ for $|\psi\rangle = \sum_j \alpha_j |\mathbf{j}\rangle$.

¹⁵Not surprisingly, this is very closely related to the inner product for operators, because $\rho^\dagger = \rho$.

measurement on the system to extract an answer. Even more importantly, any attempt to carry out such a task in a physical system will require the ability to measure, characterise and analyse success, and if necessary, to determine how to improve the outcome. This section reviews the fundamental process of measurement in quantum mechanics (more rigorous detail can be found in Ref. [1] beginning on p. 84). Working out how to extract useful information from the measurement results is a different challenge entirely and is a major focus of all the research reported in this thesis.

In quantum mechanics, a measurement is a process with a range of possible outcomes $\{m\}$, each of which has an associated measurement operator M_m and occurs with a probability $p(m)$. The measurement operator describes the effect of the measurement itself on the state of the system, making the new state

$$\rho' = \frac{M_m \rho M_m^\dagger}{p(m)}, \quad (2.41)$$

where $p(m) = \text{Tr} \{M_m \rho M_m^\dagger\}$. The denominator is included to ensure the final state is still normalised, i.e. $\text{Tr} \{\rho'\} = 1$. Since there is definitely some result no matter what the input state is ($\sum_m p(m) = 1$), it follows that $\sum_m M_m^\dagger M_m = I$, which is the *completeness relation* for measurement operators.

Normally, when attempting to measure a system capable of exhibiting quantum features such as those required for quantum computing, this process would need to couple the system with a *meter* device from which the result can be read, such as a photon counter.

Probably the easiest measurements to implement in an experiment are *projective measurements*. In terms of the general theory outlined above, projective measurements are a special case where the measurement operators are *orthogonal projectors*. In other words, they are Hermitian (like all projectors), and they must satisfy the condition $M_{m'} M_m = \delta_{m,m'} M_m$. Defining $M_m \equiv P_m$, then $p(m) = \text{Tr} \{P_m \rho\}$ and $\sum_m P_m = I$, and the state after the measurement is

$$\rho' = \frac{P_m \rho P_m}{p(m)}. \quad (2.42)$$

In the laboratory, the most common measurement is a projection onto a complete set of basis states for the system. I will refer to such a set as a *complete measurement set* or a *POVM set*¹⁶. In this case, the measurement operators are $P_m = |m\rangle\langle m|$. However, this implies that the measurement apparatus actively detects all outcomes simultaneously, whereas for practical reasons, this is not always the case. Often, the technique used is to run identical experiments many times, detecting only one outcome at a time. This requires the assumption that the experiment does not change from shot to shot, and it may take a great deal of effort to ensure this is a good approximation. This does not normally introduce an extra experimental challenge, however, because often this effort is already necessary. For example, performing the tomographic techniques required to thoroughly characterise an experiment, already requires many experimental runs to estimate measurement probabilities.

¹⁶This term arises in relation to the POVM formalism of measurement theory (*Positive Operator-Valued Measure*).

Because a set of projectors, $\{P_m\}$, must satisfy an orthogonality condition, they can be used to define a Hermitian operator, $M = \sum_m mP_m$, which is therefore an observable of the measured system. The projective measurement is a projection onto the eigenspace of this observable, with the outcomes of the measurement being the eigenvalues. Conversely, *any* observable can be expanded in such a way using its eigenvalues and eigenvectors. Because this matrix is “diagonal” with respect to the basis of eigenvectors, it is then said to be *diagonalised*.

Using this form, there is a simple way to calculate the average value of an observable’s measurement outcomes, written $\langle M \rangle$, using basic rules of probability.

$$\langle M \rangle = \sum_m mp(m) \quad (2.43)$$

$$= \sum_m m \text{Tr} \{P_m \rho\} \quad (2.44)$$

$$= \text{Tr} \{M \rho\} \quad (2.45)$$

Using this, the spread in the measurement results, $\Delta(M)$, can be calculated from the *variance*

$$\Delta(M)^2 = \langle (M - \langle M \rangle)^2 \rangle = \langle M^2 \rangle - \langle M \rangle^2. \quad (2.46)$$

2.1.6 Quantum correlations and entanglement

Alice laughed. “There’s no use trying,” she said: “one *can’t* believe impossible things.”

“I dare say you haven’t had much practice,” said the Queen. “When I was your age I always did it for half an hour a day. Why, sometimes I’ve believed as many as six impossible things before breakfast.”

Through the Looking-Glass (and what Alice found there),
Lewis Carroll

Entanglement is one of the most distinctive features of quantum mechanics, and one of the most important physical quantities in quantum information and quantum computing. But what measurements are required to prove that a system is entangled, and if it is entangled, to determine how strongly? By definition, entanglement is possessed by a joint quantum system if it is “non-separable”—i.e. its individual components do not carry definite state information if not considered in conjunction with the rest of the system. When looking at the components individually, the entanglement looks instead like mixture¹⁷.

For simplicity, consider a bipartite system ($\mathcal{H} = \mathcal{H}^A \otimes \mathcal{H}^B$). A pure state is *separable* if it can be written as a tensor product of the sub-system states, $|\psi_{\text{sep}}\rangle = |\psi^A\rangle \otimes |\psi^B\rangle$. A mixed state is separable if it can be written as a convex sum of separable states,

¹⁷In fact, mathematically, mixture in a system can be described as entanglement with some ancillary system which is then discarded; and physically, mixture arises when a system couples to the environment (by definition, unmeasured and inaccessible).

$\rho_{\text{sep}} = \sum_j p_j \rho_j^A \otimes \rho_j^B$, where $p_j \geq 0$ and $\sum_j p_j = 1$ [10]. A quantum system is *entangled* if it is *inseparable* under these criteria. However, to quantify the “amount” of entanglement possessed by a system, not only is it necessary to prove that the state is *not* separable, but also *how* “not separable” it is.

The simplest examples of entangled states are the maximally entangled Bell states for two qubits—the smallest system capable of exhibiting entanglement:

$$\begin{aligned} |\Phi^\pm\rangle &= \frac{1}{\sqrt{2}} (|\mathbf{00}\rangle \pm |\mathbf{11}\rangle), \\ |\Psi^\pm\rangle &= \frac{1}{\sqrt{2}} (|\mathbf{01}\rangle \pm |\mathbf{10}\rangle). \end{aligned} \quad (2.47)$$

As the smallest possible, completely entangled states, Bell pairs are often used as the fundamental unit of entanglement in quantum information theory. A system of two qudits can possess more entanglement than a Bell pair, because it can exhibit correlations over a larger “alphabet” of basis states. The generalised family of maximally entangled qudits is

$$|\Phi_{pq}^{(d)}\rangle = \frac{1}{\sqrt{d}} \sum_j e^{ijq2\pi/d} |\mathbf{j}, \mathbf{j} \oplus \mathbf{p}\rangle, \quad 0 \leq p, q \leq (d-1) \quad (2.48)$$

where \oplus indicates addition modulo d . Another way of storing larger amounts of entanglement is to consider systems entangled across more than two particles. In this case, however, it is not as straightforward to generalise the maximally entangled states as with bipartite qudit systems. For example, with multiple qubits, two fundamentally different types of entangled states are the GHZ- or parity-entangled states and the W states [11, 12];

$$\begin{aligned} |\psi_{\text{GHZ}}^{(3)}\rangle &= \frac{1}{\sqrt{2}} (|\mathbf{000}\rangle + |\mathbf{111}\rangle), \\ |\psi_{\text{W}}^{(3)}\rangle &= \frac{1}{\sqrt{3}} (|\mathbf{001}\rangle + |\mathbf{010}\rangle + |\mathbf{100}\rangle), \end{aligned} \quad (2.49)$$

Wherever entanglement is sought, and whatever its desired purpose may be, it would be extremely useful to have a sensible, consistent way of characterising and quantifying it. Unfortunately, this is a very difficult problem, and despite being a continuously active area of research, it remains unsolved. There is still no general, all-purpose measure of entanglement which is both meaningful and convenient to measure and calculate. The main problems arise when trying to calculate the entanglement of mixed states, multi-particle systems, and qudit-based systems. In this thesis, I only consider bi-partite systems, and I have used three different types of entanglement measures which are outlined below. Note, however, that the following discussion is not intended to be an extensive survey of entanglement measures, but rather a brief description of those that I have used. A comprehensive, conceptual overview of the general theory of entanglement measures can be found in Ref. [13].

CHSH Bell violations

The Bell inequality as formulated by Clauser, Horne, Shimony and Holt (CHSH) places constraints on the measurable correlations between two systems which can be achieved in a locally realistic theory, and they show that these can be violated by quantum mechanics

[14]. In the standard description, the CHSH Bell inequality states that $|S| \leq 2$ for all locally realistic theories, where the Bell parameter is defined by

$$S = C(a, b) - C(a, b^\perp) + C(a', b) + C(a', b^\perp), \quad (2.50)$$

where $a, a', b,$ and b' are measurements in different bases, and $C(a, b)$ is the correlation function within the “ a, b ” basis. These can be calculated from measured probabilities or count rates ($\mathcal{P} = \mathcal{N}/\overline{\mathcal{N}}$):

$$C(a, b) = \mathcal{P}(a, b) - \mathcal{P}(a, b^\perp) - \mathcal{P}(a^\perp, b) + \mathcal{P}(a^\perp, b^\perp), \quad (2.51)$$

$$= \frac{\mathcal{N}(a, b) - \mathcal{N}(a, b^\perp) - \mathcal{N}(a^\perp, b) + \mathcal{N}(a^\perp, b^\perp)}{\mathcal{N}(a, b) + \mathcal{N}(a, b^\perp) + \mathcal{N}(a^\perp, b) + \mathcal{N}(a^\perp, b^\perp)}. \quad (2.52)$$

According to quantum mechanics, however, this limit can be violated up to a maximum $S = 2\sqrt{2}$.

A simple state which achieves this maximum is the Φ^+ Bell state. By noting that

$$|\Phi^+\rangle \sim |\theta\theta\rangle + |\theta^\perp\theta^\perp\rangle \quad (2.53)$$

for all linear superposition bases ($|\theta\rangle = \cos\theta|\mathbf{0}\rangle + \sin\theta|\mathbf{1}\rangle$ and $|\theta^\perp\rangle = \sin\theta|\mathbf{0}\rangle - \cos\theta|\mathbf{1}\rangle$), it is straightforward to show that $\mathcal{P}(\theta, \phi) = \frac{1}{2}\cos^2(\phi - \theta)$, and consequently $C(\theta, \phi) = \cos 2(\phi - \theta)$. Therefore, by choosing the settings, $a=0, a'=\pi/4, b=\pi/8,$ and $b'=3\pi/8$, this gives $S = 4 \cos \pi/4 = 2\sqrt{2}$.

When performing experiments, it is useful to note that there is a unitary freedom in the choice of measurement settings that will achieve the maximum Bell violation. In particular, using the identity that $[I \otimes U]|\Phi^+\rangle = [U^T \otimes I]|\Phi^+\rangle$, it is easy to see that $P(U^*|a, U|b) = P(a, b)$. So if one set of $\{a, a', b, b'\}$ gives maximal correlations with the Φ^+ Bell state, then so will all sets of the form $\{U^*a, U^*a', Ub, Ub'\}$. Moreover, since all maximally entangled states are related to Φ^+ via $|\psi_{\max\text{-ent}}\rangle = [I \otimes V]|\Phi^+\rangle$, it is easy to find an optimal measurement set for an arbitrary entangled state.

The Bell violation test serves as an indicator of entanglement, since no separable quantum state can violate Bell’s inequality. In this sense, a Bell inequality is a form of *entanglement witness* [13, 15, 16]. However, although all entangled pure states violate a Bell inequality, introducing mixture causes problems. In fact, mixed states can be quite entangled before they are able to violate a Bell inequality [17]. In other words, a Bell violation is a sufficient, but not necessary, condition of inseparability, and thus does not provide a good quantitative measure of entanglement.

Entanglement of formation, concurrence and tangle

One of the key signatures of an entangled system is that when the reduced state of one subsystem is considered in isolation, it appears to be mixed, even if the state of the combined system is pure. In fact, this is the motivation for the *entropy of entanglement*, $E(\psi)$, which is the appropriate measure of entanglement for pure states:

$$E(\psi) \equiv S(\rho_A) = -\text{Tr} \{ \rho_A \log \rho_A \}, \quad (2.54)$$

where $\rho_A = \text{Tr}_B \{\psi\}$ is the reduced density matrix of subsystem A . This is used to define the *entanglement of formation*, $E_F(\rho)$, for mixed states [18, 19]:

$$E_F(\rho) \equiv \min \sum_j p_j E(\psi_j), \quad (2.55)$$

which is minimised over all pure-state decompositions of the mixed state, $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|$.

The entanglement of formation, $E_F(\rho)$, represents the minimum amount of entanglement (in *ebits*, the number of maximally entangled Bell states) required to create the state ρ . Although calculating this directly requires a non-trivial numerical optimisation, there is a closed-form analytical solution for the special case of two qubits. It was shown in Ref. [20] that

$$E_F(\rho) = h\left(\frac{1}{2} + \frac{1}{2}\sqrt{1 + C^2(\rho)}\right), \quad (2.56)$$

where $h(x) = -x \log x - (1-x) \log(1-x)$, and $C(\rho) = \max\{0, \lambda_1 - \lambda_2 - \lambda_3 - \lambda_4\}$ is the *concurrence* of the state. In the concurrence, the λ_j are the eigenvalues (in decreasing order) of the matrix $R \equiv \sqrt{\sqrt{\rho}\tilde{\rho}\sqrt{\rho}}$, where $\tilde{\rho}$ is the ‘‘spin-flipped’’ density matrix, $\tilde{\rho} = (\sigma_y \otimes \sigma_y)\rho^*(\sigma_y \otimes \sigma_y)$. The *tangle*, $T(\rho) \equiv C^2(\rho)$ and the concurrence are both useful entanglement measures in their own right, although their physical meaning is only obtained from their connection to the entanglement of formation. The tangle, however, is also a conserved measure of the entanglement shared between multiple qubits [21]. In this thesis, I use the tangle to describe the entanglement of two-qubit states.

It turns out that it is quite difficult to generalise the concurrence (which provides a closed-form solution of E_F) to higher dimensions [13, 22]. As a result, the entanglement of formation for higher-dimensional systems can only be calculated numerically (e.g. using an algorithm like that given in Ref. [23])—and the optimisation is difficult.

Entanglement of distillation and negativity

The *entanglement of distillation* or *distillable entanglement*, $E_D(\rho)$, is defined to be the maximum amount of entanglement (again, in *ebits*) that can be distilled from ρ using only *local operations and classical communication* (LOCC) [24]:

$$E_D(\rho) \equiv \max \lim_{n \rightarrow \infty} \frac{m}{n}, \quad (2.57)$$

where n is the number of copies of ρ used, and m is the number of pure, maximally entangled Bell pairs obtained, maximised over all possible LOCC protocols. Importantly, $E_D(\rho)$ is not necessarily the same as $E_F(\rho)$ —in fact, $E_D(\rho) \leq E_F(\rho)$, in general, and $E_D(\psi) = E_F(\psi) = E(\psi)$ for bi-partite pure states [13]. Thus, it is possible to create states with entanglement that cannot be ‘‘used’’—this is called *bound entanglement* [10, 24–27]. Unfortunately, the entanglement of distillation, though a very important physical quantity, is extremely difficult to calculate, except in very special cases.

The Peres criterion for separability says that for a separable state, the partial transpose of its density matrix is positive semi-definite, i.e. it is also a legitimate state [28]. In other words, if the partial transpose of a state is *negative*, then the state is definitely *not*

separable. The *negativity*, $N(\rho)$, is a measure of entanglement which seeks to quantify this notion, and is defined as follows [29–31].

$$N(\rho) \equiv \|\rho^{TB}\| - 1 = 2 \max\{0, -\lambda_{\text{neg}}\}, \quad (2.58)$$

where $\|X\| \equiv \text{Tr} \left\{ \sqrt{X^\dagger X} \right\}$, and λ_{neg} is the sum of the negative eigenvalues of ρ^{TB} .

The negativity is particularly useful because it is easy to calculate for all bi-partite quantum states, and because it can be related to the entanglement of distillation. In fact, $N(\rho) = 0$ is a necessary and sufficient condition for $E_D(\rho) = 0$, and in general, $N(\rho)$ can be used to place upper bounds on $E_D(\rho)$ [13]. Also, for two-qubit pure states, the negativity is equal to the concurrence [32].

2.2 Quantum information in optics

2.2.1 Fields and photons

In quantum optics, modes of the electromagnetic field are described by field operators, a_j , the quantum equivalents of the complex electric field amplitude. They are populated by discrete quanta of energy in the manner of harmonic oscillators—a single excitation of the field mode is called a *photon*. The operators, a_j , and their Hermitian conjugates, a_j^\dagger , are called the *annihilation* and *creation* operators respectively, characterised by their effect on the number of energy quanta (photons) occupying the mode, either decreasing or increasing it by one.

$$\begin{aligned} a_j |n_j\rangle &= \sqrt{n_j} |n_j - 1\rangle \\ a_j^\dagger |n_j\rangle &= \sqrt{n_j + 1} |n_j + 1\rangle \end{aligned} \quad (2.59)$$

Here, $|n_j\rangle$ is the quantum state with n_j photons in mode j , and the annihilation operator satisfies the additional requirement that $a_j |0\rangle = 0$, where $|0\rangle$ is the zero-photon or *vacuum* state (as distinct from $|\mathbf{0}\rangle$, the logical state). Several useful identities can be derived from these definitions. The operator commutation relations are

$$\left[a_j, a_k^\dagger \right] = \delta_{jk} \hat{1}, \quad \forall j, k. \quad (2.60)$$

It is also easy to see that the number states, $|n_j\rangle$, are eigenstates of the number operator, $N_j = a_j^\dagger a_j$, with eigenvalues n_j . Finally, the number states can be written in terms of the creation operators and the vacuum state:

$$|n_j\rangle = \frac{1}{\sqrt{n!}} \left(a_j^\dagger \right)^n |0\rangle. \quad (2.61)$$

The observables for the energy (the Hamiltonian) and the momentum of the field modes are

$$H = \sum_j \hbar \omega_j \left(a_j^\dagger a_j + \frac{1}{2} \right), \quad (2.62)$$

$$\mathbf{G} = \sum_j \hbar \mathbf{k}_j a_j^\dagger a_j, \quad (2.63)$$

where $\hbar\omega$ and $\hbar\mathbf{k}$ are the energy and momentum of a single photon with angular frequency ω and wave vector \mathbf{k} . Throughout this thesis, I am only interested in changes in the energy of the electromagnetic field (i.e. processes which change the number photons in a mode, such as photon detection and down-conversion), so I will ignore the zero-point energy ($\frac{1}{2}\hbar\omega_j$) in the Hamiltonian. As the system evolves, this only produces an unmeasurable global phase in the photon wave-function.

In the classical limit, many electromagnetic fields, like lasers, are well-described by *coherent states*, $|\alpha_j\rangle$, which are defined to be eigenstates of the annihilation operator a_j , where the eigenvalue α_j is the classical field amplitude.

$$a_j|\alpha_j\rangle = \alpha_j|\alpha_j\rangle \quad (2.64)$$

It is important to note that $|\alpha\rangle$ is *not* an eigenstate of the creation operator¹⁸ a_j^\dagger . To satisfy the underlying quantum principles, it must also be possible to expand the coherent state as a sum over number states, $|\alpha\rangle = \sum_n c_n(\alpha)|n\rangle$. Using the above equation gives rise to a recurrence relation for the amplitude coefficients.

$$\sqrt{n+1}c_{n+1} = \alpha c_n \quad (2.65)$$

$$c_{n+1} = \frac{\alpha}{\sqrt{n+1}}c_n \quad (2.66)$$

$$\Rightarrow c_n = \frac{\alpha^n}{\sqrt{n!}}c_0 \quad (2.67)$$

To ensure that the coherent state is normalised, $\langle\alpha|\alpha\rangle = 1$, the coefficients must also satisfy the condition,

$$1 = \sum_n |c_n|^2 = \sum_n \frac{|\alpha|^{2n}}{n!} |c_0|^2 = e^{|\alpha|^2} |c_0|^2. \quad (2.68)$$

Therefore,

$$|\alpha\rangle = e^{-|\alpha|^2/2} \sum_n \frac{\alpha^n}{\sqrt{n!}} |n\rangle. \quad (2.69)$$

Coherent states can be generated from the vacuum by a displacement operator, $D(\alpha)$:

$$\begin{aligned} |\alpha\rangle &= D(\alpha)|0\rangle, \\ D(\alpha) &= e^{\alpha a^\dagger - \alpha^* a}. \end{aligned} \quad (2.70)$$

This can be seen as follows.

$$D(\alpha)|0\rangle = e^{\alpha a^\dagger - \alpha^* a}|0\rangle \quad (2.71)$$

$$= e^{\alpha a^\dagger} e^{-\alpha^* a} e^{\frac{1}{2}|\alpha|^2 [a^\dagger, a]} |0\rangle \quad (2.72)$$

$$= e^{-|\alpha|^2/2} e^{\alpha a^\dagger} e^{-\alpha^* a} |0\rangle \quad (2.73)$$

$$= e^{-|\alpha|^2/2} e^{\alpha a^\dagger} |0\rangle \quad (2.74)$$

$$= e^{-|\alpha|^2/2} \sum_n \frac{\alpha^n (a^\dagger)^n}{n!} |0\rangle \quad (2.75)$$

$$= e^{-|\alpha|^2/2} \sum_n \frac{\alpha^n}{\sqrt{n!}} |n\rangle \quad (2.76)$$

$$= |\alpha\rangle \quad (2.77)$$

¹⁸In fact, $a_j^\dagger|\alpha_j\rangle$ is an extremely non-classical state.

Here we have made use of a special case of the Baker-Hausdorff theorem from group theory, i.e. for two non-commuting operators, A and B ,

$$e^{A+B} = e^A e^B e^{-\frac{1}{2}[A,B]}, \quad (2.78)$$

when $[A, [A, B]] = [B, [A, B]] = 0$.

2.2.2 Different pictures of evolution

In any formalism of the basic principles of quantum mechanics, such as may be found in a standard undergraduate text, one vital component is a description of the evolution of a quantum system. An example of such a description is the quantum process model introduced in Sec. 2.1.3. However, in different contexts, there have evolved many different approaches to this problem, depending on what is convenient for the requirements of individual situations, and these approaches can lead to different intuitions about the physics involved. Nevertheless, they are all subject to the overriding proviso that they must describe reality. In other words, in order to be equivalent at a fundamental level, they must all predict the same measurement statistics, e.g. $\langle A(t) \rangle$, but depending on the situation, approaches may vary greatly in the difficulty of these calculations. In this section, I introduce two pictures of evolution which are perhaps the most fundamental—the *Schrödinger picture*, which underlies the model in Sec. 2.1.3, and the *Heisenberg picture*. I will use these two methods as convenient throughout the thesis.

The Schrödinger picture assumes that states evolve with time and that observables remain constant, i.e.,

$$\begin{aligned} \langle A(t) \rangle &= \langle \psi(t) | A | \psi(t) \rangle, \\ \text{where } i\hbar \frac{\partial |\psi\rangle}{\partial t} &= H |\psi\rangle, \end{aligned} \quad (2.79)$$

whereas the Heisenberg picture assumes that observables evolve and states remain constant, i.e.,

$$\begin{aligned} \langle A(t) \rangle &= \langle \psi | A(t) | \psi \rangle, \\ \text{where } i\hbar \frac{\partial A}{\partial t} &= [A, H]. \end{aligned} \quad (2.80)$$

In optics, most system evolution consists of light in some input quantum state being transformed as it propagates through some system of optical elements (a *quantum circuit*) into some output state. In a closed quantum system, this transformation can always be represented by a unitary operation, i.e.,

$$|\psi_{\text{out}}\rangle = U |\psi_{\text{in}}\rangle. \quad (2.81)$$

For systems with time-independent Hamiltonians, the Schrödinger equation can be solved directly to give

$$|\psi(t)\rangle = e^{-\frac{i}{\hbar} H t} |\psi(0)\rangle, \quad (2.82)$$

$$\Rightarrow U(t) = e^{-\frac{i}{\hbar} H t}. \quad (2.83)$$

It is then simple to calculate the equivalent evolution equations for operators in the Heisenberg picture, since both pictures must lead to the same measurement statistics.

$$\langle \psi | A(t) | \psi \rangle = \langle \psi(t) | A | \psi(t) \rangle = \langle \psi | U^\dagger A U | \psi \rangle \quad (2.84)$$

$$\Rightarrow A(t) = U^\dagger(t) A(0) U(t) \quad (2.85)$$

Evolution of a coherent state

As an example, it is worth calculating the time-evolution of a coherent state in free space. The corresponding unitary transformation is

$$U(t) = e^{\frac{-i}{\hbar} H t} = e^{-i \omega t a^\dagger a}. \quad (2.86)$$

Therefore, for an initial state $|\alpha(0)\rangle = |\alpha\rangle$ (recall that this is *not* an eigenstate of the creation operator),

$$|\alpha(t)\rangle = U(t)|\alpha\rangle, \quad (2.87)$$

$$= e^{-|\alpha|^2/2} \sum_n \frac{\alpha^n}{\sqrt{n!}} e^{-i \omega t a^\dagger a} |n\rangle, \quad (2.88)$$

$$= e^{-|\alpha|^2/2} \sum_n \frac{\alpha^n e^{-i \omega t n}}{\sqrt{n!}} |n\rangle, \quad (2.89)$$

$$= e^{-|\alpha|^2/2} \sum_n \frac{(\alpha e^{-i \omega t})^n}{\sqrt{n!}} |n\rangle, \quad (2.90)$$

$$= |\alpha(0) e^{-i \omega t}\rangle. \quad (2.91)$$

So as expected, the classical field amplitude picks up a rotating phase factor.

Obviously, electromagnetic fields vary in space as well as time, and it is also useful to understand how the state of a field at one position relates to the state at another. For example, for a coherent state, how does $\alpha(\mathbf{x}_0) = \langle \mathbf{x}_0 | \alpha \rangle$ relate to $\alpha(\mathbf{x}_0 + \mathbf{x}) = \langle \mathbf{x}_0 + \mathbf{x} | \alpha \rangle$? It turns out that the position eigenstates, $|\mathbf{x}_0\rangle$ and $|\mathbf{x}_0 + \mathbf{x}\rangle$, are related via the *translation operator*,

$$T(\mathbf{x}) = e^{\frac{-i}{\hbar} \mathbf{x} \cdot \hat{\mathbf{p}}}, \quad (2.92)$$

$$\Rightarrow T(\mathbf{x}) |\mathbf{x}_0\rangle = |\mathbf{x}_0 + \mathbf{x}\rangle. \quad (2.93)$$

It is useful to note that $T(\mathbf{x})^\dagger = T(-\mathbf{x})$. So for a coherent state in free space with a field momentum given by $\mathbf{G} = \hbar \mathbf{k} a^\dagger a$ [Eq. (2.63)],

$$\alpha(\mathbf{x}_0 + \mathbf{x}) = \langle \mathbf{x}_0 + \mathbf{x} | \alpha \rangle \quad (2.94)$$

$$= \langle \mathbf{x}_0 | T(\mathbf{x})^\dagger | \alpha \rangle \quad (2.95)$$

$$= \langle \mathbf{x}_0 | e^{i \mathbf{x} \cdot \mathbf{k} a^\dagger a} | \alpha \rangle \quad (2.96)$$

$$= \langle \mathbf{x}_0 | \alpha e^{i \mathbf{x} \cdot \mathbf{k}} \rangle \quad (2.97)$$

$$= \alpha(\mathbf{x}_0) e^{i \mathbf{x} \cdot \mathbf{k}} \quad (2.98)$$

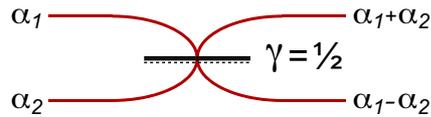


Figure 2.2: Evolution of “classical” coherent states through a 50:50 beam splitter.

Quantum circuits

When analysing optical systems it is not normally either necessary or useful to consider the detailed time-evolution of the light through the system. Instead, it is easiest to only consider the effect of the system as a whole. In this *quantum circuit model*, we only look at the relationship between the inputs and outputs, which will usually take the form of a unitary operation, $U_{\text{circuit}} = U$, i.e. for the two different pictures,

$$|\psi_{\text{out}}\rangle = U|\psi_{\text{in}}\rangle \quad \text{or} \quad A_{\text{out}} = U^\dagger A_{\text{in}} U. \quad (2.99)$$

The goal of such a calculation will normally be to determine the quantum state of the optical field exiting the circuit, but it turns out that it is extremely complicated to perform the bulk of the calculation in the Schrödinger picture. Consider the almost trivial case of an n -photon Fock state arriving at one port of a beam splitter. Logic dictates that the output must be a superposition of $n + 1$ different states: $|n, 0\rangle, |n - 1, 1\rangle, \dots, |1, n - 1\rangle, |0, n\rangle$. Keeping track of all of these states through a more complicated circuit would be extremely difficult, not to mention uninteresting. Moreover, this would only give the answer for a single input state.

Fortunately, it is much simpler to work in the Heisenberg picture. When the electromagnetic field is quantised, the classical Fourier amplitudes are replaced by annihilation operators. As a result, the Heisenberg equation of motion is a direct quantum analogue of the classical description, and the evolution equations for the field operators are exactly the same as the equations for the corresponding classical amplitudes. A rigorous justification for this can be found in the details of the canonical quantisation of the electromagnetic field (see, e.g., Ref. [33]). Here, however, I will use a simple example to illustrate the basic idea. Consider two classical fields, as represented by the coherent states¹⁹ $|\alpha_1\rangle$ and $|\alpha_2\rangle$, impinging on the two input ports of a 50:50 beam splitter [Fig. 2.2]. The input state is therefore²⁰

$$|\alpha_1\rangle \otimes |\alpha_2\rangle = e^{\alpha_1 a_1^\dagger - \alpha_1^* a_1} e^{\alpha_2 a_2^\dagger - \alpha_2^* a_2} |0\rangle. \quad (2.100)$$

After the beam-splitter, the state becomes

$$|\alpha_1\rangle \otimes |\alpha_2\rangle \rightarrow \left| \frac{1}{\sqrt{2}}(\alpha_1 + \alpha_2) \right\rangle \otimes \left| \frac{1}{\sqrt{2}}(\alpha_1 - \alpha_2) \right\rangle, \quad (2.101)$$

¹⁹Despite their explicitly quantum description, coherent states are often considered to be “classical” states, because they evolve in exactly the same way as the corresponding classical electromagnetic fields (e.g. in free space, $|\alpha\rangle \rightarrow |e^{-i\omega t}\alpha\rangle$).

²⁰Why is this a product state and not the superposition state, $1/\sqrt{2}(|\alpha_1, 0\rangle + |0, \alpha_2\rangle)$? In fact, the latter state is a highly non-classical state, closely related to the “Schrödinger cat states”. When detecting photons in both modes, this state would preclude the possibility of photons being detected simultaneously, while enforcing strong correlated $(n, 0)$ and $(0, n)$ detection events. Therefore the superposition state clearly does not match the classical scenario.

using simple beam-splitter equations for the classical fields, where the signs ensure energy is conserved. Expanding these states, and using the fact that $[a_1, a_2^\dagger] = 0$,

$$e^{\alpha_1 a_1^\dagger - \alpha_1^* a_1} e^{\alpha_2 a_2^\dagger - \alpha_2^* a_2} |0\rangle \quad (2.102)$$

$$\rightarrow e^{\frac{1}{\sqrt{2}}(\alpha_1 + \alpha_2) a_1^\dagger - \frac{1}{\sqrt{2}}(\alpha_1^* + \alpha_2^*) a_1} e^{\frac{1}{\sqrt{2}}(\alpha_1 - \alpha_2) a_2^\dagger - \frac{1}{\sqrt{2}}(\alpha_1^* - \alpha_2^*) a_2} |0\rangle, \quad (2.103)$$

$$= e^{\alpha_1 \frac{1}{\sqrt{2}}(a_1^\dagger + a_2^\dagger) - \alpha_1^* \frac{1}{\sqrt{2}}(a_1 + a_2)} e^{\alpha_2 \frac{1}{\sqrt{2}}(a_1^\dagger - a_2^\dagger) - \alpha_2^* \frac{1}{\sqrt{2}}(a_1 - a_2)} |0\rangle. \quad (2.104)$$

Comparing these terms gives the evolution equations for the field operators:

$$\begin{aligned} a_1^\dagger &\rightarrow \frac{1}{\sqrt{2}}(a_1^\dagger + a_2^\dagger), \\ a_2^\dagger &\rightarrow \frac{1}{\sqrt{2}}(a_1^\dagger - a_2^\dagger), \end{aligned} \quad (2.105)$$

as expected.

Using the Heisenberg picture

So if the operator evolution equations are generally fairly easy to calculate, are there any disadvantages to working in the Heisenberg picture? The main inconvenience is that ultimately it will still be necessary to calculate the output state (or at least specific measurement statistics) for different input states. However, it turns out that for most optical circuits this will be fairly straightforward.

This section basically follows the argument presented in Ref. [34] (p. 80–81).

Consider a quantum circuit which performs a general unitary operation, U , on k optical modes. An input state given by

$$|\psi_{\text{in}}\rangle = |n_1, \dots, n_k\rangle = \prod_j \frac{1}{\sqrt{n_j!}} (a_j^\dagger)^{n_j} |0\rangle \quad (2.106)$$

is a general input state in the sense that an arbitrary input must be a superposition of such states. The output will therefore be:

$$|\psi_{\text{out}}\rangle = U \prod_j \frac{1}{\sqrt{n_j!}} (a_j^\dagger)^{n_j} |0\rangle, \quad (2.107)$$

$$= \left(\prod_j \frac{1}{\sqrt{n_j!}} U (a_j^\dagger)^{n_j} U^\dagger \right) U |0\rangle, \quad (2.108)$$

using $U^\dagger U = I$. In the Heisenberg picture, the terms $U (a_j^\dagger)^{n_j} U^\dagger$ are the operator evolution equations for fields propagating *backwards* through the circuit—the *time-reversed* evolution equations—or in other words, the input field operators written in terms of the output operators. Since these can be calculated from the classical transformation rules, the output state can be calculated for any input state results from the Heisenberg picture. This is particularly simple for most optical circuits, where $U|0\rangle = |0\rangle$, which is true for linear optical elements (beam splitters, wave plates, etc.) and many other processes such as spontaneous parametric down-conversion. I will use this technique throughout the thesis.

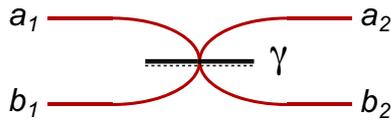


Figure 2.3: A general beam splitter is characterised by its reflectance, γ , and a reflection phase, φ .

The general beam splitter

The beam splitter is one of the most common elements in quantum optical circuits, with the simplest form of operator evolution equations being given above. However, these equations represent a very special case. In general, a beam splitter with *reflectance* γ [Fig. 2.3] is described by (except for an irrelevant global phase):

$$\begin{aligned} a_2 &= e^{i\varphi} \sqrt{\gamma} a_1 + \sqrt{\tau} b_1, \\ b_2 &= \sqrt{\tau} a_1 - e^{-i\varphi} \sqrt{\gamma} b_1, \end{aligned} \quad (2.109)$$

where $\tau = 1 - \gamma$ is the *transmittance* and φ is a general reflection phase.

2.2.3 The single photon

Throughout this thesis I will mainly be concerned with a special case of the quantum optical field—the single photon. It is perhaps the simplest example of a truly quantum state of light, fundamentally different from the “classical” coherent state. Nevertheless, in some ways, its behaviour is almost as simple as the coherent state, largely as a result of the fact that a single photon state is a *linear* function of creation operators, i.e.,

$$|\psi^{(1)}\rangle = \left(\sum_j \alpha_j a_j^\dagger \right) |0\rangle = \sum_j \alpha_j |\dots, 1_j, \dots\rangle. \quad (2.110)$$

Here, the α_j are complex amplitudes. As is the case here, however, a single photon can nevertheless populate many optical modes in the form of a superposition. I will use the term “single photon” to refer to any pure or mixed quantum state containing exactly one photonic excitation. In particular, a single photon can represent a qudit with logical states $|\mathbf{j}\rangle = a_j^\dagger |0\rangle = |\dots, 1_j, \dots\rangle$. This is called *multi-rail logic* or *multi-rail encoding*, the most common being *dual-rail logic* for a single qubit.

Using the free-space Hamiltonian for an electromagnetic field [Eq. (2.62)], it is easy to show that a single photon in an initial state $|\psi^{(1)}(0)\rangle = |1\rangle$ evolves very simply in time.

$$|\psi^{(1)}(t)\rangle = e^{i\omega t a^\dagger a} |1\rangle = e^{i\omega t} |\psi^{(1)}(0)\rangle \quad (2.111)$$

This looks very similar to the evolution of a coherent state [Eq. (2.91)], except that the rotating phase factor affects the quantum amplitude instead of the classical field amplitude. This distinction lies at the heart of the difference between single photons and coherent states.

2.3 Chapter 2 References

- [1] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, UK, 2000.
- [2] E. Knill. Non-binary unitary error bases and quantum codes. *quant-ph*, page 9608048, 1996.
- [3] E. Knill. Group representations, error bases and quantum codes. *quant-ph*, page 9608049, 1996.
- [4] M. A. Nielsen. A simple formula for the average gate fidelity of a quantum dynamical operation. *Physics Letters A*, 303:249, 2002.
- [5] I. L. Chuang and M. A. Nielsen. Prescription for experimental determination of the dynamics of a quantum black box. *Journal of Modern Optics*, 44:2455, 1997.
- [6] A. Jamiolkowski. *Reports on Mathematical Physics*, 3:275, 1972.
- [7] M.-D. Choi. Completely positive linear maps on complex matrices. *Linear Algebra and its Applications*, 10:285, 1975.
- [8] A. Gilchrist, N. K. Langford, and M. A. Nielsen. Distance measures to compare real and ideal quantum processes. *Physical Review A*, 71(6):062310, 2005.
- [9] R. Jozsa. Fidelity for mixed quantum states. *Journal of Modern Optics*, 41(12):2315, 1994.
- [10] M. Horodecki, P. Horodecki, and R. Horodecki. Mixed-state entanglement and distillation: Is there a “bound” entanglement in nature? *Physical Review Letters*, 80:5239, 1998.
- [11] W. Dür, G. Vidal, and J. I. Cirac. Three qubits can be entangled in two inequivalent ways. *Physical Review A*, 62:062314, 2000.
- [12] P. Walther, K. J. Resch, and A. Zeilinger. Local conversion of Greenberger-Horne-Zeilinger states to approximate W states. *Physical Review Letters*, 94:240501, 2005.
- [13] M. B. Plenio and S. Virmani. An introduction to entanglement measures. *e-print arXiv*, quant-ph/0504163, 2005.
- [14] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23:880, 1969.
- [15] B. Terhal. Bell inequalities and the separability criterion. *Physics Letters A*, 271:319, 2000.
- [16] A. G. White, A. Gilchrist, G. J. Pryde, J. L. O’Brien, M. J. Bremner, and N. K. Langford. Measuring two-qubit gates. *Journal of the Optical Society of America B*, 24(2):172, 2007.
- [17] W. J. Munro, K. Nemoto, and A. G. White. The Bell inequality: a measure of entanglement? *Journal of Modern Optics*, 48(7):1239, 2001.

- [18] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters. Mixed-state entanglement and quantum error correction. *Physical Review A*, 54:3824, 1996.
- [19] S. Hill and W. K. Wootters. Entanglement of a pair of quantum bits. *Physical Review Letters*, 78:5022, 1997.
- [20] W. K. Wootters. Entanglement of formation of an arbitrary state of two qubits. *Physical Review Letters*, 80:2245, 1998.
- [21] V. Coffman, J. Kundu, and W. K. Wootters. Distributed entanglement. *Physical Review A*, 61:052306, 2000.
- [22] P. Rungta, V. Bužek, C. M. Caves, M. Hillery, and G. J. Milburn. Universal state inversion and concurrence in arbitrary dimensions. *Physical Review A*, 64:042315, 2001.
- [23] K. Audenaert, F. Verstraete, and B. De Moor. Variational characterizations of separability and entanglement of formation. *Physical Review A*, 64:052304, 2001.
- [24] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters. Purification of noisy entanglement and faithful teleportation via noisy channels. *Physical Review Letters*, 76:722, 1996.
- [25] A. Peres. Collective tests for quantum nonlocality. *Physical Review A*, 54:2685, 1996.
- [26] D. Deutsch, A. Ekert, R. Jozsa, C. Macchiavello, S. Popescu, and A. Sampera. Quantum privacy amplification and the security of quantum cryptography over noisy channels. *Physical Review Letters*, 77:2818, 1996.
- [27] M. Horodecki, P. Horodecki, and R. Horodecki. Inseparable two spin-1/2 density matrices can be distilled to a singlet form. *Physical Review Letters*, 78:574, 1997.
- [28] A. Peres. Separability criterion for density matrices. *Physical Review Letters*, 77:1413, 1996.
- [29] K. Życzkowski, P. Horodecki, A. Sanpera, and M. Lewenstein. Volume of the set of separable states. *Physical Review A*, 58:883, 1998.
- [30] K. Życzkowski. Volume of the set of separable states. II. *Physical Review A*, 60:3496, 1999.
- [31] G. Vidal and R. F. Werner. Computable measure of entanglement. *Physical Review A*, 65:032314, 2002.
- [32] T.-C. Wei, K. Nemoto, P. M. Goldbart, P. G. Kwiat, W. J. Munro, and F. Verstraete. Maximal entanglement versus entropy for mixed quantum states. *Physical Review A*, 67:022110, 2003.
- [33] D. F. Walls and G. J. Milburn. *Quantum Optics*. Springer, Berlin, 1994.
- [34] H.-A. Bachor and T. C. Ralph. *A Guide to Experiments in Quantum Optics*. WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, Deutschland, 2004.

Alice had not a moment to think about stopping herself before she found herself falling down a very deep well. [...] First, she tried to look down and make out what she was coming to, but it was too dark to see anything. [...] Down, down, down. Would the fall *never* come to an end?

Down the Rabbit-Hole from *Alice's Adventures in Wonderland*, Lewis Carroll

“Curiouser and curiouser!” cried Alice (she was so much surprised, that for the moment she quite forgot how to speak good English).

Alice's Adventures in Wonderland, Lewis Carroll

Chapter 3

Quantum state and process tomography

Whenever attempting to perform an experiment, it is vital to be able to characterise the system (and the results) as completely and rigorously as possible. In general this means we must be able to make measurements on the system to extract information. However, quantum mechanics itself steps in to make this task difficult, and many of the challenges in experimental physics revolve around trying to measure quantum systems. This task is of particular importance in the field of quantum information processing, and the ultimate success or failure of the attempt to build a quantum computer may rest on our ability to perform measurements.

The main problem with quantum measurements is caused by one of the most basic ideas of quantum mechanics—the very act of making a measurement on a quantum system itself affects the state of the system being measured. One consequence of this is that it is not always possible to measure two different properties of a quantum system both precisely and simultaneously, so that there is a fundamental limit to the knowledge about its state (such incompatible observables are called *conjugate variables*). This is *Heisenberg's uncertainty principle*, and has profound implications for quantum information.

The uncertainty principle is very closely related to the *No Cloning theorem* [1, 2]. This states that it is not possible to use a single quantum system in an unknown state to generate multiple identical copies of the original system. Essentially, enforcing the uncertainty principle is equivalent to enforcing the No Cloning theorem, and vice versa. If an unknown quantum state could be cloned, then the uncertainty principle could be circumvented by creating many copies and measuring each of the conjugate variables with arbitrary accuracy. Conversely, if the uncertainty principle did not hold, then an unknown quantum

state could be cloned by just measuring the state and then creating as many copies as desired.

In order to measure the “complete” state of a quantum system¹, it is necessary to resort to techniques based on statistical reconstruction, since “single-shot” measurements alone are not sufficient, and we cannot make perfect copies of a single system. These are normally referred to as tomographic techniques—in particular, *quantum state tomography* and *quantum process tomography*. I will introduce and investigate these techniques in some detail in this chapter.

At various stages in the tomographic procedure it is necessary to make some assumptions about the noise sources in the quantum system. I will focus on photonic systems, where the stochastic nature of current sources and detectors introduces Poissonian statistical fluctuations into measured counts. In fact, I will normally assume that this is the main source of measurement noise in my experiments. Nevertheless, the general principles discussed here should transfer across to other systems with only slight modifications.

Some examples of situations where it may be useful to be able to know the quantum state of the system being studied are: checking the reliability of a source of quantum states and the ability to manipulate the state; measuring the output state from some quantum device; characterising the source of errors in an experiment; and characterising the operation of a quantum device as part of quantum process tomography. All of these cases have arisen during the work for this thesis.

3.1 Heisenberg’s uncertainty principle

The general form of Heisenberg’s uncertainty principle is

$$\Delta(A)\Delta(B) \geq \frac{1}{2}|\langle[A, B]\rangle|. \quad (3.1)$$

How does this affect measurements made on a single qubit? It is easy to show that the observables corresponding to measurements made in each of the three standard qubit bases [see Sec. 2.1.1] are the Pauli matrices:

$$\begin{aligned} \sigma_1 &= +P_+ - P_-, \\ \sigma_2 &= +P_{+i} - P_{-i}, \\ \sigma_3 &= +P_0 - P_1, \end{aligned} \quad (3.2)$$

where P_j are the relevant projectors of the form $|\mathbf{j}\rangle\langle\mathbf{j}|$. Since the Pauli matrices obey the commutation relations ($j, k, l = 1, 2, 3$): $[\sigma_j, \sigma_k] = 2i \sum_l \epsilon_{jkl} \sigma_l$, then, by the uncertainty principle, they are incompatible observables. For example, since $[\sigma_1, \sigma_2] = 2i\sigma_3$,

$$\Delta(\sigma_1)\Delta(\sigma_2) \geq |\langle\sigma_3\rangle|. \quad (3.3)$$

¹What does it mean to measure the “complete” state of a quantum system? In reality, this term must be used with qualification, since there are normally far too many variables defining a quantum system to make this possible. The idea of the complete state will generally refer to the state of the quantum system within some restricted target subspace. The ways that we can or wish to restrict the system vary greatly and will be discussed as they are relevant.

So both $\Delta(\sigma_1)$ and $\Delta(\sigma_2)$ must be strictly greater than zero unless $|\langle\sigma_3\rangle| = 0$, which is only possible for states with equal $\mathbf{0}$ and $\mathbf{1}$ populations, e.g. pure states of the form $|\mathbf{0}\rangle + e^{i\phi}|\mathbf{1}\rangle$. In fact, by direct calculation it is easy to verify that if $\Delta(\sigma_1) = 0$, then $\psi = |\pm\rangle$, and in this case², $\Delta(\sigma_2) = 1$ (n.b. this is the maximum value allowed for qubit observables³). In other words, if a precise measurement is made of σ_1 , then σ_2 is completely uncertain. Physically, this means that it is impossible to make a precise measurement of both the real and imaginary parts of an off-diagonal element of the density matrix.

It can be shown that two observables are simultaneously diagonalisable (i.e. they have the same eigenvectors) *if and only if* $[A, B] = 0$. The above discussion of the uncertainty principle shows already that this is the condition for the only case where measurement results for both observables can be simultaneously known with certainty. However, this idea allows the development of an alternative understanding of the measurement problem. If two observables have the same eigenvectors, then it does not matter which order they are measured in—after the first measurement, the system will be in an eigenstate of the first observable, and because this is also an eigenstate of the second observable, the second measurement will also obtain a definite result. If this is repeated many times, the measurement order will not change the statistics for the two sets of outcomes. But consider again the single-qubit case described above. After a σ_1 measurement, the system state will be $|\pm\rangle$. Because these are not also eigenstates of σ_2 , the statistics of the second measurement will always be the same *no matter what the initial state was*. In particular, if the system starts in the state $|+i\rangle$, then a σ_1 measurement will give results equally and randomly distributed between the two possible outcomes, as will the succeeding σ_2 measurement. However, if the order is reversed, the σ_2 measurement will always give the answer $+1$, with the σ_1 results unchanged. Clearly, only this second order gives the results expected for the initial state, and this only works in this one special case.

As discussed previously, the only way to solve this problem is to run the experiment many times under repeatable conditions and build up statistics about the quantum state. From now on, I will ignore “single-shot” measurements. A *single measurement* will refer to collecting data (e.g. photon counting) from a single projector (or occasionally a single observable) for an ensemble of repetitions of the experiment.

3.2 Linear quantum state tomography

For a d -dimensional system in a (possibly) mixed state, the density matrix is written in terms of d^2 parameters. These variables are restricted by the requirements of physicality, the most rigid being the normalisation constraint $\sum_{j=0}^{d-1} p_j = 1$. In theory, this means that only $d^2 - 1$ probabilities are needed to describe the density matrix. In practice, however, since we do not measure probabilities but a physical quantity such as a current or a count rate⁴, the overall normalisation is also a free parameter that must be measured.

²It is useful to note that $\Delta\sigma_j^2 = 1 - \langle\sigma_j\rangle^2$, because $\langle\sigma_j^2\rangle = \text{Tr}\{\sigma_j^2\rho\} = \text{Tr}\{\rho\} = 1$.

³The maximum value is always finite in a finite-dimensional Hilbert space. Contrast this with the infinite-dimensional (and continuous) position and momentum variables where the uncertainty can be infinite.

⁴Note that such a measurement should be capable of detecting a single quantum system in order to be a true quantum measurement.

Not surprisingly, these d^2 parameters exactly correspond to the d^2 elements in a basis for the space of density matrices, i.e.,

$$\rho = \sum_{j=1}^{d^2} v_j V_j, \quad (3.4)$$

where the V_j are a basis. For a given basis, then, the problem at hand is to estimate the coefficients, v_j —this is accomplished using *quantum state tomography*. If the basis elements are orthonormal under the Hilbert-Schmidt norm (i.e. $(V_j, V_k) = \text{Tr}\{V_j^\dagger V_k\} = \delta_{jk}$), then

$$v_j = (V_j, \rho) = \langle V_j^\dagger \rangle. \quad (3.5)$$

If V_j is a projector operator, then $\langle V_j^\dagger \rangle = \langle V_j \rangle = p(j)$ is the probability of measuring the system in the projected state, which can be calculated very easily from a count rate and a normalisation rate. Unfortunately, it is not possible to construct a density operator basis from projectors while still satisfying the condition for orthonormality⁵. We therefore construct another basis, with non-orthogonal elements, using only projectors, $W_j = |\psi_j\rangle\langle\psi_j|$. By definition, they are normalised ($\text{Tr}\{W_j^2\} = \text{Tr}\{W_j\} = 1$). I have given an example of such a basis in Sec. 2.1.1. The elements of this basis can now be expanded in terms of the orthonormal basis:

$$W_j = \sum_k q_{jk} V_k, \quad (3.6)$$

$$\& \quad q_{jk} = \text{Tr}\{V_k^\dagger W_j\}. \quad (3.7)$$

Since W_j must be linearly independent to form a basis, $\tilde{q} = (q_{jk})$ is therefore an invertible matrix, and both \tilde{q} and \tilde{q}^{-1} can be determined without reference to any measurement results.

The main advantage of the W basis is that the elements, W_j , can be directly related to easily measurable quantities, namely \mathcal{N}_j (e.g. currents or count rates), which form a minimal, complete set of measurements. They can be used to calculate a total, or *normalisation* count rate, $\overline{\mathcal{N}}$, and therefore also the “probabilities”,

$$\mathcal{W}_j = \frac{\mathcal{N}_j}{\overline{\mathcal{N}}}. \quad (3.8)$$

Due to noise inherent in all measurements and due to the fact that count times are finite⁶, these are actually *estimates* of the expected measurement probabilities, $w_j = \langle W_j \rangle = \text{Tr}\{W_j \rho\}$. The details of how the counts are normalised are extremely important and I will discuss this in some depth in following sections. (N.b. Throughout this chapter, I generally use uppercase letters, e.g. W_j , to represent operators; lowercase letters, e.g. w_j , to represent the corresponding probabilities or expansion coefficients; and calligraphic uppercase letters, e.g. \mathcal{W}_j , to represent experimental estimates of the lowercase quantities.)

⁵Since projectors are defined in terms of quantum states, $P_j = |\psi_j\rangle\langle\psi_j|$, there are at most d orthogonal projector operators for a d -dimensional system, and d^2 elements are required to form a basis.

⁶Because count times, and therefore counts, are finite, even the fact that a photon counter can only measure an integer number of photons introduces error into measurements.

Combining these results,

$$\mathcal{W}_j \approx w_j = \text{Tr} \{W_j \rho\}, \quad (3.9)$$

$$= \text{Tr} \{W_j^\dagger \rho\}, \quad (3.10)$$

$$= \sum_k q_{jk}^* \text{Tr} \{V_k^\dagger \rho\}, \quad (3.11)$$

$$= \sum_k q_{jk}^* v_k. \quad (3.12)$$

Since \tilde{q} is invertible, this in turn gives estimates for the v_j coefficients,

$$\underline{v} \approx \underline{\mathcal{V}} \equiv (\tilde{q}^*)^{-1} \underline{\mathcal{W}}, \quad (3.13)$$

where I have used the vector notation, $\underline{v} = (v_j)$, etc. The estimated density matrix can then be directly calculated from these coefficients,

$$\underline{\rho} = \rho_{\text{est}} \equiv \sum_j \mathcal{V}_j V_j. \quad (3.14)$$

To simplify the calculation, it is always possible to choose V_j to be Hermitian, which implies that \underline{v} and \tilde{q} are real, i.e. $\underline{v} = \tilde{q}^{-1} \underline{w}$.

3.2.1 Single qubit tomography

The simplest example of quantum state tomography is the single qubit case, first discussed by Stokes [3] in the context of the polarisation of light (as first noted in Ref. [4]). This procedure foreshadows the ideas used to define the qubit sphere [Sec. 2.1.1].

The first step is to define the two operator bases, using normalised versions of the Pauli operators and the standard measurement bases, i.e.,

$$\begin{aligned} V_j &= \frac{1}{\sqrt{2}} \sigma_j, \quad j = 0, 1, 2, 3 \\ W_j &= \{P_0, P_1, P_+, P_{+i}\}. \end{aligned} \quad (3.15)$$

With these definitions, the coefficients, v_j , can be easily identified as normalised versions of the famous Stokes parameters for polarisation [3]. The required normalisation parameter is $\overline{\mathcal{N}} = \mathcal{N}_0 + \mathcal{N}_1$, and the normalisation constraint for the v_j coefficients manifests itself in a particularly simple way, namely $v_0 = \text{Tr} \{ \frac{1}{\sqrt{2}} \sigma_0 \rho \} = \frac{1}{\sqrt{2}} \text{Tr} \{ \rho \} = \frac{1}{\sqrt{2}}$. Therefore, using these bases, the conversion matrix and its inverse are

$$\tilde{q} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} \quad \& \quad \tilde{q}^{-1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 & 0 \\ -1 & -1 & 2 & 0 \\ -1 & -1 & 0 & 2 \\ 1 & -1 & 0 & 0 \end{bmatrix}. \quad (3.16)$$

It is easy to generalise this case to multi-qubit systems by defining:

$$\begin{aligned} V_{jk\dots l}^{(d)} &= V_j^{(1)} \otimes V_k^{(1)} \otimes \dots \otimes V_l^{(1)} \\ W_{jk\dots l}^{(d)} &= W_j^{(1)} \otimes W_k^{(1)} \otimes \dots \otimes W_l^{(1)}, \end{aligned} \quad (3.17)$$

where $V_j^{(1)}$ and $W_j^{(1)}$ are the single-qubit operator bases defined above.

3.2.2 Qudit tomography

In Ref. [5], Thew et al. provide a generalised recipe for multi-*qudit* tomography, similar to the one given here. They suggest a physical operator basis which explicitly involves superpositions of all computational states. In Sec. 2.1.1 and in Ref. [6], I described another complete basis which consists of the computational states and two-state superpositions thereof. In theory, these two cases are completely equivalent, being related by a simple basis change, but in many optical systems the latter shows some strong practical advantages. In experimental systems, there is usually a “natural” choice for the computational basis states arising from physical considerations, and the two-state superpositions can often be prepared or measured far more easily than different multi-state superpositions. Moreover, these simple basis measurement sets are easily extended to consist entirely of POVM sets. For these reasons, I always use such tomographic measurement sets throughout the work in this thesis⁷.

For a qutrit, the orthonormal and physical operator bases are those given in Eqs (2.15) and (2.4), respectively:

$$\begin{aligned} V_j &= \frac{\Lambda_j}{\sqrt{\text{Tr}\{\Lambda_j^2\}}}, \\ W_j &= \Omega_j, \quad j = 0, \dots, 8. \end{aligned} \tag{3.18}$$

The normalisation parameter is therefore $\overline{\mathcal{N}} = \sum_{j=0}^2 \mathcal{N}_j$.

3.3 Maximum likelihood quantum state tomography

As mentioned above, measurement noise ensures that the measured quantities, \mathcal{W}_j , are estimates of the “theoretical” measurement probabilities, w_j , which are properties of the “actual state” of the system—the state that we are trying to determine. When these estimates are used directly in the linear tomography calculation, the main consequence is that the resultant density matrix is no longer guaranteed to satisfy the requirements of physicality. In particular, although it is straightforward to ensure that the matrix is Hermitian and normalised, it is no longer necessarily positive. Mathematically, its eigenvalues do not necessarily satisfy the constraint that $0 \leq \lambda \leq 1$.

The way we solve this problem is to use an optimisation procedure which is very similar to common data-fitting techniques. Importantly, however, this requires no assumptions about the form of the measured state except that it be physical. The basic idea is to search the entire space of allowed, physical density matrices for the one with the highest probability of generating the measurement results. This technique is normally called *maximum likelihood tomography* and involves three basic steps: (1) parametrising the physical density matrices; (2) defining an appropriate penalty function; and (3) performing a numerical optimisation. The discussion below follows the basic procedure described in Ref. [8].

⁷Similar measurements were also used in Ref. [7].

3.3.1 Parametrising the physical density matrix

The first step is to write the density matrix in a form which explicitly satisfies *all* of the physicality constraints which, to reiterate, require that it is:

$$\begin{aligned} \text{Hermitian:} & \quad \rho = \rho^\dagger, \\ \text{positive:} & \quad \langle \psi | \rho | \psi \rangle \geq 0 \quad \forall |\psi\rangle, \\ \text{\& normalised:} & \quad \text{Tr} \{ \rho \} = 1. \end{aligned} \quad (3.19)$$

The first two conditions are satisfied by choosing $\rho \sim \tau^\dagger \tau$, and the normalisation can be included explicitly, giving the final form:

$$\rho_{\text{phys}} = \frac{\tau^\dagger \tau}{\text{Tr} \{ \tau^\dagger \tau \}}. \quad (3.20)$$

Now the exact form of the τ matrix becomes important. For a d -dimensional density matrix, any useful parametrisation needs d^2 independent variables (subject to a normalisation constraint). Moreover, while the matrix described above is guaranteed to be physical, no matter what τ is used, the specific form of the τ matrix must ensure that *all* states are accessible. In other words, the parametrised form must span the space of density matrices.

One way of fulfilling this requirement is to choose a τ which is tridiagonal, e.g. for qutrits,

$$\tau = \begin{bmatrix} r_1 & s_{12} & s_{13} \\ 0 & r_2 & s_{23} \\ 0 & 0 & r_3 \end{bmatrix} = \begin{bmatrix} t_1 & t_4 + it_5 & t_6 + it_7 \\ 0 & t_2 & t_8 + it_9 \\ 0 & 0 & t_3 \end{bmatrix}, \quad (3.21)$$

with real numbers down the diagonal and complex numbers above. The density matrix can now be written in its parametrised form,

$$\rho_{\text{phys}} = \begin{bmatrix} \rho_{11} & \rho_{12} & \rho_{13} \\ \rho_{21} & \rho_{22} & \rho_{23} \\ \rho_{31} & \rho_{32} & \rho_{33} \end{bmatrix} = \tau^\dagger \tau = \begin{bmatrix} r_1^2 & r_1 s_{12} & r_1 s_{13} \\ r_1 s_{12}^* & r_2^2 + s_{12}^* s_{12} & r_2 s_{23} + s_{12}^* s_{13} \\ r_1 s_{13}^* & r_2 s_{23}^* + s_{12} s_{13}^* & r_3^2 + s_{13}^* s_{13} + s_{23}^* s_{23} \end{bmatrix}. \quad (3.22)$$

In terms of the new variables of the parametrisation, $\underline{t} = (t_j)$, the normalisation condition simplifies to $\sum_j t_j^2 = 1$. The important thing to note about the density matrix written in this form is that the elements become increasingly complex moving diagonally down and right through the matrix, and a given element depends only on parameters found above and left of it. Importantly, this makes it possible to calculate the values of t_j which give rise to a specific density matrix. In other words, because the definition of ρ_{phys} can be inverted for this choice of τ , this parametrisation must span the entire space. The easiest way to see how this inversion works is by writing down the procedure explicitly for the qutrit case above. A simple inductive argument shows that the result follows for matrices of any size.

$$\begin{aligned} r_1 &= \sqrt{\rho_{11}} & s_{12} &= \frac{\rho_{12}}{r_1} & s_{13} &= \frac{\rho_{13}}{r_1} \\ r_2 &= \sqrt{\rho_{22} - s_{12}^* s_{12}} & s_{23} &= \frac{\rho_{23} - s_{12}^* s_{13}}{r_2} \\ r_3 &= \sqrt{\rho_{33} - s_{13}^* s_{13} - s_{23}^* s_{23}} \end{aligned} \quad (3.23)$$

3.3.2 The penalty function

The next step is to develop a function characterising how well a particular guess of the density matrix fits a particular set of measurements. Associated with the guess, ρ , is a series of expected measurement probabilities $w_j(\rho) = \text{Tr}\{W_j\rho\}$, compared with the observed estimates $\mathcal{W}_j = \mathcal{N}_j/\bar{\mathcal{N}}$, which define the *expected counts*, $n_j(\rho) = w_j(\rho)\bar{\mathcal{N}}$.

The details of the penalty function are likely to depend on the underlying physics of the system being studied. As discussed previously, I am focussing here on photonic systems with Poissonian noise statistics.

The penalty function is calculated based on a physical argument answering the question: “What is the probability that a density matrix guess, ρ , will give rise to a series of measurements, \mathcal{N}_j ?” This argument will therefore vary according to exactly how we make the measurements. In particular, do we successively measure a series of count rates for different single projector operators, or do we make “complete” projective measurements? For example, for a single-qubit measurement, do we measure P_+ and then P_- ? Or do we detect both outcomes simultaneously? In the first case, we can assume that each measurement exhibits Poissonian noise statistics, whereas in the second case, only the total count will—the statistics of the individual counts will be constrained. Although the end result may not be too different, the details of the penalty function might change.

I collected all results in this thesis using successive, single-projector measurements, so I will assume for the moment that the noise on all counts is Gaussian with a variance determined by the expected count rate⁸, $\sigma_j(\rho)^2 \approx n_j(\rho)$. Therefore, the conditional probability of obtaining the observed measurements is:

$$p(\mathcal{N}|\rho) \propto \prod_j \exp\left(-\frac{(\mathcal{N}_j - n_j(\rho))^2}{2\sigma_j(\rho)^2}\right) = \prod_j \exp\left(-\bar{\mathcal{N}}\frac{(\mathcal{W}_j - w_j(\rho))^2}{2w_j(\rho)}\right) \quad (3.24)$$

The simplest way to calculate the density matrix which is most likely according to this probability is to minimise the following penalty function,

$$\Pi(t) = \sum_j \bar{\mathcal{N}}\frac{(\mathcal{W}_j - w_j(\underline{t}))^2}{w_j(\underline{t})}, \quad (3.25)$$

where $w_j(\underline{t}) = w_j(\rho_{\text{phys}}(\underline{t}))$. This is just a weighted least-squares optimisation, where the weights also depend on the input.

Recently, we made an important modification to the maximum likelihood procedure outlined in Ref. [8], when it was pointed out that the normalisation parameter can be included as a variable in the optimisation⁹, i.e. $\bar{\mathcal{N}} = \bar{\mathcal{N}}(\underline{t})$. Essentially, this allows the normalisation parameter to be estimated using *all* of the measurement results, rather than just those in the computational basis, which significantly improves the reconstruction quality.

⁸For large enough counts the Gaussian noise is a very good approximation of the Poissonian statistics describing the photon arrivals. Of course, if the state is sufficiently pure, it is possible that individual measurements may be close to zero for any integration time. However, these measurements are likely to form only a small subset of the full tomographic measurement set, so the error that this introduces should also be small.

⁹*Private communication* from Alexei Gilchrist.

3.3.3 The numerical optimisation

The numerical optimisation of this function can take any form, but not surprisingly, it is quite a complex task. Even for two qubits, the problem involves optimising 16 parameters to fit at least 16 data points (and often up to 36) using a function with a potentially very complex terrain. Originally, we used a local minimisation routine seeded with an initial guess from the linear tomography results. However, for more complex systems, inverting the density matrix reconstructed from the linear tomography to calculate the “initial guess” state parameters¹⁰ ($t_j(\varrho_{\text{lin}})$) becomes an increasingly laborious task. Moreover, a local minimisation routine is always subject to fundamental limitations which can be very sensitive to the quality of the initial guess. Therefore, as they became available, we used global optimisation routines. For tomographic analysis of more complex systems, however, these tend to be extremely slow—we began to reach the limits of our computational resources even for problems such as two-qubit process tomography and two-qutrit state tomography, which are still much smaller than may ultimately be required in quantum information.

More recently, we found a better alternative to this tricky procedure, called *convex optimisation*, which was developed by two colleagues [9]. While quite different to the maximum likelihood technique, this approach solves an equivalent problem. I will summarise it in some detail in the next section, following the discussion in Ref. [9], because I will use it for results throughout this thesis.

3.4 Convex optimisation tomography

3.4.1 Convex optimisation problems

Convex optimisation problems are a special class of optimisation problems which are now extremely well understood. A problem is a *convex optimisation* if the task is to minimise a *convex function* which is defined over a *convex set*. A function, f , is convex if it satisfies the inequality:

$$f\left(\sum_j p_j x_j\right) \leq \sum_j p_j f(x_j), \quad (3.26)$$

where p_j are probabilities which sum to one ($\sum_j p_j = 1$). A set, \mathcal{S} , is convex if

$$\sum_j p_j x_j \in \mathcal{S}, \quad \forall x_j \in \mathcal{S}. \quad (3.27)$$

It is not difficult to show that the set of physical density matrices is a convex set, a clue that convex optimisation techniques are very useful in quantum systems.

¹⁰Note: Because the linear tomography produces a non-physical density matrix, ϱ_{lin} , inverting this results in initial guess parameters, $t_j(\varrho_{\text{lin}})$, which are not necessarily real. A close, *physical* density matrix can be calculated by discarding these imaginary contributions—i.e. using $\tilde{t}_j = \text{Re}\{t_j(\varrho_{\text{lin}})\}$, and then renormalising.

Some very great advantages for finding solutions can be obtained by casting problems as convex optimisations. The most important fact with regard to the issue of quantum state tomography is that *any local minimum is also a global minimum*. Consequently, a simple gradient descent optimisation will not only find the global minimum effectively, but also very quickly—in fact, they are provably efficient in a computational sense, which means they run in polynomial time with regard to the size of the problem¹¹ [10]. This completely circumvents the problems discussed above for maximum likelihood optimisation in complex systems. Convex optimisation problems have already been the focus of a large body of work, and good routines to solve them are freely available for a variety of computational tools, such as Matlab. Moreover, the form of the solutions do not depend on the size of the system (cf. inverting the physical density matrix parametrisation in maximum likelihood tomography). Finally, casting a tomographic minimisation as a convex optimisation may allow improved insight into how to calculate errors in quantities derived from the output density matrix. This issue remains largely unsolved, except by using slow stochastic methods—but these also benefit greatly from the faster algorithms.

3.4.2 Tomography as a semidefinite programme

A *semidefinite programme* is a special type of convex optimisation of a linear function of Hermitian matrices, subject to constraints including positivity, which makes it especially useful for problems involving density matrices. This task can be written in the form:

$$\begin{aligned} & \text{minimise} && \underline{c} \cdot \underline{x}, \\ & \text{subject to:} && F_0 + \sum_j x_j F_j \geq 0, \end{aligned} \quad (3.28)$$

where the constraint equation implies that the left-hand side is a positive matrix. The vector \underline{x} are the real variables of the optimisation, \underline{c} are the real coefficients describing the function to be minimised (the *objective function*), and F_0 and F_j are square Hermitian matrices).

The maximum likelihood optimisation can be re-written in a similar form¹²:

$$\begin{aligned} & \text{minimise} && \sum_j \frac{\delta_j(\rho)^2}{w_j(\rho)}, \\ & \text{subject to:} && \rho \geq 0 \quad \& \quad \text{Tr}\{\rho\} = 1, \end{aligned} \quad (3.29)$$

where $\delta_j(\rho) = \mathcal{W}_j - w_j(\rho)$ are the *residuals*—the distance between the measured and “expected” probabilities.

Three steps are required to re-cast the maximum likelihood problem as a semidefinite programme (see Ref. [9] for details).

¹¹This is separate to the fact that the problem size, for quantum tomography, scales exponentially with the number of qubits.

¹²For the sake of simplicity, I have not included a variable normalisation in this penalty function.

- **Step 1:** The non-linear least-squares objective function can be replaced by a linear objective function with non-linear constraints, to give the equivalent optimisation:

$$\begin{aligned} & \text{minimise } y, \\ & \text{subject to: } y - \sum_j \frac{\delta_j(\rho)^2}{w_j(\rho)} \geq 0, \\ & \rho \geq 0 \quad \& \quad \text{Tr}\{\rho\} = 1, \end{aligned} \tag{3.30}$$

where y is called a *slack variable*. The density matrix which will produce a minimum under the least-squares penalty function will also make y a minimum.

- **Step 2:** The non-linear constraint,

$$y - \sum_j \frac{\delta_j(\rho)^2}{w_j(\rho)} \geq 0, \tag{3.31}$$

can be shown to be equivalent to a linear matrix inequality of higher dimension, namely (in block form):

$$\begin{bmatrix} y & \delta(\rho)^T \\ \delta(\rho) & \tilde{w}(\rho) \end{bmatrix} \geq 0, \tag{3.32}$$

where $\tilde{w}(\rho)$ is the diagonal matrix defined according to $(\tilde{w})_{jj} = w_j(\rho)$.

- **Step 3:** By expanding the density matrix ρ in a suitable basis, the c , F_0 and F_j elements of the semidefinite programme can be explicitly identified and written down (see Ref. [9]).

3.5 Quantum process tomography

As discussed in Sec. 2.1.3, quantum processes can be described very much like quantum states, so it is not surprising that quantum process tomography also operates in a very similar way to its state equivalent. The main difference is that the only resource available for state tomography is an ensemble of states prepared by a “black box” device, whereas process tomography allows, or indeed requires access to both the input and output states of a quantum process [Fig. 3.1].

The intuitive physical description is that a series of different quantum states are fed into the unknown process and state tomography is used to measure each of the corresponding output states. Provided the input states are a “complete set” (i.e. they span the input Hilbert space), these measurements comprise all accessible information about the process. Knowing the output from each of these basic input states allows the calculation of the output for an arbitrary input state.

For practical reasons, I will only consider the special case of trace-preserving operations. By the nature of our detection techniques (we filter only the results where no photons are lost), they are generally the only processes we can measure in the laboratory.

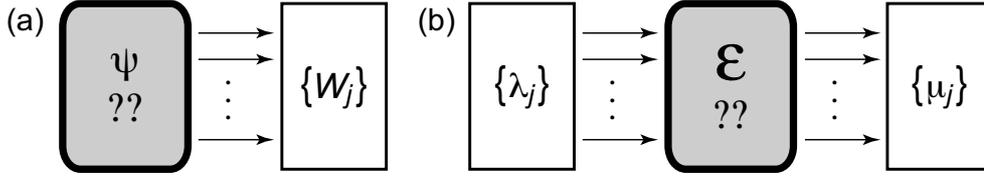


Figure 3.1: A conceptual comparison of quantum state and process tomography. (a) State tomography: an ensemble of identically prepared unknown quantum states ($|\psi\rangle$) are fed into a measurement device which integrates the signals filtered by a complete set of different measurement projectors, $\{W_j\}$ —these can be turned into probabilities. (b) Process tomography: known input states (from a complete set, $\{\lambda_j\}$) are repeatedly sent through an unknown quantum process (\mathcal{E}), and the output, $\mathcal{E}(\lambda_j)$ is fed into the state tomography measurement device (with projectors $\{\mu_j\}$).

Quantum process tomography was first introduced in Ref. [11] (see also Ref. [12]). Here, I provide a simpler description of linear process reconstruction which is the direct generalisation of the earlier description of linear state tomography. The notation is easy to implement numerically, and is more compatible for translation into the structure of the convex optimisation method of Ref. [9]. The symbols follow the earlier conventions, except that Greek letters are used to represent the operator bases.

Consider a quantum process, \mathcal{E} , with a corresponding process matrix, $\rho_{\mathcal{E}}$. Let $\{\lambda_j\}$ be the density matrices of the spanning set of (physical) input states and let $\{\mu_k\}$ be the complete set of measurement projectors. Using the Jamiolkowski representation, the probability of measuring the output state μ_k given an input state λ_j is:

$$p_{jk} = \text{Tr} \{ \mu_k \mathcal{E}(\lambda_j) \}, \quad (3.33)$$

$$= \text{Tr} \{ \mu_k \text{Tr}_a \{ (\lambda_j^T \otimes I) \rho_{\mathcal{E}} \} \}, \quad (3.34)$$

$$= \text{Tr} \{ (\lambda_j^T \otimes \mu_k) \rho_{\mathcal{E}} \}, \quad (3.35)$$

$$\approx \frac{\mathcal{N}_{jk}}{\bar{\mathcal{N}}_j} \equiv \mathcal{P}_{jk}, \quad (3.36)$$

where \mathcal{P}_{jk} is the *estimated probability* calculated from the measured counts \mathcal{N}_{jk} and $\bar{\mathcal{N}}_j$ (allowing for changes in the normalisation with input state). We next introduce an orthonormal basis set of operators γ_j , such that $\lambda_j = \sum_k l_{jk} \gamma_k$ and $\mu_j = \sum_k m_{jk} \gamma_k$, where l_{jk} and m_{jk} are both invertible, because $\{\lambda_j\}$ and $\{\mu_j\}$ are complete sets. Using the γ_j , we then construct a complete orthonormal basis to expand the process matrix—defining for convenience, $\Gamma_{jk} \equiv \gamma_j^T \otimes \gamma_k$:

$$\rho_{\mathcal{E}} = \sum_{jk} g_{jk} \Gamma_{jk}, \quad (3.37)$$

where $g_{jk} = \text{Tr} \{ \Gamma_{jk}^\dagger \rho_{\mathcal{E}} \}$. In other words, knowledge of the g_{jk} is equivalent to knowledge of $\rho_{\mathcal{E}}$, and the aim of quantum process tomography is to estimate these parameters. It is therefore necessary to relate the g_{jk} to the measured parameters, p_{jk} :

$$p_{jk} = \text{Tr} \{ (\lambda_j^T \otimes \mu_k) \rho_{\mathcal{E}} \}, \quad (3.38)$$

$$= \text{Tr} \{ (\lambda_j^T \otimes \mu_k)^\dagger \rho_{\mathcal{E}} \}, \quad (3.39)$$

$$= \text{Tr} \left\{ \left(\sum_l l_{jl} \gamma_l^T \otimes \sum_m m_{km} \gamma_m \right)^\dagger \rho_{\mathcal{E}} \right\}, \quad (3.40)$$

$$= \sum_{lm} l_{jl}^* m_{km}^* \text{Tr} \left\{ \Gamma_{lm}^\dagger \rho_{\mathcal{E}} \right\}, \quad (3.41)$$

$$= \sum_{lm} l_{jl}^* m_{km}^* g_{lm}, \quad (3.42)$$

$$\Rightarrow \tilde{p} = \tilde{l}^* \tilde{g} \tilde{m}^\dagger, \quad (3.43)$$

where I have again used the notation $\tilde{x} = (x_{jk})$. When inverted, this then gives estimates for g_{jk} :

$$\tilde{g} \approx \tilde{\mathcal{G}} \equiv (\tilde{l}^*)^{-1} \cdot \tilde{\mathcal{P}} \cdot (\tilde{m}^\dagger)^{-1}. \quad (3.44)$$

Once again, \tilde{l} and \tilde{m} depend only on the choices of operator bases, and not on any measured quantities. Moreover, choosing Hermitian matrices for the γ_j ensures that \tilde{l} , \tilde{m} and \tilde{g} are all real. Finally, as before, the estimated process matrix is

$$\rho_{\mathcal{E}} = \sum_{jk} \mathcal{G}_{jk} \Gamma_{jk}. \quad (3.45)$$

Since this is the Jamiolkowski form of the process matrix, it is written in the elementary operator basis, which is not necessarily particularly useful when it comes to visualising a measured process. The change-of-basis formula in Eq. (2.31) can be used to view the matrix in a more revealing form, e.g. in the Pauli basis.

As with quantum state tomography, noise in experimental measurements gives rise to unphysical process matrices. Using the same techniques to move from a linear to a maximum-likelihood tomographic technique, it is very easy to write down the relevant least-squares penalty function. However, the minimisation becomes much more difficult because it must be solved subject to the extra trace-related constraints for process matrices. We have not yet found a tractable way of incorporating these extra constraints directly into the parametrisation. Moreover, the fact that process tomography problems are by nature larger than their state tomography equivalents makes the minimisations more complex still. The first quantum process tomography calculations stretched the computational power we had available to us to the point of impracticality [13].

Fortunately, maximum-likelihood quantum process tomography can also be translated into a convex optimisation problem with all its associated benefits to speed and convenience. In the language of the previous section, the optimisation can be summarised:

$$\begin{aligned} & \text{minimise} && \sum_{jk} \frac{\delta_{jk}(\rho_{\mathcal{E}})^2}{p_{jk}(\rho_{\mathcal{E}})}, \\ & \text{subject to:} && \rho_{\mathcal{E}} \geq 0 \quad \& \quad \text{Tr}_b \{ \rho_{\mathcal{E}} \} = I/d, \end{aligned} \quad (3.46)$$

again defining the residuals, $\delta_{jk}(\rho_{\mathcal{E}}) = \mathcal{P}_{jk} - p_{jk}(\rho_{\mathcal{E}})$. The proceeding steps are almost identical to the state tomography case, and for a detailed discussion, I refer the reader to Ref. [9].

3.6 Calculating errors in tomographic reconstructions

There are three main sources of error that arise during a tomographic procedure: 1) errors in the measurement settings; 2) errors from statistical fluctuations in the counts; and 3) errors arising from how poorly or how well the data fit is achieved. These errors are manifested in both the matrix elements themselves, and the physical quantities derived from the density matrix, such as entanglement and linear entropy for quantum states. The former, however, can be very difficult to interpret meaningfully, since direct fluctuations in the matrix elements will often be inconsistent with physicality constraints, even if they are correlated appropriately to ensure that the matrix remains Hermitian. I will therefore ignore these errors in my analysis.

To date, there have been two main approaches taken to error calculations in quantum tomography. One relies on detailed analytical calculations using explicit formulae derived from linear tomography [8], and the other uses stochastic numerical calculations (*Monte-Carlo simulations*) to directly simulate the errors. Although neither approach is able to target the third type of error, the latter has several advantages over the former. First, the analytical calculations are extremely laborious and must be repeated for each new type of error in the measurement settings, whereas the stochastic technique involves only repeated calculations of the quantity itself, along with a simple statistical determination of the mean and standard deviation. The main discrepancy between the two methods, however, is that the analytical results significantly over-estimate the size of the errors compared with the Monte-Carlo simulations [14]. This is most likely because data-fitting techniques, such as linear regression and maximum-likelihood tomography, are generally inherently more robust against noise¹³. Although this is automatically taken into account by the stochastic calculations, its effect cannot be incorporated into the analytical results, because they are based on linear tomography (i.e. it is very difficult to see how these errors should be propagated through a data fitting process like maximum likelihood tomography where the fitting itself can partially compensate for inconsistencies in the data).

Error considerations also play an important role when deciding exactly how to perform a tomography. In particular, it is possible to choose an over-complete set of measurements (i.e. more than the d^2 elements of a basis; see Sec. 3.7 for details). There are several ways this can reduce tomographic errors. At a simplistic level, the more measurements used, the smaller will be the resulting errors. Also, extra measurements will produce some redundancy in the data, which makes the tomography more robust to outlying data points. Finally, making an appropriate choice of measurements gives rise to different ways, less prone to error, to normalise the data (discussed more in Sec. 3.7). These effects will in turn be reflected in the spread in the distributions of results generated in a Monte-Carlo simulation of errors, which is another advantage of the stochastic method of estimating errors. It would be difficult to incorporate these effects into direct analytical calculations.

In this thesis, I will take a two-pronged approach to calculations of tomographic errors. I will use Monte-Carlo simulations to estimate the errors in physical quantities derived from the density matrix, making the assumption that Poissonian fluctuations in the counts are the dominant source of error in the reconstruction [see Sec. 3.6.2]. I will then use a *fit*

¹³This is similar to the way the “standard error in the mean” is different from the standard deviation for a set of repeated measurements.

quality parameter, defined in the next section [Sec. 3.6.1], to determine how good this assumption is and give a qualitative indication of the size of any effects which are not related to count statistics.

3.6.1 Fit quality

In an ideal world, the measured data would always agree with the expected counts (as calculated from the maximum-likelihood density matrix) to within the limits expected by the intrinsic statistical fluctuations of the measurements (e.g. Poissonian fluctuations for photon counting). In practice, however, this is not necessarily the case. There are often many other experimental factors which influence the measurements. For example, as the systems being characterised become larger and more complex, the total counting time of the tomography also increases and the systems become more susceptible to instability and drift. Trying to overcome such instability is a large part of most experiments (including those in this thesis), but ultimately it is necessary to be able to characterise the success of these endeavours.

Essentially, we need to be able to answer the question: “*How well* does the maximum-likelihood density matrix fit the measured data?” In fact, a convenient indication is the final value of the penalty function after the optimisation. Recall that

$$\Pi(\rho) = \sum_j \frac{(\mathcal{N}_j - n_j(\rho))^2}{\sigma_j^2(\rho)}, \quad (3.47)$$

where the penalty function is written here in terms of counts rather than probabilities. For a given state, ρ , the measured count \mathcal{N}_j is a Poissonian-distributed random variable¹⁴ with mean, $\langle \mathcal{N}_j \rangle = n_j(\rho) \propto \text{Tr} \{W_j \rho\}$, and variance, $\sigma_j^2(\rho) \equiv \langle (\mathcal{N}_j - \langle \mathcal{N}_j \rangle)^2 \rangle (=n_j)$. Therefore, since the reconstructed density matrix, ϱ , is the best estimate of the true state, ρ , the expected size of the penalty function after the optimisation is

$$\langle \Pi(\varrho) \rangle = \sum_{j=1}^M \frac{\langle (\mathcal{N}_j - n_j(\varrho))^2 \rangle}{\sigma_j^2(\varrho)} = \sum_{j=1}^M 1 = M, \quad (3.48)$$

where M is the total number of measurements. Essentially, the minimum penalty function is just a more complex random variable with a variance defined by $(x_j \equiv \mathcal{N}_j - n_j(\varrho))$:

$$\langle \Delta \Pi^2(\varrho) \rangle \equiv \langle \Pi^2(\varrho) \rangle - \langle \Pi(\varrho) \rangle^2 = \sum_{j,k=1}^M \frac{\langle x_j^2 x_k^2 \rangle - \langle x_j^2 \rangle \langle x_k^2 \rangle}{\sigma_j^2 \sigma_k^2} = \sum_{j,k=1}^M \frac{\langle x_j^4 \rangle - \langle x_j^2 \rangle^2}{\sigma_j^4}, \quad (3.49)$$

because the noise is uncorrelated between measurements so that $\langle x_j^2 x_k^2 \rangle - \langle x_j^2 \rangle \langle x_k^2 \rangle = 0$, unless $j=k$. If the measured counts are large enough, then the \mathcal{N}_j are in fact Gaussian random variables and $\langle x_j^4 \rangle = 3n_j^2$, giving $\langle \Delta \Pi^2(\varrho) \rangle = 2M$. For many states, however, there may be particular measurements with very low counts (even with long count times)

¹⁴Recall from Sec. 3.3 that this assumption is only true if data are collected using successive, single-projector measurements. If data for all POVM elements are collected simultaneously, then only the total count will be Poisson-distributed (i.e. the normalisation, $\overline{\mathcal{N}}_j$).

that are truly Poissonian instead of Gaussian. This introduces a slight correction, since¹⁵ $\langle x_j^4 \rangle = n_j(1 + 3n_j)$ [15], which gives $\langle \Delta\Pi^2(\varrho) \rangle = 2M + \sum_j 1/n_j$.

If the optimal value of the penalty function is much greater than the expected value of M (compared to the standard deviation, $\sqrt{2M}$), then that indicates that Poissonian noise is not sufficient to explain the noise in the data. While it is not obvious how to incorporate this notion into the model used to make quantitative error estimates described above, it does at least provide a qualitative assessment of the fitting process itself, or from another perspective, an assessment of the quality of the measured data. This is similar to the way the R^2 parameter calculated from linear (or curve-fit) regression can be used to estimate how well the data fits the modelled trend.

Using this basic principle, I define the following, more intuitive *fit quality* parameter¹⁶,

$$Q(\varrho) = \sqrt{\frac{1}{M}\Pi(\varrho)}, \quad (3.50)$$

which is essentially the RMS count error (per measurement). Following the above results, it is possible to show that this parameter should have a mean of $\overline{Q} = 1$ (independent of the number of measurements), and a standard deviation of $1/\sqrt{2M}$, given Poissonian noise statistics. I will use this parameter to test the validity of the assumption that Poissonian fluctuations are the largest errors in the tomographic measurements.

There is one complication with the above calculations. The theory describes a comparison of the measured data with the target data as calculated from the “actual” state. However, the “actual” state is unknown in a real experiment and the fit quality is instead calculated using the expected data, as predicted from the reconstructed state. The presence of noise in the measurements pulls the tomographic reconstruction away from the “actual” state as the optimisation process successfully skews it towards the measured data. This results in a smaller (i.e. better) value for the fit quality, an effect which is borne out by the simulated tomographies in the next section [Figs 3.4, 3.5 and 3.6]. This is probably a less significant effect in larger systems where more random (independent) errors need to be accounted for, both as a whole and within each POVM set. Again, this is supported by comparing the one-qubit [Fig. 3.5] and two-qubit [Fig. 3.6] results that follow.

3.6.2 The Monte-Carlo approach

The main difficulty in performing Monte-Carlo simulations is that it can be computationally intensive, requiring potentially complex optimisations to be performed a large number of times. This is particularly a problem when characterising larger quantum systems. However, with the advent of faster techniques such as convex optimisation tomography, this problem has been drastically reduced, making stochastic calculations of errors practically realisable on an everyday basis.

The Monte-Carlo approach to error estimation is:

¹⁵This can be proved using the recursive differential technique described in the hint to Problem 11.9 in Ref. [15] (Ch. 11 of this reference is devoted entirely to the statistics of the Poisson distribution).

¹⁶Note that the maximum-likelihood penalty function and the fit quality parameter are closely related to, respectively, the χ^2 and reduced χ^2 distributions from probability theory (see, e.g., [15], Ch. 12).

1. From the original measurement values, $\mathcal{N} = (\mathcal{N}_j)$, generate a large number, S , of random *sample* data sets, $\{\tilde{\mathcal{N}}^s\}$, using appropriate estimates of the forms of data probability distributions. For example, to model the effect of photon counting statistics, each data point, \mathcal{N}_j , gives rise to a set of sample data points, $\{\mathcal{N}_j^s\}$, selected at random from a Poissonian distribution with variance, $\sigma_j^2 = \mathcal{N}_j$.
2. Perform the numerical tomography as normal for each of the sample sets, calculating the desired physical quantities, $f^s(\varrho) = f(\varrho_s)$, from each resulting reconstructed density matrix.
3. For each quantity, the statistical properties of the resulting distribution, $\{f^s(\varrho)\}$, can be calculated directly (e.g. variance, skew, etc.). In particular, the uncertainty, $\Delta f(\varrho)$, is just the standard deviation.

The key remaining question about the Monte-Carlo approach is how much is “enough”? Or more specifically, how many sample data sets are sufficient to obtain a good estimate of the standard deviation for a particular quantity? In later chapters, I report the results of tomographies of very large systems, and for some of these, each reconstruction takes around 6 hours of computing time. In such cases, there is an enormous difference even between performing 50, 100 or 200 reconstructions¹⁷.

To answer this question, it is first necessary to consider what determines whether an estimated error is “good”. Normally, when making repeated measurements of a physical quantity, one would decide when to stop based on the convergence of its mean. However, since the goal of these Monte-Carlo simulations is to estimate errors, I have instead considered the convergence behaviour of the standard deviation. To this end, I define the *running standard deviation*, $\Delta^r f(\varrho, s)$, of a quantity, $f(\varrho)$, to be the standard deviation of $\{f^s(\varrho)\}$ for the first s reconstructions in the Monte-Carlo simulation.

In my first test, I have considered series of 50 randomly selected pure states for one and two qubit systems (generated uniformly according to the Haar measure¹⁸), and implemented a Monte-Carlo procedure with 200 samples for each. To give comparable one- and two-qubit results, I assumed the count rate and integration times were the same (~ 670 photons per measurement). In Figure 3.2(i), I have plotted the convergence curves for 5 typical one- and two-qubit states, using the error in the fit quality, $\Delta^r Q(\varrho, s)$, as the example physical quantity, since it can be compared across all systems (other physical quantities, such as linear entropy and tangle, give similar results).

The first point to notice is the shape of the curves. Each new sample density matrix, ϱ_s , produces a fit quality from the underlying distribution $Q(\varrho, \mathcal{N})$. If this data point lies outside the range of previous samples, the standard deviation will jump sharply, but once

¹⁷The time taken to run these reconstructions would be greatly reduced if one could “seed” the Monte-Carlo tomographies with an initial guess based on the reconstruction from the original data set—a very reasonable starting point. Unfortunately, we have not yet found a way to specify an initial state for any of the available convex optimisation solvers. Solving this issue would make Monte-Carlo simulations a practical approach in systems of almost any reasonable size (at least by comparison with the time required for the initial tomography).

¹⁸To choose a pure state in a d -dimensional Hilbert space ($|\psi\rangle = \sum_j \alpha_j |\mathbf{j}\rangle$), set the real and imaginary parts of the α_j to be Gaussian random variables with mean 0 and variance 1. Then, the normalised vector, $|\psi\rangle = \sum_j \alpha_j |\mathbf{j}\rangle / \sqrt{\sum_j |\alpha_j|^2}$, is randomly selected from the Haar measure.

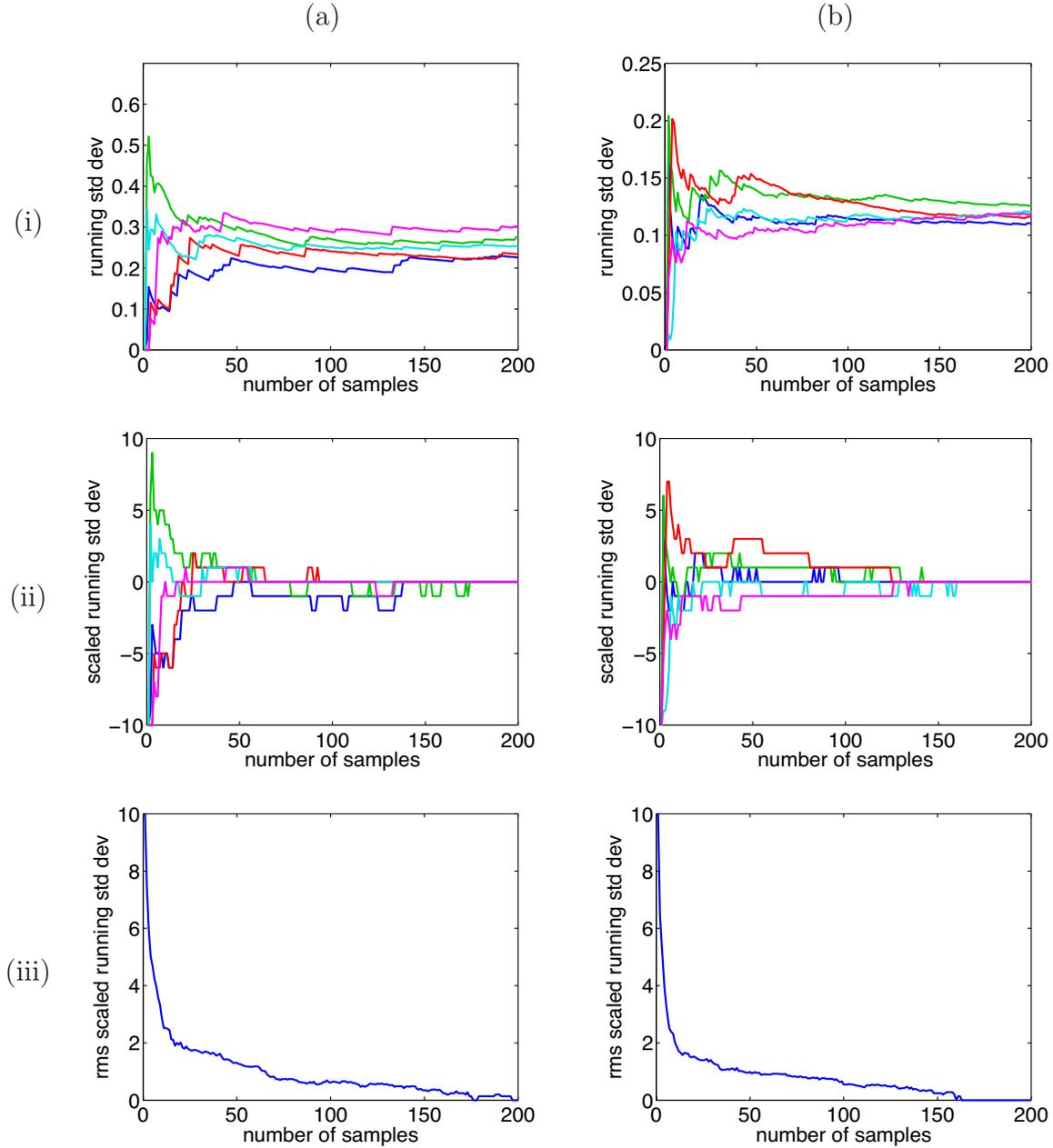


Figure 3.2: Convergence of Monte-Carlo errors in the fit quality parameter, Q , for randomly generated (a) one- and (b) two-qubit pure states. The plots illustrate: (i) the running standard deviation, $\Delta^r Q(\varrho, s)$, for 5 typical states; (ii) the scaled running standard deviation, $\Delta^{r,sc} Q(\varrho, s)$, for the same states; and (iii) the rms scaled running standard deviation, $\Delta_{\text{rms}}^{r,sc} Q(\varrho, s)$, averaged over 50 random states.

the samples spread through the entire distribution, each new point will only have a small effect on the calculated error, which will settle towards a steady value.

In all cases shown in Fig. 3.2(i), the errors reach reasonable stability somewhere in the range of 100–150 samples, but this is a very subjective assessment, and is further complicated by the fact that the curves approach different values for different states. Ultimately, however, the goal is to calculate an uncertainty which, in general, only needs to be re-

ported to one significant figure, and it is this that sets the standard for whether an error estimate is good enough.

To quantify the convergence of the errors objectively, I have defined a *scaled*, running standard deviation, $\Delta^{r,sc}f(\varrho, s)$, which is offset so that the final and best estimate of the (scaled) error is zero, $\Delta^{r,sc}f(\varrho, s=S) = 0$, and scaled to be in units of one tenth the value of the best estimate, i.e.,

$$\Delta^{r,sc}f(\varrho, s) = \frac{\Delta^r f(\varrho, s) - \Delta^r f(\varrho, s=S)}{0.1 \Delta^r f(\varrho, s=S)}. \quad (3.51)$$

In Figure 3.2(ii), I have plotted the scaled error for the same states as shown in (i), rounding the curves to the nearest integer to highlight the point where the target accuracy is reached. In fact, using the one-tenth scaling is perhaps overly harsh in some circumstances, but it is a useful all-purpose mark to aim at, and can be compared from one system or physical quantity to another. Finally, Figure 3.2(iii) shows the rms average of this rounded value for the 50 randomly selected states.

The plots in Fig. 3.2 demonstrate that the scaled, running standard deviation provides a good indication of when the Monte-Carlo have converged to sufficiently stable error estimates, and in these simulations, 200 samples have been sufficient. But how does this change when moving to higher-dimensional systems? It would not be unreasonable to expect that more samples would be required in larger systems, but this does not seem to be supported by the results in Fig. 3.2(a-iii) and (b-iii), which perhaps even show that the tomographies converge faster for two qubits than for one.

To explore this more comprehensively, I have analysed the convergence behaviour of a number of the experimental tomographies that I will report in later chapters. Figure 3.3(a) shows the scaled, rounded running errors for a number of physical quantities of a 36-dimensional quantum state. This is the largest quantum system considered in this thesis, and all of the errors converge well within 200 samples in the Monte-Carlo simulation. In Fig. 3.3(b), I have plotted the unscaled and scaled fit quality errors, $\Delta^r Q(\varrho, s)$ and $\Delta^{r,sc}Q(\varrho, s)$, for a range of quantum systems with between 4 and 36 dimensions. There does not seem to be any obvious relationship between the convergence rate and the size of the system.

This result seems quite surprising at first, but may perhaps be explained by the following argument. Each sample data set produces a density matrix estimate, ϱ_s , and both the number of Poissonian counts in the data set and the size of the Hilbert space containing the resulting density matrix grow exponentially with the size of the quantum system. Consequently, it seems quite a natural expectation that the number of samples required to estimate the tomographic errors would grow correspondingly large. Ultimately, however, each density matrix is used to produce a single value of the physical quantity of interest, sampled according to the very distribution we are trying to measure. Repeating this for each reconstructed matrix, the Monte-Carlo simulation generates a set of the random variables, $\{f^s\}$, and the spread in this set provides an estimate of the width of the underlying distribution. Therefore, the accuracy of the error estimate depends only on the number of samples, and not at all on the number of random counts or the size of the density matrix which generated them.

Based on the results illustrated in Figs 3.2 and 3.3, I have therefore shown that 200 samples

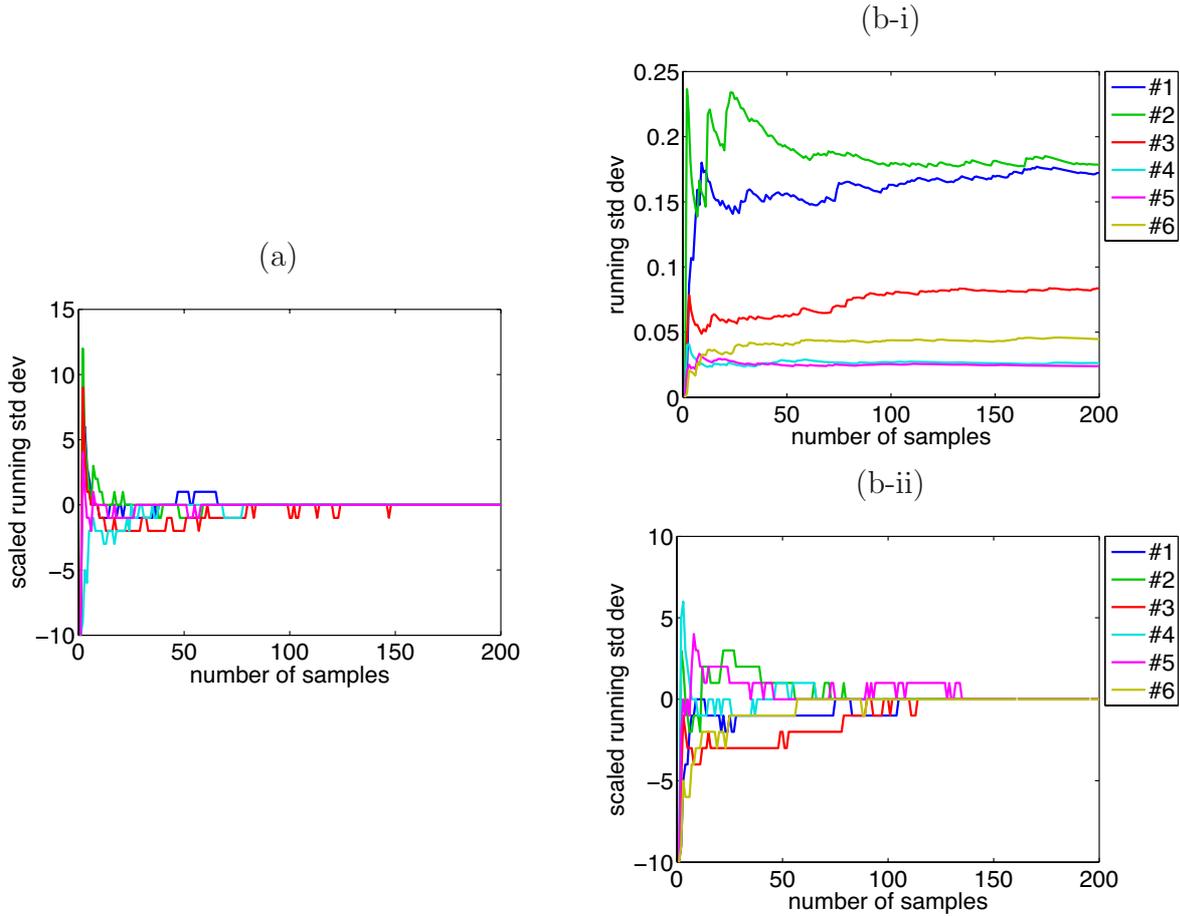


Figure 3.3: Convergence of Monte-Carlo errors for a range of experimental tomographies. (a) The scaled, running standard deviations of several physical quantities of a 36-dimensional quantum system. (b) The (i) unscaled and (ii) scaled fit quality errors ($\Delta^r Q(\varrho, s)$ and $\Delta^{r,sc} Q(\varrho, s)$, resp.), for a range of quantum systems: $d=4$ (#1, #2); $d=9$ (#3); $d=36$ (#4); $d=16$ state (#5); and $d=16$ process (#6).

are sufficient and probably necessary to reliably produce a good estimate of tomographic errors. Throughout this thesis, I have calculated all tomographic errors using Monte-Carlo simulations with this number of samples¹⁹.

As a final comment, it is well known that given N measurement outcomes sampled from an arbitrary probability distribution, the uncertainty in the estimated mean of the distribution will scale as $1/\sqrt{N}$ (see, e.g., Ref. [15]). A similar result can probably be derived for the uncertainty in the estimated standard deviation of the distribution. Such a result would provide additional support to the empirical investigations that I have described above.

¹⁹In future work, when performing tomographies with long reconstruction times, it may be sensible to monitor the convergence of the errors during the Monte-Carlo reconstruction process, so that the simulation can be stopped as soon as an acceptable accuracy has been reached.

3.7 Over-complete measurement sets

In the work in this thesis, I have made a substantial change in the standard approach to performing tomographic measurements in optical systems used in previous demonstrations (see, e.g., Refs [6, 16, 17]; cf. Refs [4, 18–20])—namely, instead of measuring the minimum number of counts required to completely characterise the density matrix, I have used over-complete sets of measurements which include all required POVM sets in full. This has two main advantages: using more measurements can provide better accuracy (as I show here); and the over-specification of the density matrix makes the result less sensitive to spurious data points. It also allows measurement normalisation within POVM sets, which provides better normalisation in the presence of noise and minimises some systematic experimental effects. In this section I will describe this technique and compare it with the standard approach.

Applying the standard approach to tomography for a single-qubit system, the measurements are $\mathcal{N}_j = \{\mathcal{N}_0, \mathcal{N}_1, \mathcal{N}_+, \mathcal{N}_{+i}\}$, and the probability estimates are calculated according to:

$$\mathcal{W}_{0,1} = \frac{\mathcal{N}_{0,1}}{\mathcal{N}_0 + \mathcal{N}_1} \quad \mathcal{W}_2 = \frac{\mathcal{N}_2}{\mathcal{N}_0 + \mathcal{N}_1} \quad \mathcal{W}_3 = \frac{\mathcal{N}_3}{\mathcal{N}_0 + \mathcal{N}_1} \quad (3.52)$$

In contrast, the over-complete set of measurements is $\{\mathcal{N}_{0,1}, \mathcal{N}_2^\pm, \mathcal{N}_3^\pm\} = \{\mathcal{N}_{0,1}, \mathcal{N}_\pm, \mathcal{N}_{\pm i}\}$, and the corresponding probability estimates are:

$$\mathcal{W}_{0,1} = \frac{\mathcal{N}_{0,1}}{\mathcal{N}_0 + \mathcal{N}_1} \quad \mathcal{W}_2^\pm = \frac{\mathcal{N}_2^\pm}{\mathcal{N}_2^+ + \mathcal{N}_2^-} \quad \mathcal{W}_3^\pm = \frac{\mathcal{N}_3^\pm}{\mathcal{N}_3^+ + \mathcal{N}_3^-} \quad (3.53)$$

So each data point is normalised using its own POVM set. The first and most simplistic consequence of these new definitions is that the probability estimates are always legal probabilities ($\mathcal{W}_j \leq 1$), even in the presence of the noise. Using the standard definitions with noisy data makes it possible that $\mathcal{N}_{2,3} > \mathcal{N}_0 + \mathcal{N}_1$.

As I described earlier, the simplest tomographic measurement basis for single-qudit systems consists of the computational basis states and the equal two-state superpositions, $|\mathbf{j}\rangle + |\mathbf{k}\rangle$ and $|\mathbf{j}\rangle + i|\mathbf{k}\rangle$. One easy way to extend this measurement set appropriately is to add the orthogonal superposition states, $|\mathbf{j}\rangle - |\mathbf{k}\rangle$ and $|\mathbf{j}\rangle - i|\mathbf{k}\rangle$. For qutrits, the corresponding measured counts are $\mathcal{N}_j = \{\mathcal{N}_0, \mathcal{N}_1, \mathcal{N}_2, \mathcal{N}_3^\pm, \dots, \mathcal{N}_8^\pm\}$, where I have extended the notation of Eqs (2.4) and (3.18) for the superposition states, $\{\Omega_3, \dots, \Omega_8\}$, to $\{\Omega_3^\pm, \dots, \Omega_8^\pm\}$.

As with qubit systems, the measurements are each normalised “internally” to give the probability estimates $\mathcal{W}_j = \mathcal{N}_j / \overline{\mathcal{N}}_j$, where $\overline{\mathcal{N}}_j$ is calculated using a POVM set containing \mathcal{N}_j , e.g. the superposition state, $|\mathbf{j}\rangle - |\mathbf{k}\rangle$, can be normalised using the set $\{|\mathbf{j}\rangle \pm |\mathbf{k}\rangle, |\mathbf{m}\rangle\}$, $\forall m \neq j, k$.

Before launching into a detailed discussion of the merits of different measurement sets, it is important to clarify here what is meant by a “measurement”. In Section 2.1.5, I briefly discussed the fact that in a true projective measurement, all possible outcomes are detected simultaneously. Thus, a “single measurement” obtains several data points—an entire POVM set, in fact. Using such a device to perform tomography automatically acquires the whole over-complete set of probabilities, and the number of “measurements”

is the same as the number of POVM sets²⁰. However, true projective measurements are often approximated by a series of successive, single-state counts/integrations, relying on an implicit *ergodic assumption*²¹. Since this is the technique used in the work of this thesis, I will adopt the convention that a “measurement” is an integrated count of a single projected state (i.e. a single \mathcal{N}_j), leading to a single probability estimate. Some of the following discussion is only meaningful in this context.

Performing quantum state tomography using an over-complete measurement set requires $2d^2 - d$ measurements for a single qudit and $(2d^2 - d)^n$ for n qudits (cf. d^2 and d^{2n} in the standard approach). Therefore, for large d , this requires almost 2^n times as many data points as the standard approach, and at best, for qubits, this is still 1.5^n times. Will this be a problem for a technique that already becomes exponentially more difficult with larger systems? To answer this question, there are two important issues that need to be considered.

The first issue is related to experimental stability. It has no fundamental significance for the theory of tomographic reconstruction, but has important practical implications, and highlights one of the most important advantages of using over-complete measurement sets. Because the measurements all form POVM sets, they can be made together in groups such that the major stability timescale reduces from the full tomography time to $O(d^n)$ measurements (for a single POVM set)—much less than both d^{2n} and $(2d^2 - d)^n$. To illustrate this, an example measurement set for the two-qubit case is given in Table 3.1. The measurements are grouped by the POVM sets with only four elements each, making the tomography less sensitive to some types of slow systematic fluctuations, such as fluctuations in the overall count rate. For example, in some of our experiments, this overall intensity oscillated by as much as 30–40% over the duration of the air-conditioning cycle of the laboratory. We believe that the slight temperature changes affected the laser pointing direction enough to cause significant changes in coupling efficiencies into single-mode fibres [see Fig. 10.6 and related text for more details].

As an aside, I note that treating a two-qubit system as a single four-dimensional qudit gives rise to $2 \times 16 - 4 = 28$ measurements in the over-complete set (for the specific choice described previously), instead of the $(2 \times 4 - 2)^2 = 36$ measurements in Table 3.1. This relates to the fact that some of these are entangled states.

The second question is more fundamental: how does using an over-complete set of measurements affect the accuracy of a tomographic reconstruction, assuming that the data collection is subject to time constraints? This issue is important even in the context of an idealised system with perfect stability. It explores the effects of the intrinsic measurement errors which arise when data is collected with finite integration times. In optics, without access to perfect detectors and sources which can provide single photons on demand, these errors are caused by Poissonian noise²², and I will explore this issue in some detail in the

²⁰It is in fact (no. of single-qudit POVM sets) ^{n} . This will depend on how these sets are constructed, which, in practice, will be somewhat constrained by experimental considerations. For the single-qudit normalisation sets above, the number will be $d(d - 1) + 1$, though this is not generally the most efficient choice.

²¹The ergodic assumption is that time-based averages give the same results as ensemble-based averages. This requires that the system is stable and that successive measurements are independent and equivalent.

²²N.b. In any system, even without any such statistical fluctuations, these errors are produced as a consequence of the fact that only a discrete number of system copies can be measured, making any

Table 3.1: Two-qubit measurement order.

j	$W_j^{(1)}$	$W_j^{(2)}$	j	$W_j^{(1)}$	$W_j^{(2)}$	j	$W_j^{(1)}$	$W_j^{(2)}$
1)	0	0	13)	+	0	25)	$+i$	0
2)	0	1	14)	+	1	26)	$+i$	1
3)	1	0	15)	-	0	27)	$-i$	0
4)	1	1	16)	-	1	28)	$-i$	1
5)	0	+	17)	+	+	29)	$+i$	+
6)	0	-	18)	+	-	30)	$+i$	-
7)	1	+	19)	-	+	31)	$-i$	+
8)	1	-	20)	-	-	32)	$-i$	-
9)	0	$+i$	21)	+	$+i$	33)	$+i$	$+i$
10)	0	$-i$	22)	+	$-i$	34)	$+i$	$-i$
11)	1	$+i$	23)	-	$+i$	35)	$-i$	$+i$
12)	1	$-i$	24)	-	$-i$	36)	$-i$	$-i$

remainder of this section.

There are two main competing considerations: more measurements will increase the total time for a tomography, but the corresponding improvements in accuracy will allow shorter integration times for each measurement. I have used a range of numerical simulations to probe different aspects of the reconstruction process, while limiting and controlling other influences. The simulations were designed to mimic a practical situation where an experimentalist has (on average) N copies of an unknown quantum state produced by a source at some (average) rate, R , and wishes to know what measurement technique will determine the unknown state with the greatest accuracy. The experimenter can choose how many and which measurements to make, and also how to divide the copies between the different measurements. I have assumed that the experimenter integrates each measurement for the same amount of time (a non-adaptive scheme).

Since maximum likelihood tomography is required only in the presence of noise, for realistic and meaningful results, this must be included in the simulation. For example, in photonic systems, the accuracy of the measured counts is fundamentally limited by Poissonian fluctuations, even if the measurement apparatus and detectors are perfect. This has some important implications which may possibly be exploited to optimise the tomographic process for optical systems.

The simulation procedure was:

1. For the target state, ρ_t , calculate the expected measurement probabilities, $w_j = \text{Tr}\{W_j\rho_t\}$.
2. For each of the M measurements, calculate the expected number of counts, $n_j = w_jN/M$.
3. From these expected counts, generate a series of “noisy” data sets and perform a tomographic reconstruction using each set (similar to Monte-Carlo procedure for

probability calculations approximate.

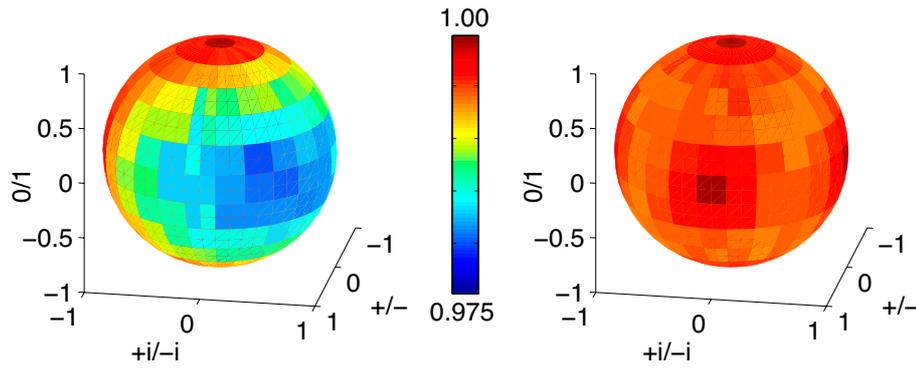


Figure 3.4: Tomography of the single-qubit pure states, analysed with the standard measurements (left) and the over-complete measurement set (right). The colour scale represents the mean fidelity of the reconstructed states with the target state based on Monte-Carlo simulations (400 sample sets per location). The “north” and “south poles” are the logical states, $|0\rangle$ and $|1\rangle$, respectively. The “equator” represents the equal superposition states, with $|+\rangle$ at the front and $|+i\rangle$ at the rightmost position.

error estimation in Sec. 3.6.2).

4. Compare the reconstructed density matrices against the target state.

The simplest comparison between the two techniques is to look at the states lying on the surface of the qubit sphere—the pure, single-qubit states. The results for this simulation are contained in Fig. 3.4 and Fig. 3.5. The simulations were performed with the standard measurement set ($M = 4$) and the over-complete set ($M = 6$), with $N = 4000$ and the number of samples for each target state, $S = 400$. Figure 3.4(left) and (right) show the results for the mean fidelity (over the ensemble) for the standard and over-complete measurement sets, respectively. For ease of visualisation, this has been flattened out and plotted again in Fig. 3.5(a), along with (b) the spread in fidelity ($\Delta F = \sigma \{F^s(\varrho, \rho_t)\}$), (c) the average linear entropy and (d) the average fit quality. These plots clearly demonstrate one of the most important advantages of using the over-complete measurement set—it provides an essentially unbiased reconstruction. In all of the plots, the standard measurements give strongly asymmetric results, showing much poorer performance in the region between and around the $|+\rangle$ and $|+i\rangle$ target states. In contrast, the over-complete reconstructions do not favour any particular region of the qubit sphere; and even where the standard measurement set performs best, they provide comparable results. Finally, the plot of average fit quality [Fig. 3.5(d)] also shows that the six-state fits are closer to the data used in the optimisation²³. This indicates that the over-complete tomography gives a “truer” reconstruction in terms of the information available to the experimentalist.

It is perhaps somewhat surprising that using the $\{0, 1, +, +i\}$ measurement set should give accurate results near $\{-, -i\}$ and not $\{+, +i\}$. This is probably a consequence of performing tomography on data which is subject to Poissonian noise, because a “near

²³These results support the discussion in Sec. 3.6.1. Although one would expect that the fit quality should be approximately 1 for both types of measurement tomography, it is smaller, because the optimisation process skews the reconstructed state away from the “actual” state and towards the measured data.

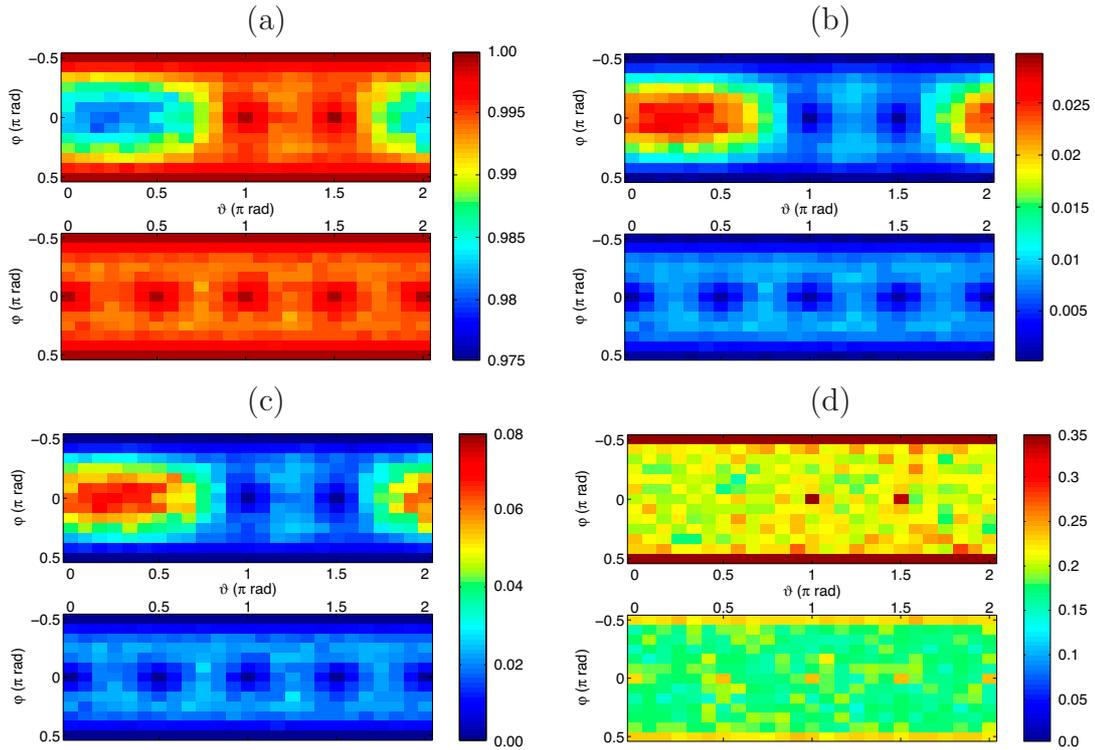


Figure 3.5: A comparison of tomography with standard measurements (upper) with the over-complete measurement set (lower) for single-qubit pure states using Monte-Carlo simulations (400 sample sets per grid point): (a) mean fidelity; (b) fidelity standard deviation; (c) mean linear entropy; (d) mean fit quality. The bottom and top rows of each image ($\varphi = \pm\frac{\pi}{2}$) are the logical states $\mathbf{0}$ and $\mathbf{1}$, respectively (the “north” and “south poles” of the qubit sphere). The states along the middle row of each image ($\varphi = 0$) are the equal superpositions, $|\mathbf{0}\rangle + e^{i\vartheta}|\mathbf{1}\rangle$ (the qubit sphere equator). Note: In this map projection, the entire bottom and top rows represent single points on the sphere. Also, the map wraps around from the right edge to the left, and the two edges ($\vartheta=0$ and $\vartheta=2\pi$) are the same points on the sphere.

zero” count is a much stronger constraint than a high count. (Although small counts have a large *relative* error, their *absolute* error is still very small.) This effect should be particularly significant in single-qubit tomography, because, for example, if the state is definitely *not* $|+\rangle$, then it *must be* $|-\rangle$. This strong “negative measurement” condition is unique to single qubits, and perhaps differentiates them from larger systems at some fundamental level. It can also be exploited for other purposes, such as designing optimal state discrimination techniques for single-qubit systems [12].

The next simulations I carried out explored how the two techniques would perform for some of the typical states that are produced in experiments, particularly those involving entanglement (requiring two-qubit states) and mixture. These two-qubit results (using $M = 4 \times 4$ and $M = 6 \times 6$) are in Fig. 3.6, and were generated using Monte-Carlo simulations with $N = 3 \times 10^4$ and $S = 400$.

To investigate the effects of mixture, I considered Werner-like states of the form, $\rho_t = p|\psi\rangle\langle\psi| + (1-p)I_4/4$ (where I_4 is the identity operator in four dimensions) with a range of

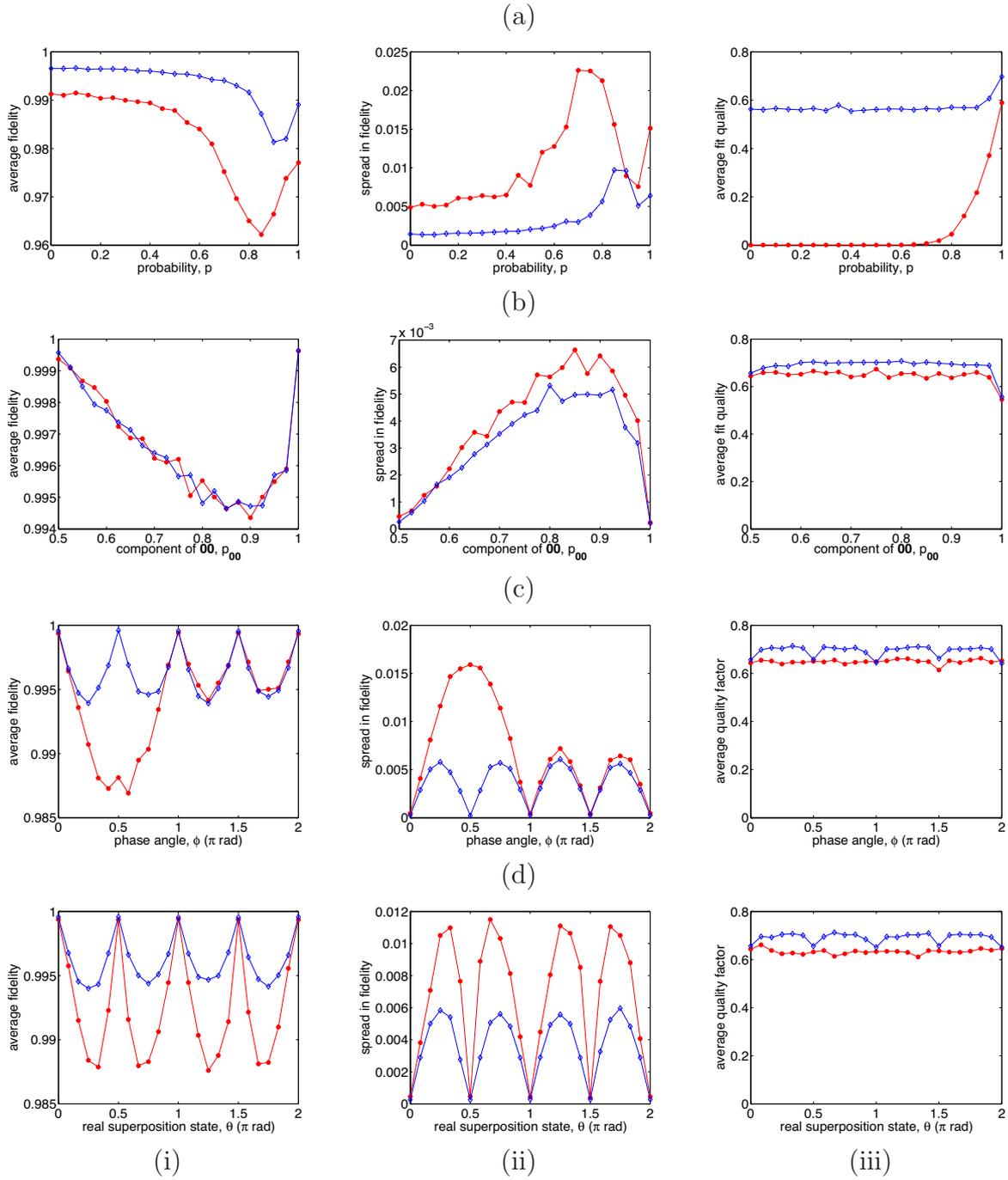


Figure 3.6: A comparison of tomography with standard measurements (—●—: red) with the over-complete measurement set (—◇—: blue) for different two-qubit states using Monte-Carlo simulations (400 sample sets per plot point): (i) mean fidelity; (ii) fidelity standard deviation; (iii) mean fit quality. The different target states are: (a) $\rho_t = p|\psi\rangle\langle\psi| + (1-p)I_4/4$ (Werner states: partially mixed), where $|\psi\rangle \sim |ab\rangle + |\bar{a}\bar{b}\rangle$ is maximally entangled—the qubit sphere coordinates for a and b are $\{\frac{2\pi}{3}, 0\}$ and $\{\frac{\pi}{4}, \frac{\pi}{6}\}$, respectively, and \bar{a} and \bar{b} are their orthogonal counterparts; (b) $|\psi_t\rangle = \sqrt{p_{00}}|\mathbf{00}\rangle + \sqrt{1-p_{00}}|\mathbf{11}\rangle$ (partially entangled and pure); (c) $|\psi_t\rangle \sim |\mathbf{00}\rangle + \exp(i\phi)|\mathbf{11}\rangle$ (maximally entangled and pure); and (d) $|\psi_t\rangle \sim |\mathbf{0}\theta\rangle + |\mathbf{1}\bar{\theta}\rangle$ (maximally entangled and pure), where $|\theta\rangle = \cos\frac{\theta}{2}|\mathbf{0}\rangle + \sin\frac{\theta}{2}|\mathbf{1}\rangle$.

“ideal” (pure) states, $|\psi\rangle$, both entangled and unentangled. Figure 3.6(a) shows a typical example of the results. Once again, the over-complete tomography performs significantly better than the standard, particularly in the region of most interest to the experimentalist, where the states are close to, but not completely pure. Perhaps the most surprising result is the plot of the average fit quality [Fig. 3.6(a-iii)], which approaches approximately 0.55 and 0 for the over-complete and standard tomography, respectively, as the target state becomes more mixed ($p \rightarrow 0$). Although, at first glance, this seems to contradict the evidence of plots (a-i) and (a-ii), it can be understood by recognising that quantum state tomography is closely analogous to simple linear regression by least-squares fitting. If a linear function is reconstructed using only two data points (cf. the minimum measurement set used in standard tomography), then the predicted line will fit the data precisely, but will not necessarily accurately reflect the *actual system* being measured. Using more data points (cf. an over-complete set) will allow the effects of noise to be diagnosed and filtered, providing a more accurate prediction. However, the “accurate” reconstruction will no longer perfectly fit the noisy data, and the size of this disparity will reflect the size of the noise. This also explains why the fit quality for the over-complete set is fairly similar ($\sim 0.55\text{--}0.70$) for all states that I investigated [see Figs 3.6(a,b,c,d-iii)].

It is important to note that this analogy seems to break down for states which are quite pure—that is, the standard tomography reconstructions only give “perfect” fits with the data in the presence of mixture [e.g. $p \lesssim 0.7$ in Fig.3.6(a-iii)]. The problem with the analogy is the unphysical predicted states which result from noisy tomographic measurements—a situation which does not normally arise in linear regression. This allows a further insight into the tomographic process. The plots in Fig. 3.6(a) show that standard tomography of states with mixture produces almost perfect data fits, but lower reconstruction fidelities, which contrasts with the over-complete alternative. Therefore, in this case, measurement noise does not tend to make a physical state unphysical, but instead converts it into another, different physical state. This can perhaps be most easily understood by considering the eigenvalues of the reconstructed density matrix. For pure states, most of the eigenvalues are close to zero, so noisy measurements may cause negative eigenvalues. For mixed states, however, the eigenvalues tend to be balanced and non-zero. Noise is then less likely to produce the negative eigenvalues which are the signature of unphysical states.

The remaining simulations [Fig. 3.6(b,c,d)] investigate several different classes of two-qubit entangled states that are commonly produced in the laboratory. Figure 3.6(b-i) suggests that varying the entanglement of the target state has a fairly small effect on the reconstruction quality. The 4×4 and 6×6 tomographies perform comparably well, although the plot in (b-ii) shows that the 6×6 reconstruction fidelities are slightly less variable with noisy data. On the other hand, the results in (c) and (d) describe two different families of maximally entangled states, and both the \overline{F} and ΔF plots show that the 4×4 tomography is strongly sensitive to which single-qubit states contribute to the two-qubit superposition. This is a natural consequence of the asymmetry observed in the single-qubit simulations [Fig. 3.5], and it again demonstrates the value of the unbiased 6×6 measurement set.

The results presented in this section suggest that there are many advantages to using over-complete measurement sets in quantum tomography. On a practical level, an appropriate choice of measurements allows them to be taken in complete POVM sets which

significantly eases the requirements on experimental stability. But there are also more fundamental benefits. The results in Fig. 3.5 show that using the over-complete measurement set is generally more consistent and accurate than using the standard tomography set, providing an unbiased, high-quality reconstruction for an arbitrary pure state. This holds also for more complex systems where entanglement is present, as demonstrated by Fig. 3.6(c) and (d). The over-complete tomography also performs significantly better in the presence of mixture where the redundancy in measurements allows the optimisation to compensate for the effects of noise and produce a reconstruction which is closer to the target state [Fig. 3.6(a)].

These simulations are not intended to be a comprehensive exploration of what constitutes the ideal measurements for tomographic reconstruction. Nevertheless, they provide clear evidence that the over-complete measurement sets which I have identified for qubit-based systems are simple and practical alternatives to the standard tomographic sets and substantially improve the performance of the reconstruction. Moreover, these over-complete sets can also be shown to be superior to the best minimum measurement sets which are based on a tetrahedral spread of measurements on the qubit sphere²⁴. There are also other interesting problems which need to be tackled. For example, as more measurements are included in a tomographic set, is there a point where the disadvantages of extra measurements outweigh the advantages? Also, it seems likely that these trends will continue in higher-dimensional systems, but this needs to be explored. Finally, what improvements can be made using adaptive tomographic techniques, where the experimenter analyses the data as it is obtained and uses the partial information to target the future measurements? In such a scheme, there are many variables that the experimenter could control—e.g. how long for each measurement; how many measurements; for a given number of measurements, which exact measurements to make, including the internal shape of the measurement set and its overall orientation in the Hilbert space.

3.8 Weighting the likelihood function

3.8.1 Poissonian errors

As discussed earlier, the weighted least-squares penalty function which must be minimised to provide *maximum likelihood reconstruction* as defined by standard parameter estimation theory is

$$\Pi(\rho) = \sum_j \overline{\mathcal{N}} \frac{(\mathcal{W}_j - w_j(\rho))^2}{w_j(\rho)}, \quad (3.54)$$

where \mathcal{W}_j and $w_j(\rho)$ are the measured and expected count probabilities, respectively. In this case, the residuals are weighted according to the expected probability variances, $\sigma_j^2(\rho) = w_j(\rho)$. A similar penalty function can be defined as

$$\Pi_{\text{FW}}(\rho) = \sum_j \overline{\mathcal{N}} \frac{(\mathcal{W}_j - w_j(\rho))^2}{\mathcal{W}_j}, \quad (3.55)$$

²⁴Private communication from Mark De Burgh.

where the residuals are now weighted by the Poissonian error in the measured counts. This has the particular advantage that the weighting factors are fixed throughout the optimisation (i.e. they do not depend on the current test state, ρ), making it less taxing on numerical resources (either memory or operations). This was particularly important when we were using older computers and slower optimisation techniques. Accordingly, we used the “fixed-weighting” (FW) penalty function, Π_{FW} , in the earlier experiments described in this thesis (in particular for the results published in Ref. [6], see Ch. 7). One disadvantage of using fixed weighting parameters is that Π_{FW} is infinite if any of the measurements give zero counts. However, this was a less important consideration when the optimisations were so slow, and on the very rare occasions when it was necessary²⁵, we arbitrarily replaced all counts of 0 with 1, which should have had very little effect on the reconstruction.

Although $\Pi_{\text{FW}}(\rho)$ does not give the traditional maximum likelihood reconstruction, it is a physically sensible penalty function which, as required, weights the fit towards the measurements which are known to within tighter bounds²⁶. In fact, one might reasonably expect that both penalty functions would give almost exactly the same tomographic reconstruction estimates (ρ). However, when recently reanalysing earlier results with our newer optimisation techniques, I found that states obtained using Π_{FW} were consistently more pure and more highly entangled [for several different states, see again Ch. 7]. Therefore, in the light of this possible trend, it was important to scrutinise the differences between the penalty functions more carefully.

For a real photon-counting experiment, where Poissonian statistics affect even an ideal measurement, one plausible argument suggests that $\Pi(\rho)$ will have more difficulty than $\Pi_{\text{FW}}(\rho)$ in reconstructing a very pure state, which could be expected to produce extremely low counts. After all, a predicted count of 1 is much “closer” (in a weighted sense) to a measurement of 9 ($\sigma = \sqrt{9} = 3 \Rightarrow |\delta| = 8 \sim 3\sigma$), than a measurement of 9 is to the prediction ($\sigma = \sqrt{1} = 1 \Rightarrow |\delta| = 8\sigma$). This argument is borne out by the results of some preliminary numerical investigations, described below.

My first test was to compare the performance of the two penalty functions for randomly selected pure, two-qubit states (generated uniformly according to the Haar measure²⁷), while varying the size of the noise [Fig. 3.7(a)— $\Pi(\rho)$: blue, \diamond ; $\Pi_{\text{FW}}(\rho)$: red, \bullet]. To quantify the performance, I have plotted the reconstruction error, $1-F$ [Fig. 3.7(a-i)], and the linear entropy of the reconstructed state [Fig. 3.7(a-ii)]. Each data point is an average over 1000 states, with the size of the Poissonian noise determined by the normalisation parameter, \overline{N} , since $n_j(\rho) = w_j(\rho)\overline{N}$ (the total number of photons = $M\overline{N}$, where $M = 36$ is the number of measurements), and the data span a range of count rates ($\overline{N} = 10-1000$). The two plots show that the FW reconstruction performs better for most of the range considered ($\overline{N} = 50$ and above). The difference is not large, but significant compared to the size of the errors ($\sim 6-12\sigma$ in this range), which are actually plotted but too small to see²⁸. These results support the argument given above, but could also be

²⁵In most experiments, the inevitable array of noise sources ensures that this almost never happens.

²⁶In photon counting (with Poissonian statistics), small counts can be measured more accurately ($\sigma \sim \sqrt{n}$), despite the fact that their *relative error* is larger ($\sim 1/\sqrt{n}$).

²⁷This procedure is explained in a footnote in Sec. 3.6.2.

²⁸The errors are the standard errors in the mean, not the standard deviation of the distribution over pure states.

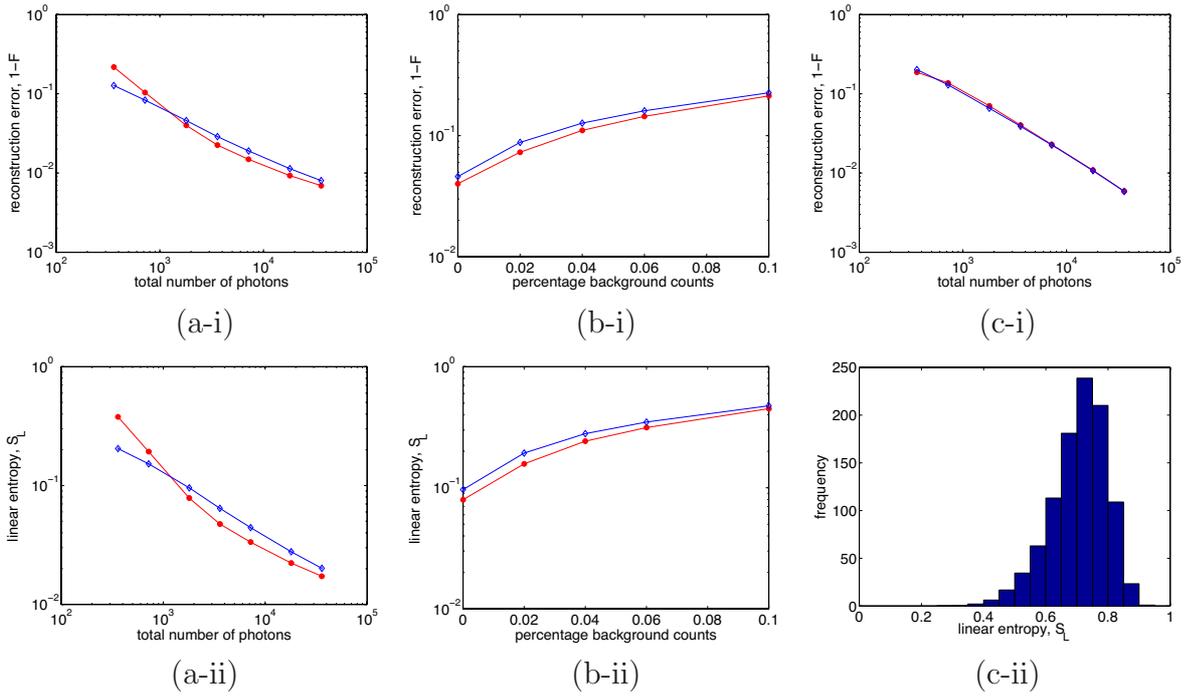


Figure 3.7: Weighting the tomographic penalty function ($\text{---}\diamond\text{---}$ (blue): maximum likelihood penalty function, $\Pi(\rho)$; $\text{---}\bullet\text{---}$ (red): fixed-weighting penalty function, $\Pi_{FW}(\rho)$). Plots of tomographic reconstruction quality for: (a) randomly selected, two-qubit pure states as a function of integration time (i.e. total photons); (b) randomly selected, two-qubit pure states in the presence of a systematic, constant background signal (target counts before noise = photons \times ideal prob. + background); and (c) randomly selected, two-qubit mixed states as a function of integration time. Reconstruction quality is quantified by (i) the reconstruction error, $1-F_t$ (F_t : target state fidelity), and (ii) for the pure states, the linear entropy, S_L . (c-ii) The histogram of linear entropy values illustrates how the mixed states are sampled.

explained by an optimisation which “pushes” unjustifiably towards pure states. However, the results in Fig. 3.7(b) suggest that is unlikely. In such a case, one might expect the FW reconstruction to perform particularly well in the presence of a uniform background signal. In these plots, as the background level (range: 0–5 photons, with $\overline{N}=50$) increases, the two optimisations give closer to the same results. In Fig. 3.7(a), the poorer performance of the FW reconstruction for low count rates may relate to the fact that counts of 0 photons are arbitrarily replaced with counts of 1 so that the penalty function is not undefined. With extremely small normalisations, counts of 0 are more likely and the change from 0 to 1 represents a larger change relative to the total number of counts. It may make more sense in the future to use a value which is a certain percentage of the normalisation parameter (e.g. replace 0s with $\overline{N}/1000$ instead of 1).

In my second main test, I compared the two optimisations for randomly selected mixed, two-qubit states, generated by tracing over half of a four-qubit pure state (chosen again according to the Haar measure) [Fig. 3.7(c-i)]. For the same range of states as in (a), the true maximum-likelihood optimisation is consistently better, although the difference between the two is much less ($\sim 0\text{--}5\sigma$). Unfortunately, as the histogram in Fig. 3.7(c-ii)

shows, the distribution of mixed states is mostly concentrated towards the high-mixture regime, and this only becomes more pronounced as the dimension of the ancilla system increases (because the random pure state becomes close to maximally entangled on average [21, 22]).

So it seems that the FW reconstruction performs better for pure states once there are sufficient counts, the standard maximum-likelihood performs slightly better for very mixed states, and there is clearly a large range of states which need to be explored more fully. These results are not yet comprehensive, but they do emphasise that the technique of tomographic reconstruction is not yet well understood. Moreover, while (or perhaps because) Π_{FW} provides more attractive results from the experimentalist's perspective, it is impossible to say impartially which is "better" purely on the basis of experimental measurements. Instead, where practical, I will generally report results based on both reconstructions.

3.8.2 Variable normalisations

In Section 3.7, I discussed the use of over-complete measurement sets in quantum tomography. In particular, I noted one important advantage, that it allows the data to be normalised within each POVM set. This has particular implications for cases where the normalisations may vary throughout the tomography, such as may result from systems where the different types of measurements may have different efficiencies, or even systems which experience some long-term drift instability. To incorporate these ideas into the optimisation requires a simple modification of the penalty function, which was originally derived using a single, constant normalisation parameter [see Sec. 3.3]. This penalty function is

$$\Pi(\rho) = \sum_j \frac{(\mathcal{N}_j - n_j(\rho))^2}{n_j(\rho)} = \sum_j \overline{\mathcal{N}}_j \frac{(\mathcal{W}_j - w_j(\rho))^2}{w_j(\rho)}, \quad (3.56)$$

where $\overline{\mathcal{N}}_j$ is the normalisation calculated using a POVM set which includes the j^{th} measurement. For numerical reasons, it may be desirable to replace this with a scaled version²⁹, $\overline{\mathcal{N}}_j/\overline{\mathcal{N}}_{\text{ave}}$. It may also be important to take this potential variation into account when performing quantum process tomography. For example, sometimes an experimentalist will have to create different input states with different brightnesses, and a quantum gate may also transmit different states with different efficiencies.

²⁹In fact, a completely general form of penalty function would incorporate an arbitrary weighting parameter.

3.9 Chapter 3 References

- [1] D. Dieks. Communication by EPR devices. *Physics Letters A*, 92:271, 1982.
- [2] W. K. Wootters and W. H. Zurek. A single quantum cannot be cloned. *Nature*, 299:802, 1982.
- [3] G. G. Stokes. On the composition and resolution of streams of polarised light from different sources. *Transactions of the Cambridge Philosophical Society*, 9:399, 1852.
- [4] A. G. White, D. F. V. James, P. H. Eberhard, and P. G. Kwiat. Nonmaximally entangled states: Production, characterization, and utilization. *Physical Review Letters*, 83(16):3103, 1999.
- [5] R. T. Thew, K. Nemoto, A. G. White, and W. J. Munro. Qudit quantum-state tomography. *Physical Review A*, 66:012303, 2002.
- [6] N. K. Langford, R. B. Dalton, M. D. Harvey, J. L. O'Brien, G. J. Pryde, A. Gilchrist, S. D. Bartlett, and A. G. White. Measuring entangled qutrits and their use for quantum bit commitment. *Physical Review Letters*, 93(5):053601, 2004.
- [7] Y. I. Bogdanov, L. A. Krivitsky, and S. P. Kulik. Statistical reconstruction of the quantum states of three-level optical systems. *Journal of Experimental and Theoretical Physics Letters*, 78:352, 2003.
- [8] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White. Measurement of qubits. *Physical Review A*, 64:052312, 2001.
- [9] A. C. Doherty and A. Gilchrist. Quantum tomography using convex optimisation. *preprint*, 2006.
- [10] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2004.
- [11] I. L. Chuang and M. A. Nielsen. Prescription for experimental determination of the dynamics of a quantum black box. *Journal of Modern Optics*, 44:2455, 1997.
- [12] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, UK, 2000.
- [13] J. L. O'Brien, G. J. Pryde, A. Gilchrist, D. F. V. James, N. K. Langford, T. C. Ralph, and A. G. White. Quantum process tomography of a controlled-NOT gate. *Physical Review Letters*, 93(8):080502, 2004.
- [14] A. G. White, A. Gilchrist, G. J. Pryde, J. L. O'Brien, M. J. Bremner, and N. K. Langford. Measuring two-qubit gates. *Journal of the Optical Society of America B*, 24(2):172, 2007.
- [15] J. R. Taylor. *An Introduction To Error Analysis*. University Science Books, Sausalito, CA, second edition, 1997.
- [16] J. T. Barreiro, N. K. Langford, N. A. Peters, and P. G. Kwiat. Generation of hyperentangled photon pairs. *Physical Review Letters*, 95(26):260501, 2005.

-
- [17] N. K. Langford, T. J. Weinhold, R. Prevedel, K. J. Resch, A. Gilchrist, J. L. O'Brien, G. J. Pryde, and A. G. White. Demonstration of a simple entangling optical gate and its use in Bell-state analysis. *Physical Review Letters*, 95(21):210504, 2005.
- [18] K. J. Resch, P. Walther, and A. Zeilinger. Full characterization of a three-photon Greenberger-Horne-Zeilinger state using quantum state tomography. *Physical Review Letters*, 94:070402, 2005.
- [19] N. Kiesel, C. Schmid, U. Weber, R. Ursin, and H. Weinfurter. Linear optics controlled-phase gate made simple. *Physical Review Letters*, 95:210505, 2005.
- [20] M. W. Mitchell, C. W. Ellenor, S. Schneider, and A. M. Steinberg. Diagnosis, prescription, and prognosis of a Bell-state filter by quantum process tomography. *Physical Review Letters*, 91:120402, 2003.
- [21] D. N. Page. Average entropy of a subsystem. *Physical Review Letters*, 71:1291, 1993.
- [22] S. Sen. Average entropy of a quantum subsystem. *Physical Review Letters*, 77:1, 1996.

Part II

Encoding and measuring information in photons

“I see nobody on the road,” said Alice.

“I only wish *I* had such eyes,” the King remarked in a fretful tone. “To be able to see Nobody! And at that distance too! Why, it’s as much as I can do to see real people, by this light!”

Through the Looking-Glass (and what Alice found there),
Lewis Carroll

Chapter 4

Encoding in polarisation

4.1 The electromagnetic field

In classical electromagnetic theory, the behaviour of an electromagnetic field is governed by Maxwell’s equations. For a medium with no free charges (and hence no currents), these are:

$$\begin{aligned}\nabla \cdot \epsilon \mathfrak{E} &= 0, \\ \nabla \cdot \mu \mathfrak{H} &= 0, \\ \nabla \times \mathfrak{E} &= -\mu \frac{\partial \mathfrak{H}}{\partial t}, \\ \nabla \times \mathfrak{H} &= \epsilon \frac{\partial \mathfrak{E}}{\partial t},\end{aligned}\tag{4.1}$$

where $\mathfrak{D} = \epsilon(\mathfrak{E})\mathfrak{E}$ and $\mathfrak{B} = \mu(\mathfrak{H})\mathfrak{H}$ are tensor equations called the *constitutive relations*, and \mathfrak{E} , \mathfrak{D} , \mathfrak{B} and \mathfrak{H} are the real vectors of electric field, electric displacement, magnetic field and magnetic intensity, respectively. The dielectric permittivity, ϵ , and the magnetic permeability, μ , can also be written in terms of the dielectric and magnetic susceptibilities, χ_e and χ_m :

$$\begin{aligned}\epsilon(\mathfrak{E}) &= \epsilon_0[1 + \chi_e(\mathfrak{E})], \\ \mu(\mathfrak{H}) &= \mu_0[1 + \chi_m(\mathfrak{H})].\end{aligned}\tag{4.2}$$

These quantities completely describe the way a medium interacts with electromagnetic fields. For light propagating in a simple dielectric, the following simplifying assumptions are normally valid—the medium is: *linear* ($\epsilon(\mathfrak{E}) = \epsilon = \epsilon_0[1 + \chi_e]$ and $\mu(\mathfrak{H}) = \mu = \mu_0[1 + \chi_m]$); *isotropic* (i.e. no birefringent effects: $(\chi_e)_{jk} = \chi_e \delta_{jk}$); *homogeneous* (ϵ and μ do not vary with position); and *non-dispersive* (ϵ and μ do not vary with frequency). For most dielectrics of interest, $\mu = \mu_0$ (they are non-magnetic).

Propagating solutions of Maxwell’s equations are called the *modes* of the electromagnetic field. To move to a quantum mechanical description, the electromagnetic field is quantised by replacing the amplitudes of these modes with quantum field operators (the *annihilation* and *creation* operators of a harmonic oscillator) [see also Sec. 2.2]. It is then possible to

derive the quantum observable for the electric field vector using Maxwell's equations, giving the result (for free-space propagation):

$$\mathfrak{E}(\mathbf{r}, t) = i \sum_j \left(\frac{\hbar\omega_j}{2\epsilon_0} \right)^{\frac{1}{2}} \left[\mathbf{E}_j(\mathbf{r})e^{-i\omega_j t} a_j - \mathbf{E}_j^*(\mathbf{r})e^{+i\omega_j t} a_j^\dagger \right]. \quad (4.3)$$

These variables can be interpreted as follows: the creation operator, a_j^\dagger , creates a photon in mode j , which describes its energy ($\hbar\omega_j$), its polarisation (the direction of $\mathbf{E}_j(\mathbf{r})$), and its spatial mode (the spatial variation of $\mathbf{E}_j(\mathbf{r})$). This also describes the different degrees of freedom of the electromagnetic field. Any or all of them can be used to store information, either in the number of photons occupying a mode, or in the field itself, through its time-frequency distribution, its polarisation or its spatial mode. Investigating these different techniques, primarily with regard to single photons, is the main focus of this thesis. In this chapter and the next, I will describe relevant background concepts for encoding information in the polarisation and spatial mode of a single photon. Encoding in its time-frequency distribution will be discussed in Chapter 8.

4.2 Polarisation and the Poincaré sphere

The polarisation of an electromagnetic wave describes the direction of its electric field vector. Under most conditions, it is perpendicular to the local direction of the wave's energy flow¹ (a transverse wave). It is thus restricted to lying in a plane, making it a discrete, binary Hilbert space—a natural qubit! If the light is *linearly polarised*, the field vector oscillates along a line; if it is *elliptically polarised*, the light traces out an ellipse in the transverse plane. For the polarisation qubit of a single-photon, the standard basis states are:

$$\begin{aligned} \text{horizontal:} & \quad |H\rangle = |1_H, 0_V\rangle \equiv \mathbf{0}, \\ \text{vertical:} & \quad |V\rangle = |0_H, 1_V\rangle \equiv \mathbf{1}, \\ \text{diagonal:} & \quad |D\rangle \equiv |H\rangle + |V\rangle = |+\rangle, \\ \text{antidiagonal:} & \quad |A\rangle \equiv |H\rangle - |V\rangle = |-\rangle, \\ \text{right-circular:} & \quad |R\rangle \equiv |H\rangle + i|V\rangle = |+i\rangle, \\ \text{left-circular:} & \quad |L\rangle \equiv |H\rangle - i|V\rangle = |-i\rangle, \end{aligned} \quad (4.4)$$

where the notation $|1_H, 0_V\rangle$ means the state with one photon in the H mode and no photons in the V mode.

These standard bases again define the axes of a sphere, called the *Poincaré sphere* [Fig. 4.1], which provides a geometrical representation of the polarisation state. This differs from the Bloch sphere (see Sec. 2.1.1) only by a matter of convention—the linear polarisation states lie on the equator of the sphere, with the circular polarisation states (instead of the logical states) at the poles. The coordinates of the polarisation state

¹The main exception to this rule arises when an electromagnetic wave is travelling through a charged medium, e.g. a plasma, where there is a longitudinal component to the polarisation. Note: the energy flow of a wave is described mathematically by its *Poynting* vector: $\mathfrak{N} = \mathfrak{E} \times \mathfrak{H}$. It is usually parallel to the wave vector of the field, \mathbf{k} , except, for example, for waves in a birefringent material.

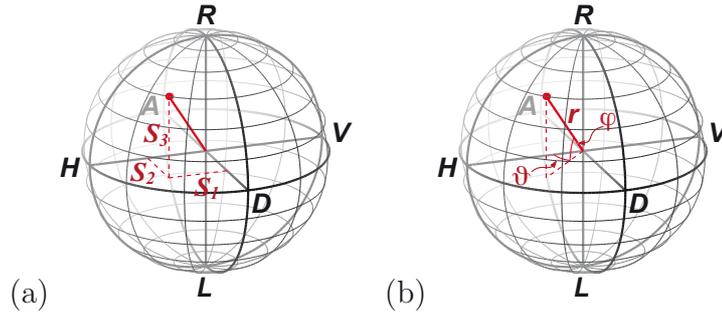


Figure 4.1: The Poincaré sphere in: (a) Stokes (Cartesian) coordinates, (S_1, S_2, S_3) ; (b) polar coordinates (r, ϑ, φ) . Note that the Poincaré form of the qubit sphere is oriented differently from the Bloch form [cf. Fig. 2.1]. In polar coordinates: r is the degree of polarisation, φ is the angle between the state and the equatorial plane of linear polarisations (S_1, S_2) , and ϑ is the angle of the state around S_3 defined relative to the horizontal polarisation.

in the Poincaré sphere representation are in fact the well-known Stokes parameters [1], $s_j = \text{Tr} \{ \rho S_j \}$, where

$$\begin{aligned}
 S_0 &= |H\rangle\langle H| + |V\rangle\langle V| = \sigma_0, \\
 S_1 &= |H\rangle\langle H| - |V\rangle\langle V| = \sigma_1, \\
 S_2 &= |D\rangle\langle D| - |A\rangle\langle A| = \sigma_2, \\
 S_3 &= |R\rangle\langle R| - |L\rangle\langle L| = \sigma_3.
 \end{aligned}
 \tag{4.5}$$

It was Stokes' great realisation that the polarisation of a beam of light could be described completely by these four parameters. In modern terms, this allows the density matrix to be rewritten,

$$\rho = \frac{1}{2} \begin{bmatrix} s_0 + s_1 & s_2 - is_3 \\ s_2 + is_3 & s_0 - s_1 \end{bmatrix}.
 \tag{4.6}$$

This was the first example of quantum state tomography (as pointed out in Ref. [2]).

4.3 Birefringence

When light is propagating through a medium which can interact with its changing electric and magnetic fields, the interactions affect the solution to Maxwell's equations. These effects can differ greatly in complexity, ranging from simple refraction to frequency mixing and down-conversion. When moving through a simple dielectric, light propagates as through a vacuum, but with a slower phase velocity, $v = c/n$, determined by the refractive index, n . This appears in the solution as a correction to the wave vector, $k = \omega n/c$. For such materials, the refractive index is uniform throughout and a wave will refract as it moves from one medium to another.

In practice, many optical media do not conform to the simple dielectric approximations, and refractive effects can be much more complicated. For example, *birefringent materials* are not isotropic, and the refractive index varies with the directions of polarisation and

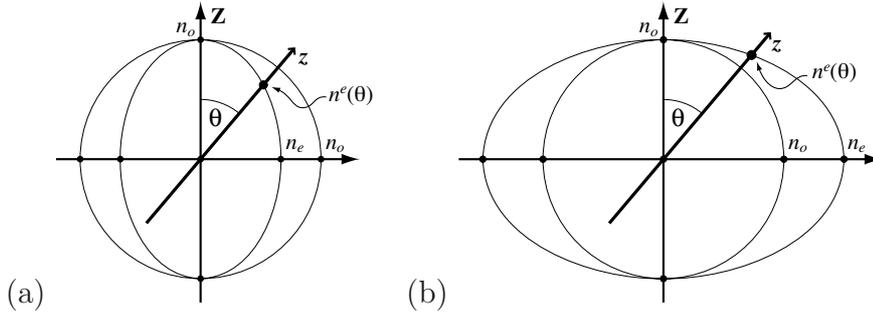


Figure 4.2: The refractive index ellipse (relative to the principal plane): (a) a negative uniaxial crystal ($n_e < n_o$); (b) a positive uniaxial crystal ($n_e > n_o$).

propagation. Understanding the consequences of this will be important in later calculations, so I present here a brief summary of the relevant ideas. In particular, I will restrict my attention to the optics of uniaxial crystals, since they are the only type used in the experiments described in this thesis. For the material in this discussion, I have mainly drawn on Dmitriev, et al., (a thorough technical overview) [3] and Hecht (an undergraduate-level introduction) [4].

The simplest birefringent crystals to understand are called *uniaxial crystals*. They have a single axis of symmetry, called the *optic axis* (the Z -axis, say), and as a result, they respond differently to light polarised along the axis (the *extraordinary ray*, or *e-ray*) and light polarised perpendicular to the axis (the *ordinary ray*, or *o-ray*). To determine how light propagates inside such a crystal, the first step is to identify these two special polarisation directions. For a transverse wave they must both be perpendicular to the direction of propagation, but by introducing the *principal plane* containing both the optic axis and the direction of propagation (the z -axis), it is easy to see that the normal to this plane always defines a polarisation which is also perpendicular to the optic axis—the ordinary ray. The extraordinary ray is then the orthogonal transverse polarisation which lies in the principal plane. Depending on the direction of propagation, this can be perpendicular to the optic axis and experience the same refractive index as the ordinary ray, n_o , or parallel, giving the extraordinary refractive index, n_e . As the direction changes between these two extremes the index variation traces out an ellipse [Fig. 4.2].

Writing the implicit equation for this ellipse as

$$\frac{x^2}{n_e^2} + \frac{y^2}{n_o^2} = 1, \quad (4.7)$$

and recognising that $x = n^e(\theta) \sin \theta$ and $y = n^e(\theta) \cos \theta$, the refractive index of the extraordinary wave can be written explicitly as a function of the angle θ between the Z and z axes:

$$n^e(\theta) = \left(\frac{\cos^2 \theta}{n_o^2} + \frac{\sin^2 \theta}{n_e^2} \right)^{-\frac{1}{2}}, \quad (4.8)$$

$$= n_o \sqrt{\frac{1 + \tan^2 \theta}{1 + n_o^2/n_e^2 \tan^2 \theta}}. \quad (4.9)$$

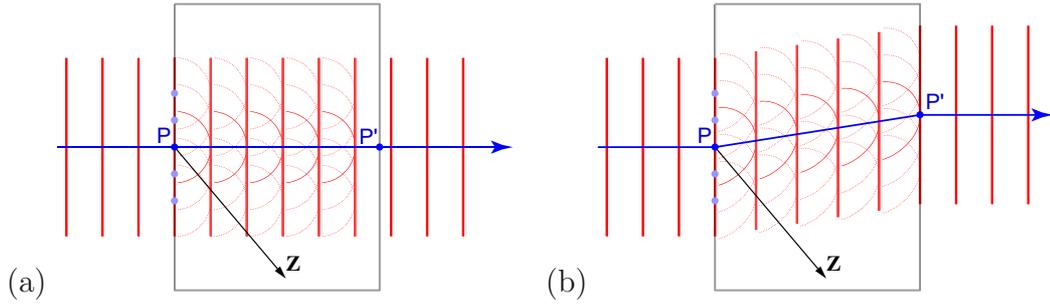


Figure 4.3: Walk-off in a negative uniaxial crystal. (a) The path of the ordinary ray. (b) The path of the extraordinary ray. The wave fronts for both rays remain parallel to the incident wave fronts throughout the crystal, but the extraordinary ray “slides” away from the optic axis. The phase of the extraordinary wave is also retarded relative to that of the ordinary wave.

Because the ordinary refractive index does not change with the direction of propagation, it traces out a circle in the principal plane.

The values of n_o and n_e depend on fundamental properties of the birefringent crystal. For a *negative uniaxial crystal* (e.g. BBO: β -Barium borate—used for our down-conversion crystals), $n_e < n_o$ and the extraordinary index ellipse lies inside the circle traced out by the ordinary refractive index [Fig. 4.2(a)]. A common example of a *positive uniaxial crystal* ($n_e > n_o$) [Fig. 4.2(b)] is quartz, which is used for our wave plates.

Because of their underlying atomic structure, crystals are also *dispersive*—their refractive indices depend on wavelength, varying according to the *Sellmeier equation*²:

$$n(\lambda) = \sqrt{A + \frac{B}{\lambda^2 - C} - D\lambda^2}. \quad (4.10)$$

The Sellmeier coefficients have been determined for many crystals to great precision by many years of careful measurements—we use these values, many of which can be found in [3].

The last effect I will discuss here is known as *birefringent walk-off* and was perhaps the first observed evidence of birefringence. It refers to the process that occurs when light, passing through an appropriately oriented birefringent crystal, splits into two distinct beams which are, respectively, ordinarily and extraordinarily polarised. The effect is sufficiently pronounced that a smallish thickness of quartz (~ 5 – 10 mm) is able to produce a double image of an object observed through the crystal³.

Walk-off occurs when a beam of light passes through a birefringent crystal which is oriented so that the optic axis is neither exactly parallel nor perpendicular to the direction of propagation as illustrated in Fig. 4.3. The effect can be understood by considering

²Other forms of the Sellmeier equation are also sometimes used, e.g. $n^2(\lambda) = A + B\lambda^2/(\lambda^2 - C) + D\lambda^2/(\lambda^2 - E)$ [5] and $n^2(\lambda) = A + B\lambda^2/(\lambda^2 - C) + D/(\lambda^2 - E)$ [3].

³This effect is in fact the origin of the term *birefringent*, since the ordinary ray *refracts* according to Snell’s law in the normal way and the extraordinary ray does not.

the Huygen's interpretation of wave propagation in terms of secondary wavelets⁴ [6]. As the light wave front travels through the crystal, the *o*-ray sees the same refractive index in all directions, so the Huygen's wavelets expand uniformly in a circle [Fig. 4.3(a)]. In contrast, the extraordinary refractive index varies with the direction of the wave vector, so the Huygen's wavelets expand with different speeds in different directions, and are thus distorted [Fig. 4.3(b)]. In both cases, the wavefronts at a later time, formed by an appropriately weighted superposition of the contributing wavelets, will remain parallel to the original wavefronts and have the same spatial profile. However, for the extraordinary beam, the energy will flow along a different direction, so the beam "slides sideways"⁵. In a negative uniaxial crystal (pictured) the phase velocity of the *e*-ray is fastest when moving perpendicularly to the optic axis ($v_o \leq v^e(\theta) \leq v_e$), and the beam bends away from the optic axis of the crystal. In a positive uniaxial crystal, the beam would bend towards the optic axis.

In addition to their anisotropy, birefringent crystals often possess a significant optical nonlinearity. In Chapter 6, I will describe how they are used in the nonlinear down-conversion process.

4.4 Wave plates

Perhaps the most important and certainly the most ubiquitous optical components made using birefringent materials are wave plates. They are primarily designed to allow precise manipulation of the polarisation of a transmitted optical field. Ideally, a wave plate is an optically flat piece of birefringent crystal (usually quartz) which is cut so that the optic axis is parallel to the front face. This means that a normally incident beam sees the maximum distinction between the ordinary and extraordinary refractive indices ($n^e(\pi/2) = n_e$). Moreover, in such a configuration, the polarisation components of a normally incident beam will not see any walk-off effects, which are potentially detrimental in sensitive interferometric experiments. In practice, this means that some care must be taken to ensure that the wave plate is aligned perpendicular to the direction of beam propagation. The difference in the refractive index for the *o*- and *e*-rays means that they experience a relative phase shift as they pass through the wave plate, and this is the means of manipulating the polarisation of the transmitted beam.

The operator describing the effect of a wave plate with the optic axis oriented horizontally

⁴This was originally a qualitative description proposed to explain these effects of reflection and refraction [6], and was later found to explain the diffraction and interference of waves passing through apertures (e.g. Young's double slit experiment). It has also been formulated into a quantitative wave theory, leading to the field of Fourier optics [7]. The basic idea is that a propagating wave front is in fact a combination of infinitesimal point-sources of secondary wavelets. At some later time, the new wave front is interpreted as the combined effect of the secondary wavelets. In the mathematical description, this sum is explicitly represented by an integral over the surface of the original wave front.

⁵In more mathematical terms, the wave front (or surface of equal phase) is, by definition, locally perpendicular to the direction of the wave vector, \mathbf{k} , which is a parameter defined in propagating solutions of Maxwell's equations. However, for a wave travelling through a dielectric, the solutions are waves of the electric displacement, \mathfrak{D} , which is not, in this case, parallel to the electric field, \mathfrak{E} . The energy flow, defined by the Poynting vector, $\mathfrak{N} = \mathfrak{E} \times \mathfrak{H}$, is perpendicular to the electric field and no longer parallel to the wave vector.

(in the lab frame) is

$$U_{\text{wp}}(\phi) = \begin{bmatrix} e^{i\frac{\phi}{2}} & 0 \\ 0 & e^{-i\frac{\phi}{2}} \end{bmatrix}, \quad (4.11)$$

where ϕ is the relative phase shift between the o and e polarisations. Here, I have ignored the global phase shift which always results from light passing through an optically dense material—this is only relevant in some experiments involving precise interference with a well-determined local oscillator. I have chosen the specific configuration with the phase shift split equally between the two polarisations for later convenience.

In general, the thickness of a wave plate must be carefully engineered to ensure precision in the relative phase shift. There are three main designs of wave-plates. *True zero-order wave plates* are thin pieces of crystal with a thickness engineered to produce the desired phase shift directly—they are expensive and easy to break. *Multi-order wave plates* are thick optical elements which separate the two polarisations by many wavelengths in phase, but produce the desired phase shift *modulo* 2π . Though thicker and easier to handle, the optical path lengths are no longer matched. We use exclusively *zero-order wave plates*, which are compound optical elements consisting of two orthogonally oriented, multi-order wave plates (cemented together or air-spaced) with thicknesses engineered to produce the same *nett* phase-shift difference as a true zero-order wave plate without being so fragile. Obviously, it would be impractical to produce or buy a different wave plate every time a different phase shift is required. Fortunately, however, that only two different types of wave plates (with $\phi = \pi$ and $\phi = \pi/2$) are required to enable arbitrary manipulation of polarisation.

The flexibility of wave plates is built in by placing them in a mount which can be rotated around an axis along the direction of normal incidence. Thus a real wave plate can be mounted with its optic axis aligned at an angle θ to horizontal:

$$U_{\text{rwp}}(\phi, \theta) = R^{\{2\}}(-\theta)U_{\text{wp}}(\phi)R^{\{2\}}(\theta), \quad (4.12)$$

where $R^{\{2\}}(\theta)$ is the matrix that rotates the reference frame by an angle θ from the horizontal, i.e. it converts vectors into coordinates of the new reference frame, so the wave plate can still be described by $U_{\text{wp}}(\phi)$. The superscript is used to indicate that $R^{\{2\}}(\theta)$ describes a rotation in a 2-dimensional *physical* space and distinguish it from the Euler rotations introduced in Sec. 2.1.2 [Eq. (2.12)], which describe rotations in qubit space as represented by the Bloch sphere or the Poincaré sphere⁶. The matrix for such a physical rotation is:

$$R^{\{2\}}(\theta) = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}. \quad (4.13)$$

The effect of the U_{rwp} operation can be understood in terms of the unitary operations discussed in Sec. 2.1.2. In fact U_{rwp} describes a subset of unitary operations, where the eigenstates lie on the equator of the Poincaré sphere and the rotation angle is the relative phase shift of the plate. In other words, a real wave plate rotates the Poincaré sphere—every state on or inside the Poincaré sphere—by an angle ϕ around the linear polarisation at an angle θ to H (ϕ is a phase, and therefore an angle on the sphere; θ is a physical angle).

⁶It is particularly important to note that angles in physical space are not the same as angles on the Poincaré sphere. For example, H and V are separated by 90° in physical space and 180° on the Poincaré sphere.

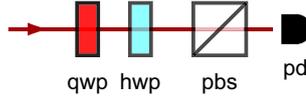


Figure 4.4: The arbitrary polarisation analyser—a polarisation filter followed by a detector (possibly a photo-diode or a photon counter). qwp: quarter-wave plate; hwp: half-wave plate; pbs: polarising beam splitter; pd: photodiode.

Therefore, the matrix for a general half-wave plate (HWP: $\phi = \pi$) is

$$U_{\text{hwp}}(\theta) = U_{\text{rwp}}(\pi, \theta) = e^{i\frac{\pi}{2}} \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}, \quad (4.14)$$

and the matrix for a quarter-wave plate (QWP: $\phi = \pi/2$) is

$$U_{\text{qwp}}(\theta) = U_{\text{rwp}}(\frac{\pi}{2}, \theta) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 + i \cos 2\theta & i \sin 2\theta \\ i \sin 2\theta & 1 - i \cos 2\theta \end{bmatrix}. \quad (4.15)$$

Obviously, half- and quarter-wave plates rotate the Poincaré sphere by 180° and 90° respectively. Because wave plates are physical objects, I will always give the optic axis setting as an angle in physical two-dimensional space, not as an angle on the qubit sphere (which is twice as big).

With the wave plate definition I have used here, successive applications of a quarter-wave plate with optic axis aligned horizontally ($\theta = 0$) transform the superposition states around the loop $D \rightarrow L \rightarrow A \rightarrow R \rightarrow D$. In other words, when looking at the Poincaré sphere with the optic axis pointing towards you, a wave plate rotates states clockwise around the sphere. Note, however, that these descriptions are only a matter of convention—any such set of definitions, consistently applied, is equivalent.

4.4.1 Using wave plates to perform polarisation analysis

There are many different optical elements which are designed to perform polarisation projections. Two common types are polarising filters (e.g. polaroid sheets) and polarising beam splitters (PBSs), which are suitable for different situations. Polarising filters operate by transmitting only the target polarisation and absorbing the orthogonal component, whereas PBSs generally act as true projectors, allowing access to both output polarisations. The most convenient PBSs for this purpose are PBS cubes, with perpendicularly oriented output beams, and polarising beam displacers, with displaced, parallel output beams. The main advantages of polarising filters are that they are easy to mount in a rotating mount and are thinner optical elements, giving rise to less beam steering. Both polarising filters and PBSs are available with high polarising quality, depending largely on the available budget. In my experiments, I have mostly used PBS cubes⁷ as polarising filters, i.e. only using the transmitted beam (polarising efficiency $\sim 99.7\%$), which provided better transmission and projection quality than the linear polarising laminated film (from

⁷Melles-Griot 03PBB012: broadband cube PBSs, $\sim 95\% : 0.3\%$ polarisation ratio in transmission at 702nm, $\sim 5\% : 99.7\%$ in reflection

target state	QWP setting	HWP setting
H	0°	0°
V	0°	45°
D	45°	22.5°
A	45°	67.5°
R	45°	45°
L	45°	0°

Table 4.1: Typical wave plate settings for polarisation analysis (assuming PBS is set to filter H).

Edmund Optics: transmission efficiency $\sim 25\%$ for unpolarised light, polarising efficiency $> 99\%$). Note that cube PBSs often have poorer polarising properties in the reflected beam.

While it is possible (though sometimes impractical) to rotate a polarisation filter to analyse different linear polarisation bases (e.g. DA instead of HV), a completely different optical element is required if one wishes to filter an elliptical polarisation state. This problem is easily overcome using wave plates. Figure 4.4 shows the polarisation analyser used throughout this thesis, consisting of a polarisation filter and a detector. A little thought is enough to see that this polarisation filter is capable of projecting any arbitrary pure state—the QWP is used to convert the (potentially elliptical) target state into a linear state which the HWP then maps onto the filtered state (here, H).

It is worthwhile to note here that great care needs to be taken to ensure that wave plates are never mounted in an experiment *backwards*, because this flips the position of the optic axis⁸ ($\theta_{OA} \rightarrow -\theta_{OA}$). This can effectively change the definitions of the reference frame orientation or phase direction.

4.5 Real wave plates in quantum state tomography

4.5.1 Real wave plates are not ideal

There are two main ways that real wave plates can behave non-ideally. There may be inaccuracies in the determined position of the optic axis, or in the size of the engineered phase shift. The first problem can cause extremely inconvenient systematic errors if undiagnosed. It can, however, be almost entirely eliminated with careful measurement techniques, especially with the precision and repeatability offered by computer-controlled rotation mounts. The second problem is trickier to deal with, because the experimentalist has no way of fixing it. Still, we can characterise these errors very precisely, and in many cases, including quantum tomography measurements, that is sufficient.

⁸The direction of rotation of an optical mount (either manual or computer-controlled) is generally labelled according to its physical orientation instead of the propagation direction of the beam.

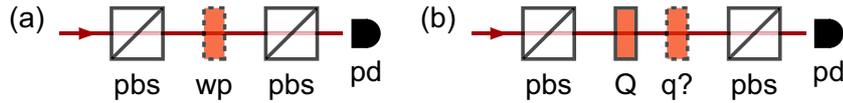


Figure 4.5: Experimental apparatus for (a) determining the optic axis position and relative phase shift for a wave plate, and (b) standardising relative alignment of all quarter-wave plates. pbs: polarising beam splitter; wp: wave plate; pd: photodiode; Q: reference quarter-wave plate; q?: quarter-wave plate being oriented.

Figure 4.5(a) illustrates a simple experimental set-up that can be used to measure both the position of the optic axis of a mounted wave plate and the relative phase shift it imparts. Figure 4.6 shows the power fringes for a numerical model of this experiment for ideal and “real” half- and quarter-wave plates. The mounted position of the optic axis (here 20°) can be determined by finding the position of the first maximum. It is very important to note that these fringes do not reveal the position of the optic axis unambiguously—they cannot differentiate between $\theta_{\text{OA}} = 20^\circ$ and $\theta_{\text{OA}} = 20^\circ + 90^\circ = 110^\circ$. While this difference is meaningless for HWPs, it is very important when characterising QWPs, because choosing the optic axis position is equivalent to choosing a definition for R and L . Thus it is vital that all quarter-wave plates are defined consistently in one way, even though it does not actually matter which way that is. Figure 4.5(b) shows a simple way to compare two quarter-wave plates. When their optic axes are oriented parallel to each other, the combined effect will look like a half-wave plate, whereas if they are oriented perpendicularly, they will act to cancel each other out. Thus, one QWP can be used as a reference to orient all others.

The relative phase shift induced by the wave plate can be determined from the depth of the fringes, normally using a numerical fit to one or more fringes. This quantity is very sensitive to any background signal produced by imperfect PBSs. Therefore, to find a reliable value for ϕ , the optimum crossed-polariser extinction must be carefully measured with no wave plates included in the set-up. There is also another ambiguity in the fringes, because they do not distinguish between ϕ and $2\pi - \phi$ phase shifts. For quarter-wave plates, this has no effect on the measured size of ϕ [Fig. 4.6(b)], essentially because only one of the two values is close to 90° . For half-wave plates, however, the two values give rise to the same fringes [Fig. 4.6(a)]. One way to resolve the ambiguity is to repeat the measurements at two different wavelengths, because the value of ϕ is wavelength-dependent. This dependence is easily calculated using the Sellmeier equations [Eq. (4.10)]. Only one value of ϕ will give meaningful results at both wavelengths.

4.6 Arbitrary single-qubit operations using wave plates

In applications of quantum information, it is vital to be able to perform arbitrary local (i.e. one-qubit) unitary rotations. How they are performed can vary greatly, however. Polarisation-based optical approaches rely on wave plates, but in many other implementations (e.g. ion-trap quantum computing), Pauli-based Bloch rotations are used. While both methods are widely understood to be able to implement arbitrary single-qubit unitaries, they operate in fundamentally different ways, which depend on what the experi-

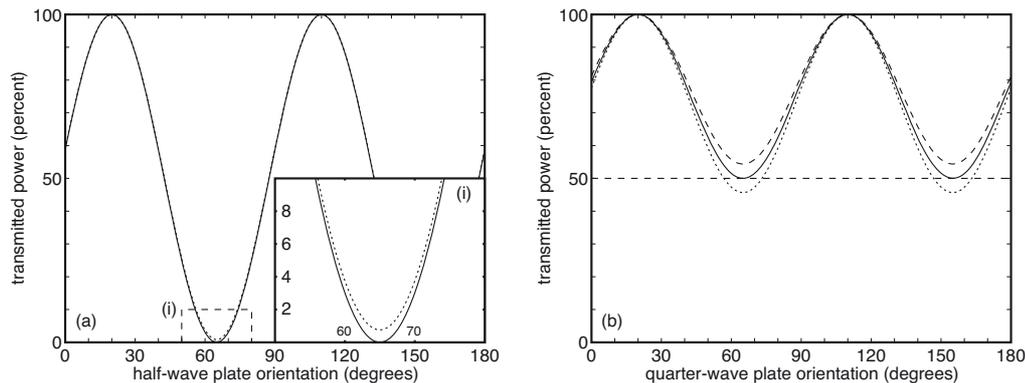


Figure 4.6: Polarisation fringes for the apparatus in Fig. 4.5 (calculated with optic axes mounted at 20°): (a) an ideal half-wave plate (solid) and a “real” HWP (dotted) with $\phi = 170^\circ$ (or $\phi = 190^\circ$); (b) an ideal quarter-wave plate (solid) and “real” QWPs with $\phi = 85^\circ$ (dashed) and $\phi = 95^\circ$ (dotted). The horizontal dashed line shows the expected visibility.

menter is able to control.

In implementations such as ion-trap quantum computing, the axes of rotations available are often fixed by external conditions such as the alignment of strong magnetic fields, but the amount of rotation can be adjusted—for example, by changing the interaction time of a laser beam used to address the ion. This is exactly the sort of situation described by the Euler decomposition [Eq. (2.12)], so it is immediately obvious how such operations can be used to perform an arbitrary unitary rotation. In particular, since three parameters are required to describe an arbitrary single-qubit unitary (ignoring the global phase), it is not surprising that the general Euler decomposition involves three successive Bloch rotations.

In polarisation implementations, the experimenter will generally only have access to quarter- and half-wave plates, but the optic axis (the axis of rotation) can be set parallel to any linear polarisation. It is not immediately obvious here how this may be used to produce arbitrary rotations—or even if it is possible. I believe this is suggestive of a wider disparity that exists between the perspectives of quantum information and the original applications in the field of optics. They are quite different paradigms.

Once the behaviour of wave plates is understood, it only takes a few moments’ thought to see that with two quarter-wave plates and a half-wave plate (in any order), any given pure state can be transformed into any desired (pure) target state⁹. However, this task is somehow less complex than performing an arbitrary single-qubit rotation, because for any specified pair of initial and target states, there are many unitary rotations which can complete the transformation (a continuous set, in fact). Each, however, has a different effect on other states on the Poincaré sphere. Thus, in the standard optics paradigm, where the initial polarisation state is known, a more simplistic understanding of how to use wave plates is sufficient. This distinction is not widely appreciated, even in the optical quantum information community, and it will lead to some profound implications for the

⁹This is easiest to understand using the Poincaré sphere. The argument follows very similar lines to that of Sec. 4.4.1 for the arbitrary polarisation analyser.

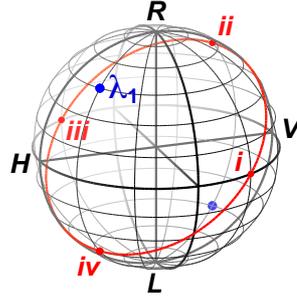


Figure 4.7: An arbitrary single-qubit unitary operation. The blue spots mark the eigenstates which are the poles of the rotation axis. Equal superpositions of these states define the corresponding equator (the red line): (i) $|\lambda_1\rangle + |\lambda_2\rangle$; (ii) $|\lambda_1\rangle + i|\lambda_2\rangle$; (iii) $|\lambda_1\rangle - |\lambda_2\rangle$; (iv) $|\lambda_1\rangle - i|\lambda_2\rangle$.

problem at hand.

In the following sections, I will propose a method which implements arbitrary single-qubit unitary operations with wave plates. I will discuss this technique in the context of Berry's phase and identify explicit prescriptions for performing this task. Finally, I will describe how these results can be used to controllably manipulate two-qubit entangled states.

4.6.1 Unitary operations

To begin with, it is important to understand more precisely the interpretation of single-qubit unitaries as rotations of the Poincaré sphere.

Consider an arbitrary unitary operation U and define its eigenvectors to be $|\lambda_k\rangle$, $k = 1, 2$. Since $U^\dagger U = I$, its eigenvalues all have a modulus of 1, i.e.,

$$U|\lambda_k\rangle = e^{i\phi_k}|\lambda_k\rangle, \quad \langle\lambda_j|\lambda_k\rangle = \delta_{jk} \quad (4.16)$$

Therefore for an arbitrary state $|\psi\rangle = \alpha_1|\lambda_1\rangle + \alpha_2|\lambda_2\rangle$, the two eigenbasis components pick up a relative phase shift under the operation of the unitary U :

$$U|\psi\rangle = e^{i\phi_1}(\alpha_1|\lambda_1\rangle + \alpha_2 e^{i\Delta_{12}}|\lambda_2\rangle) \quad (4.17)$$

where $\Delta_{12} = \phi_2 - \phi_1$.

Now, imagine that the two orthogonal eigenstates represent “poles” on the Poincaré sphere [Fig. 4.7]. The corresponding “equator” is then the set of all equal superpositions of the two eigenvectors, $\{\frac{1}{\sqrt{2}}(|\lambda_1\rangle + e^{i\phi}|\lambda_2\rangle)\}$. In other words, the relative phase shift introduced by the unitary operation causes all the states on the equator to rotate by an angle Δ_{12} around the axis defined by the poles. In fact, this is true for any line of “latitude” defined relative to the eigenstates. This illustrates the effect of an arbitrary single-qubit unitary as a rotation of the Poincaré sphere around an axis passing through the eigenstates [Eq. (2.12)]¹⁰.

¹⁰It is useful to note here that if Δ_{12} is positive, then when looking at $|\lambda_1\rangle$, the unitary rotates the sphere anticlockwise. This is opposite to a wave plate, which has a negative Δ_{12} in the definition used here.

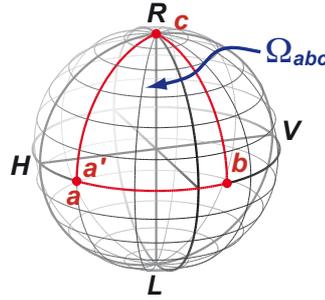


Figure 4.8: Pancharatnam’s phase. A polarisation state traversing a closed loop, abc , on the Poincaré sphere (implemented by wave plates) will accrue a geometrical phase shift related to the solid angle enclosed by the loop, Ω_{abc} .

This interpretation provides an interesting insight into the different methods of implementing local unitaries. In the Bloch rotation method, the experimenter can control the amount of rotation around several fixed rotation axes, whereas wave plates allow control over the rotation axis, but with only two available rotation angles.

4.6.2 Berry’s or Pancharatnam’s phase

In order to compare the phase of two differently polarised beams of light, Pancharatnam recognised that the natural way was to look at the phase of the overlap between the two polarisations [8]. He defined two different polarisation states to be *in phase* if the overlap (or scalar product) of the two vectors was real and positive, which Berry called *Pancharatnam’s connection* [9]. He showed that this connection was intransitive, i.e. that “ $|a\rangle$ is in phase with $|b\rangle$ ” and “ $|b\rangle$ is in phase with $|c\rangle$ ” does not imply “ $|a\rangle$ is in phase with $|c\rangle$ ”. Consider the states: $|a\rangle = |H\rangle$, $|b\rangle = |D\rangle$ and $|c\rangle = e^{-i\frac{\pi}{4}}|R\rangle$. It is easy to verify that $\langle a|b\rangle = \langle b|c\rangle = \frac{1}{\sqrt{2}}$, but that $\langle a|c\rangle = \frac{1}{\sqrt{2}}e^{-i\frac{\pi}{4}}$. He further showed that if we define a new state $|a'\rangle \propto |a\rangle$ that is *in phase* with $|c\rangle$, then:

$$\langle a|a'\rangle = e^{-i\frac{\Omega_{abc}}{2}}, \quad (4.18)$$

where Ω_{abc} is the solid angle enclosed on the Poincaré sphere by the geodesic triangle with vertices abc [see Fig. 4.8].

Later, Berry formulated a general theory describing the geometric phase accumulated by a quantum system as it evolves adiabatically around an enclosed circuit in its Hilbert space (Berry’s phase) [10], and subsequently showed Pancharatnam’s phase to be a special case of this [9]. In Berry’s original work, the system’s energy, given by its Hamiltonian, $H(\lambda)$, varies as it travels along the path, λ , and the system must be transported adiabatically to ensure “that at any instant [it] is an eigenstate of the instantaneous [Hamiltonian] H .” In other words, it must travel slowly compared to timescales determined by the rate of energy change so it remains in a definite state along the entire path. An example of such a system is a qubit encoded in a spin- $\frac{1}{2}$ ion in an external magnetic field. What makes the polarisation case unusual is that different polarisation states have the same energy. As a result, because a photon’s energy does not change as its state evolves, the evolution

does not need to be slow, and can instead be accomplished almost instantaneously with wave plates.

As an aside, it is fairly simple to show that the Berry's phase is reversed if the path is traversed in the opposite direction. For a state travelling along the path $|a\rangle \rightarrow |b\rangle \rightarrow |c\rangle \rightarrow |a'\rangle$, the phase shift is $\exp(-i\frac{\Omega_{abc}}{2})$. Consider now the extended path $|a\rangle \rightarrow |b\rangle \rightarrow |c\rangle \rightarrow |a'\rangle \rightarrow |c\rangle \rightarrow |b\rangle \rightarrow |a\rangle$. As Berry explains in Ref. [10], the geometrical phase shift is zero for a circuit which retraces itself and so encloses no area. Therefore, the phase shift for the second half of the circuit must be $\exp(i\frac{\Omega_{abc}}{2})$ to bring the final state back into phase with the initial state.

To be consistent, I will from now on define the *area enclosed by the path* on the Poincaré sphere (i.e. and thus the solid angle) to be the area on the left of an ant¹¹ walking along the path on the surface of the sphere [as shown in Fig. 4.8].

Because Berry's phase is related to the solid angle, it is also called a *geometric phase*. This is fundamentally different from the *dynamical phase* which describes the standard evolution of an eigenstate with the passing of time [10]. A simple example of a dynamical phase is a single half-wave plate. With its optical axis aligned horizontally, incident H and V optical components pick up different phases (π phase difference) even though both travel along trivial paths (paths which do not enclose any area).

4.6.3 Using Berry's phase to implement Bloch rotations

By definition, a single qubit unitary will leave its eigenstates unchanged (barring perhaps a phase factor). Therefore, in order to implement a particular unitary rotation, the required sequence of wave plates must leave the eigenstates of the rotation unchanged. In addition, to implement a rotation about this axis, the path traversed by these eigenstates must enclose a non-zero area on the Poincaré sphere, the size of which controls the angle of rotation.

The first interesting result which arises from these ideas is to use wave plates to implement the Bloch rotations [defined in Eq. (2.13)]. For such a method to be as flexible as in other systems (e.g. ion-trap experiments), it is particularly important that the amount of rotation can be controlled.

For the Poincaré (Bloch) sphere, rotations around the S_1 (z), S_2 (x) and S_3 (y) axes can be performed with 3, 3 and 2 wave plate operations respectively (minimum). The following example wave plate combinations can be verified using the wave plate operations defined earlier [summarised in Fig. 4.9]:

$$\begin{aligned}
 S_1 : \quad R_z(\theta) &= U_{\text{qwp}}\left(\frac{\pi}{4}\right) U_{\text{hwp}}\left(-\frac{\pi}{4}-\frac{\theta}{4}\right) U_{\text{qwp}}\left(\frac{\pi}{4}\right) \\
 &= \begin{bmatrix} e^{-i\frac{\theta}{2}} & 0 \\ 0 & e^{i\frac{\theta}{2}} \end{bmatrix}
 \end{aligned} \tag{4.19}$$

¹¹Note that an ant is not a physically privileged insect.

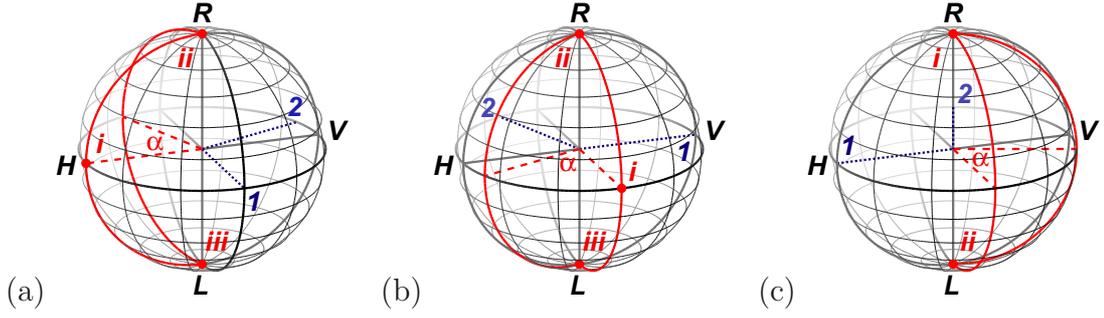


Figure 4.9: Implementing Bloch rotations with waveplates. (a,b) Waveplate posns, 1→2→1; eigenstate path, $i \rightarrow ii \rightarrow iii \rightarrow i$; (c) Waveplate posns, 1→2; eigenstate path, $i \rightarrow ii \rightarrow i$. In all cases, the central angle between the eigenstate paths is $\alpha = \theta/2$.

$$S_2 : \quad R_x(\theta) = U_{\text{qwp}}\left(\frac{\pi}{2}\right) U_{\text{hwp}}\left(-\frac{\theta}{4}\right) U_{\text{qwp}}\left(\frac{\pi}{2}\right) \\ = \begin{bmatrix} \cos \frac{\theta}{2} & -i \sin \frac{\theta}{2} \\ -i \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{bmatrix} \quad (4.20)$$

$$S_3 : \quad R_y(\theta) = U_{\text{hwp}}\left(\frac{\pi}{2} + \frac{\theta}{4}\right) U_{\text{hwp}}(0) \\ = \begin{bmatrix} \cos \frac{\theta}{2} & -\sin \frac{\theta}{2} \\ \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{bmatrix} \quad (4.21)$$

These are exactly the same as the Bloch rotations defined in Eq. (2.13). It is important to remember that the arguments to the wave plate operations are physical angles.

Not surprisingly, the wave plate settings for S_1 and S_2 are special cases of a general algorithm for a rotation about any linear polarisation, i.e. for a linear polarisation angle α , the unitary rotation about that axis is given by:

$$R_\alpha(\theta) = U_{\text{qwp}}\left(\alpha + \frac{\pi}{4}\right) U_{\text{hwp}}\left(\alpha - \frac{\pi}{4} - \frac{\theta}{4}\right) U_{\text{qwp}}\left(\alpha + \frac{\pi}{4}\right) \\ = \begin{bmatrix} \cos \frac{\theta}{2} - i \sin \frac{\theta}{2} \cos 2\alpha & -i \sin \frac{\theta}{2} \sin 2\alpha \\ i \sin \frac{\theta}{2} \sin 2\alpha & \cos \frac{\theta}{2} + i \sin \frac{\theta}{2} \cos 2\alpha \end{bmatrix} \quad (4.22)$$

In all of these cases, the solid angle enclosed by the paths of the two eigenstates can be used to calculate the phase shifts as follows:

$$\phi_1 = -\frac{1}{2} \frac{\theta}{2\pi} 4\pi = -\frac{\theta}{2} \quad (4.23)$$

$$\phi_2 = -\frac{1}{2} \frac{2\pi - \frac{\theta}{2}}{2\pi} 4\pi = -2\pi + \frac{\theta}{2} \equiv \frac{\theta}{2} \quad (4.24)$$

In other words, the matrix for the S_1 rotation would be:

$$R_z = \begin{bmatrix} e^{-i\frac{\theta}{2}} & 0 \\ 0 & e^{i\frac{\theta}{2}} \end{bmatrix}, \quad (4.25)$$

which agrees with the matrix calculated directly from the wave plate matrices.

For experimentalists, it may also be useful to be able to implement the R_y rotation with the same three-wave-plate combination used for R_x and R_z , i.e.,

$$R_y(\theta) = U_{\text{qwp}}\left(\frac{\pi}{2} + \frac{\theta}{2}\right) U_{\text{hwp}}\left(\frac{\theta}{4}\right) U_{\text{qwp}}\left(\frac{\pi}{2}\right). \quad (4.26)$$

In this configuration, there is an extra dynamical phase difference of π between the R and L eigenstates even when the solid angle is zero. This results from the half-wave plate operation and is easily compensated for by an extra geometric phase. Essentially, the dynamical phase adds an initial condition that the wave plate protocol must satisfy, i.e. $R_y(\theta)$ must produce a zero phase shift when θ is zero.

It should be noted that there is also a global phase which arises according to how the basic wave plate operation is defined, e.g.,

$$U_{\text{wp}}(\phi) = \begin{bmatrix} 1 & 0 \\ 0 & e^{-i\phi} \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} e^{i\phi} & 0 \\ 0 & 1 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} e^{i\frac{\phi}{2}} & 0 \\ 0 & e^{-i\frac{\phi}{2}} \end{bmatrix} \quad (4.27)$$

Since the basic building block of the S_j rotations is two quarter-wave plates and one half-wave plate, if either of the first two forms are used to define the basic wave plate operation, then the resulting S_j rotations experience an extra $\pm\pi$ global phase shift compared to that predicted by Berry's phase.

4.6.4 Optimal method for implementing arbitrary unitaries

Once we can implement the Bloch rotations using wave plates, it follows immediately that we can also implement arbitrary single-qubit unitary rotations using the Euler decomposition. However, using this method directly without any further thought requires at least seven wave plates (if rotation axes are y , x/z and y). If this number were in fact necessary, it would be quite inconvenient, but fortunately it can be reduced easily.

The optimal method with regard to the number of elements is the ‘‘standard’’ combination of a half-wave plate sandwiched between two quarter-wave plates (QHQ), which has always been *assumed* to be sufficient, and indeed, it *is* sufficient. However, I will show that it has some strong practical disadvantages, which can be overcome at the cost of some extra elements.

So far, I have described how to use the QHQ combination of wave plates to perform Bloch rotations and rotations around any linear polarisation axis. I now need to extend this to include rotations around elliptical polarisation states. Using a simple numerical search, it is easy to satisfy oneself that this can be done, but such a solution has limited usefulness to the experimentalist, who would prefer a more prescriptive answer. Instead, I will investigate the following question: ‘‘For a given (elliptical) rotation axis, can the eigenstates be transported through a closed path in such a way that the enclosed area can be controlled to provide the required range of phase shifts?’’ For convenience, I will define the following intermediate states: $|\psi_1\rangle = U_{\text{qwp}}(q_1)|\psi_0\rangle$, $|\psi_2\rangle = U_{\text{hwp}}(h)|\psi_1\rangle$ and $|\psi_3\rangle = U_{\text{qwp}}(q_2)|\psi_2\rangle$.

The first step in answering this question is to look at the effect of the first quarter-wave plate on an elliptical polarisation state. Figure 4.10 describes this effect for a range of elliptical states: $\vartheta_0 = \pi/4$ and $\varphi_0 = \{0, \dots, 5\}\pi/12$. The blue dots mark the initial states (λ_1) and the red curves trace out the states accessible by a single quarter-wave plate operation, $|\psi_1\rangle$, as q_1 varies between $\vartheta_0/2$ and $\vartheta_0/2 + \pi$ (in physical space); the red dots mark the states accessed by the major QWP positions, $\vartheta_0/2 + n\pi/4$ (blue dashed lines).

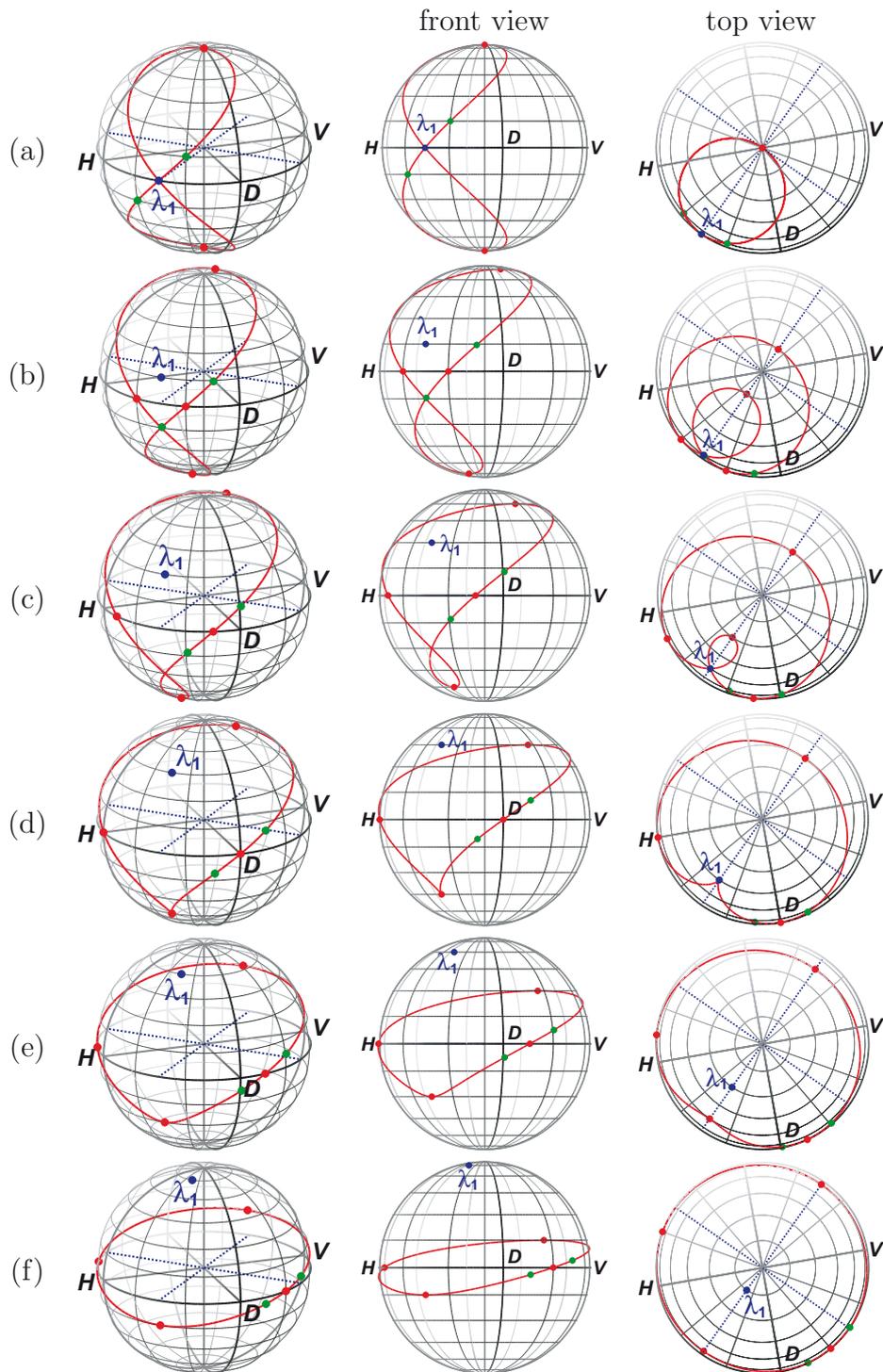


Figure 4.10: The QWP step (see text for details).

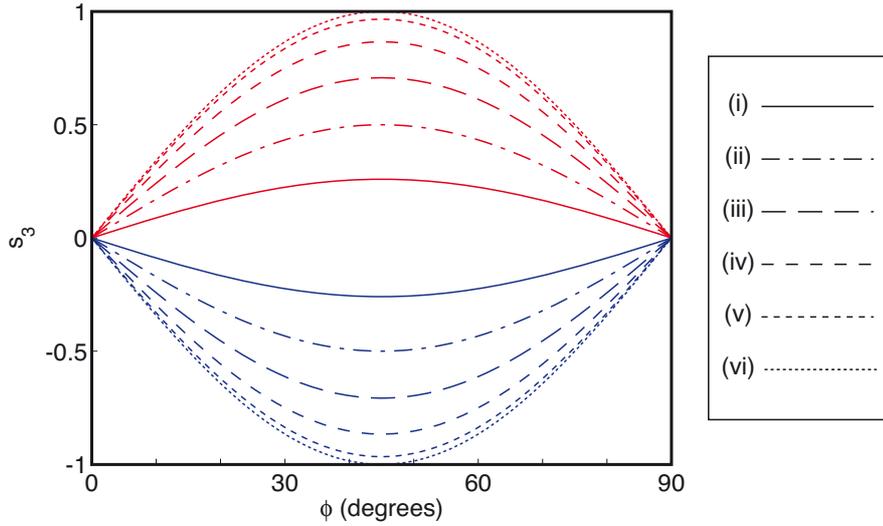


Figure 4.11: The S_3 component of $|\psi_1\rangle$, the state after the first QWP step of the optimal unitary rotation protocol. For an initial state, $|\psi_0\rangle$ at (ϑ_0, φ_0) : $|\psi_1\rangle = U_{\text{qwp}}(\vartheta_0/2 + \phi)|\psi_0\rangle$. The value of s_3 is plotted vs ϕ (red) and $\pi - \phi$ (blue): $\vartheta_0 = \frac{\pi}{4}$; $\varphi_0 = \{\frac{5\pi}{12}, \frac{4\pi}{12}, \frac{3\pi}{12}, \frac{2\pi}{12}, \frac{\pi}{12}, 0\}$ (i–vi).

The key to understanding this technique is to realise that these curves also represent the set of states which *have access to* the initial states using a single QWP—mathematically, the wave plate operation can be inverted,

$$|\psi_0\rangle = U_{\text{qwp}}\left(q_1 + \frac{\pi}{2}\right)|\psi_1\rangle. \quad (4.28)$$

In other words, if the initial state is an eigenstate of the QHQ operation, then the final quarter-wave plate must return it to its starting position. So the state $|\psi_2\rangle$ must also be in the set, $\{|\psi_1\rangle\}$, in order that $|\psi_3\rangle$ can equal $|\psi_0\rangle$. Thus, for a specified rotation axis, $\hat{n} = (\vartheta_0, \varphi_0)$, and initial QWP position, $q_1 = \vartheta_0/2 + \phi$, the HWP position (h) is immediately constrained to produce a state which also lies on the red curve.

The effect of a half-wave plate is much easier to understand. By moving its optic axis, we can select any azimuthal position (ϑ) for the output state, but the angle φ is always flipped to $-\varphi$ (i.e. $s_3 \rightarrow -s_3$). Looking at the front views in Fig. 4.10, the “northernmost” and “southernmost” points arise from $q_1 = \vartheta_0/2 \pm \pi/4$, respectively, and that they lie at equivalent “latitudes” (relative to the poles, R and L). In other words, for any ϕ , there is at least one half-wave plate position that allows the possibility that $|\psi_3\rangle = |\psi_0\rangle$. To highlight this, Fig. 4.11 plots the S_3 components of $|\psi_1\rangle$ against ϕ (in red) and $\pi - \phi$ (in blue) for all of the input states. This shows that there are in fact two possible solutions: given ϕ , the target states for the HWP operation are either $|\psi_2\rangle = |\psi_1(\pi - \phi)\rangle$ or $|\psi_2\rangle = |\psi_1(\phi + \pi/2)\rangle$. It turns out that this second possibility is the trivial solution where $q_1 = q_2 = h$ —this produces the identity operation, i.e. it does not rotate the Poincaré sphere. In other words, because there is only one interesting solution, specifying both the rotation axis and q_1 immediately specifies the other two optic axis positions (h and q_2) as well.

A summary of the optimal procedure is given below. For each initial state in Fig. 4.10, I have highlighted an example loop with $\phi = 7.5^\circ$ by marking the intermediate states with

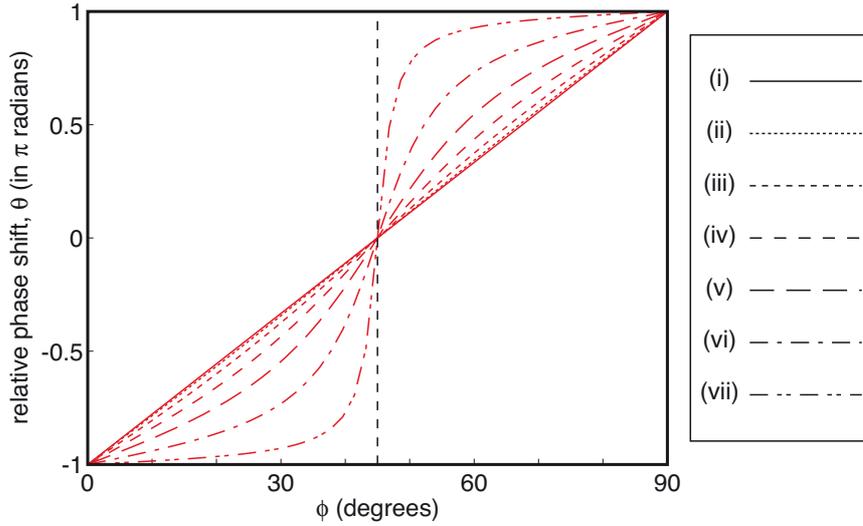


Figure 4.12: The phase shift, θ , from the optimal unitary rotation protocol, plotted against ϕ : $\vartheta_0 = \frac{\pi}{4}$; $\varphi_0 = \{\frac{\pi}{2}, \frac{5\pi}{12}, \frac{4\pi}{12}, \frac{3\pi}{12}, \frac{2\pi}{12}, \frac{\pi}{12}, \frac{\pi}{48}\}$ (i–vii).

green dots. The enclosed solid angle is the geodesic triangle with the blue and green dots as vertices.

- **Rotation Axis:** $\hat{n} = (\vartheta, \varphi)$ defines two eigenstates, $|\lambda_1\rangle \equiv |\psi_0\rangle$ and $|\lambda_2\rangle$.

- **Step 1:** $|\psi_1\rangle = U_{\text{qwp}}(\vartheta_0/2 + \phi) |\psi_0\rangle$

$$\begin{aligned} \Rightarrow s_1^{(1)} &= -\cos \varphi_0 \sin(2\phi + \vartheta_0) + \cos 2\phi \cos(2\phi + \vartheta_0) \sin \varphi_0 \\ s_2^{(1)} &= \cos \varphi_0 \cos(2\phi + \vartheta_0) + \cos 2\phi \sin(2\phi + \vartheta_0) \sin \varphi_0 \\ s_3^{(1)} &= \sin 2\phi \sin \varphi_0 \end{aligned} \quad (4.29)$$

- **Step 2:** $|\psi_1\rangle = U_{\text{hwp}}(h) |\psi_1\rangle$

$$\begin{aligned} \Rightarrow s_1^{(2)} &= \cos \varphi_0 \sin(2\phi - \vartheta_0) + \cos 2\phi \cos(2\phi - \vartheta_0) \sin \varphi_0 \\ s_2^{(2)} &= \cos \varphi_0 \cos(2\phi - \vartheta_0) - \cos 2\phi \sin(2\phi - \vartheta_0) \sin \varphi_0 \\ s_3^{(2)} &= -\sin 2\phi \sin \varphi_0 \end{aligned} \quad (4.30)$$

- **Step 3:** $|\psi_3\rangle = U_{\text{qwp}}(\vartheta_0/2 - \phi + \pi/2) |\psi_2\rangle = |\psi_0\rangle$

Given the two intermediate states, $|\psi_1\rangle$ and $|\psi_2\rangle$, the required HWP position is half-way between their two azimuthal coordinates, i.e. $h = \frac{1}{4}(\vartheta_1 + \vartheta_2)$. Since $\tan \vartheta_j = s_2^{(j)}/s_1^{(j)}$, this is a complicated function of the three unconstrained variables, ϑ_0 , φ_0 and ϕ , but it is easy to calculate numerically.

In the final stage of this analysis, I have determined the required h numerically and directly calculated the relative phase shift, θ , of the resulting QHQ operation as a function of ϕ , the position of the initial quarter-wave plate [plotted in Fig. 4.12]. There are several

important conclusions to draw from this plot. The first and most obvious is that this protocol does fulfil its objectives by providing a full 2π range of rotation around an arbitrary elliptical rotation axis which can be controlled by adjusting the position of the QWP optic axis, q_1 (and hence also h and q_2). However, it also provides the clue to the shortcomings associated with performing arbitrary rotations with just three wave plates.

The first problem arises because, for elliptical rotation axes, the phase shift exhibits a complex dependence on ϕ and φ_0 . Coupled with the complicated functional form of h , this makes it difficult to determine the wave plate settings required to produce a specified rotation. The best that can be managed is to find the required wave plate positions numerically—not an ideal solution for the experimentalist.

The second problem is most pronounced when the rotation axis is close to the linear polarisation “equator”, where the relative phase shift is particularly sensitive to the wave plate settings—most of the rotations are produced over a small range of ϕ . In fact, the shape of the phase shift curve, $\theta(\phi)$, asymptotes towards a step function. This effect is most easily understood by looking at the example “eigenloops” marked by the blue and green points in Fig. 4.10. The protocol actually breaks down for linear polarisation axes, because the area enclosed by the wave plate loop is identically zero for all ϕ [there is no geometric phase shift; see Fig. 4.10(a)], a dilemma which is only avoided by the special nature of purely linear and circular polarisations [see earlier example, Eq. (4.22)]. In many situations, this problem will not be too significant. Even for a reasonably small angle of $\varphi_0 = 15^\circ$ [Fig. 4.12(vi)], the curve is not too steep. However, in a field such as quantum information where fine control over precise operations are required, this effect could be very important, particularly since the closer the axis is to the equator, the less precise the rotation would be.

At the beginning of this section, I showed that the task of transforming between arbitrary initial and target states is inherently less complex than performing arbitrary single-qubit unitaries, since a continuous set of different unitaries can be used to produce the same transformation. I have now shown that, when using only this combination of three wave plates, the desired unitary operation can be very sensitive to the positions—the task is *ill-conditioned*, quite unlike the results for arbitrary state transformation. This contrast can be understood as a direct consequence of the difference in complexity of the two problems.

So far I have only discussed performing single-qubit unitaries with *perfect* half- and quarter-wave plates. I have not considered the problems that may arise from using real, imperfect wave plates. While I will leave a detailed consideration of these issues for future study, it is worth making a couple comments here. I showed above how arbitrary unitaries performed with the QHQ combination would be particularly sensitive to errors in the wave plate positions (though these can generally be made very small with computer control). However, accepting such errors as unavoidable, it may be possible to design approximate protocols more easily which might perform the desired unitary to within some tolerance. For example, a rotation around an axis very close to the equator of the Poincaré sphere (most sensitive to errors) could be approximated by a rotation around the closest linear polarisation (sensitivity to errors is only linear).

The other problem with real wave plates is that they do not normally produce a perfect

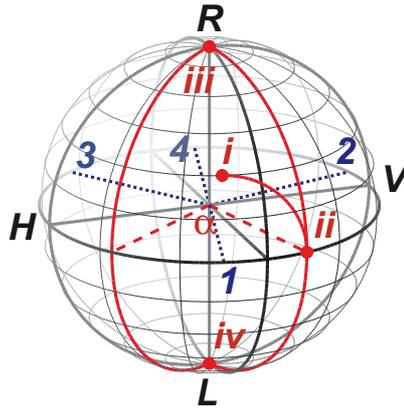


Figure 4.13: A simple, practical protocol for implementing an arbitrary unitary rotation: waveplate posns, $1 \rightarrow 2 \rightarrow 3 \rightarrow 2 \rightarrow 4$; eigenstate path, $i \rightarrow ii \rightarrow iii \rightarrow iv \rightarrow ii \rightarrow i$. Once again, the central angle between the main eigenstate path segments is $\alpha = \theta/2$.

half- or quarter-wave phase shift. Since these phases can be precisely characterised experimentally [see Sec. 4.5], they can be compensated. The main problem will be that, although this will not produce inaccurate unitaries, there will be some unitaries which are inaccessible to a given combination of real wave plates. This problem is most easily illustrated by considering some extreme (ridiculous) cases. For example, imagine that the central half-wave plate is so imperfect that it is actually a quarter-wave plate. There is no sequence of QQQ wave plate positions which will transport a linearly polarised initial around a non-trivial closed loop (one which encloses a non-zero area). Therefore, such a combination will not be able to produce any unitary rotation with an axis lying on the equator of the Poincaré sphere, except those that can be performed by a single quarter-wave plate. Alternatively, imagine that the quarter-wave plates were both so inaccurate as to be half-wave plates. Since a half-wave plate turns all states lying above the equator into states below the equator (and vice versa), all HHH combinations will have eigenstates lying on the equator, and even these will traverse trivial paths and will only experience dynamic phase shifts. In fact, any such combination is exactly equivalent to a single half-wave plate operation, which is again a very limited subset of the possible single-qubit unitaries.

4.6.5 Simple, practical method for implementing arbitrary unitaries

Both of the practical problems with the optimal rotation protocol can be overcome using extra wave plates. Equation (4.22) describes how to produce a controlled rotation around any linear polarisation axis. This can be extended to the completely general case by adding a quarter-wave plate before and after, i.e. for a rotation around the unit vector defined by the polar angles¹² $\hat{n} = (\vartheta, \varphi)$, the recipe is [see Fig. 4.13]:

$$R_{\hat{n}}(\theta) = U_{\text{qwp}}\left(\frac{\vartheta}{2} + \frac{\pi}{2}\right) R_{\frac{\vartheta}{2} + \frac{\pi}{2}}(\theta) U_{\text{qwp}}\left(\frac{\vartheta}{2}\right). \quad (4.31)$$

¹²These angles are Bloch sphere angles, not physical angles

The outside QWPs define the azimuthal position of the rotation axis (ϑ), the inside QWPs define its elevation (φ), and the HWP defines the relative phase shift. The above equation directly specifies the exact wave plate positions required for a particular operation, and also shows that they are related in a simple linear fashion to the target variables, $(\vartheta, \varphi, \theta)$.

4.6.6 Non-local Berry's phase effects

When two systems are strongly entangled, their states are so closely correlated that it becomes difficult to tell as an outside observer which system an operation is interacting with. This is illustrated by the identity¹³,

$$(U \otimes I)|\Phi\rangle = (I \otimes U^T)|\Phi\rangle, \quad (4.32)$$

where $|\Phi\rangle = 1/\sqrt{d} \sum_j |\mathbf{j}\rangle \otimes |\mathbf{j}\rangle$ is a maximally entangled state of two qudits. This has some interesting implications for results I have described above. Equations (4.14,4.15) show that $U_{\text{hwp}}^T(\theta) = U_{\text{hwp}}(\theta)$ and $U_{\text{qwp}}^T(\theta) = U_{\text{qwp}}(\theta)$, so it follows immediately that

$$\begin{aligned} (U_{\text{hwp}}(\theta) \otimes I)|\Phi^+\rangle &= (I \otimes U_{\text{hwp}}(\theta))|\Phi^+\rangle, \\ \& \quad (U_{\text{qwp}}(\theta) \otimes I)|\Phi^+\rangle &= (I \otimes U_{\text{qwp}}(\theta))|\Phi^+\rangle. \end{aligned} \quad (4.33)$$

Obviously there are similar identities for the other Bell states.

When implementing a unitary rotation on one half of an entangled state using Berry's phase, this leads to a quirky result—the component wave plate operations can be performed on different parts of the system, e.g.,

$$(I \otimes R_y(\theta))|\Phi^+\rangle = (U_{\text{hwp}}(\frac{\theta}{4}) \otimes U_{\text{hwp}}(\frac{\pi}{2}))|\Phi^+\rangle. \quad (4.34)$$

A useful consequence of this is the following result, which I will need later in Ch. 10:

$$[U_{\text{qwp}}(-\frac{\pi}{4}) U_{\text{hwp}}(\frac{\phi}{8}) \otimes U_{\text{qwp}}(\frac{\pi}{4}) U_{\text{hwp}}(-\frac{\phi}{8})] |\Phi^+\rangle = \frac{1}{\sqrt{2}} (|HH\rangle + e^{i\phi}|VV\rangle). \quad (4.35)$$

¹³This is easy to verify by direct calculation, using the observation that $A|\mathbf{j}\rangle = \sum_k A_{kj}|\mathbf{k}\rangle$.

4.7 Chapter 4 References

- [1] G. G. Stokes. On the composition and resolution of streams of polarised light from different sources. *Transactions of the Cambridge Philosophical Society*, 9:399, 1852.
- [2] A. G. White, D. F. V. James, P. H. Eberhard, and P. G. Kwiat. Nonmaximally entangled states: Production, characterization, and utilization. *Physical Review Letters*, 83(16):3103, 1999.
- [3] V. G. Dmitriev, G. G. Gurzadyan, and D. N. Nikogosyan. *Handbook of Nonlinear Optical Crystals*, volume 64 of *Springer Series in Optical Sciences*. Springer, Berlin, Heidelberg, second, revised and updated edition, 1997.
- [4] E. Hecht. *Optics*. Addison-Wesley Publishing Company, Reading, Massachusetts, second edition, 1987.
- [5] K. J. Resch. *Making photons talk to each other—nonlinear optics in the quantum domain*. PhD thesis, The University of Toronto, Toronto, Canada, 2003.
- [6] C. Huygens. *Traité de la lumière (Treatise on light)*. University of Chicago Press, 1678. Presented to the Royal Academy of Science in Paris in 1678. English translation by S. P. Thompson published in 1912.
- [7] J. W. Goodman. *Introduction to Fourier Optics*. The McGraw-Hill Companies, New York, second edition, 1996.
- [8] S. Pancharatnam. Generalized theory of interference, and its applications—i. *Proceedings of the Indian Academy of Sciences A*, 44:247, 1956.
- [9] M. V. Berry. The adiabatic phase and pancharatnam’s phase for polarized light. *Journal of Modern Optics*, 34:1401, 1987.
- [10] M. V. Berry. Quantal phase factors accompanying adiabatic changes. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 392:45, 1984.

“Who did you pass on the road?” the King went on [...]
 “Nobody,” said the Messenger.
 “Quite right,” said the King: “this young lady saw him too. So of course Nobody walks slower than you.”
 “I do my best,” the Messenger said in a sullen tone.
 “I’m sure nobody walks much faster than I do!”
 “He can’t do that,” said the King, “or else he’d have been here first.”

Through the Looking-Glass (and what Alice found there),
 Lewis Carroll

Chapter 5

Encoding in transverse spatial modes

The spatial mode of an electromagnetic wave describes the spatial profile of the intensity of the electric field, containing no information about the vector nature of the field (the polarisation). As a result, it occupies an infinite-dimensional Hilbert space inside three spatial dimensions. There are many different ways of describing the spatial mode of light relating to how one solves the spatial dependence of Maxwell’s equations—two examples are plane waves and Gaussian modes. They are all equivalent in the sense that each is a complete description of the electromagnetic field, so any field in one picture can also be written in another. However, each gives rise to a different way of encoding information, needing its own distinct experimental techniques and having different practical advantages and disadvantages in the same way that different mathematical approaches are more suited to some situations than others.

The most common method of encoding spatial information is a “which path?” technique, taking most of its features from the plane wave description of the field, which is probably the simplest, and normally the first to be discussed in any basic introduction to waves. In this method, information is stored in the direction of wave propagation (the \mathbf{k} vector) and is manipulated using *phase shifters* and *interference* at beam-splitters. Its main disadvantage is that any non-trivial circuit normally involves the alignment of extremely complicated networks of interferometers which all need to be stabilised and locked. On the other hand, in some situations this problem can be significantly reduced using cleverly designed passive stabilisation [1–3]. This method is also ideally suited to micro-optics and guided optics technologies, though it will still take some effort to completely solve the problems of stability.

More recently another method has been attracting interest, in which information is encoded into the transverse spatial profile of a propagating wave, and this is usually described in the Gaussian mode picture. In this method, the different transverse spatial modes travel along the same longitudinal path, which eliminates many of the stability problems but also introduces new challenges involved with how to manipulate the modes

individually. Investigating these new issues forms a large part of two of the main experiments contributing to this thesis [Chs 7 and 8], and I will focus on the transverse spatial approach in this chapter. Many of the relevant concepts in this approach are closely related to ideas arising in the field of quantum imaging.

One important advantage of any technique involving the spatial encoding of information is that the Hilbert space is not restricted to two dimensions as with polarisation. Thus, spatial modes can encode information in the form of qudits.

In this chapter, I will give an overview of the background concepts required to use the spatial degree of freedom in quantum information. My main reference sources have been Siegman (a thorough description of modes and beam propagation) [4], Goodman (a rigorous text on Fourier optics) [5], and the course notes from my undergraduate electromagnetic theory subject [6].

5.1 The paraxial wave equation

Recall from Eq. (4.1) that Maxwell's equations for an electromagnetic field in a dielectric are:

$$\begin{aligned}\nabla \cdot \epsilon \mathbf{E} &= 0, \\ \nabla \cdot \mu \mathbf{H} &= 0, \\ \nabla \times \mathbf{E} &= -\mu \frac{\partial \mathbf{H}}{\partial t}, \\ \nabla \times \mathbf{H} &= \epsilon \frac{\partial \mathbf{E}}{\partial t}.\end{aligned}\tag{5.1}$$

For a simple dielectric (i.e. linear, isotropic, non-dispersive and homogeneous), they can be solved to give the wave equations (using the vector identity, $\nabla \times (\nabla \times \mathbf{F}) = \nabla(\nabla \cdot \mathbf{F}) - \nabla^2 \mathbf{F}$):

$$\begin{aligned}\nabla^2 \mathbf{E} &= \mu \epsilon \frac{\partial^2 \mathbf{E}}{\partial t^2}, \\ \nabla^2 \mathbf{H} &= \mu \epsilon \frac{\partial^2 \mathbf{H}}{\partial t^2}.\end{aligned}\tag{5.2}$$

In other words, the propagation of light through a simple dielectric medium is completely described by a single wave equation:

$$\nabla^2 \mathcal{U} = \mu \epsilon \frac{\partial^2 \mathcal{U}}{\partial t^2},\tag{5.3}$$

where \mathcal{U} represents any electric or magnetic field component.

For a propagating, monochromatic wave, the real-valued fields can be written in terms of complex phasors, i.e. $\mathcal{U}(\mathbf{r}, t) = \text{Re}\{U(\mathbf{r}) \exp(-i\omega t)\}$. As a result, the wave equation simplifies to (the *Helmholtz equation*):

$$(\nabla^2 + k^2)U(\mathbf{r}) = 0,\tag{5.4}$$

where $k = n\omega/c$ is the wave vector, $c = (\mu_0\epsilon_0)^{-\frac{1}{2}}$ is the speed of light in a vacuum, and $n = c\sqrt{\mu\epsilon}$ is the refractive index. As an aside, also note the simplified form of Maxwell's

equations:

$$\begin{aligned}\nabla \cdot \epsilon \mathbf{E} &= 0, \\ \nabla \cdot \mu \mathbf{H} &= 0, \\ \nabla \times \mathbf{E} &= i\omega \mu \mathbf{H}, \\ \nabla \times \mathbf{H} &= -i\omega \epsilon \mathbf{E},\end{aligned}\tag{5.5}$$

where, e.g. $\mathfrak{E}(\mathbf{r}, t) = \text{Re}\{\mathbf{E}(\mathbf{r}) \exp(-i\omega t)\}$.

The wave equation can be further simplified by assuming that the field is a travelling wave moving in the z direction. Such a wave will display a $\exp(ikz)$ phase factor that can be included explicitly by defining

$$U(\mathbf{r}) \equiv u(\mathbf{r})e^{ikz},\tag{5.6}$$

so that the transverse field is described by $u(\mathbf{r})$. Substituting this into the wave equation gives:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} + 2ik \frac{\partial u}{\partial z} = 0.\tag{5.7}$$

The *paraxial approximation* is expressed conceptually by the idea that the main z dependence of the complex field $U(\mathbf{r})$ is contained in the rapidly rotating $\exp(ikz)$ phase term. Consequently, $u(\mathbf{r})$ will vary only slowly with z , changing mainly due to diffraction effects. Mathematically, this can be expressed by the inequalities:

$$\left| \frac{\partial^2 u}{\partial z^2} \right| \ll \left| 2k \frac{\partial u}{\partial z} \right|, \left| \frac{\partial^2 u}{\partial x^2} \right|, \left| \frac{\partial^2 u}{\partial y^2} \right|.\tag{5.8}$$

Neglecting this term as small then gives the *paraxial wave equation*,

$$\nabla_t^2 u(\mathbf{s}, z) + 2ik \frac{\partial u(\mathbf{s}, z)}{\partial z} = 0,\tag{5.9}$$

where \mathbf{s} is a two-dimensional vector describing the position in the transverse plane, e.g. (x, y) or (r, θ) , and ∇_t^2 acts on the variables in this plane. The paraxial approximation can be shown to hold for any beam for which most of the components in its plane wave decomposition make an angle of $\theta \lesssim 30^\circ$ with the z axis¹ [4]. This makes it an excellent model for almost all practical situations involving a travelling wave, i.e. most optics experiments, especially in quantum information applications.

5.2 The Gaussian mode families

In general, finding a good solution to the paraxial wave equation involves finding a whole family of solutions, because of the infinite size of the spatial Hilbert space. Although there are many ways to do this, each with different merits, the Gaussian spatial mode families are an ideal choice in most practical situations. Not only are they complete in the sense that any paraxial beam can be written as a superposition of Gaussian modes², but they

¹In fact, in the context of Huygen's integral, analytical solutions to the paraxial wave equation arise directly out of a small angle approximation.

²This is true for all beams which satisfy vanishing boundary conditions, i.e. $u(\mathbf{s}, z) \rightarrow 0$ as $\mathbf{s}, z \rightarrow \infty$.

are also analytical solutions of the paraxial wave equation, so that any superposition is guaranteed to satisfy the necessary small-angle constraints.

The Gaussian mode families are all built on the plain (lowest-order) Gaussian beam³:

$$u_g(\mathbf{s}, z) = e^{i\psi(z)} \frac{\sqrt{2}}{w(z)\sqrt{\pi}} \exp \left[\frac{ik s^2}{2R(z)} - \frac{s^2}{w^2(z)} \right] \quad (5.10)$$

where $\psi(z) = \arctan(z/z_r)$ is the fundamental Gouy phase shift [4], $2w(z) = 2w_0\sqrt{1+z^2/z_r^2}$ is the mode diameter ($1/e^2$ point in intensity), $R(z) = z(1+z_r^2/z^2)$ is the radius of curvature, $z_r = \pi w_0^2/\lambda$, w_0 is the mode diameter at the beam waist, and $s^2 = |\mathbf{s}|^2$. The first term in the main exponent describes the radial phase dependence of the beam which is caused by the curvature of the wave fronts, and the second is the standard Gaussian envelope. One of the most important properties of the Gaussian beam is the Gouy phase shift, which is the extra phase⁴ accumulated by the beam as it passes through a focus. Higher-order modes are calculated using a *separation of variables* technique to find solutions to the paraxial wave equation of the form⁵

$$u(\mathbf{s}, z) = u_g(\mathbf{s}, z) f(s_1) g(s_2) e^{i\phi(z)}. \quad (5.11)$$

When substituting this into the paraxial wave equation, different families of solutions will arise from different parametrisations of the transverse plane. There are several main families of Gaussian modes, described by the following mode equations:

(a) Hermite-Gauss modes $|HG_{mn}\rangle$:

$$u_{mn}^{HG}(x, y, z) = u_g(x, y, z) e^{i(n+m)\psi(z)} A_{mn}(x, y, z) \quad (5.12)$$

where:

$$A_{mn}(x, y, z) = \frac{1}{\sqrt{n!m!} 2^{(n+m)}} H_m \left(\frac{\sqrt{2}x}{w(z)} \right) H_n \left(\frac{\sqrt{2}y}{w(z)} \right)$$

(b) Laguerre-Gauss non-vortex modes $|LGN_{pl}\rangle$:

$$\begin{aligned} l > 0 : & \quad u_{pl}^{LGN}(r, \theta, z) = u_g(r, \theta, z) e^{i(2p+|l|)\psi(z)} A_p^{|l|}(r, z) \cos(|l|\theta) \\ l = 0 : & \quad u_{pl}^{LGN}(r, \theta, z) = u_g(r, \theta, z) e^{i(2p+|l|)\psi(z)} \frac{1}{\sqrt{2}} A_p^{|l|}(r, z) \\ l < 0 : & \quad u_{pl}^{LGN}(r, \theta, z) = u_g(r, \theta, z) e^{i(2p+|l|)\psi(z)} A_p^{|l|}(r, z) \sin(|l|\theta) \end{aligned} \quad (5.13)$$

where:

$$A_p^{|l|}(r, z) = \left(\frac{\sqrt{2}r}{w(z)} \right)^{|l|} \sqrt{\frac{2p!}{(|l|+p)!}} L_p^{|l|} \left(\frac{2r^2}{w^2(z)} \right)$$

³In most situations, where people are not specifically concerned with higher-order spatial modes, this is often just called a Gaussian beam. I will generally specify the lowest-order mode by the term ‘‘plain Gaussian’’ to distinguish it explicitly from the extended families which I will just call Gaussian modes.

⁴This phase is ‘‘extra’’ to the standard $\exp(ikz)$ propagation phase. Importantly, it is even observed for the field along the z axis ($s = 0$).

⁵For convenience, the transverse \mathbf{s} variables in $f(s_1)$ and $g(s_2)$ are normally replaced with scaled versions, $f(\xi_1)$ and $g(\xi_2)$, to incorporate the natural scaling of the propagating plain Gaussian beam. For example, with Cartesian coordinates, where $\mathbf{s} \equiv (x, y)$, then $\xi_j \equiv \sqrt{2}s_j/w(z)$.

In the above equations, H_m and $L_p^{|l|}$ are the well-known Hermite and Laguerre polynomials, which lend their names to their respective mode families. Since these modes are all derived from the plain Gaussian mode, they also obey the standard Gaussian beam propagation relations. There is also a complex-sum form of the Laguerre-Gauss family.

(c) Laguerre-Gauss vortex modes $|LGV_{pl}\rangle$:

$$u_{pl}^{LGV}(r, \theta, z) = u_g(r, \theta, z) e^{i(2p+|l|)\psi(z)} \frac{1}{\sqrt{2}} A_p^{|l|}(r, z) e^{il\theta} \quad (5.14)$$

The Gaussian mode families satisfy several practical conditions that make them very useful for quantum information purposes. Mathematically, they each form an infinite, discrete, orthonormal basis set for the space of paraxial waves, i.e.,

$$|\psi_{\text{parax}}\rangle = \sum_j c_j |u_j\rangle, \quad (5.15)$$

$$\& \quad \langle u_j | u_k \rangle = \iint d\mathbf{s} u_j^*(\mathbf{s}, z) u_k(\mathbf{s}, z) = \delta_{jk}. \quad (5.16)$$

Physically, all of the basis elements are *self-similar*, which means that the intensity pattern of the field does not change as the beam propagates⁶ (barring the overall size scaling which is a natural part of paraxial propagation). I will discuss this in detail below.

It is important to note that the higher-order Gaussian modes have the same beam parameters (spot size and radius of curvature) and as such, they obey the same Gaussian propagation relations, i.e. they focus at the same point and diverge and converge at the same rate. However, because of the shape of the special polynomials which define the mode families, the energy densities of the higher-order modes tend to be concentrated at larger distances from the axis of propagation. At the single photon level, this means that there is a larger probability of detecting a photon further away from the propagation axis.

Each of these mode families exhibits different forms of spatial symmetries in their intensity distributions [see Fig. 5.1]. The Hermite-Gauss (HG) family ($\mathbf{s} = (x, y)$) possesses rectangular symmetry, whereas the Laguerre-Gauss (LG) families ($\mathbf{s} = (r, \theta)$) possess cylindrical symmetry. In fact, the most general mode family is the Ince-Gauss (IG) family [7–9] which arises from choosing an elliptical coordinate system. It varies smoothly from displaying rectangular symmetry to cylindrical symmetry, reducing to the HG and LG mode families at the two different extremes. In all cases, the lowest-order mode is the plain Gaussian mode. This is normally a very good approximation to the output of a well-aligned laser, although lasers can also produce higher-order modes if the resonating cavity is slightly asymmetrical, whether by design or misalignment.

The symmetries of the Gaussian mode families are most easily described in terms of the phase distributions of the modes, because the phase discontinuities (π phase jumps) define the lines (or points) of zero intensity. These phase discontinuities occur at the zeroes of the corresponding special polynomial function (e.g. the Hermite polynomial). The HG mode numbers ($m, n > 0$) represent the number of linear phase jumps in the x and y directions respectively, and in the LG modes, p is the number of ring-shaped phase jumps ($p > 0$) and l is the number of radial lines of discontinuous phase (l any integer).

⁶Mathematically, each basis mode is its own Fourier transform.

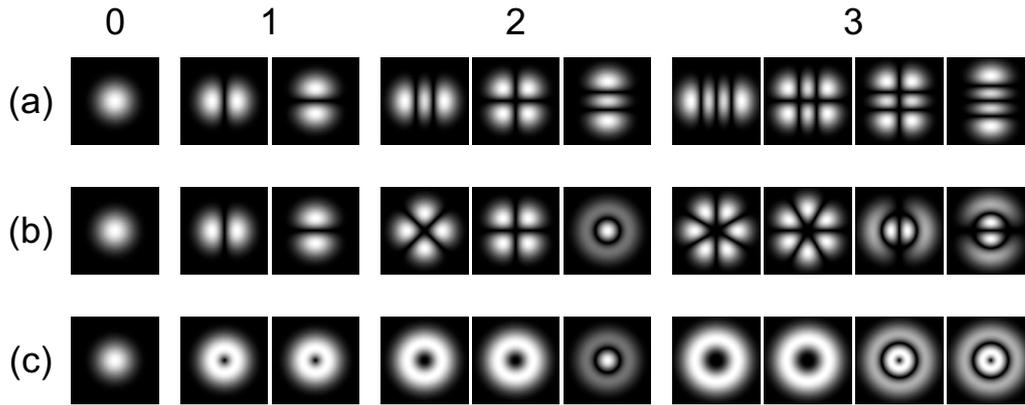


Figure 5.1: The spatial mode families for a Gaussian beam: (a) the Hermite-Gauss (HG) family—rectangular symmetry; (b) the Laguerre-Gauss non-vortex (LGN) family—cylindrical symmetry, no phase singularities; (c) the Laguerre-Gauss vortex (LGV) family—cylindrical symmetry, with phase singularities and orbital angular momentum. The mode families are divided into *mode generations*—all modes which have the same Gouy phase shift and approximately the same spatial extent.

The meaning of this last mode number defines the difference between the two Laguerre-Gauss mode families. The LGV family contains a set of modes, known as the vortex modes (VM) or optical vortices (OV) with $l \neq 0$. These contain a phase singularity ($e^{il\theta}$) and therefore possess orbital angular momentum, in the same way that circularly polarised light possesses spin angular momentum [10, 11]. This arises because the wave-fronts of the beam exhibit a helicoidal⁷ structure, spiralling around its propagation axis. The angular momentum of such a beam can be used to exert a torque on an absorbing object [12]. Unlike circularly polarised light, however, which carries at most \hbar of angular momentum per photon [13], the torque produced by a vortex mode depends on the strength of the phase singularity (called its charge). This torque is in the same sense as the helicity of the beam and has a well-defined value given by (for linearly polarised light) $l\hbar$ per photon [12, 14, 15].

5.2.1 Self-similarity and the Gaussian mode generations

As mentioned above, one of the most important properties of the Gaussian mode families is that the individual basis elements are all self-similar. But can the same be said of a beam which is a superposition of several of these basis modes? The answer to this question relies on understanding the properties of the Gouy phase shift.

Combining the mode equations of Eq. (5.10) and Eqs (5.12–5.14), it is easy to see that the Gouy phase shift for higher-order Gaussian modes explicitly depends on the mode numbers. More specifically, the Gouy phase shift is $\psi_o(z) = (o + 1)\psi(z)$, where o is the *mode order*—for HG modes, $o = m + n$; for LG modes, $o = 2p + |l|$. The mode order is a natural way to divide the Gaussian mode families into smaller groups. From now on, I will refer to the subset of $(o + 1)$ modes in a family which have the same mode order as

⁷A helicoid is a surface which follows the path of a helix, e.g. a spiral staircase.

a mode *generation*.

It is a good rule-of-thumb that modes from the same generation will have similar spatial extent. Also, when a basis element from the o_j th generation of one mode family is described in terms of another mode family (a different basis), all of the contributing elements from the new mode family will also possess the same mode order, o_j . In other words, the generational relationships of spatial modes are not tied to a single family, but remain true across all Gaussian mode families.

Now consider a beam of light which is in the following superposition state at its waist,

$$|u(0)\rangle = \frac{1}{\sqrt{2}}(|u_j\rangle + |u_k\rangle), \quad (5.17)$$

where the basis states are elements of the same mode family with mode orders o_j and o_k [define $\psi_j(z) \equiv (o_j + 1)\psi(z)$]. After propagating some distance, z , the state will evolve to

$$|u(z)\rangle = \frac{1}{\sqrt{2}}(e^{i\psi_j(z)}|u_j\rangle + e^{i\psi_k(z)}|u_k\rangle), \quad (5.18)$$

$$= \frac{1}{\sqrt{2}}e^{i\psi_j(z)}(|u_j\rangle + e^{i(o_k - o_j)\psi(z)}|u_k\rangle), \quad (5.19)$$

since, for self-similar mode elements, in the scaled coordinate system all other properties remain the same. So unless the two basis states belong to the same mode generation ($o_k - o_j = 0$), the relative phase of the superposition state will evolve as it propagates, and so will its intensity distribution. This means that such a beam will *not* be self-similar—I will call these *nondegenerate*⁸ superpositions. If, however, all elements of a superposition do belong to the same mode generation (a *degenerate* superposition), the resulting state will be self-similar.

5.2.2 Example: nondegenerate vortex mode superpositions

As an example of nondegenerate superpositions, it is particularly useful to investigate a beam which is a superposition of the plain Gaussian mode ($u_{0,0}^{LGV}$) with a vortex mode ($u_{0,m}^{LGV}$). Using the fact that $L_0^{|m|}(x) = 1$, the field equations for these modes are:

$$\begin{aligned} u_{0,0}^{LGV}(r, \theta, z) &= e^{i\psi(z)} \frac{\sqrt{2}}{w(z)\sqrt{\pi}} \exp \left[\frac{ikr^2}{2R(z)} - \frac{r^2}{w(z)^2} \right], \\ u_{0,m}^{LGV}(r, \theta, z) &= e^{i(|m|+1)\psi(z)} \frac{\sqrt{2}}{w(z)\sqrt{\pi}} \left[\frac{\sqrt{2}r}{w(z)} \right]^{|m|} \frac{1}{\sqrt{|m|!}} e^{im\theta} \\ &\quad \times \exp \left[\frac{ikr^2}{2R(z)} - \frac{r^2}{w(z)^2} \right]. \end{aligned} \quad (5.20)$$

⁸I gratefully acknowledge Paul Kwiat for suggesting this term.

Consider now the following nondegenerate superposition:

$$u(r, \theta, z) = \frac{1}{\sqrt{1+\gamma^2}} (u_{0,0}^{LGV}(r, \theta, z) + \gamma e^{i\phi} u_{0,m}^{LGV}(r, \theta, z)), \quad (5.21)$$

$$\begin{aligned} &= \frac{1}{\sqrt{1+\gamma^2}} e^{i\psi(z)} \frac{\sqrt{2}}{w(z)\sqrt{\pi}} \left(1 + \gamma e^{i\phi} e^{i|m|\psi(z)} \left[\frac{\sqrt{2}r}{w(z)} \right]^{|m|} \frac{1}{\sqrt{|m|!}} e^{im\theta} \right) \\ &\quad \times \exp \left[\frac{ikr^2}{2R(z)} - \frac{r^2}{w(z)^2} \right]. \end{aligned} \quad (5.22)$$

At the beam waist ($z = 0$), this is greatly simplified because $\psi(0) = 0$, $w(0) = w_0$ and $R(0) = \infty$:

$$u(r, \theta, 0) = \frac{1}{\sqrt{1+\gamma^2}} \frac{\sqrt{2}}{w_0\sqrt{\pi}} \left(1 + \gamma e^{i\phi} \left[\frac{\sqrt{2}r}{w_0} \right]^{|m|} \frac{1}{\sqrt{|m|!}} e^{im\theta} \right) \exp \left[-\frac{r^2}{w_0^2} \right]. \quad (5.23)$$

As before, the easiest way to understand the structure of a spatial mode is to look for the zero points in intensity. The trivial answer is that all Gaussian modes (superposition states included) must have zero intensity at $r = \infty$, a result of the $\exp(-r^2/w_0^2)$ term in the superposition. All other zeroes in intensity arise from π phase discontinuities. For this example, a non-trivial zero in intensity (amplitude) only occurs when

$$\left(1 + \gamma e^{i\phi} \left[\frac{\sqrt{2}r}{w_0} \right]^{|m|} \frac{1}{\sqrt{|m|!}} e^{im\theta} \right) = 0. \quad (5.24)$$

The solutions to this equation occur at angles which satisfy

$$\phi + m\theta = \pi + n2\pi, \quad (5.25)$$

$$\Rightarrow \theta = \frac{\pi}{m} + \frac{n}{m}2\pi - \frac{\phi}{m}, \quad (5.26)$$

and at a distance r from the axis:

$$1 = \gamma \left[\frac{\sqrt{2}r}{w_0} \right]^{|m|} \frac{1}{\sqrt{|m|!}}, \quad (5.27)$$

$$\Rightarrow r = \left(\frac{\sqrt{|m|!}}{\gamma} \right)^{\frac{1}{|m|}} \frac{w_0}{\sqrt{2}}. \quad (5.28)$$

Thus, this superposition displays m zeroes in intensity—charge ± 1 phase singularities, in fact—which are evenly spaced around the beam axis at a distance that increases with the size of the plain Gaussian contribution [Eq. (5.28)]. The initial orientation of the singularities is set by the phase ϕ . Fig. 5.2 shows the first few examples of this type of superposition.

But what happens as the beam propagates? After all, the superposition is nondegenerate. From Eq. (5.22), it is clear that the Gouy phase shift is the only relevant effect. The parameters $w(z)$ and $R(z)$ only have a global effect on the superposition (they both appear in the fundamental Gaussian mode, Eq. (5.10), common to all mode elements, and

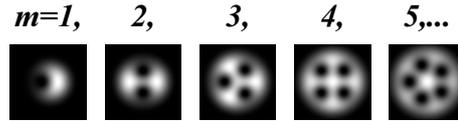


Figure 5.2: The spatial superposition $u_{0,0}^{LGV} + u_{0,m}^{LGV}$ for $m = 1 \dots 5$.

otherwise $w(z)$ is always part of a r/w ratio producing a mode-independent scaling). As discussed above, however, the Gouy phase shift introduces a relative phase shift between nondegenerate modes. In this example, this causes the positions of the phase singularities to change under propagation:

$$\theta(z) = \frac{\pi}{m} + \frac{n}{m}2\pi - \frac{\phi}{m} - \frac{|m|}{m}\psi(z). \quad (5.29)$$

Therefore, as the superposition beam propagates from $z = 0$ to $z = \infty$, the singularities will rotate by 90° in a direction determined by the sign of the charge of the vortex mode. Unless otherwise specified, I will generally assume $z = 0$ for most spatial-mode calculations.

5.3 Spatially encoded qudits

Distinguishing between degenerate and nondegenerate superpositions is vitally important when using spatial modes in quantum information applications, where the relative phase of a superposition often plays a pivotal role in the operation of a given quantum circuit. Of course, since the Gouy phase shift is well-defined and deterministic, depending on the application, it may be possible to correct for its effects.

This is one of several factors that must be considered when selecting which spatial modes are to form the basis states of an experimental quantum system. Other factors include how many basis elements are required, how they need to be manipulated, and how much the elements need to interact. Because of the size of the spatial Hilbert space, there is freedom to choose the basis states with the most suitable properties.

In the experiments contributing to this thesis, we have for practical reasons restricted ourselves to basis states belonging to the zeroth- and first-order generations, and hereafter I will use the following definitions:

$$\begin{aligned} |g\rangle &\equiv |HG_{00}\rangle = |LGN_{0,0}\rangle = |LGV_{0,0}\rangle, \\ |h\rangle &\equiv -|HG_{01}\rangle = -|LGN_{0,-1}\rangle, \\ |v\rangle &\equiv |HG_{10}\rangle = |LGN_{0,+1}\rangle, \\ |d\rangle &\equiv |h\rangle + |v\rangle, \\ |a\rangle &\equiv |h\rangle - |v\rangle, \\ |r\rangle &\equiv i|LGV_{0,+1}\rangle = |h\rangle + i|v\rangle, \\ |l\rangle &\equiv -i|LGV_{0,-1}\rangle = |h\rangle - i|v\rangle, \end{aligned} \quad (5.30)$$

where h , v , d and a refer to horizontal, vertical, diagonal and antidiagonal lines of phase discontinuity, respectively, and r (right) and l (left) describe the handedness of the optical

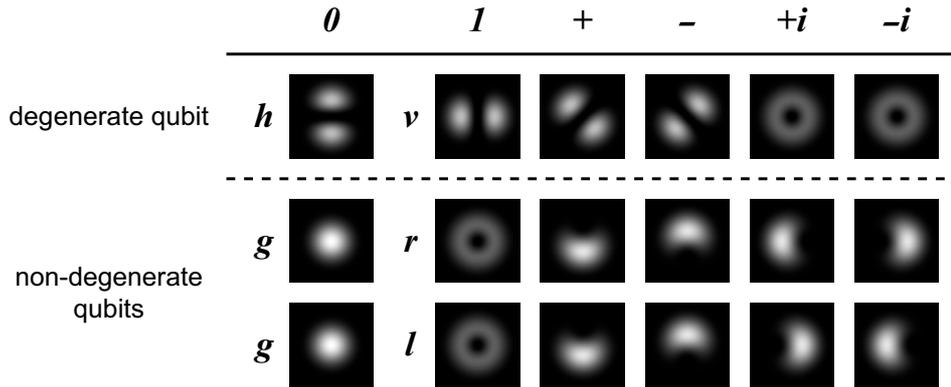


Figure 5.3: The standard bases of the spatial-mode qubits. Recall that the standard superposition basis states are: $|\pm\rangle = |\mathbf{0}\rangle \pm |\mathbf{1}\rangle$ and $|\pm i\rangle = |\mathbf{0}\rangle \pm i|\mathbf{1}\rangle$.

vortex. I have used this notation both to highlight the analogy between the first-order spatial subspace and polarisation, and to distinguish these states from their polarisation counterparts⁹.

These basic states are sufficient for studying many interesting facets of spatially encoded quantum information. We can investigate both qubit and qutrit quantum systems, and we can also explore the differences between degenerate and nondegenerate qubits. In all cases, these are the simplest examples possible with spatial modes.

5.3.1 The degenerate qubit

The degenerate spatial qubit has the following basis states [11]:

$$|\mathbf{0}\rangle \equiv |h\rangle \quad |\mathbf{1}\rangle \equiv |v\rangle. \quad (5.31)$$

The other standard basis states are therefore $|d\rangle$, $|a\rangle$, $|r\rangle$ and $|l\rangle$ [see Fig. 5.3]. As with polarisation, these six basis states are the most important qubit states, because of the role that measuring them plays in quantum tomography [see Ch. 3]. All states of this degenerate qubit belong in the 1st mode generation and therefore experience the same Gouy phase shift.

5.3.2 The nondegenerate qubits

There are two convenient nondegenerate qubits from the first two mode generations, with the basis states:

$$gr \text{ qubit:} \quad |\mathbf{0}\rangle \equiv |g\rangle \quad |\mathbf{1}\rangle \equiv |r\rangle, \quad (5.32)$$

$$gl \text{ qubit:} \quad |\mathbf{0}\rangle \equiv |g\rangle \quad |\mathbf{1}\rangle \equiv |l\rangle. \quad (5.33)$$

⁹This will be useful later, when we will study polarisation and spatial properties simultaneously.

According to the calculation at the end of Sec. 5.2, the equal superposition states which form the other two standard bases will all contain a single displaced phase singularity at a radius of $r = w_0/\sqrt{2}$ and an angle determined by Eq. (5.26) [see Fig. 5.3]. In these nondegenerate superpositions, these phase singularities will revolve around the axis of the beam as it propagates.

5.3.3 The nondegenerate qutrit

For our experiments, there was only one convenient way to define a nondegenerate qutrit using only zero- and first-order modes. The main basis states are:

$$|0\rangle \equiv |l\rangle \quad |1\rangle \equiv |g\rangle \quad |2\rangle \equiv |r\rangle. \quad (5.34)$$

For a qutrit, the simplest set of states that must be measured for tomographic characterisation are the computational basis states and the standard two-state equal superpositions of pairs of these basis states. Note that these are exactly the same as the fifteen states required to perform spatial tomographies in the different qubit subspaces.

5.4 Holograms

There are many different ways of creating and manipulating higher-order Gaussian modes [11, 16–19] (e.g. they can be actively generated inside laser cavities), but we focussed on techniques using holograms, and in particular, computer generated holograms [20–22].

The main challenge when dealing with spatial modes is that a complete description of the optical field must contain both the phase and intensity distributions. Unfortunately, image recording media, which absorb photons, store only intensity information. One way to obtain both parts of the field is by interfering the unknown (*subject*) field with a known reference beam. A hologram is a recording of this interference pattern. Conveniently, it is also fairly easy to produce a copy of the original field at a later time by shining a copy of the reference beam on the image.

Historically, many different types of holograms have been produced for a large variety of applications. In particular, using computer-generated (CG) images allows an enormous amount of flexibility to the experimenter, and in this section I discuss the main properties one can control when producing holograms. I also analyse in some more detail the basic types which are most relevant to the task of producing and manipulating higher-order Gaussian modes.

For the work in this thesis, we mainly used sinusoidal or blazed, off-axis, transmission, phase holograms—phase holograms for good diffraction efficiency; transmission holograms, because they can be produced easily on standard holographic plates; the off-axis configuration in order to obtain spatially separated diffraction orders; and sinusoidal patterns for simplicity, or blazed patterns to maximise diffraction efficiency. Apart from offering great flexibility, CG holograms also obviate the need for interferometrically stable recording apparatus. The production process involves several steps: image generation;

slide printing; holographic exposure; development; and bleaching. I have included a detailed description of the procedure for producing CG holograms and the issues involved with the generation of the images in a later experimental chapter [Sec. 7.2].

At this point it is necessary to introduce two important practical parameters which categorise the operation of a hologram—the *diffraction efficiency* and the *mode-conversion efficiency*. The diffraction efficiency is the fraction of power which is transmitted into the desired diffraction order. This is often defined relative to the total incident power, which is normally the most relevant quantity in theoretical analysis, but in some practical situations, it is more convenient to describe the efficiency in relation to the *total transmitted power*, thus neglecting reflection and leftover absorption losses. The mode-conversion efficiency describes the proportion of the diffracted power which is converted into the desired target mode.

The diffraction efficiency is used widely throughout the literature, whereas the mode-conversion efficiency is normally over-looked. In many practical situations, however, this latter quantity is vitally important, and ignoring it has led to some confusion in the literature, issues which I will discuss further in the next few sections.

5.4.1 Different types of holograms

The information in a holographic image can be stored in either two or three dimensions, called *thin* or *thick (volume)* holograms respectively. In thick holograms, the interference pattern varies with depth as well as the transverse directions. Essentially, a volume hologram is a complicated Bragg grating, the analysis of which, even in simple cases, is difficult [23]. In contrast, a thin hologram only modifies the incident field at each point of its transverse profile individually. In other words, the field leaving the hologram is the point-wise product of the input field and the transmittance (a complex transfer function describing the interference pattern),

$$u_{\text{out}}(\mathbf{s}, z) = \tau(\mathbf{s}) u_{\text{in}}(\mathbf{s}, z). \quad (5.35)$$

This allows the calculation of the field at some arbitrary point after the hologram by solving Huygen's Integral. I will focus exclusively on thin holograms, because with current technology, it is very difficult to make CG thick holograms, which means therefore that making thick holograms requires complex, interferometrically stable recording equipment.

There are several ways of categorising thin holograms. The most obvious categories are reflection (transmission) holograms which produce the reconstructed beam in the reflected (transmitted) spatial mode. High-efficiency transmission holograms are made easily using standard holographic plates.

The second type of category relates to the angle between the reference and subject beams used to create the original interference pattern. For an on-axis hologram, the reference, subject and reconstructed beams are collinear (and usually perpendicular to the image plane) which makes alignment for the reconstruction process fairly simple, but it can then be difficult to separate out the reconstructed mode(s). A Fresnel zone plate is a simple

example of an on-axis hologram¹⁰, where the diffracted modes are focussed with different focal lengths.

For off-axis holograms, the reference and subject beams are aligned at an angle, thus superimposing linear fringes over the broad features of the interference pattern. As a result, the different beams are diffracted at different angles and the reconstructed beam can be easily isolated after a sufficient distance of propagation.

The final major categories describe whether the interference pattern of the hologram is stored in the transmittance function as a phase or amplitude modulation. By their very nature, amplitude, or absorption, holograms have a low theoretical maximum diffraction efficiency for two reasons. Even for ideal amplitude holograms, which have the strongest possible modulation, there is, on average, an unmodulated component corresponding to half the incident amplitude (a quarter of the incident energy), so 25% of the light passes through undiffracted. Moreover, the absorbing and transparent regions of such holograms would have to be roughly equal in size (for an average transmittance of 50%), so they would absorb at least half the incident energy ($\frac{5}{8}$, in fact, for ideal sinusoidal holograms; see below).

Phase holograms have neither of these problems. Theoretically, they can be made transparent and can be designed to diffract 100% of the incident power into the desired diffracted mode¹¹.

5.4.2 Example: the sinusoidal plane-wave hologram

The simplest useful hologram is the off-axis, plane-wave hologram, which is created from the interference of two plane waves. For convenience, the reference beam is tilted by an angle α (in the x direction, say) to the subject beam (the z -axis), and the information about the subject mode is stored in its transverse phase profile, $\phi(\mathbf{s})$. The intensity in the interference plane ($z = 0$) is therefore:

$$I(\mathbf{s}) \sim \left| \frac{1}{\sqrt{2}} (e^{ik_x x} + e^{i\phi(\mathbf{s})}) \right|^2 = 1 + \cos[k_x x - \phi(\mathbf{s})], \quad (5.36)$$

where $k_x = k \sin \alpha$. When the interference pattern is recorded onto a photographic plate, it produces an amplitude hologram where the transmittance is related to the intensity via

$$\tau(\mathbf{s}) = 1 - \frac{1}{2}\varepsilon I(\mathbf{s}), \quad (5.37)$$

where ε is an exposure parameter (for these calculations, I will assume that the photographic emulsions have a linear exposure response¹²). Consider now a plain Gaussian

¹⁰An interference pattern between a plane wave and a spherical wave.

¹¹This *diffraction efficiency* describes only the ability of the hologram to diffract the incoming mode in the same direction as the desired output mode, not its ability to convert the mode's transverse spatial profile as desired. Thus a hologram may have an extremely good *diffraction efficiency* but a poor *mode-conversion efficiency*, and vice versa [for examples of each, see Sec. 5.5].

¹²This approximation could, if necessary, be corrected for by making careful measurements of the emulsion's actual response function and adjusting (reverse engineering) the computer-generated image to produce the correct final transmittance profile.

mode which is transmitted through the hologram (at its beam waist). The output field is [23]:

$$u_{\text{out}}(\mathbf{s}) = \tau(\mathbf{s})u_g(\mathbf{s}, 0), \quad (5.38)$$

$$\begin{aligned} &= (1 - \frac{1}{2}\varepsilon)u_g(\mathbf{s}, 0) \\ &\quad - \frac{1}{2}\varepsilon u_g(\mathbf{s}, 0)e^{i[k_x x - \phi(\mathbf{s})]} - \frac{1}{2}\varepsilon u_g(\mathbf{s}, 0)e^{-i[k_x x - \phi(\mathbf{s})]}. \end{aligned} \quad (5.39)$$

For small angles, $u_g(\mathbf{s}, 0) \exp(ik_x x)$ is a plain Gaussian beam travelling at an angle, α , to the z -axis. So the output field consists of a straight-through unmodified mode, and two diffracted modes at $\theta_d = \pm\alpha$ with phase profiles that are modified by the conjugate phase shifts, $\exp[\pm i\phi(\mathbf{s})]$, exactly as if they had passed through a simple phase plate. Not surprisingly, the best diffraction efficiency arises from the maximum-contrast transmittance ($\varepsilon = 1$), with 6.25% ($\frac{1}{16}$) of the incident power being diffracted into each of the first orders, and 25% passing straight through. The remaining $\frac{5}{8}$ of the incident beam is absorbed by the hologram.

Alternatively, the amplitude hologram can be converted to a phase hologram via a bleaching process, giving a *transparent* transmittance (ignoring an unimportant global phase shift),

$$\tau(\mathbf{s}) = e^{i\frac{1}{2}\theta_\varepsilon \cos[k_x x - \phi(\mathbf{s})]}, \quad (5.40)$$

where θ_ε is the appropriate exposure parameter. Assuming that $\phi(\mathbf{s})$ is a slowly varying function which is essentially constant on the length scale of the linear fringes ($2\pi/k_x$), the transmittance can be interpreted as a periodic function of a new composite, angular variable, $y = k_x x - \phi(\mathbf{s})$. Expanding this as a Fourier series gives:

$$\tau(\mathbf{s}) = e^{i\frac{1}{2}\theta_\varepsilon \sin(\frac{\pi}{2} - y)}, \quad (5.41)$$

$$= \sum_{n=-\infty}^{\infty} J_n\left(\frac{1}{2}\theta_\varepsilon\right) i^n e^{-iny}, \quad (5.42)$$

where J_n is a Bessel function of the first kind, order n , and I have used the identity [5],

$$e^{ia \sin y} = \sum_{n=-\infty}^{\infty} J_n(a) e^{iny}. \quad (5.43)$$

Thus, the output field for a plain Gaussian input beam is:

$$u_{\text{out}}(\mathbf{s}) = \sum_{n=-\infty}^{\infty} J_n\left(\frac{1}{2}\theta_\varepsilon\right) i^n e^{in\phi(\mathbf{s})} u_g(\mathbf{s}, 0) e^{-ink_x x}. \quad (5.44)$$

So the output from a sinusoidal phase hologram is a series of different beams, where the n^{th} order is diffracted at an angle, $\theta_d = n\alpha$, with an efficiency $J_n^2\left(\frac{1}{2}\theta_\varepsilon\right)$, and experiences a transverse phase modulation, $i^n \exp[in\phi(\mathbf{s})]$.

5.4.3 Example: engineered plane-wave holograms

Because we use CG holograms, we are not limited to these simple, sinusoidal holograms [22]. In practice, we can engineer the transmittance to be any periodic function of

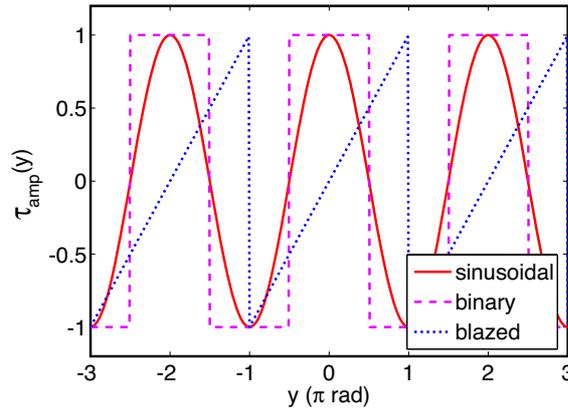


Figure 5.4: The fringe profiles for sinusoidal, blazed and binary holograms.

$$y = k_x x - \phi(\mathbf{s}):$$

$$\tau_{\text{amp}}(\mathbf{s}) = \frac{1}{2} + \frac{1}{2}\varepsilon f(y) \quad \& \quad \tau_{\text{pha}}(\mathbf{s}) = e^{i\frac{1}{2}\theta_\varepsilon f(y)}, \quad (5.45)$$

where $f(y)$ is a periodic function that varies between -1 and $+1$ with a period, 2π . These transmittance functions can then be expanded in a Fourier series:

$$\tau(y) = \sum_{n=-\infty}^{\infty} c_n e^{iny}, \quad (5.46)$$

$$\text{where } c_n = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(y) e^{-iny} dy. \quad (5.47)$$

Once again, this hologram can be interpreted as an optical element which splits an incident paraxial beam into a series of different modes with efficiencies $|c_n|^2$, where each order diffracts at an angle $\theta_d = n\alpha$ and accrues a transverse phase shift, $\exp[-in\phi(\mathbf{s})]$.

There are two main variations to the simple, sinusoidal functions above—*binary holograms* and *blazed holograms* [see Fig. 5.4]:

$$f_{\text{sin}}(y) = \cos y, \quad (5.48)$$

$$f_{\text{bla}}(y) = \frac{1}{\pi} [y + \pi \pmod{2\pi}] - 1, \quad (5.49)$$

$$f_{\text{bin}}(y) = \begin{cases} -1 : & -\pi < y < -\frac{\pi}{2} \\ +1 : & -\frac{\pi}{2} < y < +\frac{\pi}{2} \\ -1 : & +\frac{\pi}{2} < y < +\pi \end{cases}. \quad (5.50)$$

The amplitude coefficients of the Fourier series decompositions are summarised in Table 5.1, with the results for both phase- and amplitude-modulated holograms. For amplitude holograms, maximum diffraction efficiencies are only achieved for full-contrast fringes, the best being the binary structure with only around 10% into each of the $\pm k_x x$ modes. In all cases, however, the phase holograms perform much better. For binary holograms, the maximum diffraction efficiency occurs under the same conditions for all higher

	$\tau_{\text{amp}}(\mathbf{s})$	c_0	c_n	$\max c_1^2 $
	amplitude	$f_{\text{sin}}(y)$	$\frac{1}{2}$	$\frac{1}{4}\varepsilon\delta_{ n 1}$
$f_{\text{bla}}(y)$		$\frac{1}{2}$	$\frac{i}{2n\pi}\varepsilon(\cos n\pi - \text{sinc } n\pi)$	2.53% @ $\varepsilon = 1$
$f_{\text{bin}}(y)$		$\frac{1}{2}$	$\frac{1}{2}\varepsilon \text{sinc } \frac{n\pi}{2} + \frac{1}{2}(1 - \varepsilon) \text{sinc } n\pi$	10.1% @ $\varepsilon = 1$
	$\tau_{\text{pha}}(\mathbf{s})$	c_0	c_n	$\max c_1^2 $
	phase	$f_{\text{sin}}(y)$	$J_0(\frac{1}{2}\theta_\varepsilon)$	$(-i)^n J_{-n}(\frac{1}{2}\theta_\varepsilon)$
$f_{\text{bla}}(y)$		$\text{sinc } \frac{1}{2}\theta_\varepsilon$	$\text{sinc } \pi(n + \frac{1}{2}\theta_\varepsilon)$	100% @ $\theta_\varepsilon = 2\pi$
$f_{\text{bin}}(y)$		$\cos \frac{1}{2}\theta_\varepsilon$	$e^{-i\frac{1}{2}\theta_\varepsilon} \text{sinc } n\pi + i \sin \frac{1}{2}\theta_\varepsilon \text{sinc } \frac{n\pi}{2}$	40.5% @ $\theta_\varepsilon = \pi$

Table 5.1: Engineered phase and amplitude plane-wave holograms—Fourier series decomposition and diffraction efficiencies.

orders, $\theta_\varepsilon = \pi \pmod{2\pi}$. At this point, the straight through mode disappears entirely, because the rays emerging from adjacent fringes are always exactly out of phase.

The best diffraction efficiency is obtained with a blazed phase hologram, where the light can be selectively directed into a single, chosen diffraction order by setting $\theta_\varepsilon = -n2\pi$. For small angles, satisfying this condition is equivalent to ensuring that the refraction angle (set by the blazing angle¹³) matches the diffraction angle (set by the fringe spacing, $\theta_d = n\alpha$) for the desired order. The asymmetrical fringe pattern is required to diffract light to only one side of the straight through mode¹⁴.

5.4.4 An intuitive description of off-axis holograms

To understand how off-axis holograms work in a simple, intuitive way, their operation can be broken up into the effects of three different regions of the spatial frequency spectrum [Fig. 5.5]. This is reflected elegantly in the mathematics of the analytical solutions given above.

The *medium spatial frequency* is just the fringe spacing of the interference pattern, determined by k_x , and controls the overall diffraction pattern of the output, i.e. the positions of the diffracted orders ($\theta_d = n\alpha$). Mathematically, this is just the overall periodicity of the transmittance function in terms of y .

The *high spatial frequencies* relate to the shape of the individual fringe, mathematically represented by $f(y)$. Fine-tuning these frequencies gives control over the diffraction efficiencies into the various spatial orders in the form of an envelope that is superimposed

¹³In our experiments, the phase modulation is produced by a chemically induced swelling of the photographic emulsion, the size of which increases with the level of exposure. Thus, the thickness profile of the blazed phase grating is a saw-tooth pattern. See Sec. 7.2 for details.

¹⁴A sinusoidal hologram cannot distinguish between a subject beam tilted at $+\alpha$ with a phase, $-\phi(\mathbf{s})$, and one at $-\alpha$ with $+\phi(\mathbf{s})$ —the interference patterns are identical.

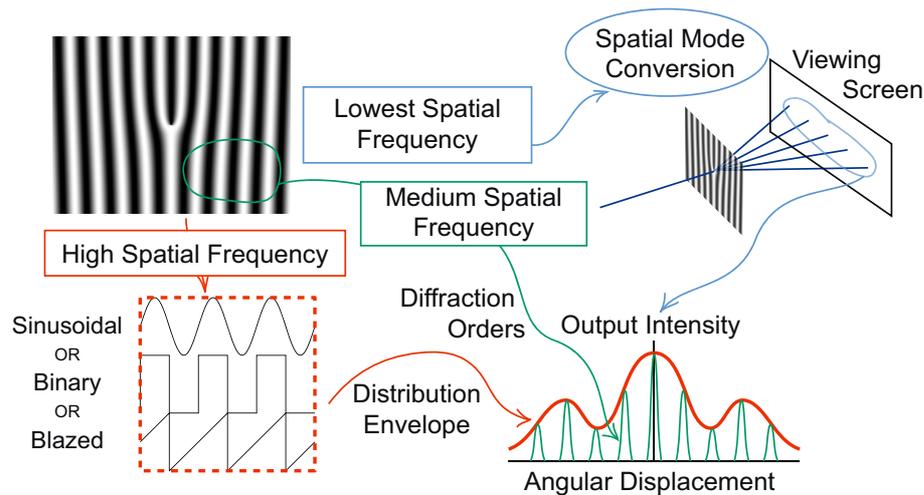


Figure 5.5: An intuitive picture of how off-axis holograms work. Their effect can be separated into three main spatial frequency regimes. The medium spatial frequency (the fringe spacing) produces diffraction orders in the output field. The high spatial frequencies (the shape of the fringes) determine the diffraction efficiencies into the different orders. The low spatial frequencies (the slow variation of the fringes across the hologram) provide the mode conversion properties of the hologram.

onto the distribution of diffracted orders which, as described above, are determined independently by the medium spatial frequency.

The *low spatial frequencies* describe the information which is stored in the phase and amplitude features of the subject beam, characterised by $\phi(\mathbf{s})$ in the above analysis. These control the *mode conversion* properties of the hologram.

This interpretation relies on the assumption that $\phi(\mathbf{s})$ is a “slowly varying” spatial distribution, although this is not actually necessary for sinusoidal amplitude holograms (see analysis above). Are there any important limitations on this assumption? For practical as well as analytical reasons, it is generally desirable to operate in this regime when using holograms. For example, when the fringe spacing is sufficiently small compared to the subject mode, the amplitude and phase information is easy to extract separately—the amplitude distribution from the peak heights of the fringes, and the phase distribution by counting fringes. Another practical advantage of operating in this regime is that the fringe profiles can be modified to fine-tune the diffraction efficiencies without having any effect on the diffraction conditions.

But how slow is “slowly varying”? Essentially, the variation of the phase distribution in the direction of the fringes should not significantly change the basic periodic structure. Depending on the subject mode, this is often not too strict a condition. In our work, the purpose of our holograms is to create or manipulate low-order Gaussian spatial modes. At the worst, these generally involve amplitude and phase distributions which are changing only a little faster than the overall Gaussian envelope (i.e. on the length scale of the spot size, not the wavelength). Typical experimental parameters give more than 12 clear fringes across the beam—enough that there is no obvious change in the fringe profiles.

The only “fast” features on these holograms are π radian discontinuities or singularities in the phase. These, however, have a limited effect for two main reasons. Because the pixels are small even compared to the fringe size (~ 18 pixels/fringe in many of our experiments), the phase jumps only have an effect in their immediate vicinity. But this is such a small area in proportion to the mode size, and the amplitude is so low, that it only has a small effect in terms of the amount of energy occupying spurious modes. Moreover, these modes will be at such high orders that they will soon diffract away from the region of interest. Essentially, these discontinuous features are too sharp to affect the diffraction properties and only appear in the “phase plate” part of the hologram’s behaviour.

I would note here that there is another useful parameter that is available when optimising the effect of holograms. In all of the examples in Table 5.1, for a specific diffraction order, the diffraction efficiency depended only on the parameter describing the fringe contrast (e.g. θ_ε). In fact, the analysis would work in exactly the same way if the fringe contrast also varied slowly in the transverse plane, i.e. $\theta_\varepsilon = \theta_\varepsilon(\mathbf{s})$. This would make the diffraction efficiency position-dependent, so that the amplitude profile in the diffracted order could be modified at the expense of some reduction in the overall diffraction efficiency.

5.5 Creating higher-order Gaussian modes

5.5.1 Using plane-wave holograms

In our experiments, we used plane-wave holograms to create higher-order Gaussian modes. These holograms incorporate only the phase distribution of the target beam, without storing the amplitude variation, i.e. $\phi(\mathbf{s}) = \angle [u_t(\mathbf{s}, z=0)]$, where $\angle \equiv \text{angle of}$ ¹⁵. For the first-order modes we used in our experiments, there are two basic hologram types—vortex holograms (for r, l) and (phase) jump holograms (for h, v, d and a). Their phase distributions are [see Fig. 5.6(a,b)]:

$$\text{vortex:} \quad \phi(\mathbf{s}) = m\theta, \quad (5.51)$$

$$\text{jump:} \quad \phi(\mathbf{s}) = \begin{cases} 0 & : y \cos \beta - x \sin \beta > 0 \\ \pi & : y \cos \beta - x \sin \beta < 0 \end{cases}, \quad (5.52)$$

where θ is the standard cylindrical coordinate and β is the angle that the line of phase discontinuity makes with the x axis (e.g. for d , $\beta = 45^\circ$).

It is important, however, to be careful when using these plane-wave holograms because of the approximation involved. When acting on a plain Gaussian mode, they do indeed create a spatial mode with the correct phase symmetry, but because there is no amplitude modulation, they do not distinguish between different higher-order modes with the same symmetry. For example, the first-order output of a charge +1 vortex mode hologram (VM_{+1}) will be a superposition of all Laguerre-Gauss modes with a charge 1 phase singularity at the centre:

$$u_{\text{out}}^{(1)} = \left(\sum_p c_p u_{p,+1}^{LGV} \right) e^{-ik_x x}. \quad (5.53)$$

¹⁵N.b. $\angle(\) \equiv \text{Im}(\log(\))$

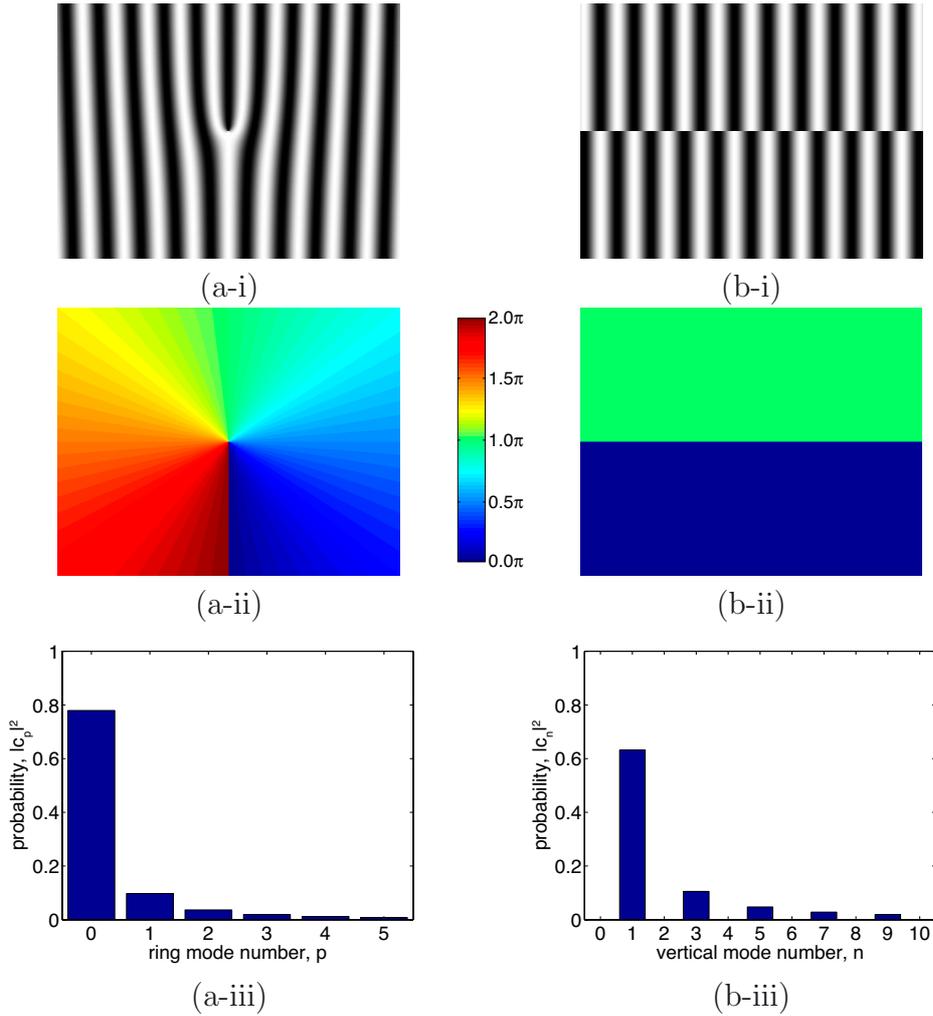


Figure 5.6: Plane-wave holograms. (i) The holograms: (a) the VM_{+1} hologram ($\eta_{mc} \sim 78\%$); and (b) the horizontal phase-jump hologram ($\eta_{mc} \sim 63\%$). (ii) The phase shifts applied by the holograms to the output modes in the first diffraction order. (iii) The contributions of the lowest-order occupied transverse modes to the first diffraction order.

Similarly, a horizontal phase jump hologram will create in the first diffraction order a superposition of all modes which have a horizontal phase discontinuity at $y = 0$ (odd functions of y), and no vertical phase discontinuities:

$$u_{\text{out}}^{(1)} = \left(\sum_{n \text{ odd}} c_n u_{0,n}^{HG} \right) e^{-ik_x x}. \quad (5.54)$$

The probabilities for the first several occupied orders of the outputs of these holograms have been plotted in Fig. 5.6(c,d).

From Figure 5.6, the mode-conversion efficiencies for the vortex hologram is around 78%, and for the jump hologram is around 63%. Sometimes this multi-mode output can be used as is and does not cause any problems, but depending on the application, it can also have a major impact on results. In particular, these output states are *nondegenerate* superpositions which means they evolve under propagation—this can be particularly important to consider when using these holograms in quantum information experiments.

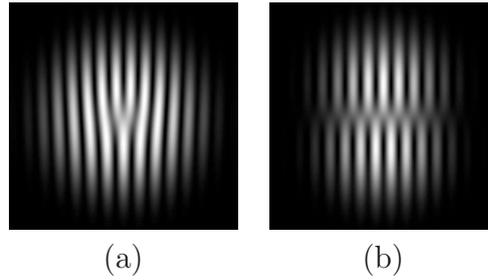


Figure 5.7: Gaussian mode holograms: (a) the VM_{+1} hologram; and (b) the horizontal phase-jump hologram.

5.5.2 Optimising mode-conversion efficiency

If it is particularly important to have a hologram with a high mode-conversion efficiency, then it must be refined to include the amplitude information which was ignored by the plane-wave holograms. For thin holograms, this is possible only at the expense of the diffraction efficiency. The first way of achieving this goal is to tailor the absorbance of the hologram across the beam width, but this is not ideal, because it is actually quite difficult to create holograms with different patterns encoded into the phase and amplitude of the interference pattern. Fortunately, the same effect can be mimicked by varying the fringe contrast across the hologram and thus creating a position-dependent diffraction efficiency (as I discussed in the previous section).

As an example, consider modifying the VM_{+1} hologram described above with the goal that a plain Gaussian reference beam produces only the lowest-order, charge +1 *LGV* mode in the first diffraction order, i.e. in the interference plane,

$$u_{0,+1}^{LGV}(\mathbf{s}, z=0) = \frac{\sqrt{2}r}{w_0} u_g(\mathbf{s}, 0) e^{-i[k_x x - \theta]}. \quad (5.55)$$

Therefore, the phase contrast distribution, $\theta_\varepsilon(\mathbf{s})$, must be engineered so that

$$\frac{\sqrt{2}r}{w_0} = |c_1[\theta_\varepsilon(\mathbf{s})]| = \text{sinc} \left[\pi + \frac{1}{2}\theta_\varepsilon(\mathbf{s}) \right]. \quad (5.56)$$

Although this will achieve the desired result, generating the CG image is computationally intensive, because inverting functions like the *sinc* and *Bessel* functions is difficult. Fortunately, Gaussian mode holograms are simpler to calculate and can produce comparable results.

5.5.3 Using Gaussian mode holograms

Gaussian mode holograms are produced by interfering two Gaussian modes, which can obviously also be simulated numerically. For example, the interference of a plain Gaussian

with a +1 vortex mode gives:

$$u(\mathbf{s}, 0) = u_g e^{ik_x x} + u_{0,+1}^{LGV} \quad (5.57)$$

$$= u_g \left(e^{ik_x x} + \frac{\sqrt{2}r}{w_0} e^{i\theta} \right) \quad (5.58)$$

The intensity pattern of this hologram is therefore [see Fig. 5.8(a)],

$$I(\mathbf{s}) = |u(\mathbf{s})|^2 = A |u_g(\mathbf{s})|^2 \left[1 + \frac{2r^2}{w_0^2} + \frac{2\sqrt{2}r}{w_0} \cos(k_x x - \theta) \right], \quad (5.59)$$

where I have included a real normalisation factor $A = [\max_{\mathbf{s}} I(\mathbf{s})]^{-1}$ to ensure that the intensity varies between 0 and 1. This is then converted to a phase hologram, giving the transmittance function, $\tau_{\text{pha}}(\mathbf{s}) = \exp(i\theta_{\varepsilon}^0 I(\mathbf{s}))$. There is an “unmodulated” (background) phase shift component which is a slowly varying function of position [red dashed line, Fig. 5.8(a)],

$$\phi_{\text{bg}}(\mathbf{s}) \sim \left[1 + \frac{2r^2}{w_0^2} \right] e^{-2r^2/w_0^2}. \quad (5.60)$$

This does not affect the diffraction properties of the hologram, but behaves essentially like a thin lens, modifying the propagation parameters of the output modes. In the modulated part of the transmittance, the standard periodic function is multiplied by another function which acts as a slowly varying phase contrast,

$$\theta_{\varepsilon}(\mathbf{s}) \sim \theta_{\varepsilon}^0 \frac{2\sqrt{2}r}{w_0} e^{-2r^2/w_0^2}. \quad (5.61)$$

This does not seem to be what is required, since there is an undesirable extra Gaussian envelope (e^{-2r^2/w_0^2}), and in any case, the efficiency does not depend linearly on the contrast, but in the sinusoidal case is $|c_n| = J_n[\frac{1}{2}\theta_{\varepsilon}(\mathbf{s})]$.

I used this information to calculate the amplitude distribution of the first diffraction order, and compare it with the target mode (the original vortex mode used to produce the hologram) in Fig. 5.8(b). I chose $\theta_{\varepsilon}^0 = 1.2\pi$, to make the diffraction efficiency (η_d) near-optimal. Using the original reference beam and expanding the output in terms of *LGV* modes with the same spot size, I plotted the results in Fig. 5.8(b-i) ($\eta_d \approx 25\%$). Although the fit is clearly poor (mode-conversion efficiency: $\eta_{\text{mc}} \approx 49\%$), the shapes are similar. The output mode looks to be smaller in diameter than expected, which is mainly due to the extra $\exp(-2r^2/w_0^2)$ variation produced by the hologram¹⁶. In fact, with the same input beam, the output shows a much closer fit to a beam with a spot size, $w_{\text{out}} = w_0/\sqrt{3}$ ($\eta_{\text{mc}} \approx 88\%$) [Fig. 5.8(b-ii)].

Further improvement is possible by noting from Eq. (5.61) that there is an extra factor of 2 in the phase contrast which can be absorbed into the spot size of the input beam, $w_{\text{in}} = w_0/2$. Figure 5.8(b-iii) shows the new fit, with a corresponding mode-conversion efficiency of $\approx 98\%$ (for $w_{\text{out}} = w_{\text{in}}/1.3$) and a similar diffraction efficiency ($\approx 26\%$). This is approximately the same amount of power as was contained in the *correct* first-order spatial mode for the plain-wave hologram ($\sim 78\% \times 34\%$). Thus, by sacrificing some overall

¹⁶ $|u_{\text{out}}^{(1)}(\mathbf{s})| \sim |c_1| |u_{\text{in}}| \approx \exp(-2r^2/w_0^2) \exp(-r^2/w_0^2)$

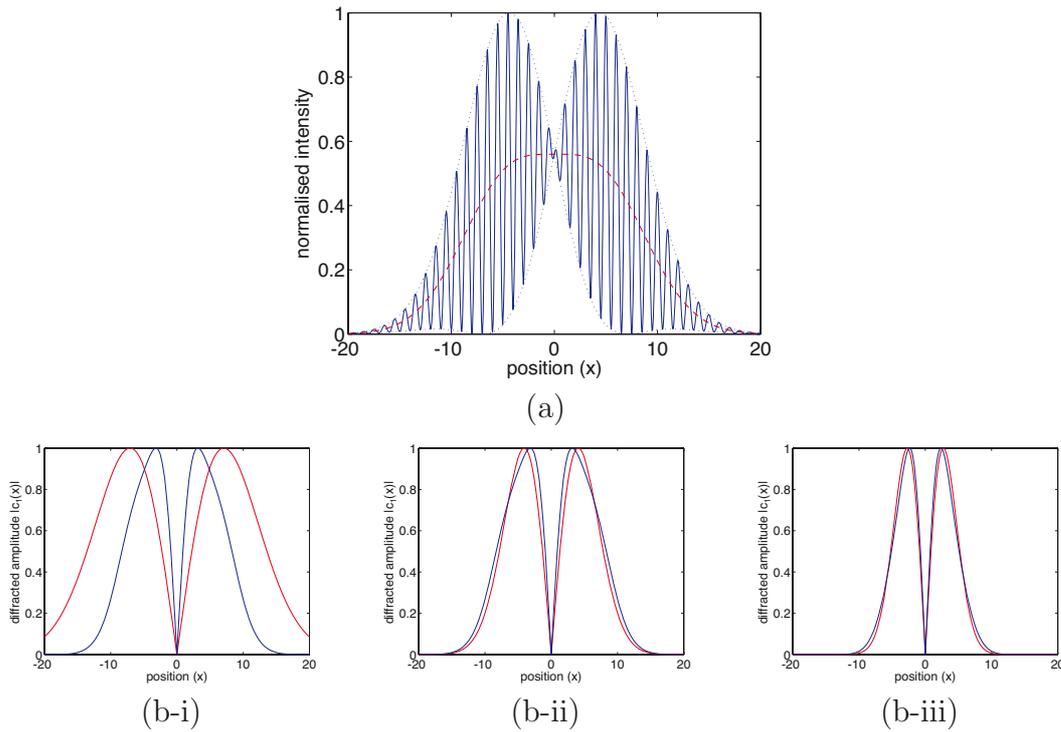


Figure 5.8: Optimising the VM_{+1} Gaussian mode hologram. (a) The x dependence of the interference pattern ($y=0$). I have ignored the $e^{i\theta}$, which is irrelevant in these calculations. (b) A comparison of the first diffraction order’s calculated amplitude variation (blue) with the “target” pattern (red) for the phase hologram in (a) and a range of beam parameters: (i) $w_{\text{in}}=w_0$, $w_{\text{out}}=w_{\text{in}}$, $\eta_{\text{mc}}\approx 49\%$; (ii) $w_{\text{in}}=w_0$, $w_{\text{out}}=w_{\text{in}}/\sqrt{3}$, $\eta_{\text{mc}}\approx 88\%$; (iii) $w_{\text{in}}=w_0/2$, $w_{\text{out}}=w_{\text{in}}/1.3$, $\eta_{\text{mc}}\approx 98\%$. w_{in} is the spot size of the incident mode (the blue curves), and w_{out} is the spot size of the “target” beam (the red curves).

diffraction efficiency, the Gaussian mode hologram produces a much closer approximation of the desired higher-order mode. However, this improvement also involves an increase in experimental complexity—not only must the hologram be mode-matched to the input beam, but it also modifies the underlying mode propagation parameters, making mode-matching even more complex.

There is one further complication to using Gaussian mode holograms—they are not “reversible” in the sense that a hologram which creates a particular spatial mode with high mode-conversion efficiency does not generally analyse the same spatial mode with equal, or even comparable efficacy. This is demonstrated in Fig. 5.9 using the example of VM_{+1} holograms. Figure 5.9(a) shows the output of a sinusoidal, plane-wave VM_{+1} hologram decomposed into the different diffraction orders (n) and higher-order modes (displaying intensities/probabilities). For ease of visualisation, I have only included the occupied vortex modes (i.e. $l = n$), plotting only the higher-order ring modes (p). As expected, this hologram creates a superposition of ring modes in each diffraction order ($\eta_{\text{mc}}\sim 78\%$). Figure 5.9(b) shows the output of the optimised Gaussian mode hologram described above ($w_{\text{in}}=w_0/2$, $w_{\text{out}}=w_{\text{in}}/1.3$), where around 96% of the power is contained in the lowest-order ring modes and the mode-conversion efficiency is $\eta_{\text{mc}}\sim 98\%$. In Figs 5.9(c,d), however, the same hologram is used as mode analyser for the two modes $LGV_{0,-1}$ and $LGV_{1,-1}$, respec-

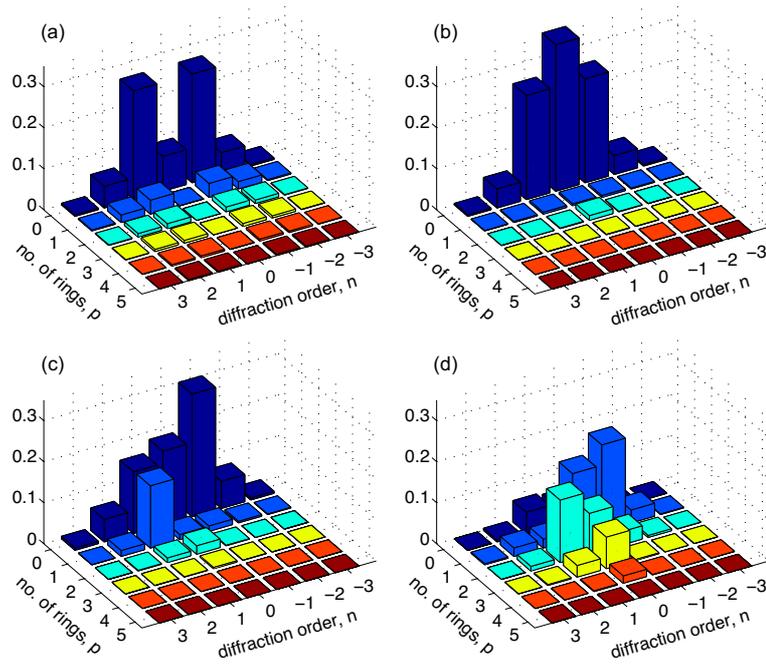


Figure 5.9: VM_{+1} holograms: (a) the mode-decomposition of the output from a sinusoidal, plane-wave hologram ($w_{\text{in}} = w_{\text{out}} = w_0$); (b,c,d) the output from a Gaussian mode hologram ($w_{\text{in}}=w_0/2$, $w_{\text{out}}=w_{\text{in}}/1.3$) for the input states $LGV_{0,0}$, $LGV_{0,-1}$ and $LGV_{1,-1}$, respectively.

tively. In these cases, the charge 0 modes emerge in the +1 diffraction order ($l=-1+n$). Unfortunately, the $LGV_{0,-1}$ input mode produces not only the plain Gaussian in the first order, but also a large component of the ring mode, $LGV_{1,0}$ [see (c)]. Moreover, the $LGV_{1,-1}$ also diffracts a significant amount of power into the first-order plain Gaussian mode. For spatial mode analysis (see below), this means that a spatial filter will see contributions from a superposition of several modes. In fact, it is impossible to use a thin diffractive element to convert most higher-order modes into another mode with 100% mode-conversion efficiency.

In all of our experiments, we used plane-wave holograms because they are easy to make and to align. It does not seem that Gaussian mode holograms provide any advantage in analysing spatial modes, although they offer significant improvements when preparing them. Indeed, it may ultimately even prove that plane-wave holograms produce the highest-efficiency measurements (in terms of mode-conversion) that can be made with thin holograms.

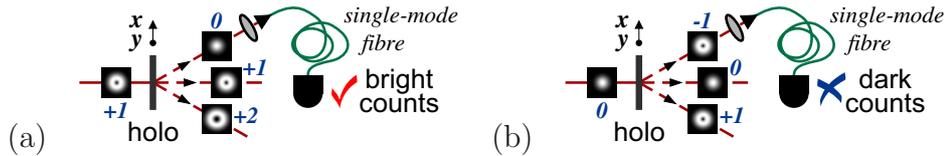


Figure 5.10: The Fourier transform recognition technique for spatial mode analysis. (a) The target mode (here, VM_{+1}) is converted into the lowest-order Gaussian mode which is filtered by a single-mode fibre. (b) Any other mode (e.g. VM_0) is converted into a higher-order mode and is interferometrically rejected by the single-mode fibre.

5.6 Spatial mode analysis

5.6.1 Detecting higher-order modes

In order to analyse the spatial mode of an optical beam experimentally, it is necessary to build a component which can detect higher-order spatial modes. A number of methods have been used to approach this task both qualitatively and quantitatively. In Ref. [17], the amount of HG_{01} was distinguished from HG_{10} by placing a localised detector on the phase discontinuity (the dark line) of the HG_{10} mode and in the centre of a lobe of the HG_{01} mode. Others (Refs [16, 19, 21, 22]) have analysed vortex modes by combining them either with themselves or a plain Gaussian mode in an interferometer and analysing the resulting intensity patterns (e.g. a vortex mode gives rise to a split fringe in the interference pattern). The intensity distributions of various modes have also been analysed directly and compared with numerical simulations to determine the combination of component beams [18, 24]. In contrast, the authors of Ref. [25] used a special interferometer involving Dove prisms which sends vortex modes of opposite chirality through two separate exit ports, distinguishing, for example, between the $VM_{\pm 1}$ modes. They then expanded this idea to build a vortex mode splitter, which they have shown to separate the $\{VM_0, VM_1, VM_2, VM_3\}$ modes of a coherent state [26]. However, most of these examples require either interferometers of varying complexity or significant amounts of post-processing (or both), while the scheme in Ref. [17] only works for one particular pair of states. Moreover, these techniques generally suffer from the same problem that plane-wave holograms have in creating higher order modes—i.e. they cannot distinguish between different higher order modes with the same phase symmetry.

The technique we used in this work is built on Fourier transform recognition, as described in Ref. [22] and implemented in Ref. [20]. The analyser is broken into two components—a spatial-mode converter and single-mode filter (cf. the polarisation analyser). We use a hologram to convert the target spatial mode into a plain Gaussian mode, which we then filter using a single-mode fibre (SMF).

As an example, consider the plane-wave VM_{+1} hologram [Fig. 5.10]. This is a phase component which modifies the vorticity of an incoming spatial mode, adding $\pm m$ quanta of orbital angular momentum to the $\pm m^{\text{th}}$ order output of the grating. Thus, an incident VM_{+1} mode will produce a plain Gaussian output in the $m=-1$ diffraction order¹⁷,

¹⁷Obviously, this also generalises so that a VM_{+m} mode produces a plain Gaussian component in the

whereas all other modes will produce a higher-order mode in this output (e.g. a plain Gaussian input $\rightarrow VM_{-1}$). Furthermore, an incident HG_{01} beam, which is a superposition of both $VM_{\pm 1}$ modes, gives rise to a weaker Gaussian component in both of the $m=\pm 1$ modes. The plain Gaussian component can then be filtered by a single-mode fibre designed for the appropriate wavelength range.

5.6.2 Single-mode fibres—the spatial filters

In a true single-mode fibre, guiding conditions (i.e. propagation without loss) occur only for the lowest order mode (the plain Gaussian), and all other modes will couple strongly to the external environment by radiating into the fibre cladding [27]. So if the fibre is sufficiently long, then the output beam will be a plain Gaussian mode *and* all the emerging light must have come from the Gaussian component of the input mode. To test the quality of this higher-order mode rejection, I performed a simple experiment using the output of a 670nm diode laser, strongly filtered with an initial SMF to give a clean spatial mode.

After appropriate mode-matching, I coupled this beam through a second SMF¹⁸ with an efficiency of $\sim 70\%$ (after $> 1\text{m}$ path length). Inserting a VM_{+1} hologram just after the first SMF output, I steered the beam to couple the first diffraction order into the second fibre, achieving a $\sim 65\%$ coupling efficiency with the hologram displaced far to one side of the singularity (7.5mm compared to $w_0 \sim 1\text{mm}$) so that it was effectively a plain diffraction grating. I then scanned the hologram singularity across the centre of the input beam and measured the coupling efficiency as a function of position¹⁹ [see Fig. 5.11].

Since the VM_{+1} hologram, when correctly aligned to the centre of the beam, produces a beam with no plain Gaussian component, the coupling efficiency should be zero at this point. In these measurements, the extinction ratio was around 1.3%, but typically, the observed ratios were better than 0.1%. The disparity arose because the mode rejection is extremely sensitive to the vertical alignment of the singularity (~ 10 microns), and in these early experiments, the height of the hologram was only aligned by eye.

Theoretically, a beam with a displaced charge +1 phase singularity can be decomposed into a superposition of a plain Gaussian (VM_0) and a VM_{+1} mode, where the position of the singularity can be used to calculate the weighting of the superposition. Therefore, using an offset VM_{+1} hologram effectively creates such a superposition. In fact, using the example in Sec. 5.2.2, the SMF coupling as a function of hologram position is,

$$\eta_{\text{disp}}(x) = \eta_0 \left(\frac{1}{1 + \gamma(x)^2} \right), \quad (5.62)$$

where $\gamma(x) = \frac{w_0/\sqrt{2}}{\sqrt{(x - x_0)^2 + \delta_y^2}},$

¹⁷ $-m^{\text{th}}$ order.

¹⁸Throughout this thesis, I used the Thorlabs KT110/M free space fibre coupler with the 4.5mm focal length C230TM-B coupling lens. The fibres were 2m long with a 620nm cut-off wavelength (Thorlabs P1-3224-FC-2). This corresponds to millions of wavelengths—much greater than the $\sim 10^3$ wavelengths calculated in Ref. [27] to be the minimum fibre length required for good spatial mode filtering.

¹⁹Similar results were reported in Ref. [28].

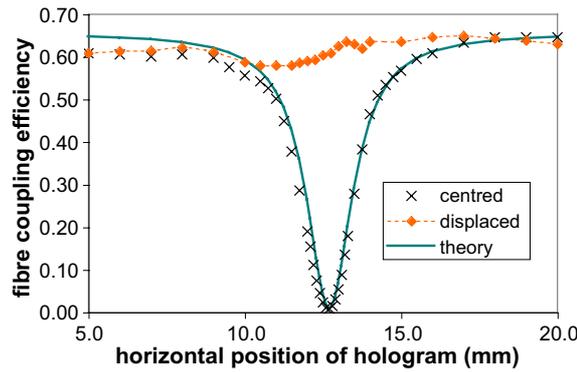


Figure 5.11: Demonstrating higher-order mode rejection by a SMF. The graph is the coupling efficiency of the first order of a VM_{+1} hologram scanned horizontally through the phase singularity (*centred*) and displaced vertically several beam widths from the singularity. The theoretical curve was calculated with parameters chosen to fit the curve to the right-hand side of the dip, which had the smallest observed deviation from the expected trends (in both *centred* and *displaced* curves): $\eta_0 = 0.646$, $w_0 = 1.25\text{mm}$, $x_0 = 12.72\text{mm}$, and the vertical misalignment, $\delta_y = 0.1\text{mm}$.

where η_0 is the best measured coupling efficiency, x_0 is the position of the phase singularity and δ_y is the misalignment of the hologram in the y direction. The data in Fig. 5.11 agrees well with the theory, except for a slight overall asymmetry and a small deviation on one edge of the dip. However, these features were also observed in the reference curve where the hologram scan was displaced a significant distance above the singularity, so they were probably the result of beam-steering caused by some slight non-uniformity in the thickness of the holographic plate.

It is often claimed (see, e.g., Refs [20, 28–30]) that the mode rejection properties of a SMF are just a consequence of the smaller spatial extent of the plain Gaussian mode (i.e. in beams of equal power, its energy will be closer to the centre than a higher order mode). This very simplistic view would mean that many higher-order modes, such as HG_{20} and LGV_{10} , would couple a significant amount of energy through a single-mode fibre, because they have a bright spot at the centre of the beam—but they do not, even though in the single-mode regime, the fibre is “large enough” to accept them²⁰. A single-mode fibre is a true mode *selector*, not merely an effective aperture for the input beam [see Fig. 5.12], because guiding conditions are determined by the field distribution of a mode. This *interferometric* rejection of higher-order modes (i.e. based on amplitude and not intensity information) is a much stronger condition than one suggested by a pin-hole analogy.

Figure 5.11 also provides very strong evidence that a SMF is an interferometric mode selector. Assuming that the 65% efficiency for the coupling of the plain Gaussian beam results from the iris effect of a pin-hole, the radius of such a pin-hole would be $\sim w_0$. The amount of energy of the corresponding VM_{+1} mode that would pass through such a pin-hole is $\sim 26\%$ —far greater than the observed minimum value of around 2%.

²⁰The fibres we use have a mode field diameter (MFD) of $4.0\mu\text{m}$ —several times the wavelength of the field.

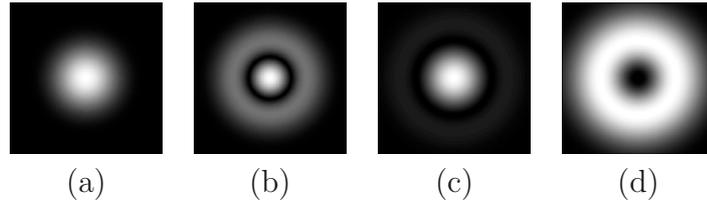


Figure 5.12: Filtering with a single-mode fibre. (a) LGV_{00} ; (b) LGV_{10} ; (c) & (d) the equal superpositions $|LGV_{00}\rangle \pm |LGV_{10}\rangle$, respectively. The plain Gaussian (a) couples into a SMF no matter what other modes are present. Therefore, the superpositions in (c) and (d) both couple with equal efficiency, despite the fact that one is bright and dark at the centre, respectively. This is very different from the effect of a small pin-hole.

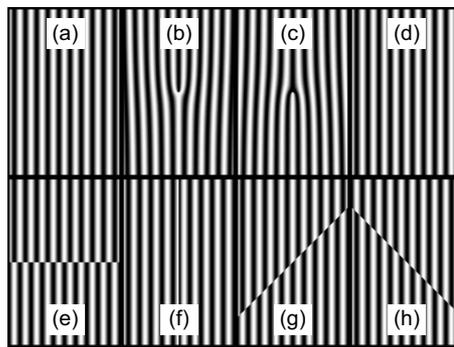


Figure 5.13: The spatial tomography hologram. Measured states: (a) g ; (b) r ; (c) l ; (d) g ; (e) h ; (f) v ; (g) d ; (h) a . The gr and gl superposition measurements can be made by displacing the appropriate vortex hologram, (b) or (c).

5.6.3 Measuring the tomographic spatial modes

The compound hologram we used to perform spatial quantum state tomography is illustrated in Fig. 5.13. Because the spatial mode filtering is very sensitive to alignment, we mounted the holographic plate on a micrometer-driven x – y translation stage to access the different segments of the hologram with precision and repeatability. These positions need to be found before performing a spatial tomography. Using a compound hologram ensures that no holograms need to be swapped during an experiment, which is important because the position-sensitivity of the spatial filtering would necessitate significant realignment of the apparatus and would make consistent measurements almost impossible. The other major advantage is that all holographic segments are then developed under the most similar conditions possible (e.g. giving the most consistent diffraction efficiencies).

The tomographic measurement set for the degenerate qubit is very straightforward, since it is completely analogous to the polarisation case. In other words, the required states are h , v , d , a , r and l , with the corresponding hologram segments (e), (f), (g), (h), (b) and (c), respectively, in Fig. 5.13.

The computational basis measurements for the nondegenerate qubit are the g and r/l states—the (a) and (b/c) segments, respectively. To measure the equal superposition states, $|g\rangle + e^{i\phi}|r/l\rangle$ ($\phi = 0, \pi/2, \pi, 3\pi/2$), the corresponding vortex hologram (b/c) is

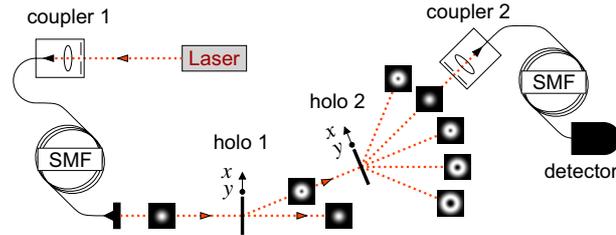


Figure 5.14: The experimental layout for spatial tomography using coherent states. The spatial mode of the bright laser beam is cleaned up by a single-mode fibre (SMF) and then passed through a hologram to prepare the input state. This is then analysed by a second hologram and another SMF.

displaced by $w_0/\sqrt{2}$ [see Eq. (5.28)] away from the centre of the beam. The direction of the displacement determines the phase of the superposition according to Eq. (5.26), i.e. $\theta = \pm(\pi - \phi)$ for r/l respectively, with the hologram positions for orthogonal superpositions always on opposite sides of the beam. Note that the phases for the gr and gl superpositions accrue by rotating the singularity in different directions.

From Figure 5.2, it is easy to see that this displacement trick only works for the lowest-order charge ± 1 holograms, and not for any other higher modes, since those superpositions exhibit several displaced, singly charged singularities. Although this result has been known for some time, e.g. [24], we pointed out in Ref. [31] that there has been some confusion in quantum information literature. For example, a number of recent papers (see Refs [20, 29, 30]) have reported that a displaced charge 2 phase singularity is just a superposition of the VM_0 and VM_2 modes. In fact, a displaced charge 2 phase singularity is much more complicated. This is one of the main advantages of the nondegenerate spatial Hilbert spaces that we focussed on in our experiments.

Finally, the measurements required for tomography of the nondegenerate spatial qutrit are just the combination of all the measurements required for the one degenerate and two nondegenerate qubits—15 measurements in total, all of which can be made with the hologram illustrated in Fig. 5.13.

5.6.4 Spatial mode tomography with coherent states

Before applying the spatial tomography technique to unknown systems, I tested the tomographic reconstruction under controlled conditions using coherent states [Fig. 5.14]. The source was a continuous-wave HeNe laser which had been spatially purified using a SMF and gave around 2.5mW of intensity stabilised power²¹. The beam passed through one hologram for state *preparation*, and then into a spatial mode analyser consisting of a second hologram and a SMF (after beam propagation and some mode-matching optics, coupling efficiency $\sim 40\%$ when aligned without the holograms).

The holograms were generated with 18 pixels per fringe, designed to produce a diffrac-

²¹The laser was stabilised using a noise-eater, Thorlabs CR200A, at the cost of some deterioration in spatial mode quality.

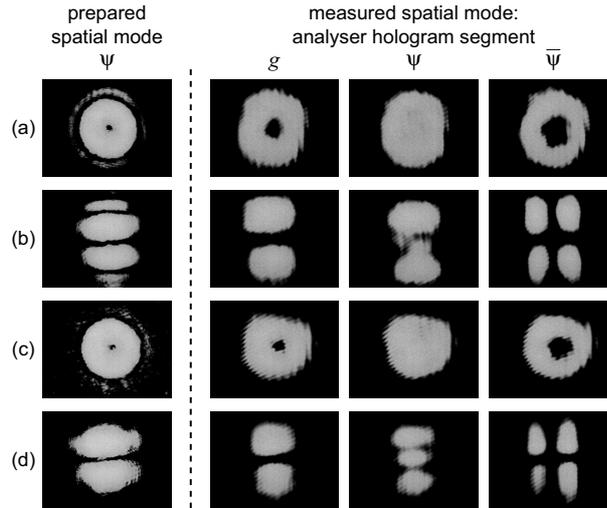


Figure 5.15: Prepared and reconstructed spatial modes used for tomography with coherent states—unclipped/clipped input states: (a/c) $\psi = r$, (b/d) $\psi = h$. $\bar{\psi}$ is the spatial mode orthogonal to ψ .

tion angle of around 0.28° at 670nm (a little smaller at 633nm). The two holograms I used had first-order diffraction efficiencies of around 20%. The beam spot size (w_0) was 1.19mm at the preparation hologram and 0.91mm at the analysis hologram (pixel size $\sim 10\mu\text{m} \times 10\mu\text{m}$).

I found the hologram singularities by looking for the positions with the maximum extinction in coupling into the second fibre (with the other hologram aligned to the plain diffraction grating segment). I consistently measured extinctions of $\lesssim 0.1\%$ (compared with the power coupled with both holograms aligned to plain diffraction gratings). When the system was well-aligned, these maximum extinction positions produced spatial modes which looked very symmetrical at the fibre input—this visual check was a good indication of misalignment.

To check visually that the apparatus was operating as expected during the tomographic measurements, I imaged the modes prepared and reconstructed by the two holograms²². The results confirmed the data obtained using the spatial filtering of the SMF. Some of these images are included in Fig. 5.15. The extra rings and lobes in the prepared states arise from the imperfect mode-conversion properties of plane-wave holograms. As expected, this effect is stronger in the jump holograms than in the vortex holograms, the latter also giving higher-quality reconstructed Gaussian modes (third column). These higher-order effects were partially eliminated by the slightly crude but simple expedient of clipping the input state with either an iris or a pair of razor blades [see (c,d)].

The differences between these holograms are reflected in the measurements of power coupled into the fibre [Tab. 5.2]. There is a clear disparity between the reconstruction efficiencies for the different holographic segments. However, although clipping some of the higher-order features of the input states had a marked effect on the mode images (partic-

²²Using an Electrim EDC-1000 CCD camera. This is not a single-photon-counting CCD camera—it only measures photocurrent.

prepared spatial mode	measured spatial mode:	
$ \psi\rangle$	$\langle\psi $	$\langle\bar{\psi} $
g	49.0	0.9
r	28.8	0.1
h	16.4	0.03
r (clipped)	28.2	0.09
h (clipped)	16.0	0.05

Table 5.2: Hologram reconstruction efficiencies. The power at the input to the fibre was $\sim 100\mu\text{W}$. For $\psi = g$, $\bar{\psi} = r$.

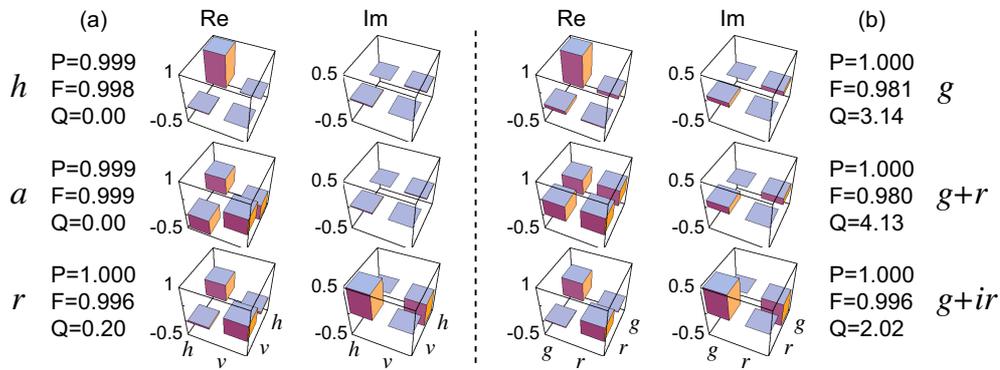


Figure 5.16: Reconstructed single-“qubit” density matrices from spatial tomography using coherent states: (a) the degenerate qubit; (b) the nondegenerate qubit. (Because the “qubit” states are classical states, this system is not a true qubit, although it has some similar behaviour [see Sec. 2.2.3].)

ularly for the h inputs and some of the other measurement combinations not shown), it had very little effect on the coupled power. This demonstrates the robust nature of the spatial filtering analysis technique.

The main potential difficulty to arise from the differing reconstruction efficiencies is that it might be expected to have a major impact on the output of the tomography. Fortunately, as it turns out, this problem is already largely solved by using the six-state normalisation procedure discussed in Sec. 3.7. Because the r/l and h/v holograms give similar reconstruction efficiencies, probabilities which are calculated by normalising the data internally within the different bases will not display the same trends. This is yet another compelling indication of the advantages which can arise from using over-complete measurement sets. For the nondegenerate encodings, however, the effect will not be completely eliminated, because POVM sets like $\{g, r\}$ will still be biased towards the g measurement. Thus, the degenerate tomographies should be somewhat more reliable.

Figure 5.16 shows several representative density matrices measured in both the degenerate (a) and nondegenerate (b) qubit subspaces. Overall, the very high purities and fidelities with the target states demonstrate the value of using spatial modes to encode and measure information.

In the measured degenerate states, the observed quality factors were excellent. In fact, for the two linear (HG -type) states, the quality factor was identically zero, because the reconstructed state was already physically valid before the maximum likelihood procedure. This is not particularly surprising in itself, since using bright coherent states allows very accurate measurements, but it does indicate that the measurements are actually doing what they should. It also verifies that the normalisation procedure is indeed largely compensating for the different reconstruction efficiencies [Tab. 5.2].

The reconstructed nondegenerate states also agreed very well with expectations. In all cases, tomographies of a range of pure states (including several unbalanced superpositions) exhibited purities and fidelities above 0.95 and generally above 0.98. Nevertheless, these results are not quite as good as the degenerate case. In particular, the larger quality factors indicate that the estimated measurement uncertainties were too small. These estimates only included the fluctuations in the power meter readings, and did not take into account any errors in the measurement settings. For the nondegenerate states, the measurements are very sensitive to any misalignment in the apparatus, e.g. very small hologram displacements ($\sim 0.1\text{mm}$) can make a significant difference to the population balance in a superposition measurement. At the time of these measurements, the experiment had been operating without a major realignment for some time and was no longer exhibiting optimal behaviour (e.g. extinction ratios were $\sim 3\text{--}5\%$). These effects probably also contributed to the lower fidelities.

The sensitivity of the nondegenerate measurements to alignment also highlights the importance of accurately measuring the spot size of the beam which is imaged by the SMF.

5.6.5 Theory: using plane-wave holograms in degenerate QST

So far in this section, I have described how to perform spatial QST with plane-wave holograms, but I have largely ignored the fact that plane-wave holograms do not analyse only a single Gaussian mode. My only concession has been to acknowledge that in the experiment, different pairs of holographic segments gave different reconstruction efficiencies [see Tab. 5.2]. Our proposed solution was to claim that this and any related problems were mainly eliminated by normalising the measured data within convenient POVM sets of an over-complete measurement set. Indeed, this notion was strongly supported by the results of the test measurements made with a coherent source. I now investigate the validity of this claim theoretically, and I show that, given an appropriate redefinition of the logical states²³, this normalisation trick circumvents the problems associated with plane-wave analysis of Gaussian modes, at least to a good approximation.

I will begin with the degenerate qubit and consider the relationship between the actual modes produced and analysed by the plane-wave holograms. When it is important to distinguish these from the ideal first-order Gaussian modes [Eq. (5.30)], I will label the pseudo-first-order modes with quotes (e.g. “ r ”, “ h ”). Table 5.3 shows the relative fidelities (mode overlap probabilities) for the pseudo-first-order modes²⁴. These show that the

²³A similar redefinition was used in Ref. [32].

²⁴The results are based on direct, numerical calculations. The non-unit probabilities down the main diagonal result from the numerical error associated with discretisation of the transverse plane and its size

$ \langle\psi_1 \psi_2\rangle ^2$	$ “h”\rangle$	$ “v”\rangle$	$ “d”\rangle$	$ “a”\rangle$	$ “r”\rangle$	$ “l”\rangle$
$\langle“h” $	0.992	0.000	0.248	0.248	0.402	0.402
$\langle“v” $		0.992	0.248	0.248	0.402	0.402
$\langle“d” $			0.992	0.000	0.402	0.402
$\langle“a” $				0.992	0.402	0.402
$\langle“r” $					0.992	0.000
$\langle“l” $						0.992

Table 5.3: Theoretical fidelities for the pseudo-first-order, degenerate basis states. These “pseudo”-states are defined to be the modes produced by the different plane-wave hologram segments.

pseudo-bases are not related like the standard qubit logical states, i.e. $|“h”\rangle \neq |“d”\rangle + |“a”\rangle \neq |“r”\rangle + |“l”\rangle$. Nor are the overlaps between different bases all the same size. However, they are balanced within each basis, as is necessary for the normalisation trick to be successful.

Based on these results, I define the new logical states, $\mathbf{0} \equiv “r”$ and $\mathbf{1} \equiv “l”$, along with their corresponding $|\pm\rangle$ and $|\pm i\rangle$ states. With these definitions, the superposition states all have the same fidelity with the corresponding linear “jump” mode, e.g. $|\langle+|“h”\rangle|^2=0.804$. Moreover, the non-corresponding overlaps are also the same, e.g. $|\langle+|“d”\rangle|^2=0.402$. I can use these to define the error modes, φ , e.g.,

$$|“h”\rangle = \alpha_{+,h}|+\rangle + \sqrt{1-|\alpha_{+,h}|^2}|\varphi_+\rangle. \quad (5.63)$$

Not surprisingly, the φ_+ error mode has no overlap with either “r”, “l” or “v”. However, it does overlap with the d/a jump modes: $|\langle\varphi_+|“d”\rangle|^2=0.026$ (or $|\langle\varphi_+|\varphi_{+i}\rangle|^2=0.139$). This small error is the source of one problem which can slip through the normalisation trick.

There is an obvious disadvantage to redefining the logical states in this way—the basis states are no longer nondegenerate (even individually), which can have significant consequences, particularly in experiments involving sequential operations (as in Ref. [29]). Therefore, this interpretation needs to be applied with care. However, it is useful, particularly when characterising spatial quantum states which are produced by another source, such as the spontaneous down-conversion used in all experiments in this thesis.

The plane-wave tomography needs to be tested in two important scenarios. The first case is when characterising quantum states which lie wholly inside the new qubit Hilbert space $\{“r”, “l”\}$, which should be a good approximation of the coherent test case described above. I considered two groups of pure states, $|\psi\rangle \sim \cos\alpha|\mathbf{0}\rangle + \sin\alpha|\mathbf{1}\rangle$ and $|\psi\rangle \sim |\mathbf{0}\rangle + e^{i\phi}|\mathbf{1}\rangle$, which sweep out two orthogonal great circles on the Poincaré sphere, and calculated the “h” measurement probabilities [Fig. 5.17]. Because the “h” and “v” measurements are balanced, the normalised probabilities (red circles) agree with the theory perfectly. Thus, the normalisation trick will work for any state in the computational space.

relative to the beam’s ($< 1\%$ error). I used a 512×512 pixel spatial grid ($9.0\text{mm}\times 9.0\text{mm}$), placed at the beam waist of a Gaussian mode with $w_0=1.5\text{mm}$.

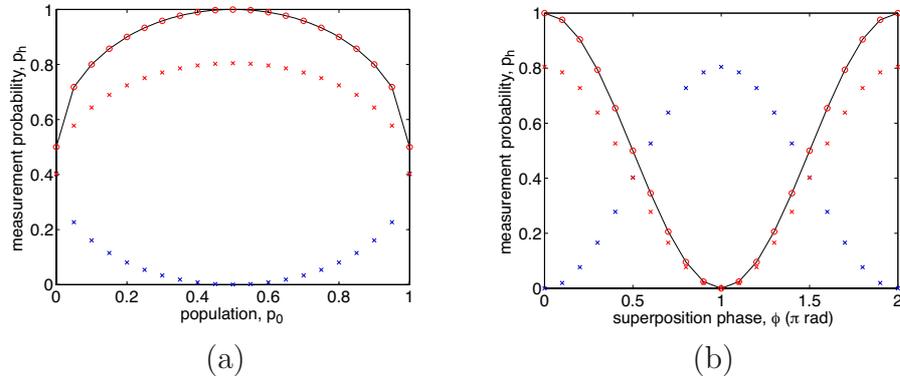


Figure 5.17: Comparison of the degenerate pseudo-basis measurement probabilities with theory (black curves): the unnormalised p_h and p_v (red & blue crosses, resp.); the normalised p_h ($\equiv p_h/[p_h+p_v]$) (red circles). The target pure states: (a) $|\psi\rangle \sim \cos \alpha |\mathbf{0}\rangle + \sin \alpha |\mathbf{1}\rangle$ ($p_0 = \cos^2 \alpha$); (b) $|\psi\rangle \sim |\mathbf{0}\rangle + e^{i\phi} |\mathbf{1}\rangle$.

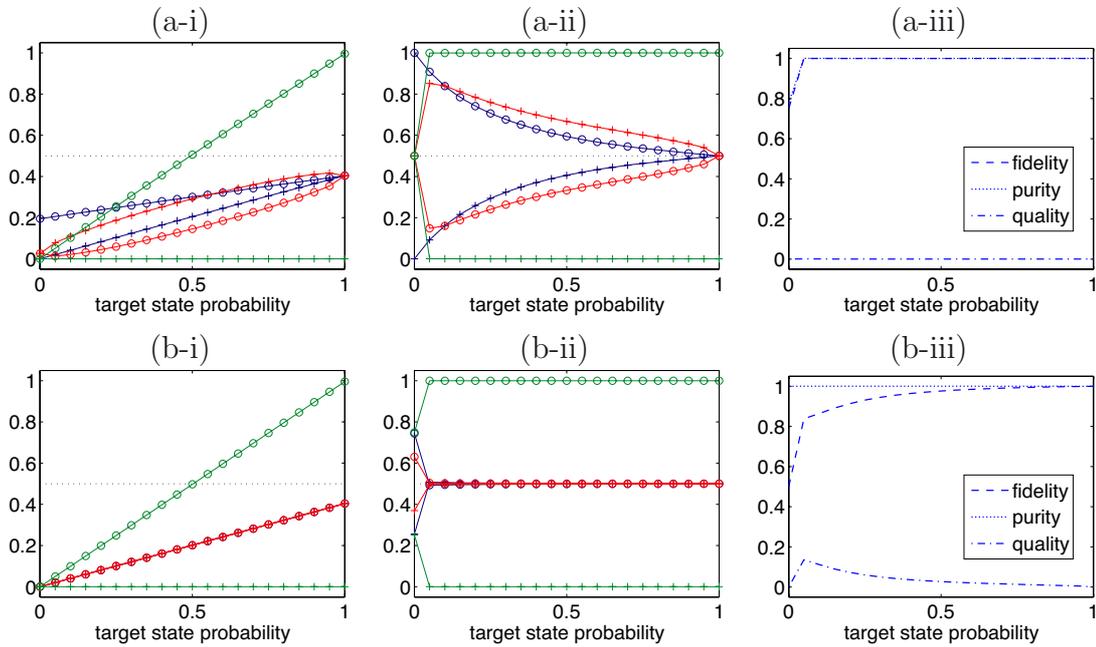


Figure 5.18: Testing the degenerate spatial mode tomography technique with $|\text{target}\rangle = |r\rangle$ and: (a) $|\text{error}\rangle = |e^{i\phi_{\text{sto}}(\mathbf{s})}\rangle$; and (b) $|\text{error}\rangle = i|\varphi_+\rangle$. (i) theoretical measurement probabilities; (ii) normalised probabilities; (iii) reconstructions: purity, fit quality, and target state fidelity. In (i) and (ii), blue: “ h/v ”; red: “ d/a ”; and green: “ r/l ” [$\circ/+$, resp.]

The second scenario is when the target qudit is only a subspace of the incoming quantum state, which is particularly relevant to our experiments in characterising the output state from a down-conversion source. The complete output was a superposition of many higher-order modes and we only considered a small subspace. The problem in this situation is that the plane-wave spatial filtering cannot distinguish photons in the computational basis from those from the error states, φ .

To break the problem down, I considered states of the following form:

$$|\psi\rangle = \sqrt{p}|\text{target}\rangle + \sqrt{1-p}|\text{error}\rangle. \quad (5.64)$$

In the first example, I set $|\text{target}\rangle = |“r”\rangle$ with an uniform, spatially incoherent background field as the error state, characterised by a stochastic phase, $\exp[i\phi_{\text{sto}}(\mathbf{s})]$ [Fig. 5.18(a)]. I have plotted the theoretically calculated probabilities, the normalised probabilities as would be measured in an experiment, and also the relevant properties of the reconstructed density matrices (purity, fit quality, and fidelity with the target state). Because I did not include any Poissonian noise in these simulations, I used an unweighted fit quality based on probabilities [cf. Eq. (3.50)]:

$$Q = \frac{1}{M} \sum_j |\mathcal{W}_j - w_j(\varrho)|. \quad (5.65)$$

Figure 5.18(a) shows that the normalisation trick works perfectly in this case, because a spatially incoherent background has a negligible overlap with the error states²⁵. So as the target state component decreases, the measurement probabilities all decrease proportionately, and the tomography predicts the correct state. These results are the same if a coherent plane-wave background is used.

To test the limits of the normalisation trick, I considered situations with the same target state, but where $|\text{error}\rangle$ was a combination of the plane-wave error states, φ . Figure 5.18(b) shows a typical set of results ($|\text{error}\rangle = i|\varphi_+\rangle$). As the error component increases, it unbalances both the $\{\pm\}$ and $\{\pm i\}$ measurement bases, giving normalised data which predict unphysical states lying outside the Poincaré sphere—in reality, a state cannot exhibit strong fringes in more than one basis. The maximum likelihood tomography then finds the closest physical state, which will naturally lie on the sphere’s surface (hence, $P = 1$). These imbalances are reflected by a loss in the fidelity, although even with quite a large error component, it is promisingly small ($F > 0.975$ for $p = 0.5$).

There are several reasons why the errors caused by this tomography technique are only moderate. In the first case, the analysis holograms do not overlap strongly with the error modes ($|\langle\varphi_+|“h”\rangle|^2 \sim 0.2$ and $|\langle\varphi_+|“d”\rangle|^2 \sim 0.03$). Moreover, the error modes are, on average, higher-order modes than the basis states, so in systems like the down-conversion source, they will make a smaller contribution [33]. Alternatively, if the error modes are created by other plane-wave holograms (such as in the bright laser tests) they will be limited to a small component ($\sim 20\%$). Finally, the “ r ” and “ l ” measurements have no overlap with the error modes, so that this measurement basis is completely reliable. One might even consider weighting these more heavily in the optimisation code, although we chose not to so that we brought no assumptions to our measurements.

²⁵Because the transverse spatial degree of freedom is infinite-dimensional, this applies to all states.

$ \langle \psi_1 \psi_2 \rangle ^2$	$ g\rangle$	$ “r”\rangle$	$ “g+r”\rangle$	$ “g-r”\rangle$	$ “g+ir”\rangle$	$ “g-ir”\rangle$
$\langle g $	0.992	0.000	0.501	0.501	0.501	0.501
$\langle “r” $		0.992	0.258	0.258	0.258	0.258
$\langle “g+r” $			0.992	0.049	0.357	0.357
$\langle “g-r” $				0.992	0.357	0.357
$\langle “g+ir” $					0.992	0.049
$\langle “g-ir” $						0.992

Table 5.4: Theoretical fidelities for the pseudo-first-order, nondegenerate basis states.

Another advantage of this tomography technique is that it provides some signature clues when the problems associated with error modes are becoming overwhelmingly significant. For example, the loss in fidelity is accompanied by a loss in fit quality [see Fig. 5.18(b-iii)] as the normalised measurements predict unphysical results. Also, because of the small overlap of the error modes with the measured states, large error components lead to surprisingly small measured counts rates compared to the total incoming power (as might be observed with a multi-mode fibre, say), particularly in the reliable “ r , l ” modes. Therefore, if the count rates in “ r ” and “ l ” are sensible in comparison to the (\pm) and ($\pm i$) measurements and total count rate, then the error modes are playing a fairly minor role.

Finally, while this tomography might predict a slightly rotated state, there does not seem to be any way it could result in predicting non-existent non-local correlations—any entanglement such as we see in the down-conversion measurements must be genuine. The worst case scenario is that it would predict a poorer-quality state than it should, with higher mixture and lower entanglement. These measurements should therefore provide a lower bound on the entanglement quality.

5.6.6 Theory: using plane-wave holograms in nondegenerate QST

The pseudo-basis fidelities for the nondegenerate qubit are given in Tab. 5.4. Now there is an additional complication which was absent from the degenerate case. The mode created by displacing a VM_{+1} plane-wave singularity $w_0/\sqrt{2}$ away from the centre of the beam is not an equal superposition of g and “ r ”. So now the g /“ r ” basis is unbalanced, and this does not even take into account the observed differences in reconstruction fidelities. Obviously, this also means that “ $g\pm r$ ” and “ $g\pm ir$ ” are not orthogonal pairs. It turns out equal superpositions can be created or measured by displacing the holograms by $\sim w_0/1.77$ from the beam centre. The results calculated for this case are given in Tab. 5.5. With these *balanced* settings, the nondegenerate tomography behaves qualitatively exactly like the degenerate case, although the numbers are slightly different.

Before we completely understood these effects, we measured the unbalanced states (displacing by $w_0/\sqrt{2}$). We used these measurements throughout the experiments in this thesis.

$ \langle\psi_1 \psi_2\rangle ^2$	$ g\rangle$	$ \text{“}r\text{”}\rangle$	$ \text{“}g+r\text{”}\rangle$	$ \text{“}g-r\text{”}\rangle$	$ \text{“}g+ir\text{”}\rangle$	$ \text{“}g-ir\text{”}\rangle$
$\langle g $	0.992	0.000	0.370	0.370	0.370	0.370
$\langle \text{“}r\text{”} $		0.992	0.371	0.371	0.371	0.371
$\langle \text{“}g+r\text{”} $			0.992	0.000	0.329	0.329
$\langle \text{“}g-r\text{”} $				0.992	0.329	0.329
$\langle \text{“}g+ir\text{”} $					0.992	0.000
$\langle \text{“}g-ir\text{”} $						0.992

Table 5.5: Theoretical fidelities for the pseudo-first-order, nondegenerate basis states—*balanced* measurements ($\Delta r = w_0/1.77$).

As above, I define new logical states, $\mathbf{0} \equiv g$ and $\mathbf{1} \equiv \text{“}r\text{”}$, the related $|\text{“}\pm\text{”}\rangle$ and $|\text{“}\pm i\text{”}\rangle$, and the error states, φ . In Figure 5.19, I have plotted the measurement comparison for states lying inside the computational Hilbert space with both the balanced and unbalanced measurements, using the same two groups of pure states [cf. Fig. 5.17]. Once again, the normalisation trick works perfectly with balanced measurement settings, and the unbalanced settings give rise to asymmetrical results, although they are still reasonably close to the expected theory curves.

The next step is to test nondegenerate tomography in the context of states spanning a space which extends further than the computational space. This time I set $|\text{target}\rangle = |g\rangle$, and Fig. 5.20 shows the example of $|\text{error}\rangle = i|\varphi_+\rangle$. On the whole, both balanced and unbalanced results lead to the same conclusions as with the degenerate tomography; that the tomography functions well even for quite large error components; that these error contributions are most likely to be fairly small anyway; and if not, that the results provide certain tell-tale signatures, such as low count rates.

The main difference arises when using unbalanced measurement positions. Because the error modes overlap somewhat with all of the measured states, the normalised data is more significantly affected, particular in the main computation basis $\{g, \text{“}r\text{”}\}$. The results show that the reconstruction fidelity drops off slightly quicker than before, and is accompanied by another tell-tale signature—in this case, higher estimated levels of mixture²⁶. Thus, once again, the worst case scenario is that the actual state is even better quality than the measurements suggest.

²⁶Since this brings the linearly reconstructed state back inside the Poincaré sphere, the maximum likelihood optimisation has no effect and the fit quality does not suffer.

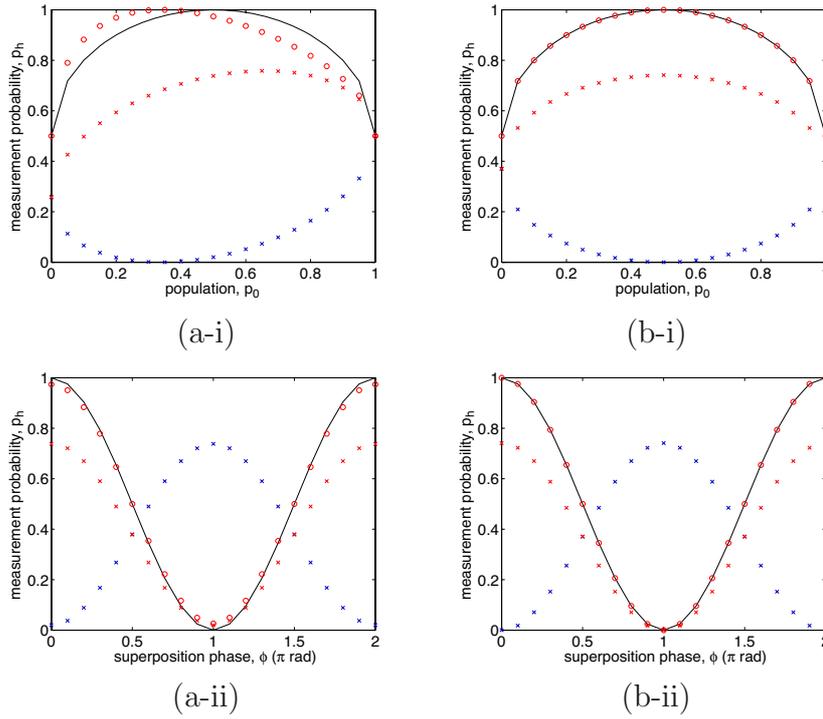


Figure 5.19: Comparison of the nondegenerate pseudo-basis measurement probabilities with theory (black curves): the unnormalised p_h and p_v (red & blue crosses, resp.); the normalised p_h ($\equiv p_h/[p_h+p_v]$) (red circles). The target pure states: (i) $|\psi\rangle \sim \cos\alpha|\mathbf{0}\rangle + \sin\alpha|\mathbf{1}\rangle$ ($p_0 = \cos^2\alpha$); (ii) $|\psi\rangle \sim |\mathbf{0}\rangle + e^{i\phi}|\mathbf{1}\rangle$. The results for the unbalanced and balanced measurement settings are in (a) and (b), respectively.

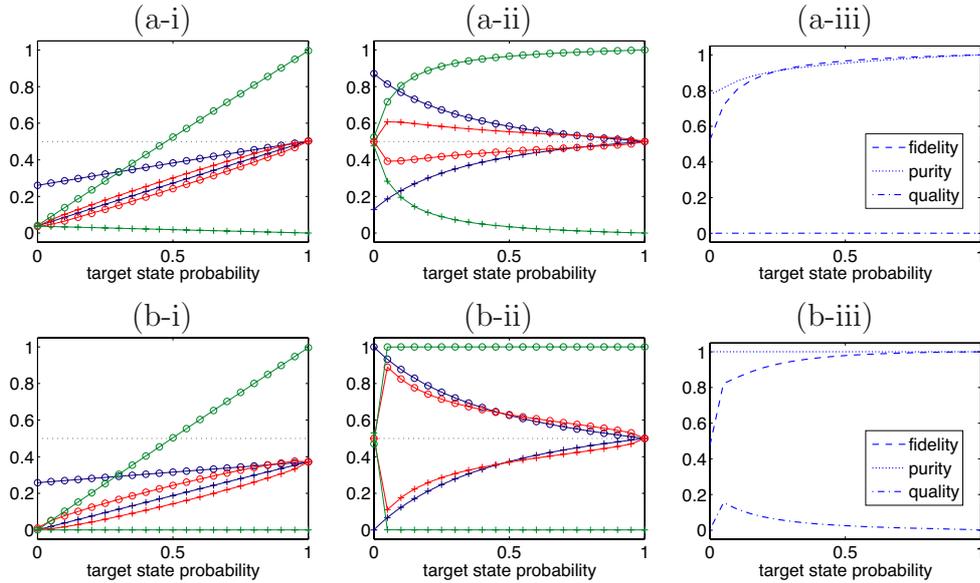


Figure 5.20: Testing the nondegenerate spatial mode tomography technique with $|\text{target}\rangle = |r\rangle$ and $|\text{error}\rangle = i|\varphi_+\rangle$: (a) unbalanced measurements; (b) balanced measurements. (i) theoretical measurement probabilities; (ii) normalised probabilities; (iii) reconstructions: purity, fit quality, and target state fidelity. In (i) and (ii), blue: “ h/v ”; red: “ d/a ”; and green: “ r/l ” [$\circ/+$, resp.]

5.7 Chapter 5 References

- [1] J. L. O'Brien, G. J. Pryde, A. G. White, T. C. Ralph, and D. Branning. Demonstration of an all-optical quantum controlled-NOT gate. *Nature*, 426:264, 2003.
- [2] J. L. O'Brien, G. J. Pryde, A. Gilchrist, D. F. V. James, N. K. Langford, T. C. Ralph, and A. G. White. Quantum process tomography of a controlled-NOT gate. *Physical Review Letters*, 93(8):080502, 2004.
- [3] K. J. Resch, K. L. Pagnell, R. Prevedel, A. Gilchrist, G. J. Pryde, J. L. O'Brien, and A. G. White. Time-reversal and super-resolving phase measurements. *quant-ph*, page 0511214, 2005.
- [4] A. E. Siegman. *Lasers*. University Science Books, Mill Valley, CA, 1986.
- [5] J. W. Goodman. *Introduction to Fourier Optics*. The McGraw-Hill Companies, New York, second edition, 1996.
- [6] K. L. Jones. 3rd year electromagnetic theory course notes, 1999.
- [7] M. A. Bandres and J. C. Gutiérrez-Vega. Ince-Gaussian beams. *Optics Letters*, 29:144, 2004.
- [8] M. A. Bandres and J. C. Gutiérrez-Vega. Ince-Gaussian modes of the paraxial wave equation and stable resonators. *Journal of the Optical Society of America A*, 21:873, 2004.
- [9] U. T. Schwarz, M. A. Bandres, and J. C. Gutiérrez-Vega. Ince-Gaussian beams. *Optics Letters*, 29:1870, 2004.
- [10] L. Allen, M. W. Beijersbergen, R. J. C. Spreeuw, and J. P. Woerdman. Orbital angular momentum of light and the transformation of Laguerre-Gaussian laser modes. *Physical Review A*, 45(11):8185, 1992.
- [11] M. J. Padgett and J. Courtial. Poincaré-sphere equivalent for light beams containing orbital angular momentum. *Optics Letters*, 24(7):430, 1999.
- [12] H. He, M. E. J. Friese, N. R. Heckenberg, and H. Rubinsztein-Dunlop. Direct observation of transfer of angular momentum to absorptive particles from a laser beam with a phase singularity. *Physical Review Letters*, 75(5):826, 1995.
- [13] R. A. Beth. Mechanical detection and measurement of the angular momentum of light. *Physical Review*, 50:115, 1936.
- [14] M. E. J. Friese, J. Enger, H. Rubinsztein-Dunlop, and N. R. Heckenberg. Optical angular-momentum transfer to trapped absorbing particles. *Physical Review A*, 54(2):1593, 1996.
- [15] N. B. Simpson, K. Dholakia, L. Allen, and M. J. Padgett. Mechanical equivalence of spin and orbital angular momentum of light: an optical spanner. *Optics Letters*, 22(1):52, 1997.

-
- [16] J. M. Vaughan and D. V. Willets. Temporal and interference fringe analysis of tem_{01}^* laser modes. *Journal of the Optical Society of America*, 73(8):1018, 1983.
- [17] C. Tamm and C. O. Weiss. Bistability and optical switching of spatial patterns in a laser. *Journal of the Optical Society of America B*, 7(6):1034, 1990.
- [18] G. Machavariani, N. Davidson, E. Hasman, S. Blit, A. A. Ishaaya, and A. A. Friesem. Efficient conversion of a Gaussian beam to a high purity helical beam. *Optics Communications*, 209:265, 2002.
- [19] A. G. White, C. P. Smith, N. R. Heckenberg, H. Rubinsztein-Dunlop, R. McDuff, C. O. Weiss, and C. Tamm. Interferometric measurements of phase singularities in the output of a visible laser. *Journal of Modern Optics*, 38(12):2531, 1991.
- [20] A. Mair, A. Vaziri, G. Weihs, and A. Zeilinger. Entanglement of the orbital angular momentum states of photons. *Nature*, 412:313, 2001.
- [21] N. R. Heckenberg, R. McDuff, C. P. Smith, and A. G. White. Generation of optical phase singularities by computer-generated holograms. *Optics Letters*, 17(3):221, 1992.
- [22] N. R. Heckenberg, R. McDuff, C. P. Smith, H. Rubinsztein-Dunlop, and M. J. Wegener. Laser beams with phase singularities. *Optical and Quantum Electronics*, 24:S951, 1992.
- [23] H. Kogelnik. Hologram efficiency and response. *Microwaves*, 6:68, November 1967.
- [24] M. S. Soskin, V. N. Gorshkov, M. V. Vasnetsov, J. T. Malos, and N. R. Heckenberg. Topological charge and angular momentum of light beams carrying optical vortices. *Physical Review A*, 56(5):4064, 1997.
- [25] J. Courtial, K. Dholakia, D. A. Robertson, L. Allen, and M. J. Padgett. Measurement of the rotational frequency shift imparted to a rotating light beam possessing orbital angular momentum. *Physical Review Letters*, 80(15):3217, 1998.
- [26] J. Leach, M. J. Padgett, S. M. Barnett, S. Franke-Arnold, and J. Courtial. Measuring the orbital angular momentum of a single photon. *Physical Review Letters*, 88(25):257901, 2002.
- [27] O. Wallner, W. R. Leeb, and P. J. Winzer. Minimum length of a single-mode fibre spatial filter. *Journal of the Optical Society of America A*, 19(12):2445, 2002.
- [28] A. Vaziri, G. Weihs, and A. Zeilinger. Superpositions of the orbital angular momentum for applications in quantum experiments. *Journal of Optics B: Quantum and Semiclassical Optics*, 4:S47, 2002.
- [29] A. Vaziri, G. Weihs, and A. Zeilinger. Experimental two-photon, three-dimensional entanglement for quantum communication. *Physical Review Letters*, 89(24):240401, 2002.
- [30] A. Vaziri, J.-W. Pan, T. Jennewein, G. Weihs, and A. Zeilinger. Concentration of higher dimensional entanglement: Qutrits of photon orbital angular momentum. *Physical Review Letters*, 91:227902, 2003.

- [31] N. K. Langford, R. B. Dalton, M. D. Harvey, J. L. O'Brien, G. J. Pryde, A. Gilchrist, S. D. Bartlett, and A. G. White. Measuring entangled qutrits and their use for quantum bit commitment. *Physical Review Letters*, 93(5):053601, 2004.
- [32] G. Molina-Terriza, A. Vaziri, J. Řeháček, Z. Hradil, and A. Zeilinger. Triggered qutrits for quantum communication protocols. *Physical Review Letters*, 92:167903, 2004.
- [33] G. Molina-Terriza, J. P. Torres, and L. Torner. Orbital angular momentum of photons in noncollinear parametric downconversion. *Optics Communications*, 228:155, 2003.

*Yesterday, upon the stair,
I met a man who wasn't there.
He wasn't there again today,
I wish that man would go away.*

Hugo Means

Chapter 6

The Type-I down-conversion process

6.1 Nonlinear optical materials

The classical behaviour of electromagnetic radiation as it propagates through a simple dielectric is fairly straightforward. As described in Sections 4.1 and 5.1, under these conditions (which include linearity), Maxwell's equations give rise to solutions in the form of a simple wave equation. However, when electromagnetic fields interact with nonlinear materials, the behaviour becomes significantly more complex. The nonlinearity of the materials is reflected in the *dielectric polarisation* vector, \mathfrak{P} , which becomes a nonlinear function of the electric field of the form¹,

$$\mathfrak{P}(\mathfrak{E}) = \epsilon_0 [\chi_e^{(1)}\mathfrak{E} + \chi_e^{(2)}\mathfrak{E}^2 + \chi_e^{(3)}\mathfrak{E}^3 + \dots], \quad (6.1)$$

where $\chi_e^{(1)}$ ($\equiv \chi_e$) is the standard linear coefficient of dielectric susceptibility [see Eq. (4.2)]. The electric displacement vector, which appears in the first of Maxwell's equations, $\nabla \cdot \mathfrak{D} = 0$, is related to the dielectric polarisation via the definition, $\mathfrak{D}(\mathfrak{E}) = \epsilon_0\mathfrak{E} + \mathfrak{P}(\mathfrak{E})$. Although all crystalline materials possess a nonzero $\chi_e^{(3)}$ coefficient, only anisotropic crystals with no symmetry centre give a nonzero square term in the polarisation expansion ($\chi_e^{(2)} \neq 0$) [1]. As a result, all optical materials with a $\chi_e^{(2)}$ -type nonlinearity are also birefringent [see Sec. 4.3]. Normally, materials are only weakly nonlinear, i.e. $\chi_e^{(1)} \gg \chi_e^{(2)}\mathfrak{E} \gg \chi_e^{(3)}\mathfrak{E}^2$, etc.

Materials which possess a $\chi_e^{(2)}$ -type nonlinearity give rise to three-wave mixing effects. In this case, when two waves interact with the crystal, the dielectric polarisation will contain terms like

$$\cos \omega_1 t \times \cos \omega_2 t \sim \cos(\omega_1 + \omega_2)t + \cos(\omega_1 - \omega_2)t, \quad (6.2)$$

and will thus produce new waves with the combination frequencies $\omega_1 + \omega_2$ (sum-frequency generation) and $\omega_1 - \omega_2$ (difference-frequency generation). The reverse process to sum-frequency generation, where a pump field (ω_p) can generate two new optical fields ($\omega_1 +$

¹Of course, for anisotropic crystals the dielectric susceptibility coefficients are in fact tensors of the appropriate ranks.

$\omega_2 = \omega_p$), is called *parametric down-conversion*, or *spontaneous parametric down-conversion* (SPDC) when the pump beam is a travelling wave propagating freely through the crystal.

All of the experiments in this thesis rely on spontaneous parametric down-conversion in a $\chi_e^{(2)}$ -type nonlinear crystal called β -barium borate (BBO) to create nonclassical states of light. In this chapter, I will give a brief summary of this process in a quantum mechanical picture, and some calculations which may be useful when building such experiments. At various stages, I draw on ideas from Dmitriev, et al., [1] and Kevin Resch's PhD thesis [2].

6.2 Spontaneous parametric down-conversion

Down-conversion is a three-mode optical process where a high-energy pump photon in one mode (ω_p) splits into two lower-energy daughter photons in different modes ($\omega_{1,2}$), where energy is conserved, i.e. $\omega_1 + \omega_2 = \omega_p$. Not surprisingly, down-conversion can be described by the following interaction Hamiltonian,

$$H_{\text{int}} = \gamma a_1^\dagger a_2^\dagger a_p + \gamma^* a_1 a_2 a_p^\dagger, \quad (6.3)$$

where $\gamma \propto \chi_e^{(2)}$ is a coupling constant, and the second term describes the reverse process where two low-energy photons combine to produce a single higher-energy photon.

In most experiments, the pump photons will be provided by a bright laser beam which is close to a coherent state of the optical field (indeed this is normally vital to produce a reasonable down-conversion count rate). So at an initial time ($t = 0$), the ‘‘input’’ state is $|\psi_{\text{in}}(0)\rangle = |0\rangle_1 |0\rangle_2 |\alpha\rangle_p$. After some interaction time, the output state can then be calculated by solving Schrödinger's equation²,

$$|\psi_{\text{out}}(t)\rangle = U(t)|\psi_{\text{in}}(0)\rangle, \quad (6.4)$$

$$= \exp\left[\frac{-i}{\hbar} H t\right] |0\rangle_1 |0\rangle_2 |\alpha\rangle_p, \quad (6.5)$$

$$= A \sum_{\nu} \frac{1}{\nu!} \left(\frac{-it}{\hbar} H\right)^{\nu} |0\rangle_1 |0\rangle_2 |\alpha\rangle_p, \quad (6.6)$$

$$= A \sum_{\nu} [1 + O(|\Gamma|^2/\nu^2) n_1 n_2 n_p + \dots] \frac{1}{\nu!} \Gamma^{\nu} (a_1^\dagger)^{\nu} (a_2^\dagger)^{\nu} a_p^{\nu} |0\rangle_1 |0\rangle_2 |\alpha\rangle_p, \quad (6.7)$$

$$= A \sum_{\nu} [1 + O(|\Gamma|^2 |\alpha|^2) + \dots] \Gamma^{\nu} \alpha^{\nu} |\nu\rangle_1 |\nu\rangle_2 |\alpha\rangle_p, \quad (6.8)$$

²The steps linking Eqs (6.6) and (6.8) are nontrivial. In Eq. (6.6), due to the form of the Hamiltonian, each element inside the sum over ν involves many terms, each with a different number of down-conversion photon pairs. In Eq. (6.8), the like terms are collected together so that each element inside the sum contains all terms with the same number (ν) of pairs, and direct algebraic manipulation gives the general form of the higher-order terms. A more complete proof of this result is provided in App. 6.A. Equation (6.7) provides a conceptual link between Eqs (6.6) and (6.8) (n.b. $a^\dagger a |\alpha\rangle \approx |\alpha|^2 |\alpha\rangle \approx a a^\dagger |\alpha\rangle$ for $|\alpha|^2 \gg 1$). After the simple zeroth order term, the next most significant terms to contribute ν photon pairs to the output will have an extra annihilation and creation operator for each mode ($\rightarrow n_1 n_2 n_p$) and will come from *two* orders higher in the Taylor expansion of Eq. (6.6), giving an extra $|\Gamma|^2 / (\nu+1)(\nu+2) = O(|\Gamma|^2 / \nu^2)$ in the expansion coefficient.

$$\approx A [|00\rangle + \Gamma\alpha |11\rangle + (\Gamma\alpha)^2 |22\rangle + \dots] \otimes |\alpha\rangle_p, \quad (6.9)$$

where $\Gamma = -it\gamma/\hbar$ and, in most cases, $|\Gamma\alpha| \ll 1$, and where $n_j = a_j^\dagger a_j$ is the number operator. So in the limit that the pump beam is not significantly changed by the interaction with the crystal (the weak nonlinearity limit), the down-conversion process actually produces a number-entangled state in the output modes. Note that this is also in a product state with the pump mode (i.e. the state is separable), as is expected for a “classical” pump beam.

So to summarise, the output state from a down-conversion source is a number-entangled state

$$|\psi_{dc}\rangle \sim |00\rangle + \eta|11\rangle + \eta^2|22\rangle + \dots, \quad (6.10)$$

where $|\nu_1\nu_2\rangle$ describes the number of photons in each of the output modes³, and $\eta = -it\gamma\alpha/\hbar$ is an overall efficiency parameter which is related to the pump power, the non-linear coupling constant, and the thickness of the crystal (via the interaction time).

In the standard approach to this problem, the pump beam is treated as a non-depleting classical field and the pump operators in the interaction Hamiltonian are replaced with classical electric field amplitudes. Using this simplification, it is then straightforward to derive the form of the output state of the down-conversion photons. The above derivation shows that the same result can be obtained using a truly quantum treatment of the system. However, the complex approximations involved [see App. 6.A] highlight that the situation is much more complicated than the simpler calculation involving a classical pump field would suggest. It would be an interesting problem to explore what happens in the down-conversion regimes where those approximations do not hold true.

When coupled with photodetection, a down-conversion source can be used as a probabilistic (spontaneous) source of single-photon pairs produced at a rate determined by η , i.e. $|\psi_{dc}\rangle \approx \eta|11\rangle$. In current experiments, the data consists only of events where a detector “click” is recorded (i.e. a photon is detected)—the “no click” events never register, so we generally do not even see the vacuum ($|00\rangle$) term. Moreover, although most detectors cannot distinguish between one-photon and many-photon events, and so do not eliminate the higher number states, in general, $|\eta|$ is so small that these higher-order components are negligible⁴. Thus, the single-pair $|11\rangle$ term normally makes the only significant contribution to the data.

Note that for sufficiently small interaction times, dt , the first-order expansion is sufficient, i.e.,

$$|\psi_{out}\rangle = dU(dt)|\psi_{in}\rangle \approx \left[|00\rangle - \frac{i\gamma\alpha dt}{\hbar}|11\rangle \right] \otimes |\alpha\rangle_p. \quad (6.11)$$

In addition to the effects described above, the efficiency of a real down-conversion process is also affected by the *phase-matching conditions*, which are determined by the geometry of the system. In the ideal (thick-crystal) limit, these phase-matching conditions ensure that momentum (as well as energy) is conserved in the photon “collision”, and this determines

³Historically, these are often called the *signal* and *idler* photons.

⁴This is always true for cw-pumped crystals, which I use in my experiments, but with pulsed pumps, the peak powers can be sufficiently high for these to make significant contributions.

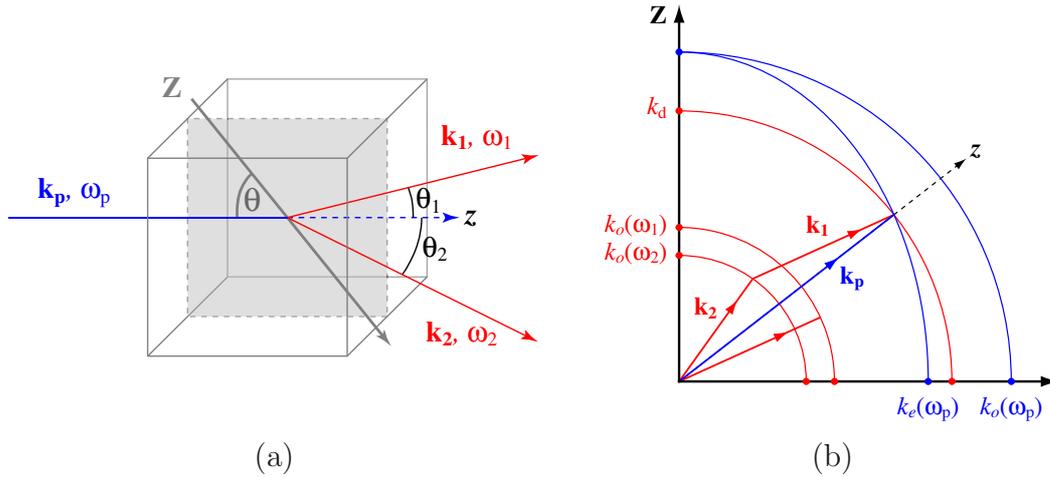


Figure 6.1: Type-I phase matching: (a) physical geometry— \mathbf{Z} is the optic axis of the crystal, \mathbf{z} is the propagation direction for the pump beam, and θ , the phase-matching angle, is the angle between these two directions; (b) geometry of momentum conservation—the blue refers to the high-energy pump beam, and the red to the lower-energy daughter photons.

the propagation angles of a given pair of output photons, ω_1 and $\omega_2 = \omega_p - \omega_1$. In finite-width crystals, the output is centred at these ideal angles, but momentum conservation is enforced less strictly, resulting in an angular spread of the output photons which increases with decreasing crystal thickness [further discussion in Sec. 6.4].

There are several different classes of phase-matching which are possible in different crystal geometries, but I will focus on Type-I phase-matching for negative uniaxial crystals⁵, since this is what we use for all the experiments in this thesis. In the next section, I will describe Type-I phase matching in the ideal limit, and after that, I will use the simple example of collinear Type-I down-conversion to highlight the main ideas behind the thick-crystal approximation for ideal phase-matching.

6.3 Ideal, noncollinear Type-I phase matching

In type-I phase matching, an extraordinarily polarised photon (ω_p) splits into two ordinarily-polarised daughter photons (ω_1 and ω_2). Consider the system illustrated in Fig. 6.1(a). To begin with, I will assume that I wish to calculate the phase-matching angle, θ , that is required to produce a particular complementary pair of photons (i.e. a particular ω_1 and ω_2) at a specific opening angle, $\theta_{\text{op}} = \theta_1 + \theta_2$. Conservation of energy and momentum gives

$$\omega_p = \omega_1 + \omega_2 \quad (6.12)$$

$$\mathbf{k}_p = \mathbf{k}_1 + \mathbf{k}_2. \quad (6.13)$$

⁵This is often written as Type-I⁽⁻⁾.

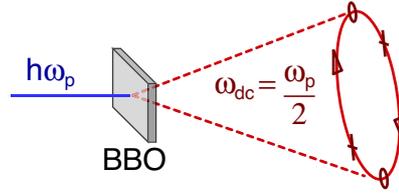


Figure 6.2: Type-I phase matching produces a cone of energy-degenerate down-conversion photon pairs.

The conservation of momentum can then be broken down into two independent component equations:

$$k_p = k_1 \cos \theta_1 + k_2 \cos \theta_2, \quad (6.14)$$

$$0 = k_1 \sin \theta_1 - k_2 \sin \theta_2. \quad (6.15)$$

Because the down-conversion photons are ordinarily polarised, the magnitude of their wave vector does not depend on their direction of propagation, i.e. $k_{1,2} = n_o(\omega_{1,2}) \omega_{1,2}/c$. Given the opening angle and output photon frequencies, the transverse momentum equation allows the calculation of the individual angles, θ_1 and θ_2 , via

$$\tan \theta_1 = \frac{k_2 \sin \theta_{op}}{k_1 + k_2 \cos \theta_{op}}, \quad (6.16)$$

which in turn determines the total *longitudinal* momentum of the down-conversion photons, $k_1 \cos \theta_1 + k_2 \cos \theta_2 \equiv k_d$. Since the pump momentum depends on the phase-matching angle, $k_p(\theta) = n^e(\theta)\omega_p/c$, the remaining task is to find the phase-matching angle that satisfies the requirement for conservation of longitudinal momentum. To do this, recall that for an extraordinarily polarised ray propagating through a birefringent material, the refractive index is [Eq. (4.9)]

$$n^e(\theta) = n_o \sqrt{\frac{1 + \tan^2 \theta}{1 + n_o^2/n_e^2 \tan^2 \theta}}. \quad (6.17)$$

Using this equation, it is easy to show that

$$\tan^2 \theta = \frac{k_d^2 / [n_o^2(\omega_p)k_p^2] - 1}{1 - k_d^2 / [n_e^2(\omega_p)k_p^2]}. \quad (6.18)$$

The geometry of momentum conservation in this phase-matching calculation is illustrated in Fig. 6.1(b). Because the wave vector is proportional to the refractive index, the momentum of light in a birefringent material is described by an ellipse [cf. the refractive index ellipse, Fig. 4.2]. Phase matching occurs at the intersection between the curves for the momentum of the extraordinarily polarised pump and the total momentum of the down-conversion photons, the radius of which is in turn determined by the opening angle.

As an example, this calculation allows one to calculate the threshold for the phase-matching angle, below which Type-I down-conversion is impossible. For a UV laser at 351.1nm pumping a BBO crystal, the phase-matching angle must be at least 33.53° before any energy-degenerate photon pairs are produced in the output modes—this conditions corresponds to *collinear* Type-I phase-matching. As θ increases, the opening angle

increase also and the output photons travel in different directions. Note, however, that momentum conservation only affects the angle between the propagation directions of each of the output photons and the pump beam. The calculation does not change if the output photons are rotated around the pump axis. Therefore, the energy-degenerate photon pairs are emitted on opposite sides of a cone with the opening angle, θ_{op} [Fig. 6.2].

The refractive indices in this calculation were determined from the following Sellmeier equations for BBO ([1], p. 99, λ in μm):

$$n_o^2 = 2.7359 + \frac{0.01878}{\lambda^2 - 0.01822} - 0.01354\lambda^2, \quad (6.19)$$

$$n_e^2 = 2.3753 + \frac{0.01224}{\lambda^2 - 0.01667} - 0.01516\lambda^2. \quad (6.20)$$

Assume now that I instead wish to calculate the opening angle for a particular pair of output photons and given phase-matching angle (perhaps determined by the cut-angle of a crystal). Using the momentum conservation equations, it is not too difficult to show that

$$\cos \theta_1 = \frac{k_p^2(\theta) - k_2^2 + k_1^2}{2k_1k_p(\theta)}, \quad (6.21)$$

where again, $k_p(\theta) = n^e(\theta)\omega_p/c$. This is the opening angle inside the crystal—the angle outside the crystal will be slightly larger and must be calculated using Snell’s Law. For example, with the UV laser above pumping a crystal cut at 33.9° , the degenerate photon pairs are produced with a half-opening angle of around 1.78° inside the crystal, and 2.95° when they exit the crystal.

6.4 Imperfect, collinear Type-I phase matching

Phase matching is in fact an interference effect. As it turns out, ideal phase-matching conditions are only created inside an infinitely thick crystal. In a real, finite-width crystal, momentum conservation is not strictly enforced—that is, some of the optical momentum may be transferred to (or from) the crystal. To understand this effect more fully, I return now to a more detailed analysis of the down-conversion process. In general, this is done using a semi-classical approach, where the high-frequency pump laser is treated as a classical field and the down-conversion photons are treated as quantum systems (see, e.g., Ref. [3]). Here, however, I present a fully quantum mechanical treatment where the pump laser is instead treated as a quantum field mode in a coherent state [as in Sec. 6.2]. To do this, I will use the ideas introduced in Sec. 2.2.2 to describe the spatial evolution of optical field modes. In fact, this calculation provides an elegant example of how the translation operator, $T(\mathbf{r})$, can be used to advantage in quantum optics. Unfortunately, I have not been able to complete the down-conversion calculation for the full, non-collinear phase-matching configuration. Instead, I consider the simpler collinear configuration which displays most of the salient features of the non-collinear case, and I will comment on how these features apply in the more general context.

Consider a case where down-conversion in a crystal of length L is pumped by a plane-wave coherent state $|\alpha\rangle_p$ [see Fig. 6.3(a)]. The goal is to calculate the output quantum state at

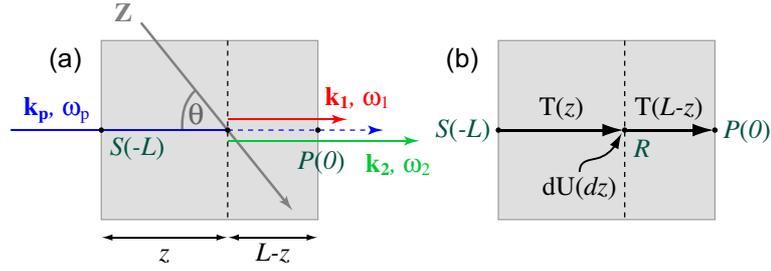


Figure 6.3: A quantum treatment of collinear, Type-I phase matching in a finite-width crystal: (a) the physical geometry for one quantum evolution path; (b) the same path described in quantum operations.

the point $P(z=0)$, the exit face of the crystal. To do this, the crystal can be broken up longitudinally into infinitesimal slices (thickness dz). As the pump propagates through each slice the evolution can be modelled by nonlinear interaction, $dU(dz)$, followed by a spatial propagation, $dT(dz)$. The interaction operator can be calculated by applying the interaction Hamiltonian for a time $dt = dz n^e(\theta)/c$ (the time it takes the pump beam to propagate a distance dz through the crystal) [cf. Sec. 6.2], giving

$$dU(dz) = 1 - \frac{i\gamma n^e(\theta) dz}{\hbar c} \left[a_1^\dagger a_2^\dagger a_p + a_1 a_2 a_p^\dagger \right]. \quad (6.22)$$

The spatial propagation is then defined by the translation operator, $T(\mathbf{r}) = \exp\left(\frac{-i}{\hbar} \mathbf{r} \cdot \mathbf{G}\right)$, where⁶ $\mathbf{r} = dz \hat{\mathbf{z}}$, and \mathbf{G} is the total momentum operator for the three optical modes [see Eq. (2.63)], i.e.,

$$\mathbf{G} = \hbar \mathbf{k}_1 a_1^\dagger a_1 + \hbar \mathbf{k}_2 a_2^\dagger a_2 + \hbar \mathbf{k}_p a_p^\dagger a_p. \quad (6.23)$$

Thus, the output state can be written

$$|\psi_{\text{out}}\rangle = dT dU \dots dT dU dT dU |\psi_{\text{in}}\rangle. \quad (6.24)$$

Given its form, successive applications of dU in this way will give rise to many interfering terms in the output, each of which corresponds to a specific path which the quantum state can evolve along as the light propagates through the crystal. Alternatively, the evolution paths can be interpreted as the different *histories*⁷ which are possible for the photons arriving at point P .

In this calculation I will assume that nonlinearity is sufficiently weak that *at most one* interaction element dU gives rise to a down-conversion event. Under this condition, the form of the output state is greatly simplified, and each contributing path is completely described by the point $R(-L+z)$ —the position of the down-conversion event relative to the point P . This is illustrated in Fig. 6.3(b). The output state now breaks up into an integral in the propagation direction

$$|\psi_P\rangle = \left[T(L) - \int T(L-z) d\tilde{U}_R(dz) T(z) \right] |\psi_S\rangle, \quad (6.25)$$

where $|\psi_S\rangle = |0\rangle_1 |0\rangle_2 |\alpha\rangle_p$ is the input state at point S , and $d\tilde{U}$ is the nonidentity component of the interaction operator ($dU \equiv 1 - d\tilde{U}$) which acts on the state at point R . This

⁶ $\hat{\mathbf{z}}$ is the unit vector in the z direction.

⁷This is essentially a Feynman path integral.

integral describes the total output state at point P as a sum of all possible positions of the down-conversion event inside the crystal. This is a quantum version of the semi-classical derivation given by Kevin Resch in App. A.3.3 of his PhD thesis [2].

Some direct calculation now gives

$$|\psi_P\rangle = \left[|00\rangle - \left(\int_0^L dz e^{-i\Delta kz} \right) e^{-i(k_1+k_2)L\eta} |11\rangle \right] \otimes |\alpha e^{-ik_p L}\rangle_P, \quad (6.26)$$

$$= \left[|00\rangle - \left[\frac{e^{-i\Delta kz}}{-i\Delta k} \right]_0^L e^{-i(k_1+k_2)L\eta} |11\rangle \right] \otimes |\alpha e^{-ik_p L}\rangle_P, \quad (6.27)$$

$$= \left[|00\rangle - L \operatorname{sinc} \left(\frac{\Delta k L}{2} \right) e^{-ik_{\text{av}} L \eta} |11\rangle \right] \otimes |\alpha e^{-ik_p L}\rangle_P, \quad (6.28)$$

where $\eta = i\gamma n^e(\theta)\alpha/(\hbar c)$, and I have used the following definitions for the momentum parameters,

$$\Delta k \equiv k_p - k_1 - k_2, \quad (6.29)$$

$$k_{\text{av}} \equiv (k_p + k_1 + k_2)/2. \quad (6.30)$$

In other words, the amplitude of the d.c. photon-pair component is proportional to η , an efficiency parameter related to phase-matching angle (θ) and properties of the nonlinear crystal [cf. Eq. (6.10)]; to L , the length of the crystal; and is also a function of Δk , a parameter describing how much momentum is lost from the optical system as it passes through the crystal. This last parameter characterises the down-conversion process in a real, finite-length crystal. In a crystal of infinite length, the only significant contribution to the d.c. amplitude is for $\Delta k=0$, and momentum is conserved. However, if L is finite, then for a given phase-matching angle and pump wavelength, there are a range of energy-conserving, photon-pair wavelengths which will exhibit non-zero collinear down-conversion power.

In the context of non-collinear, Type-I down-conversion, this basic idea can also be interpreted in a slightly different way. In this case, for a given phase-matching angle and given optical wavelengths (for pump *and* energy-conserving d.c. photon pairs), non-negligible down-conversion will occur over a range of non-collinear angles (i.e. θ_1 and θ_2). The peak probability of the angular spectrum of the down-conversion will correspond to the condition for perfect momentum conservation.

6.A Appendix: Proof—down-conversion expansion

The purpose of this appendix is to provide the details linking Eq. (6.6), the Taylor expansion of the unitary down-conversion operator, and Eq. (6.8), which collects together all terms in the expansion with the same number of down-conversion photon pairs. By direct substitution,

$$|\psi_{\text{out}}(t)\rangle = A \sum_{\nu} \frac{1}{\nu!} \left(\frac{-it}{\hbar} H \right)^{\nu} |0\rangle_1 |0\rangle_2 |\alpha\rangle_P, \quad (6.31)$$

$$= A \sum_{\nu} \frac{1}{\nu!} \left(\frac{-it}{\hbar} \gamma a_1^{\dagger} a_2^{\dagger} a_p + \frac{-it}{\hbar} \gamma^* a_1 a_2 a_p^{\dagger} \right)^{\nu} |0\rangle_1 |0\rangle_2 |\alpha\rangle_P. \quad (6.32)$$

Focussing on and expanding the expression in brackets,

$$\begin{aligned} \left(\dots \right)^\nu = \left(\frac{-it}{\hbar} \right)^\nu & \left(\gamma^\nu b_{\nu,0} (a_1^\dagger a_2^\dagger a_p)^\nu + \gamma^{\nu-2} |\gamma|^2 \underline{b_{\nu,1} (a_1^\dagger a_2^\dagger a_p)^{\nu-1} (a_1 a_2 a_p^\dagger)} \right. \\ & \left. + \dots + \gamma^{*\nu-2} |\gamma|^2 \underline{b_{\nu,\nu-1} (a_1^\dagger a_2^\dagger a_p) (a_1 a_2 a_p^\dagger)^{\nu-1}} + \gamma^{*\nu} b_{\nu,\nu} (a_1 a_2 a_p^\dagger)^\nu \right), \end{aligned} \quad (6.33)$$

where $b_{\nu,j}$ are expansion coefficients and $b_{\nu,0} = b_{\nu,\nu} = 1$. Note that each term in the expansion produces a state with *two fewer* down-conversion photon pairs than the preceding term.

In fact, this simple binomial expansion is an oversimplification, because the terms in the Hamiltonian contain non-commuting operators. If they were only complex numbers, then this simple expansion would be correct and the coefficients would be the standard numbers from combinatorics, i.e. $b_{\nu,j} = {}^\nu C_j = \nu!/[j!(\nu-j)!]$. However, the non-commutation of the operators complicates the counting argument for the underlined terms. I treat this expansion more rigorously once I have determined which terms need to be calculated.

In order to determine the final form of the output state [Eq. (6.8)], I need to collate all terms with the same number of photon pairs in the output modes. To produce ν photon pairs, the simplest process corresponds to the application of ν creation operators and no annihilation operators. There is only one such term which arises in the ν^{th} order of the Taylor expansion. The next-simplest ν -photon contributions come from terms in the $(\nu+2)^{\text{th}}$ expansion order, which apply an extra annihilation *and* creation operator, and then from the $(\nu+4)^{\text{th}}$ order, and so forth. The new output state is:

$$\begin{aligned} |\psi_{\text{out}}(t)\rangle = & A \sum_{\nu} \frac{1}{\nu!} \Gamma^\nu \left[(a_1^\dagger a_2^\dagger a_p)^\nu - \frac{|\Gamma|^2}{(\nu+1)(\nu+2)} \underline{b_{\nu+2,1} (a_1^\dagger a_2^\dagger a_p)^{\nu+1} (a_1 a_2 a_p^\dagger)} + \dots \right] |0\rangle_1 |0\rangle_2 |\alpha\rangle_p, \end{aligned} \quad (6.34)$$

where $\Gamma = -it\gamma/\hbar$. Equation (6.34) is a direct rearrangement of Eq. (6.32) where the underlined terms result from the binomial expansion in Eq. (6.33).

Before examining the underlined terms more carefully, I need to verify the following operator identities:

$$aa^{\dagger\mu}|0\rangle = \left(\mu a^{\dagger\mu-1} + a^{\dagger\mu} a \right) |0\rangle = \mu a^{\dagger\mu-1} |0\rangle, \quad (6.35)$$

$$aa^\dagger|\alpha\rangle \approx |\alpha|^2 |\alpha\rangle, \quad (6.36)$$

$$a^\mu a^\dagger |\alpha\rangle \approx \alpha^{\mu-1} (\mu - 1 + |\alpha|^2) |\alpha\rangle \approx \alpha^{\mu-1} |\alpha|^2 |\alpha\rangle, \quad (6.37)$$

where the approximations hold for large values of α . The first identity is a direct consequence of the commutation properties for annihilation and creation operators [see

Sec. 2.2.1]. The second identity can be justified as follows:

$$|a_p a_p^\dagger |\alpha\rangle|^2 = \langle \alpha | a_p a_p^\dagger a_p a_p^\dagger |\alpha\rangle, \quad (6.38)$$

$$= \langle \alpha | (1 + a_p^\dagger a_p) (1 + a_p^\dagger a_p) |\alpha\rangle, \quad (6.39)$$

$$= 1 + 2|\alpha|^2 + \langle \alpha | a_p^\dagger (1 + a_p^\dagger a_p) a_p |\alpha\rangle, \quad (6.40)$$

$$= 1 + 3|\alpha|^2 + |\alpha|^4, \quad (6.41)$$

$$\approx |\alpha|^4 \quad (\text{large } \alpha), \quad (6.42)$$

$$\Rightarrow a_p a_p^\dagger |\alpha\rangle \approx |\alpha|^2 |\beta\rangle. \quad (6.43)$$

To find $|\beta\rangle$, note that

$$\langle \alpha | \beta \rangle \approx \langle \alpha | a a^\dagger |\alpha\rangle / |\alpha|^2 = \langle \alpha | (1 + a^\dagger a) |\alpha\rangle / |\alpha|^2 = (1 + |\alpha|^2) / |\alpha|^2 \approx 1. \quad (6.44)$$

Therefore, $|\beta\rangle \approx |\alpha\rangle$, as required. The third identity then follows directly from the operator commutation relations and the second identity.

The underlined terms of the binomial expansion can now be calculated more rigorously. For example,

$$\begin{aligned} & \underline{b_{\nu+2,1} (a_1^\dagger a_2^\dagger a_p) \nu+1 (a_1 a_2 a_p^\dagger)} |0\rangle_1 |0\rangle_2 |\alpha\rangle_p = \\ & \left[(a_1^\dagger a_2^\dagger a_p) \nu+1 (a_1 a_2 a_p^\dagger) + (a_1^\dagger a_2^\dagger a_p)^\nu (a_1 a_2 a_p^\dagger) (a_1^\dagger a_2^\dagger a_p) \right. \\ & \quad \left. + \dots + (a_1 a_2 a_p^\dagger) (a_1^\dagger a_2^\dagger a_p) \nu+1 \right] |0\rangle_1 |0\rangle_2 |\alpha\rangle_p. \quad (6.45) \end{aligned}$$

Using the above identities and some direct algebraic manipulation, this can be simplified greatly for large α to give:

$$\underline{b_{\nu+2,1} (a_1^\dagger a_2^\dagger a_p) \nu+1 (a_1 a_2 a_p^\dagger)} |0\rangle_1 |0\rangle_2 |\alpha\rangle_p \approx \frac{1}{2}(\nu+1)(\nu+2) |\alpha|^2 \alpha^\nu (a_1^\dagger a_2^\dagger)^\nu |0\rangle_1 |0\rangle_2 |\alpha\rangle_p. \quad (6.46)$$

Therefore, as required, the final form of the output state is:

$$|\psi_{\text{out}}(t)\rangle \approx A \sum_{\nu} \left[1 - \frac{1}{2} |\Gamma|^2 |\alpha|^2 + \dots \right] \Gamma^\nu \alpha^\nu |\nu\rangle_1 |\nu\rangle_2 |\alpha\rangle_p. \quad (6.47)$$

6.2 Chapter 6 References

- [1] V. G. Dmitriev, G. G. Gurzadyan, and D. N. Nikogosyan. *Handbook of Nonlinear Optical Crystals*, volume 64 of *Springer Series in Optical Sciences*. Springer, Berlin, Heidelberg, second, revised and updated edition, 1997.
- [2] K. J. Resch. *Making photons talk to each other—nonlinear optics in the quantum domain*. PhD thesis, The University of Toronto, Toronto, Canada, 2003.
- [3] M. H. Rubin. Transverse correlation in optical spontaneous parametric down-conversion. *Physical Review A*, 54:5349, 1996.

Part III

Applications in quantum information

“I should see the garden far better,” said Alice to herself, “if I could get to the top of that hill: and here’s a path that leads straight to it—at least, no, it doesn’t do that—” (after going a few yards along the path, and turning several sharp corners), “but I suppose it will at last. But how curiously it twists! It’s more like a corkscrew than a path!”

Through the Looking-Glass (and what Alice found there),
Lewis Carroll

Chapter 7

Measuring entangled qutrits and their use for quantum bit commitment

Acknowledgement

The work in this chapter was first published in Ref. [1], and where appropriate, I have incorporated some of the text of that paper. The experiments were performed with Rohan Dalton, and contributed to the research project in his Honours degree in Physics at the University of Queensland in 2003.

I would particularly and most gratefully like to acknowledge Carl Warner of the now-defunct UQ Photography unit for his assistance in the developmental stages of printing the hologram images. His knowledge and technical expertise were invaluable in producing the high quality images.

7.1 Introduction

Many two-level quantum systems, or *qubits*, have been used to encode information, but using d -level systems, or *qudits*, enables access to larger Hilbert spaces, which can provide significant improvements over qubits such as increased channel capacity in quantum communication [2]. Furthermore, when entangled, *qutrits* ($d=3$) provide the best known levels of security in other quantum communication protocols such as bit-commitment and coin-flipping, which cannot be matched using qubit-based systems [3] [see Sec. 7.5 for more details]. The ability to completely characterise entangled qudits is critical for these applications. This is only possible using quantum state tomography [4, 5].

Entangled qudits have been realised in few physical systems, and only indirect measurements have been made of the quantum states of these systems. Qutrit entanglement has been generated between the arrival times of correlated photon pairs, where fringe measurements were used to infer features such as fidelities with specific entangled states and to estimate a potential Bell violation¹ [6]. Transverse spatial modes have also been used to encode higher-dimensional quantum information, but once again, there have only been measurements demonstrating, but not quantifying, spatial mode entanglement in parametric downconversion [7], including fringe measurements [8, 9] and the violation of a two-qutrit Bell inequality² [11].

In this experiment, we used quantum state tomography to completely characterise entangled, photonic qudits (both $d = 2$ and 3) encoded in transverse spatial modes, measuring the amount of entanglement and the degree of mixture. We used tomographic measurement sets which only involved two-state superpositions [see Sec. 3.2.2], which had many practical advantages, and we investigated both degenerate and nondegenerate spatial Hilbert spaces. We then show how to use the qutrit system in a quantum bit-commitment protocol and investigate the experimental requirements for achieving the best known security [3]. Finally, I explore the implications on these spatial tomography techniques of more recent theoretical work in understanding the detailed operation of plane-wave holograms, and I use this information to improve the original tomographic analysis. This work constitutes the most complete characterisation of spatially encoded qubits and qutrits to date and the first quantitative measurement of entangled qutrit states.

7.2 Making computer generated holograms

This section provides a brief technical overview of the procedure for making computer-generated holograms [Fig. 7.1], along with some of the important practical steps which are required when first setting up the process. Though not of major scientific interest, this may be of some use to the person who actually has to do the work.

7.2.1 The computer generated images

Figure 7.2 shows sinusoidal holograms which could be used to create two higher-order modes. The simple split-fringe pattern for the VM_{+1} mode (a) becomes much more complicated for the VM_{+7} mode (b), and greater resolution is required to map out the central region.

There are also several other important considerations to bear in mind when generating the interference patterns for CG holograms. The angular offset between the reference and subject beams (which determines the fringe spacing) should ensure at least several pixels

¹Fringe visibilities beyond a certain value can be used to predict the size of a Bell violation without actually performing the Bell correlation measurements.

²Violating Bell's inequality is not a measure of entanglement, as states with different amounts of entanglement and mixture can give the same Bell violation [10].

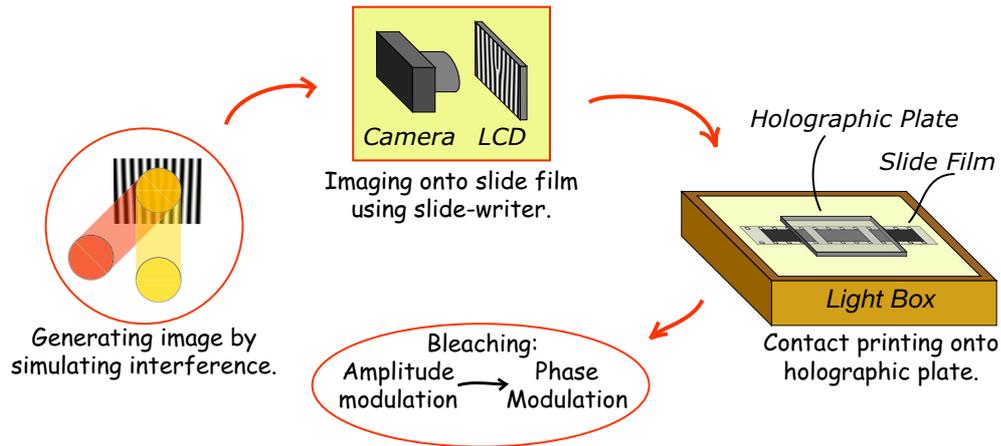


Figure 7.1: A schematic overview of the process for making CG holograms.

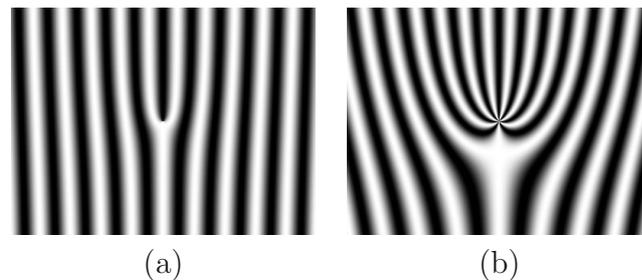


Figure 7.2: Sinusoidal holograms designed to convert a lowest-order Gaussian input beam into (a) a VM_{+1} mode; (b) a VM_{+7} mode.

per fringe. The angular offset should give an exact, even number of pixels per fringe, to ensure that fringes have a pure spatial frequency. This effect is most obvious in binary images, where a mismatch in the angular offset can give rise to an interference pattern where adjacent black and white fringes can vary by up to two pixels in size [Fig. 7.3(a)]. When this is the case, the interference pattern contains more than one spatial frequency, which may give rise to several modes being reconstructed at slightly different diffraction angles, where only one mode is desired.

The final effect is particularly relevant to jump holograms (again, it is most obvious in binary holograms). The easiest way to produce a binary hologram is to scale/shift the intensity of the interference pattern so that it is half positive and half negative, and set the pixel value (0 or 255) accordingly. If the offset of the $x-y$ grid is such that the intensity of a particular point is analytically zero, then the resulting image will be subject to machine precision error. This can produce single-pixel displacement errors in the phase discontinuity at the origin of the $x-y$ plane [Fig. 7.3(b)]. This can be fixed by shifting the $x-y$ grid half a pixel to one side.



Figure 7.3: Typical errors in binary computer-generated holograms. (a) An example of a holographic pattern with varying fringe spacing. Most of the fringes are six pixels wide, whereas the three circled fringes are seven pixels wide. (b) An example of a holographic pattern with a single-pixel displacement error in the phase discontinuity line resulting from a poorly chosen offset for the x - y grid alignment.



Figure 7.4: Two test images for characterising the slide film and the Polaroid slide writer: (a) a smoothly varying grey scaling—to check the response of the film and the exposure levels required; (b) image of a square cross (2048×2048 pixels)—allows accurate measurement of pixel size and astigmatism.

7.2.2 Making the master images

In the slide printing step, we exposed Kodak Technical Pan black and white (B&W) slide film with the CG images using a Polaroid (ProPalette 7000) slide writer, and developed the slide film using Ilford ID11 developer. It is important to use B&W film, because it only has a single layer of photographic emulsion, essential for good, low-distortion optical elements. The slide writer consists of a 4096×3072 pixel, colour LCD, which is imaged directly onto the slide film using a camera. The LCD has 8-bit greyscale resolution (256 shades between black and white) and the camera images around 4000×2700 of the central pixels, corresponding to an image size of $35\text{mm} \times 23.7\text{mm}$, which gives a physical pixel size of around $10\mu\text{m}$ square. The image should not be scaled in any way to fit the film, because this can produce undesirable aliasing effects. Although the slide writer does not allow explicit control over the exposure time of the image, this can be partially simulated by converting the greyscale image into an RGB image and adjusting the relative brightness of the red, green and blue levels all by the same amount.

We ran several tests to characterise the quality of the slide writer. To check the quality of the greyscale resolution, we printed an image with a smoothly varying ramp from white to black [Fig. 7.4(a)]. While this did not allow us to compare the transmittance between two adjacent grey levels, it did demonstrate that the greyness scale was very smooth, to

the point where the steps were indistinguishable by eye. This pattern also served as a tool for characterising and comparing different types of slide film, and for determining appropriate exposure levels. One could also quantitatively measure the absorption as a function of position on the pattern to see if a greyness scale that looks smooth by eye gives a linear absorption gradient—we did not.

We then tested the pixel resolution of the imaging system of the slide writer in combination with the film quality. We printed an image with a 3×8 array of segments. Each segment contained black and white stripes (horizontal, vertical or diagonal), varying in thickness from 1 to 8 pixels. We found that stripes were still distinguishable at the three pixel wide level.

Finally, we checked for any astigmatism in the LCD or the imaging optics by printing an image containing an X pattern with equal height and width [Fig. 7.4(b)]. By measuring the relative width and height of the X, we determined that the imaging system of the slide writer possessed an astigmatism of no more than $\sim 0.1\%$.

I generated the images as 8-bit greyscale TIFF files, using Matlab to calculate the interference patterns. We have mainly used off-axis, plane-wave holograms so far.

7.2.3 Producing the holograms

Once the image had been developed on slide film, we then used contact printing to transfer it to much higher quality holographic plates. With a good quality slide image, we can make as many holograms as required. Our holographic plates were Slavich PFG-01 plates ($6.3\text{cm} \times 6.3\text{cm}$), which were coated with a red-sensitive emulsion with a maximum sensitivity of $100\mu\text{J}/\text{cm}^2$ at 633nm and a resolution of more than 3000 lines/mm. For the contact printing, we used a “light box” which contained a diffuse, white-light source operated by a timing switch. The holographic plate was then pressed against the slide film image using a mount designed to fix their relative positions, and placed on the light box so that the slide acted as a “shadow mask” for the holographic plate (to minimise diffraction effects, the two emulsions should be in contact). The plate was then illuminated for a time set by the switch. Varying the exposure time adjusts the contrast of the hologram.

A phase hologram can be modulated in two different ways—either by changing the refractive index of the emulsion or its thickness, or both. The ideal depth of this modulation depends on the type of hologram and the desired outcome [see Sec. 5.4]. A sinusoidal phase hologram has a maximum theoretical efficiency into the first order of 33.9% for a peak-to-peak modulation depth of around 1.2π [12]. For a binary phase hologram, the optimum diffraction efficiency is around 40% with a phase contrast of π between the dark and light bands (i.e. the undiffracted mode experiences complete destructive interference) [13]. However, for a blazed phase hologram, the maximum efficiency in the first order is 100% for a phase shift of 2π between the maximum and minimum points of the blazed pattern.

In order to achieve the greatest diffraction efficiencies, we performed the contact printing with a range of exposure times for a single slide image (a VM_{+1} hologram). After the

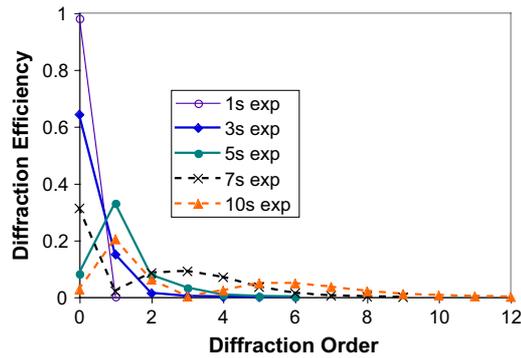


Figure 7.5: A VM_{+1} hologram was contact printed with various exposure times. These diffraction efficiency measurements were taken using a 670nm diode laser coupled through a single-mode fibre to spatially filter the elliptical laser mode.

bleaching step, different exposures give different depths of phase modulation. We measured the diffraction efficiency with a 670nm diode laser for a range of exposure times up to 10s (exposing for 15s caused complete saturation of the image) [Fig. 7.5]. The best coupling into the first order was achieved with 5s exposure, giving a 33% diffraction efficiency, compared with the maximum of 33.9% [12]. It is possible that some saturation may have occurred to alter the fringe shape more towards a binary hologram, but this is still very good for a cheap, symmetrical (i.e. not blazed) hologram. The efficiencies were calculated as the power in the desired mode relative to the total power *transmitted* through the hologram (i.e. removing the effects of absorption and reflection).

The final steps in making phase holograms are to develop and then bleach the emulsion. The plates were developed using Agfa chemicals (fixer—Agfa G354; developer—Agfa G153) and then bleached using an aqueous solution of mercuric chloride (HgCl_2). Developing the plate converts the (still reactive) exposed plate into a stable absorption image (i.e. amplitude modulating). The bleaching step converts the amplitude modulating image into a phase modulating image, by substituting the silver salt (responsible for the absorption) with the larger mercuric salt. As a result, the emulsion expands and regions of high absorption are converted into regions of thicker emulsion. It is important that the HgCl_2 solution be sufficiently concentrated (it is very difficult to dissolve HgCl_2 in water and requires heating and significant stirring), because if the bleaching process is too slow, it is hard to achieve an evenly bleached image.

There are several possible ways of improving the holographic production. At the moment, the resolution of the final hologram (and hence the quality of the beam reconstruction) is limited by the resolution of the slide film (up to 320 lines/mm), and even more by the slide writer (potentially up to ~ 200 lines/mm, depending on the model). It would therefore be beneficial to circumvent these steps. One possibility is to print the image in extremely large format (e.g. A0 size) and to photograph the image directly onto the holographic plate. A $1\text{m} \times 0.75\text{m}$ image printed at 600dpi would correspond to an image of around 23600 pixels \times 17700 pixels. Shrinking this to the size of a 35mm slide film image would give a resolution of around 670 lines/mm—already a significant improvement, and better quality printing should be possible. Although there are usually many problems with performing such extreme reductions without any optical distortion,

there is specially designed (old) poster printing equipment with mounted cameras which could be compatible with the holographic plates. Obviously, this is a technique which would be useful for printing holograms where resolution is a particular issue, and once all other problems have been ironed out. However, this would probably be impractical for the everyday requirements of experiments.

The major problem with our current hologram production process is that mercuric chloride is an extremely toxic chemical. There are two other possible types of holographic plates which we have not tested yet. The Slavich PFG-03M plate is also sensitive to 633nm—it has a much lower sensitivity ($1500\mu\text{J}/\text{cm}^2$), but better resolution (>5000 lines/mm). The lower sensitivity should not be a problem, as one could simply expose longer with a brighter light source. The major advantage is that the recommended development procedure contains an inherent bleach, so that mercuric chloride is no longer required. An alternative is a grainless dichromated gelatine plate (Slavich PFG04) which uses a non-toxic development process involving heating and alcohol baths, and produces a phase hologram inherently by inducing a direct change in the refractive index of the emulsion. Although it is the least sensitive emulsion ($250000\mu\text{J}/\text{cm}^2$), it is sensitive to green light, normally a much stronger component of any white light source than red. Its grainless structure gives it extremely high resolution and also a much higher potential diffraction efficiency (75% as opposed to 40 – 45% for a hologram recorded directly in an interferometer). It also has a much thicker emulsion, perhaps making a larger phase change achievable.

7.3 Experimental design

*“Come, listen, my men, while I tell you again,
The five unmistakable marks,
By which you may know, wheresoever you go,
The warranted genuine Snarks.*

*“Let us take them in order. The first is the
taste,
Which is meagre and hollow, but crisp:
Like a coat that is rather too tight in the waist,
With a flavour of Will-o’-the-wisp.”*

*The Hunting of the Snark, an Agony in Eight
Fits, Lewis Carroll*

The main practical goal of this experiment was to characterise the spatial degree of freedom of the two-photon output state from spontaneous parametric down-conversion using tomography, and to investigate its usefulness as a source of spatial entanglement. To perform the tomography, we used the measurement tools I described in Ch. 5. The experimental apparatus [Fig. 7.6] can be broken up into several stages—the down-conversion source, the mode-matching optics, the photon-collection optics, and the photon-counting electronics.

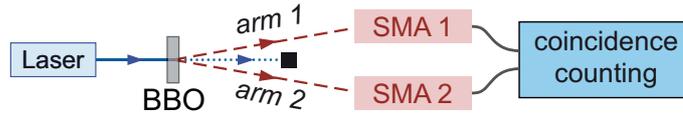


Figure 7.6: A conceptual schematic showing the main stages of the experimental apparatus. The laser pumps the BBO nonlinear crystal to produce the photon pairs, the spatial mode analysers (SMAs) filter and collect the photons, and the coincidence counting electronics detect and count the pairs that arrive simultaneously. The other stage includes the mode-matching optics (not shown) in the pump and down-conversion beams which help to optimise the collection efficiency of the SMAs.

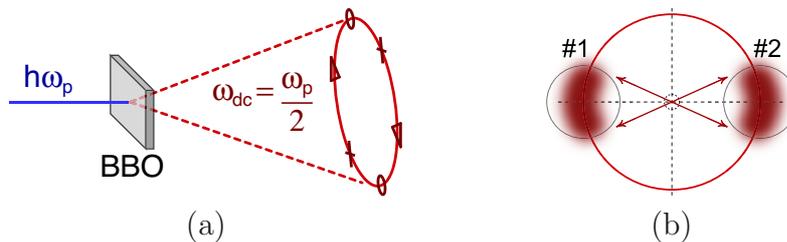


Figure 7.7: Down-conversion with Type-I phase matching: (a) energy-degenerate photon pairs exit the crystal on opposite sides of a cone; (b) the mode-matching and photon-collection optics are designed to image spatial modes with significant higher-order Gaussian components.

The down-conversion source was a 0.5mm thick BBO (β -barium borate) crystal which was cut for Type-I phase matching (optical axis at 28.7°) and pumped with a continuous-wave (cw), blue diode laser (411nm, ~ 21 mW at the crystal) to produce pairs of single photons. To conserve momentum, the 822nm, energy-degenerate photon pairs (selected with 10nm filters, FWHM) exited the crystal on opposite sides of a cone with a half-opening angle of $\sim 2.5^\circ$. The Type-I down-conversion cone and complementary spatial modes of the degenerate photon pairs are illustrated in Fig. 7.7.

The next two stages of the apparatus optimised the source to inject the down-conversion output into the photon collection optics. As we moved progressively through the alignment of the experiment, we used three different collection configurations. Initially, to align and optimise the d.c. beams, we used two free-space (FS) detector assemblies³ which consisted of a 35mm focal length, achromatic lens mounted in front of an avalanche photo-diode (APD). Using these, we were able to locate the d.c. beams and optimise the crystal alignment. With the beam positions marked precisely with irises, we could then insert flipper mirrors⁴ and align new beam paths to inject the d.c. photons into fibres, which were then plugged directly into fibre-coupled APDs. Using these flipper mirrors allowed us to switch precisely and relatively easily between the well-characterised free-space bucket detectors and the more sensitive fibre coupling arrangement.

³These “bucket” detectors are designed to collect photons over a large (cross-sectional) area of the incoming beam without performing any significant transverse spatial filtering.

⁴NewFocus 1” Flipper Mount, Model 9891

The fibre couplers⁵ were initially aligned using multimode fibres (MMFs), before switching to single-mode fibres⁶ (SMFs). Although some realignment is always required when switching between the single-mode and multimode fibres, the latter provided another interim stage of alignment which was less sensitive to drift and which could be used for comparison purposes as the experiment progressed.

Because the goal of this experiment was to use single-mode fibres to filter and analyse the two-photon spatial quantum state, there were two overriding practical imperatives when we were optimising the mode matching and collection of the d.c. source: (1) we needed to couple as many photons into the SMFs as possible; and (2) we needed to maximise the occupation of higher-order spatial modes (other than g) [Fig. 7.7(b)]. Consequently, the following considerations are important when performing such an experiment.

- (i) The divergence properties of the d.c. beam should closely match those of a simple Gaussian beam to maximise the power coupled into SMF (for higher-order modes, with the aid of a hologram). Two main factors affect the divergence of the down-conversion output—the divergence of the pump beam (a tightly focussed beam will produce more divergent output beams), and uncertainty in the phase-matching (affected by the thickness of the non-linear crystal and the filtering bandwidth of the d.c. photons). An added complication is that, because d.c. photons are twice the wavelength of the pump photons, a d.c. beam will diverge more quickly than a pump beam with the same spot size.
- (ii) A large section of the down-conversion cone should be imaged [Fig. 7.7(b)] so that the imaged mode is as “non-Gaussian” (i.e. contains as much of the higher-order modes) as possible. To do this, the first mode-collection optics need to be as close to the crystal as possible, while ensuring no overlap between the two output arms. The pump beam must be focussed loosely enough that the arms are resolved by the cone opening angle⁷.
- (iii) At the crystal, the pump power should be concentrated inside the spot imaged by the SMF.
- (iv) The imaged mode should have a beam waist at the SMF (to optimise coupling efficiency).
- (v) The pump beam should have a beam waist at the crystal—this is closest to a plane-wave, so that the mode-matching conditions and cone opening angle will be centred on the expected value.
- (vi) The imaged mode should not be strongly diverging or converging at the hologram (ideally, collimated) so that the mode-conversion behaviour is not affected.

⁵Thorlabs KT110/M free space fibre launcher with C230TM-B coupling lens (4.5mm effective focal length)

⁶Thorlabs P1-4224-FC-2: designed for 820nm with a 780nm cut-off wavelength

⁷A tightly focussed pump beam could produce overlapping down-conversion beams, because their consequent divergence could be greater than the small opening angle produced by the phase-matching conditions.

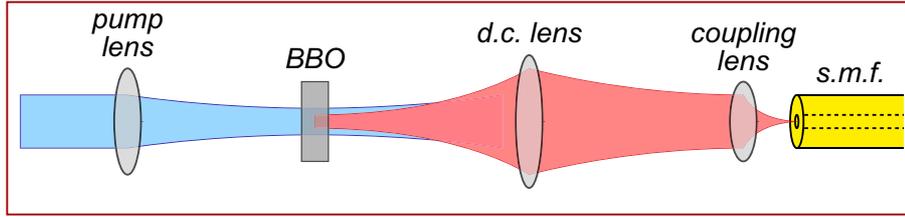


Figure 7.8: The basic conceptual design for the down-conversion mode-matching optics showing the longitudinal mode profiles pump (blue) and d.c. (red) beams. (For illustrative purposes, the beam paths have been flattened into a one-dimensional alignment. In reality, of course, the d.c. beams are not collinear with the pump beam, but exit the crystal along opposite sides of a cone, each making an angle of $\sim 2.5^\circ$ with the pump beam.) The d.c. lens is placed just over one focal length from the crystal to collect a strongly diverging spatial mode and transform it into a slowly converging beam. The beam then propagates through a beam waist and, just after the waist, is injected by a small, short-focal-length coupling lens into a single-mode fibre. A hologram (not shown) is placed at a convenient distance before the second lens as determined by the beam size and the diffraction angle of the hologram. The pump beam is focussed to concentrate the pump power in the area occupied by the imaged d.c. beams. Note, however, that the pump waist size is made larger than the waist of the imaged d.c. mode. This allows for the extra divergence of d.c. beams which results from the phase-matching conditions.

- (vii) The imaged beam should be big enough at the hologram compared to the fringe and pixel spacings, but small enough to fit within one hologram segment (even when the r/l segments are displaced) [Fig. 5.13].
- (viii) The hologram needs to be far enough from the SMF so that the diffracted modes are resolved, but close enough that the first diffraction orders are not separated too much to be accessible to a fibre coupler mounted on a translation stage.

Satisfying these conditions simultaneously involves attempting to achieve often competing outcomes, and finding a working balance between the different mode-matching requirements can be quite tricky in practice, although the calculations themselves (using standard Gaussian beam propagation rules) are quite straightforward. We sought an effective compromise using a compound mode-matching scheme which is illustrated conceptually in Fig. 7.8 and summarised below. Since they are of little scientific interest, I have omitted the details of the calculations that I used to determine the values of the parameters reported below.

We chose to image a mode with a $20\mu\text{m}$ (w_0) beam waist at the crystal, because this had a similar divergence to the mode captured by the free-space lens-detector assemblies⁸. To match this divergence with the d.c. beams, we focussed the pump beam to a beam waist approximately $40\mu\text{m}$ in size, slightly larger than the $20\mu\text{m}$ imaged spot (discussed in more detail below). The main mode-matching lens [*d.c. lens* in Fig. 7.8] was placed a

⁸We roughly measured the size of the d.c. beams using irises which were centred on the d.c. counts, by closing them down to the point where they began to clip the beams. We were able to calculate the approximate divergence using a simple geometric argument involving the size of the beam and the distance of the iris from the crystal.

little more than one focal length away from the crystal (determined by the mode-matching calculations) to produce a slowly converging beam which was steered along a well-defined optical path using a pair of mirrors. For practical reasons, the beam paths in the two arms were different [see Fig. 7.9 for a detailed diagram of the apparatus]. In arm #1 (arm #2), a 300mm (400mm) focal length lens was placed $\sim 403\text{mm}$ ($\sim 553\text{mm}$) from the BBO crystal to produce a beam waist around 1.18m (1.45m) from the lens. The fibre coupler was then placed at the appropriate position (in our case, with the coupling lens $\sim 10\text{cm}$ behind the beam waist) to give a $2\mu\text{m}$ waist size at the front of the fibre—this matches the mode field diameter of the SMF.

Full transverse spatial analysis of phase-matching conditions in down-conversion is extremely complicated [see Ch. 6 and, e.g., Ref. [14]]. As a result, a rigorous calculation of the full mode-matching requirements for the d.c. spatial modes is impractical, and, in general, these calculations will require a certain amount of educated guesswork⁹. Accepting this, we adopted an approach based on rough preliminary measurements of the beam characteristics and justified *a posteriori* by bright fibre-coupled count rates. In our experiment, although we imaged a d.c. beam with a $20\mu\text{m}$ waist at the crystal, we focussed the pump to a $\sim 40\mu\text{m}$ waist—an *ad hoc* correction designed to compensate for the extra divergence of the down-conversion beams which results from uncertainty in the phase-matching conditions [item (i) in the above list].

To complement this approach and compensate for any remaining mode mismatch, we then used the flexibility which is already built into the fibre couplers as a response to the precise requirements of single-mode filtering. The Thorlabs KT110/M fibre launchers allow x – y (i.e. transverse) positioning of the coupling lens relative to the fibre and we mounted the entire assembly on 25mm x – y translation stages. This combination allowed us to optimise both the angular and positional alignment of the incoming d.c. beam relative to the fibre. The z (longitudinal) position of the fibre can then be adjusted and aligned to the focus of the beam emerging from the coupling lens. As with any fibre coupling, the overall process involves some careful coarse-scale alignment to find some initial signal, followed by a slightly tedious optimisation using fine-tuning adjustments¹⁰.

The hologram was mounted on a micrometer-driven, x – y translation stage with x and y tilt freedom, so that the hologram could be positioned and retroreflected with good precision. It could be rotated manually (around the beam propagation axis) to align the diffracted output to the horizontal. The separate x – y translation control of the fibre coupler and the coupling lens allowed us to move from one diffraction order to the next, and also adjust the effective angle that the fibre was “looking” in.

Visible and infrared (IR) alignment lasers were used at various stages throughout the experiment. Visible (red) diode lasers were made to “piggy-back” the d.c. path (as defined by the free-space irises) to assist when aligning the coarse optical elements (mirrors, mode-matching lens, etc.). IR lasers are not much use for this sort of work, but were more useful later in the process, since they were designed to be close to the wavelength of the d.c. photons. Unfortunately, although nominally “820nm”, they were actually

⁹For example, in Ref. [15] the authors make an *ad hoc* correction to their detailed *a priori* calculations of the beam divergence (for a Type-II polarisation-entangled source).

¹⁰Warning: The fine-scale actuators used in the Thorlabs KT110/M fibre couplers are very poor quality. This made the optimisation stage quite unpredictable and frustrating.

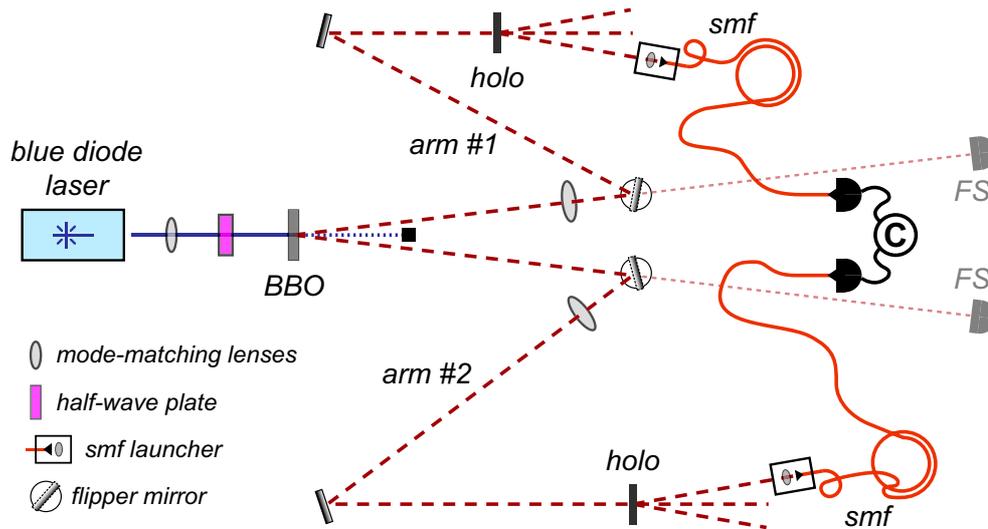


Figure 7.9: A detailed schematic of the experimental apparatus.

sufficiently different from their specifications ($\sim 808\text{nm}$ & 810nm) that they did not pass significantly through the $820\text{nm} \pm 5\text{nm}$ interference filters, which reduced their usefulness. However, they were close enough that they mimicked the propagation behaviour of the d.c. mode. By coupling them backwards through the aligned SMFs, we could directly measure the exact mode imaged by the SMFs at the hologram position by scanning a razor blade across the beam [Fig. 7.10]. In particular, from this we calculated quite precisely the mode spot sizes at the holograms (0.86mm & 0.88mm , resp.), needed for the nondegenerate spatial superposition measurements. These spot sizes are in the desired range for the holograms we used (cf. fringe size $\sim 0.18\text{mm}$; i.e. > 18 fringes across the beam cross-section between $\pm 2w$). Similar scans performed directly with the d.c. beams proved to be far more inaccurate, as they were significantly distorted by diffraction effects¹¹.

The final stage of the experimental apparatus was the counting electronics. The single photons were detected using silicon APDs¹², which produce a single voltage pulse per detection event. Coincidence detections (simultaneously arriving photons) were registered using the TAC/SCA counting module¹³ with a coincidence window of 10ns in these experiments. The single and coincidence signal pulses passed into separate channels of the Quad Counter/Timer¹⁴, which then sent the output to a LabView-controlled computer to be recorded.

Table 7.1 contains a summary of typical count rates at significant stages through the

¹¹A critical factor in the spatially entangled down-conversion source is that the d.c. beams are not just a single lowest-order Gaussian mode. In this step, we needed to measure the size of the beam as imaged by the single-mode fibre filters. Therefore, because the hologram was quite a distance from the fibre ($\sim 35\text{cm}$ (42cm) for arm #1 (#2)), the diffraction effects caused by the edge of the razor blade had a significant, complicated effect on the counts coupled into the SMFs, making the data unusable.

¹²EG&G SPCM-AQR, output: TTL logic pulse (0 to +5V), ~ 50 – 200 dark counts per second, quantum efficiency $\sim 65\%$, actively quenched (giving lower dark times and higher maximum count rates)

¹³Time-to-Amplitude Converter/Single channel Analyser, EG&G Ortec Model 567, SCA output: TTL logic pulse, TAC input: 0 to -5V pulse (N.b. the TAC terminates the signal in a 50Ω path to ground, a matched load for the detector outputs—otherwise signal reflections can give spurious counts.)

¹⁴EG&G Ortec Model 974, inputs: positive logic pulses

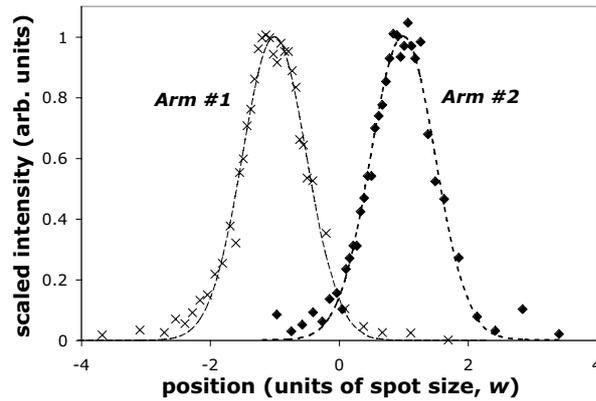


Figure 7.10: Modes imaged by the SMFs in each arm of the down-conversion (displaced for clarity). The spot sizes of the imaged modes at the hologram locations were $\sim 0.86\text{mm}$ and 0.88mm in the two arms, respectively. The data, initially collected by a photodiode, was converted into a linear power density (a form of intensity) with units of $\mu\text{W}/\text{mm}$, which was then fitted by a least-squares numerical optimisation to a Gaussian spatial profile. Because of subtle differences in the interference filters and coupler alignment, the laser power reaching the photodiode was quite different between the two arms (arm #1 max. $\sim 22\mu\text{W}/\text{mm}$; arm #2 max. $\sim 450\mu\text{W}/\text{mm}$). For comparison, I have therefore plotted here scaled intensities (max. 1).

alignment procedure. We obtained the best single arm efficiencies¹⁵ using two SMFs with no filter in one arm, giving $\eta_2 = 0.28$ —this is quite reasonable and is one indication that our mode-matching was quite successful. In fact, efficiencies of up to $\eta_2 \sim 0.35$ were achieved in this configuration in different alignment runs¹⁶ (the data in Table 7.1 was taken from a single alignment to ensure it could be compared meaningfully). When we performed the final measurement tomographies, we had further optimised the mode-matching, with $\sim 100\text{cc}/\text{s}$ coupled into the SMFs from the first diffraction orders of the holograms.

There are several comments relating to the data in Table 7.1 which are worth making at this point. The data in the first two rows shows that, as desired, we captured a large section of the down-conversion cone, since the count rate for the MMF bucket detector ($\sim 16.5\text{kcc}/\text{s}$) was significantly larger than with the free-space detector ($\sim 14\text{kcc}/\text{s}$) relative to the size of the Poissonian noise ($\sim 130\text{cc}/\text{s}$). From the next two rows, we estimated that the SMF coupling efficiency in arm #2 was around 70% ¹⁷ ($= 1100/1500$), which is comparable to the coupling efficiencies that are commonly achievable with a HeNe laser. To arrive at this value, we assume that the FS bucket detector in arm #2 is able to collect almost all of the companions to the photons coupled into the SMF in arm #1.

¹⁵The cumulative single arm efficiencies (i.e. with everything combined) can be calculated from the singles and coincidences. Given a certain photon pair production rate, R , the singles count rates are $\eta_1 R$ and $\eta_2 R$, and the coincidence rate is $\eta_1 \eta_2 R$. Therefore, for example, $\eta_1 = C/S_2$. All efficiencies given here were calculated from the raw data, without corrections for accidental coincidence counts ($A = S_1 S_2 \times \text{coincidence window}$) or background singles counts, which need to be measured experimentally.

¹⁶The main causes of this variation were probably long-term alignment drift and the difficulties associated with the poor quality of the fine-scale actuators in the Thorlabs KT110/M fibre launchers.

¹⁷The arm #1 efficiency should be similar, but we cannot calculate it directly from this data.

detector configuration		singles counts (s^{-1})		coincidence	efficiency
det #1	det #2	det #1	det #2	counts (s^{-1})	η
FS	FS	120k [‡]	150k [‡]	14k	0.12
MMF	FS	270k	150k	16.5k	0.11
SMF	FS	7900	150k	1500	0.19
SMF	SMF	8000	7700	1100	0.14
SMF	SMF (no filter)	8000	420k	2250	0.28
SMF (with holo)	SMF (n.f.)	1700	600k	360	0.21
SMF (w.h.)	SMF (n.f. & w.h.)	1300	130k	77	0.06

Table 7.1: Typical count rates during the alignment process (see text for details). ([‡]We were unable to determine why the FS singles counts were different, but we did confirm that it was not a result of different ambient background light levels.)

hologram segment	hologram position		coupled power	coincidences
	x (mm)	y (mm)	(μ W)	(in 100s)
g	—	—	38 ± 1	8533
l	8.144	10.843	0.76 ± 0.01 (2%)	2438 (29%)
l	8.312	10.989	—	494 (6%)

Table 7.2: Mapping out positions of hologram singularities in arm #1. The first l position of the hologram was found by optimising the extinction (2%) with the IR alignment laser and gave a poor extinction for the d.c. beam (29%). The second l position was found by optimising the extinction directly with the d.c. beam (6%). Despite the large difference in extinction, the discrepancy in the location (only ~ 0.2 mm in size) is less than the precision which could be expected for matching the IR alignment laser to the d.c. beams. This shows the extreme positional sensitivity of the spatial mode filtering technique.

Finally, from the data in the last three rows, we calculated that the hologram diffraction efficiencies were $\sim 16\%$ ($\sim 21\%$) in arm #1 (#2), which compares quite well with the values of $\sim 20\%$ ($\sim 22\%$) measured directly using the 810nm diode laser. Once again, this confirms that we were able to couple the d.c. beams effectively into the SMFs and that the components of the apparatus were behaving as expected.

The final step required before we could perform the spatial mode tomographies was to find the positions of the holographic singularities which defined our measurement positions. It is very important to do this accurately, because as our earlier tests already showed [Ch. 5], the coupling efficiencies are very sensitive to any misalignment.

We first tried to determine the positions of the singularities using the IR diode lasers. Having carefully aligned irises to the d.c. beam in the hologram beam path, we used these to position the IR alignment lasers. We were then able to find the singularities by looking for the maximum extinction points in the power coupled through the SMF onto a photodiode. Unfortunately, however, this method is not nearly accurate enough. The position sensitivity of the spatial filtering technique is such that even a small discrepancy between the diode and d.c. beam paths has a large effect on the tomographic measurement results which manifests itself in the form of very high mixture levels in the reconstructed

state. This is illustrated by the data in Table 7.2. The power measurements show the optimum extinction achieved for the l hologram singularity in arm #1 (2%). Using these same hologram positions to analyse the d.c. beam, and setting the arm #2 hologram to the plain diffraction grating segment (g), the extinction was only 29%. However, minimising the d.c. coincidence counts directly, the extinction could be greatly improved (6%) even with less than 0.2mm displacement in each of the x and y directions.

These results led us to the method we ultimately used to align the hologram singularities. The reason we observed a minimum in the d.c. coincidences when the singularities were aligned to the centre of the imaged mode, is provided by the phase-matching conditions. Just as the total, optical linear momentum is (approximately) conserved in the down-conversion process, so too is the orbital angular momentum (at least within the imaged modes). This is a direct consequence of the higher-order vortex modes being superpositions of plane-wave modes (with conserved linear momentum), and leads to correlations in the angular momentum of the d.c. photons, as was demonstrated in Ref. [8]. For example, since the pump was close to a zeroth-order Gaussian beam¹⁸, when filtering the $|g\rangle$ ($l = 0$) component in one arm, there is no $|r/l\rangle$ ($l = \pm 1$) component in the other arm. We were also able to locate the linear phase singularities in the other hologram segments using this method. In fact, because these are different superpositions of the $|r\rangle$ and $|l\rangle$ states, this is evidence that the spatial profiles of the d.c. photons are not just correlated in the classical sense, but also entangled. After all, if the photons were only correlated, then we would only have observed strong extinctions (i.e. fringes) in one spatial measurement basis—only entangled states can also produce strong extinctions in superposition bases. Typically we observed extinction ratios in the d.c. coincidences of $\sim 2\text{--}10\%$ of the measured gg rate (cf. $\sim 0.5\text{--}3.5\%$ with the bright IR beam). The range of results observed suggests some variation in the quality of the hologram segments (e.g. for the d.c. beam in arm #2, optimal observed extinctions were $5.7 \pm 0.2\%$ and $4.0 \pm 0.2\%$ for the r and l segments respectively, and common sense would suggest that these should be exactly the same).

One final possible way to have found the singularities was to have used the back-coupled IR laser beam, which has the strong advantage that we would have been directly and impartially positioning the hologram relative to the imaged mode (instead of relying on an assumption about the form of the d.c. spatial quantum state which is verified *a posteriori*). However, this was not possible in our measurements, because the back-coupled beam was expanding as it passed through the hologram, and the diffraction orders could not be resolved.

7.4 Results

In principle, the spatial quantum state produced by a down-conversion source will be a superposition of the whole family of Gaussian spatial modes, with the occupation population decreasing for higher orders. In our experiment, it was both impractical and unnecessary to attempt a comprehensive tomographic analysis of the entire Hilbert space, or even

¹⁸Being a diode laser, the pump was not a perfect zero-order Gaussian, but it did not contain any phase singularities. Thus the pump beam as seen by the imaging optics *was* a zeroth-order Gaussian mode.

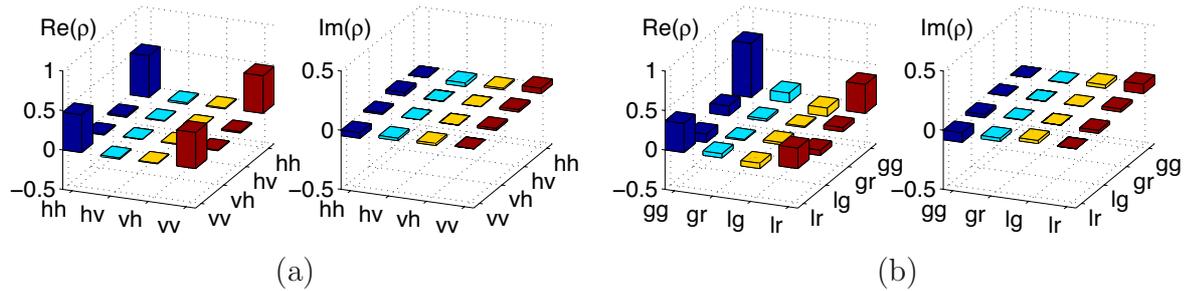


Figure 7.11: Measured, two-photon density matrices for the spatial quantum state within the (a) degenerate and (b) nondegenerate qubit subspaces.

just the significantly populated segment. Instead, we chose to restrict our attention to a characterisation of several different low-order subspaces of the spatial degree of freedom. Primarily, we wished to investigate the effects of degenerate versus nondegenerate information encoding, as well as quantitatively measuring the entanglement properties of the spatial modes in both the two qubit cases and a higher-dimensional, qudit-based subspace [see Sec. 5.3].

Before reporting these results, I wish to comment briefly on some technical issues in regard to our tomographic reconstructions. For the results we published in Ref. [1], we used optimisation code which has been greatly superseded by the technique described in Ch. 3. There were two main differences—we used a least-squares penalty function weighted by fixed (measured) error estimates, which we optimised using a local search routine (MATLAB’s `fminsearch` routine) seeded by an initial guess determined from the linear tomography. I have since reanalysed the data with our latest tomographic techniques which utilise convex optimisation to provide a global minimum. In some cases, this true minimum leads to new results which differ slightly from the published data (although qualitatively the analysis stays much the same) and I will report both for comparison purposes. Also, the use of different weightings in the penalty function can produce somewhat different results [see Sec. 3.8 for details]. Where appropriate, I will report results based on both “fixed weights” (FW) and standard maximum likelihood reconstructions [the latter in square brackets].

The simplest degenerate spatial encoding is the degenerate qubit introduced as an analogy to the Poincaré sphere in Ref. [16]: $\mathbf{0} \equiv h$, $\mathbf{1} \equiv v$, with the other standard basis states d , a , r and l [see Sec. 5.3 for definitions]. According to the ideas developed in Sec. 5.6.5, to circumvent the problems associated with using plane-wave-based spatial filtering, we actually *interpret* our measurements as a tomographic analysis of the qubit subspace defined by $\{|“r”\rangle, |“l”\rangle\}$, related to the states produced by a VM_{+1} plane-wave hologram¹⁹. We used the over-complete, 36-measurement tomographic set, integrating for 100s for each count. Figure 7.11(a) shows the two-photon state of the down-converter in the degenerate qubit basis: the state is highly entangled and very pure. The fidelity with the maximally entangled ϕ^+ Bell state is $F_{\phi^+} = 0.970 \pm 0.001$ [0.968 ± 0.001], and the degree of entanglement can be quantified by both the *tangle*²⁰, $T = 0.899 \pm 0.006$ [0.895 ± 0.005],

¹⁹I use this interpretation throughout the experiments in this thesis, and I will generally omit the quotes, for simplicity.

²⁰A measure of the entanglement required to *create* the state.

and the *negativity*²¹, $N=0.947\pm 0.003$ [0.945 ± 0.003]. The mixture of the measured state is quantified by the *linear entropy*, $S_L=0.063\pm 0.004$ [0.066 ± 0.004]. (In Ref. [1], $F_{\phi^+}=0.97$, $T=0.90$, and $S_L=0.06$.) The errors in these quantities were estimated from a Monte-Carlo simulation of 200 sample states [see Sec. 3.6.2].

Compared with a *gg* measurement of ~ 11000 (in 100s), the normalisations for the $\{h, v\}^{\otimes 2}$, $\{d, a\}^{\otimes 2}$ and $\{r, l\}^{\otimes 2}$ POVM sets were ~ 10100 , 7700 and 11600, respectively, with corresponding visibilities of $\sim 96\%$, 96% and 95%. Since these count rates are all comparable, the plane-wave error modes [see Sec. 5.6.5] did not seem to play a major role in the tomographic results. This conclusion is further supported by the fit quality for this data, $Q=2.7\pm 0.2$ [2.7 ± 0.2], which indicates that although Poissonian fluctuations are, as expected, not the only source of measurement error, nevertheless there is quite a reasonable fit between the data and the reconstructed state²². This suggests that other effects are not too critical. These results seem to agree well with my earlier predictions about tomography with plane-wave holograms [Sec. 5.6.5].

The simplest nondegenerate spatial encoding is the nondegenerate qubit described in Sec. 5.3: $\mathbf{0} \equiv g$ and $\mathbf{1} \equiv r$ or l . We made the superposition measurements by displacing the appropriate hologram segment a distance $w_0/\sqrt{2}$ from the centre of the beam [i.e. unbalanced states; see Sec. 5.6.6]. The measured nondegenerate, two-qubit state [Fig. 7.11(b)] has a lower entanglement ($T=0.605\pm 0.006$ [0.567 ± 0.007], and $N=0.769\pm 0.004$ [0.745 ± 0.005]), reflecting the larger component of g in the down-conversion beam. This state has a high fidelity, $F_\varepsilon=0.938\pm 0.002$ [0.925 ± 0.002], with a non-maximally entangled state of the form $|gg\rangle + \varepsilon|lr\rangle$ for $\varepsilon=0.59$. The nondegenerate state is also somewhat more mixed than the state in the degenerate subspace ($S_L=0.146\pm 0.004$ [0.178 ± 0.005]). (In Ref. [1], $T=0.65$, $S_L=0.11$, and $F_\varepsilon=0.95$ with $\varepsilon=0.60$.)

The normalisations for the $\{g, l\} \otimes \{g, r\}$, $\{+, -\}^{\otimes 2}$ and $\{+i, -i\}^{\otimes 2}$ POVM sets were ~ 16500 , 17000 and 15300, respectively. Once again, these count rates were not too small, suggesting that the plane-wave error modes did not significantly distort the resulting reconstruction. Nevertheless, they clearly have some effect, as indicated by the larger fit quality parameter, $Q=7.7\pm 0.2$ [7.8 ± 0.2], and are, in fact, the main cause of the reduced entanglement and increased mixture. I will return to this issue in Sec. 7.6.

For the purpose of comparison with other experiments in the literature, the results for both degenerate and nondegenerate qubit subspaces indicate that a Bell inequality could be violated [10].

Finally, we measured the nondegenerate qutrit which is a combination of these two qubit spaces [11]: $\mathbf{0} \equiv l$, $\mathbf{1} \equiv g$ and $\mathbf{2} \equiv r$; the resulting measured two-qutrit state is shown in Fig. 7.12(a). This state is somewhat mixed, with linear entropy $S_L=0.199\pm 0.003$ [$S_L=0.344\pm 0.003$], which is not surprising given the long duration of the tomography. Each of the 225 measurements was 100 seconds long, and since we were manually repositioning the holograms using translation stages in between each measurement, there was

²¹A measure of the entanglement that can be *distilled* from the state.

²²Recall from Sec. 3.6.1, that if Poissonian statistics were the only noise source, then we would expect $Q \lesssim 1$. (If $Q=x$, then on average, each measurement result is x standard deviations away from the value predicted from the reconstructed state. The minimum value for the fit quality is $Q=0$ which would indicate a perfect fit between the data and the reconstruction.)

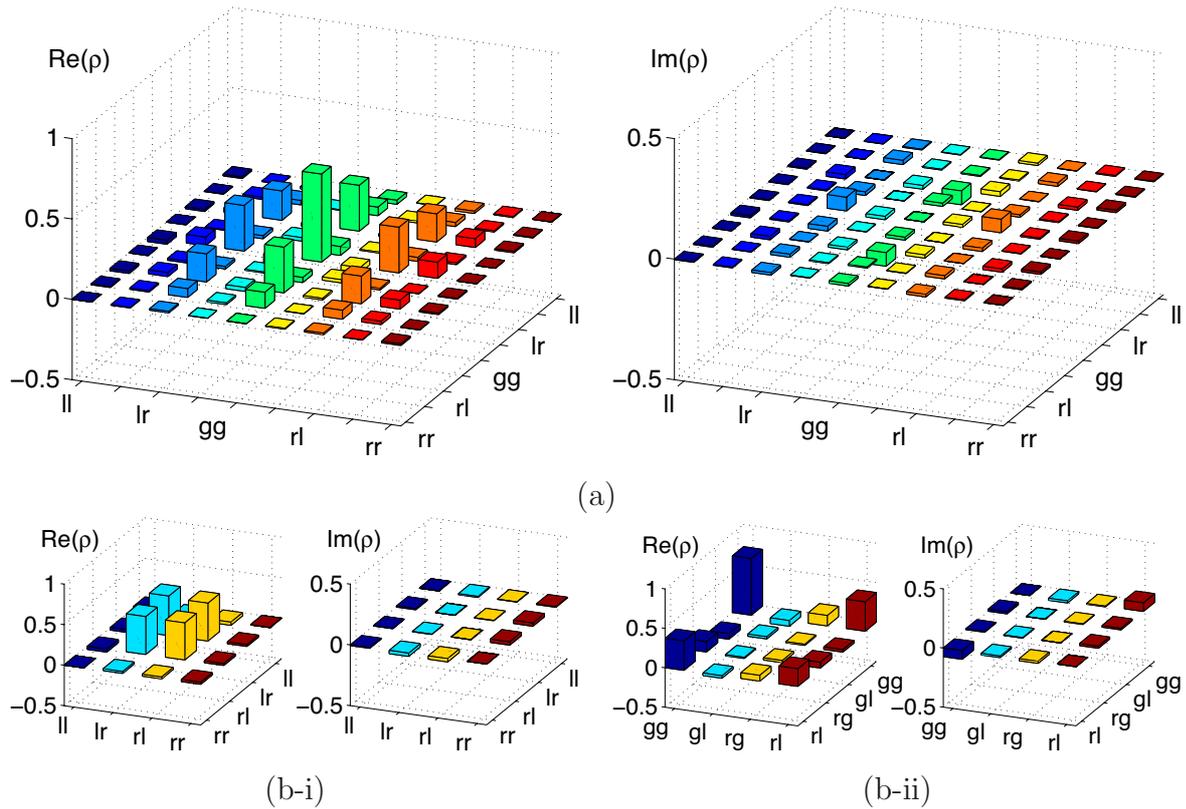


Figure 7.12: (a) Measured, two-photon density matrices for spatial quantum state within the nondegenerate qutrit subspace. (b) Extracted two-qubit density matrices for the (i) $\{l, r\}^{\otimes 2}$ and (ii) $\{g, r\} \otimes \{g, l\}$ subspaces.

actually a bit over 3 mins between consecutive count starting times. As a result, the tomography took around 13 hours from start to finish. There would no doubt have been some element of drift in the alignment of the experiment over that time period, made significant by the high positional sensitivity of the spatial filtering technique. This would produce fluctuations in the data which would appear in the reconstruction as an increase in the mixture and fit quality parameter.

The state is also highly entangled, which can be seen qualitatively by inspection of the form of the density matrix. As expected, it has three (and only three) main diagonal population components (lr , gg and rl), and strong coherences between each of those elements. However, calculating the entanglement between two qutrits when in a mixed state is less straightforward than for two qubits [see Sec. 2.1.6]. Finding a single, universally preferred candidate for this task is still an open problem [17, 18], but there are several meaningful approaches that we can take to quantifying the entanglement of this state. Given the relative populations of the basis states, we expect a non-maximally entangled state of the form, $|lr\rangle + \varepsilon|gg\rangle + |rl\rangle$; for $\varepsilon=1.80e^{0.06i\pi}$, found using numerical optimisation, the fidelity between this “ideal” and the measured nonmaximally entangled states is $F_\varepsilon=0.872\pm 0.002$ [0.800 ± 0.001].

Alternatively, we can investigate the entanglement of the whole system by looking at the entanglement in the different two-qubit subsystems. There are two ways of extracting this information. First, we can take the final two-qutrit state and project onto a given

two-qubit subspace	projected subspaces		reconstructed subspaces	
	tangle, T	negativity, N	tangle, T	negativity, N
$\{l, r\} \otimes \{l, r\}$	0.795 ± 0.008	0.891 ± 0.005	0.895 ± 0.005	0.945 ± 0.003
$\{g, l\} \otimes \{g, r\}$	0.545 ± 0.005	0.731 ± 0.003	0.566 ± 0.006	0.735 ± 0.004
$\{g, r\} \otimes \{g, l\}$	0.601 ± 0.005	0.763 ± 0.003	0.591 ± 0.006	0.763 ± 0.004

Table 7.3: Entanglement for the two-qubit subspaces extracted from the two-qutrit spatial tomography.

two-qubit subspace, e.g.,

$$\varrho_{\text{proj}}^{2 \otimes 2} = (|g\rangle\langle g| + |l\rangle\langle l|) \otimes (|g\rangle\langle g| + |r\rangle\langle r|) \varrho^{3 \otimes 3} (|g\rangle\langle g| + |l\rangle\langle l|) \otimes (|g\rangle\langle g| + |r\rangle\langle r|). \quad (7.1)$$

Figure 7.12(b) shows the density matrices for two example subspaces which display strong entanglement. The second way to examine the two-qubit subspaces is to return to the original measurements and directly reconstruct the extracted two-qubit density matrices based on the appropriate subsets of those measurements. Table 7.3 summarises the entanglement (in the form of tangle and negativity) obtained using both of these methods for each of the entangled, two-qubit subspaces.

More directly, we can calculate the negativity of the two-qutrit state²³, $N_{\text{sc}} = 0.746 \pm 0.002$ [$N_{\text{sc}} = 0.650 \pm 0.002$]. (In Ref. [1], $S_L=0.18$, and $F_\varepsilon=0.88$ with $\varepsilon=1.79e^{-0.07i\pi}$. We also reported an upper bound to the measured *entanglement of formation*²⁴ of 0.74 (scaled) [21]. Rather than repeat this slightly over-simplified calculation of the entanglement of formation here, I have chosen to focus on the negativity as a quantitative measure of two-qutrit entanglement.)

Not surprisingly, the fit quality parameter was higher for the nondegenerate qutrit state, $Q=14.09 \pm 0.08$ [12.60 ± 0.06], than for the nondegenerate qubit. This is a result of both the plane-wave holographic measurements and the extremely long duration of the tomography.

7.5 Quantum bit commitment

One advantage that entangled qutrits offer over qubits is increased security in cryptographic protocols such as quantum bit commitment (BC) and coin flipping which are designed to mediate interactions between different parties which may not trust each other. In such protocols, the two parties wish to communicate in such a way that each party has some protection against dishonest behaviour by the other. For example, in a quantum BC protocol, the first party (Alice) wishes to send a message (a bit or bit string) to the second party (Bob) such that he cannot determine what the message is until Alice chooses to reveal it at some later stage. The complication is that Bob also wishes to ensure that

²³For comparison purposes, I have reported a scaled version of the negativity where $N_{\text{sc}}=1$ for a maximally entangled, two-qutrit state. Normally, $N_{\text{max}}^{3 \otimes 3}=2$.

²⁴The pure-state, qutrit entanglement of formation, EOF, is $-\text{Tr} \{ \rho_A \log_3 \rho_A \}$, where $\rho_A = \text{Tr}_B \{ \rho_{AB} \}$. Our EOF was calculated following the minimisation algorithm given in [19] using the parametrisation for unitaries given in [20]. Only 9×9 unitaries were searched, so the figure quoted is an upper bound. Once again, this is scaled so that $\text{EOF}_{\text{sc}}=1$ for a maximally entangled state. Normally, $\text{EOF}_{\text{max}}^{3 \otimes 3} = \log_2 3$.

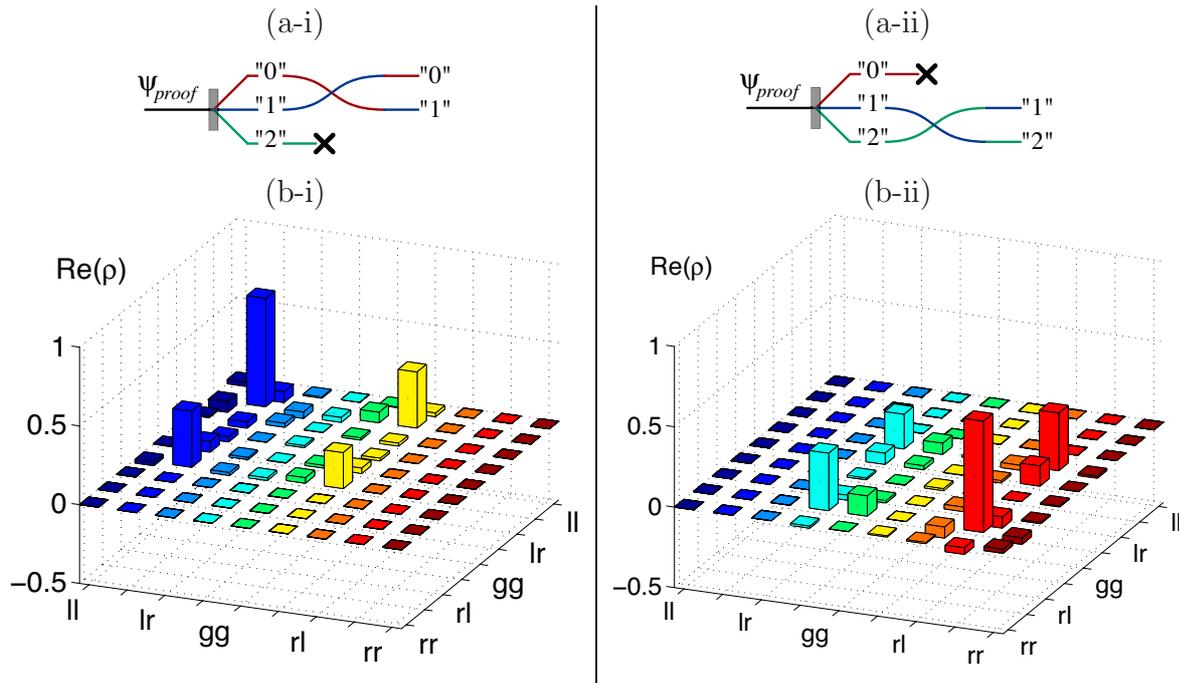


Figure 7.13: Step 1) Logical state preparation: (a) Alice’s state preparation procedure; (b) The simulated logical states prepared by Alice, given our original measured state.

Alice is unable to change her message after she has sent it (the *commitment*). BC protocols are the basis for the most secure known strong quantum coin-flipping protocols [3]. While BC protocols with unconditional security have been proved to be impossible [22, 23], they can be partially secure [3]. Indeed, the best BC protocols known are purification protocols, where Alice supplies the entire quantum system (which consists of two parts) used in the protocol. She sends the *token* subsystem to Bob to commit her bit and the *proof* subsystem later to reveal it. Maximum security in such protocols can be achieved by using two entangled qutrits (or larger) for the token and proof, but not using qubit-based systems [see [3] for details].

To explain how this works, I now outline one procedure that we developed for using our measured qutrit state, which is an entangled state of the form $|0\mathbf{2}\rangle + \varepsilon|1\mathbf{1}\rangle + |\mathbf{20}\rangle$, to implement a potentially optimal, purification BC protocol.

In the first step, depending on her choice of bit, Alice should prepare two qutrits in one of the orthogonal logical states $|0\rangle_L = \sqrt{\lambda}|\mathbf{12}\rangle + e^{i\phi}\sqrt{1-\lambda}|\mathbf{01}\rangle$ or $|1\rangle_L = e^{i\phi}\sqrt{1-\lambda}|\mathbf{21}\rangle + \sqrt{\lambda}|\mathbf{10}\rangle$, where λ is a parameter characterising the security of the protocol. To prepare such states using our system, Alice needs to implement the procedure outlined in Fig. 7.13(a). Starting with our measured state, she must post-select the entangled states that have no photons in one of the basis modes of one subsystem. For example, consider the proof subsystem in arm 1: measuring zero photons in the “2” basis mode yields $|0\mathbf{2}\rangle + \varepsilon|1\mathbf{1}\rangle$, while measuring zero photons in the “0” mode yields $\varepsilon|1\mathbf{1}\rangle + |\mathbf{20}\rangle$. In principle, manipulating the individual modes of the proof subsystem can be accomplished using a holographic interferometer in that arm. Post-selection would then require either perfect detectors or spatial-mode quantum nondemolition (QND) measurements. Here, however, we begin with the measured two-qutrit state and simulate a perfect preparation

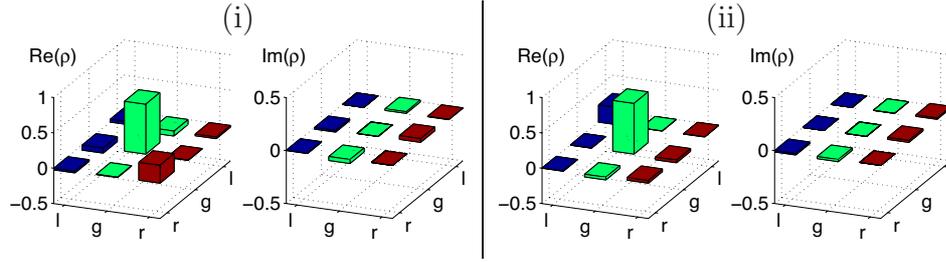


Figure 7.14: Step 2) The commitment: Alice sends one half of her prepared state (the *token*) to Bob. These are the simulated token states sent to Bob, given our original measured state.

process theoretically²⁵. The logical states are then created by swapping the remaining proof subsystem modes. Mathematically, Alice chooses to implement one of the following two operators on the proof subsystem:

$$A_0 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{or} \quad A_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad (7.2)$$

and the output state is therefore $\rho_j^{AB} = A_j \rho^{AB} A_j^\dagger / \text{Tr}\{A_j \rho^{AB} A_j^\dagger\}$. Figure 7.13(b) shows the two-qutrit logical states that result from this simulated state preparation step. In this analysis, since the only imperfections in the protocol arise from the initial state, it gives a bound for the usefulness of our entangled qutrits.

After preparing the appropriate state, Alice then sends the token qutrit to Bob. Because of the entanglement (quantified by λ), the reduced token state possessed by Bob, ρ_j^B , is mixed—this lies at the heart of the security of the purification protocol. The fact that *orthogonal* two-qutrit logical states produce *non-orthogonal* token states provides some security against Bob cheating. His maximum knowledge gain, $K \equiv D(\rho_0^B, \rho_1^B)$, is limited by the distinguishability of these states and quantified by the trace distance between them. However, it is this partial distinguishability which in turn limits Alice’s ability to cheat and change her bit after her commitment. Her maximum control, $C \equiv \sqrt{F(\rho_0^B, \rho_1^B)}$, can be quantified by the square root of the fidelity between the token states. Details can be found in²⁶ Ref. [3]. The reduced density matrices for the token states that Bob would receive if Alice used our measured two-qutrit state are shown in Fig. 7.14.

The protocol is concluded by Alice sending the proof qutrit to Bob, who performs the appropriate orthogonal, two-qutrit projective measurement, and either decodes the bit $\{|0\rangle_L \langle 0|, |1\rangle_L \langle 1|\}$, or catches Alice cheating.

Figure 7.15 shows a plot of C vs K , where the bottom left corner represents unconditional security and the top right corner represents no security. The ideal token states for this scheme give $K = \lambda/2$ and $C = (1 - \lambda)/2$, and varying λ produces the best known

²⁵We used an alternative method to simulate the perfect state preparation in Ref. [1]. We simply set to zero the measurements associated with the zero-photon mode, renormalised the remainder appropriately, and then reconstructed the new states.

²⁶Note that Spekkens and Rudolph use the other commonly used (but less well motivated) definition of fidelity in Ref. [3] [see Sec. 2.1.4 for details]. Therefore, to keep the definitions consistent within the bit commitment theory, I have used a square root in the definition of C here.

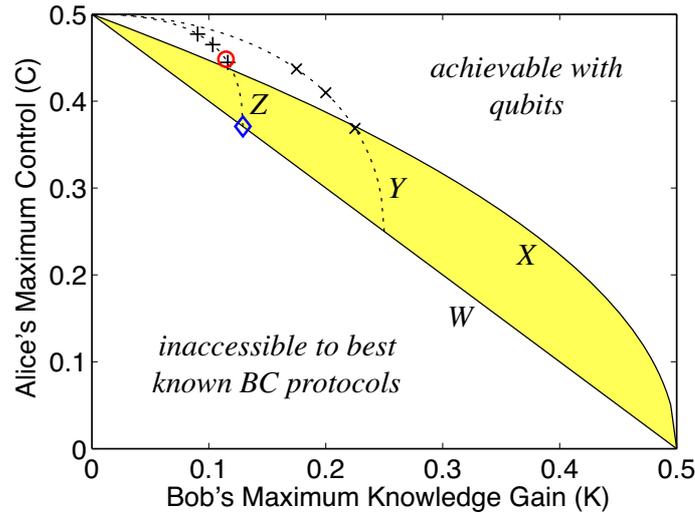


Figure 7.15: The overall security of the purification BC protocol: a plot of Alice's Control vs Bob's Knowledge Gain. \circ (red): the measured protocol; \diamond (blue): the closest ideal protocol. W and X : the best qutrit and qubit protocols known. Y & Z : Imperfect purification protocols with token states of the form, $\rho_{0,1} = p/3 I + (1-p) \rho_{0,1}^{ideal}$, where Y is $\lambda=0.5$ and Z is $\lambda=0.27$. The positions for $p = 0.1, 0.2, 0.3$ are marked with \times (Y) and $+$ (Z).

Alice-supplied²⁷ security curve (W). The shaded region between W and X highlights the area inaccessible to qubit-based, but accessible to qutrit-based BC protocols. The token states resulting from our measured state are closest to ideal states with $\lambda=0.26$ ($F \sim 0.96$). However, in spite of this high fidelity, if we determine C and K directly from the measured token states, the protocol lies just inside the area accessible to qubits ($K=0.1145 \pm 0.0007$, $C=0.4484 \pm 0.0006$): a direct result of the slight (~ 0.04) residual population in the other mode of Bob's token subsystems, originating from the defects of Alice's original state. In other words, a two-qutrit state with residual populations of < 0.01 is required to surpass the qubit boundary (X).

To implement this BC protocol, Alice must be able to perform deterministic post-selection (e.g. using QND measurements). This is hard. Even if she achieves this perfectly, we have shown that the protocol still lies in the qubit-accessible regime. In our simulation, the only differences between our protocol and the ideal resulted from imperfections in the initial state. This result demonstrates that the requirements on the initial two-qutrit entangled state are extremely stringent, and that future theoretical work in this area should consider the critical role of even small amounts of mixture.

As a final note, I would like to make a comment on the theoretical investigation described above. The quantum information community has long known that some measures for evaluating the quality of measured states such as the fidelity can be quite forgiving. The above simulated procedure confirms this by showing that the requirements for actually doing something useful can be much more stringent, suggesting that relying on measures such as the fidelity can be quite misleading. However, I suggest that these results also offer

²⁷This means that Alice supplies all of the original quantum systems, no matter how complicated the subsequent protocol.

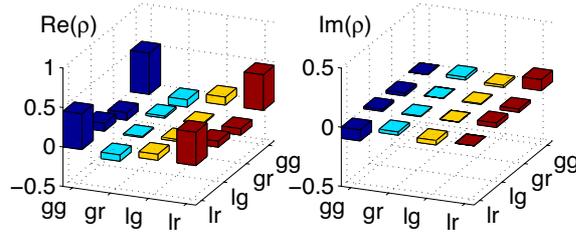


Figure 7.16: Measured, two-photon density matrices within the nondegenerate qubit subspace, reanalysed with compensation for non-optimal measurement settings.

an alternative—that measurements can be more effectively evaluated by benchmarking them against some desired outcome using a simulation such as described above.

7.6 Tomography results with unbalanced measurements

In Section 5.6.6, I investigated the effects of performing nondegenerate spatial tomography with plane-wave holograms. In particular, I showed that the standard technique for measuring nondegenerate superpositions by displacing a vortex-mode hologram analyses unbalanced superpositions of the computational basis states (plus extra error modes). Among other things, this means that “orthogonal” measurements are not actually orthogonal, thus reducing the observed visibilities in superposition bases for entangled states. The results in this experiment seem to agree with the theoretical predictions, with poorer quality reconstructed states for the nondegenerate encodings than for the degenerate qubit encoding.

In the theoretical modelling, I showed that the nondegenerate results could be dramatically improved (to be comparable with the degenerate results) by modifying the displaced-hologram measurements to measure equal superpositions (again plus error modes). Indeed, we are intending to carry out such experiments in the near future to verify this fact. Until then, however, there is an alternative approach which does not require any new measurements, because we compensate for the unbalanced superpositions within the numerical reconstruction, not with the experimental settings. In the theory for linear tomography [see Sec. 3.2], the only requirement on the measurement set is that they are complete [this allows the inversion of \tilde{q} in Eq. (3.13)]. Consequently, it does not matter what measurements are used, as long as they are known, and they form a complete set of observables.

From the modelling in Sec. 5.6.6, I was able to calculate that our nondegenerate spatial superposition measurements actually analysed states of the form:

$$|\psi\rangle \approx (0.7075|g\rangle + 0.5076|r, l\rangle) / \sqrt{0.7582}.$$

Using the corresponding observables, I reanalysed the data²⁸ for the nondegenerate qubit

²⁸Technically, this should probably also be accounted for when calculating the normalisation parameters within the POVM sets, because the observables are no longer orthogonal. However, for the sake of simplicity, I used the same data normalisations, thus introducing a systematic (but small) mis-estimate

spatial encoding and reconstructed the state in Fig. 7.16. This state is significantly more entangled

$$\begin{aligned} (T=0.901\pm 0.007 [0.854\pm 0.007]; \\ N=0.944\pm 0.004 [0.910\pm 0.004]; \\ F_{\text{opt}}=0.972\pm 0.002 [0.955\pm 0.002]) \end{aligned}$$

and less mixed ($S_L=0.047\pm 0.005 [0.0875\pm 0.005]$) than the state in Fig. 7.11(b). In fact, its quality is comparable with the measured *degenerate* two-qubit state, which further supports the theoretical model. The reanalysed nondegenerate state differs mainly because, as expected, it is rotated slightly away from the standard ϕ^+ Bell state (e.g. the main imaginary coherence is larger). A quantitative indication of this is that the optimal fidelity with a maximally entangled state ($\sim 0.97 [0.95]$) is significantly greater than the fidelity with ϕ^+ ($F_{\phi^+}\sim 0.93 [0.91]$). Finally, the fit quality parameter for the reanalysed state, $Q=8.9\pm 0.1 [8.9\pm 0.1]$, is slightly higher than for the original state ($Q\sim 7.8$). This is perhaps somewhat surprising, since one might have expected the fit to improve with more accurate observables being used in the reconstruction.

I also reanalysed the tomography for the nondegenerate two-qutrit subspace [Fig. 7.17(a)], and once again the state is significantly more entangled ($N=0.865\pm 0.002 [0.740\pm 0.002]$) and slightly less mixed ($S_L=0.157\pm 0.003 [0.316\pm 0.002]$) than with the standard observable set, and the fit quality parameter ($Q=14.20\pm 0.08 [12.88\pm 0.06]$) is roughly the same. I then used the reanalysed state in the simulated purification BC protocol described above [Fig. 7.17(b)]. The calculated token states are closest to ideal states with $\lambda=0.40$ ($F>0.95$). However, despite the increased entanglement (reflected in the larger value for λ), the token states still possess slight residual populations in the “empty” mode (~ 0.045), and consequently, the protocol still lies just inside the region accessible to qubits ($K=0.1812\pm 0.0008$, $C=0.4056\pm 0.0009$). Comparing the two reconstructed two-qutrit states directly [Fig. 7.12(a) and Fig. 7.17(a)], the non-optimal measurement tomography predicts a state with populations more evenly balanced between the three main components, and this is the main reason for the substantial increase in entanglement; the purity improves only slightly. Once again, this highlights the stringent requirements on states to perform interesting quantum tasks.

7.7 Conclusions

In this chapter, I have investigated techniques for encoding, measuring and manipulating information in the transverse spatial modes of optical quantum systems. The spatial degree of freedom is one of only three degrees of freedom of a single photon, and allows this important quantum system access to the higher-dimensional encoding of qudits. With larger alphabets and larger Hilbert spaces, qudits can provide significant improvements over qubits in many quantum communication protocols.

In these experiments, we performed the first full characterisation of entangled, spatially encoded quantum states, and achieved the first complete measurement of an entangled, two-qutrit state in *any* encoding. We used a novel quantum tomography technique which only requires two-state superpositions [described in detail in Sec. 5.6], and explored degen-

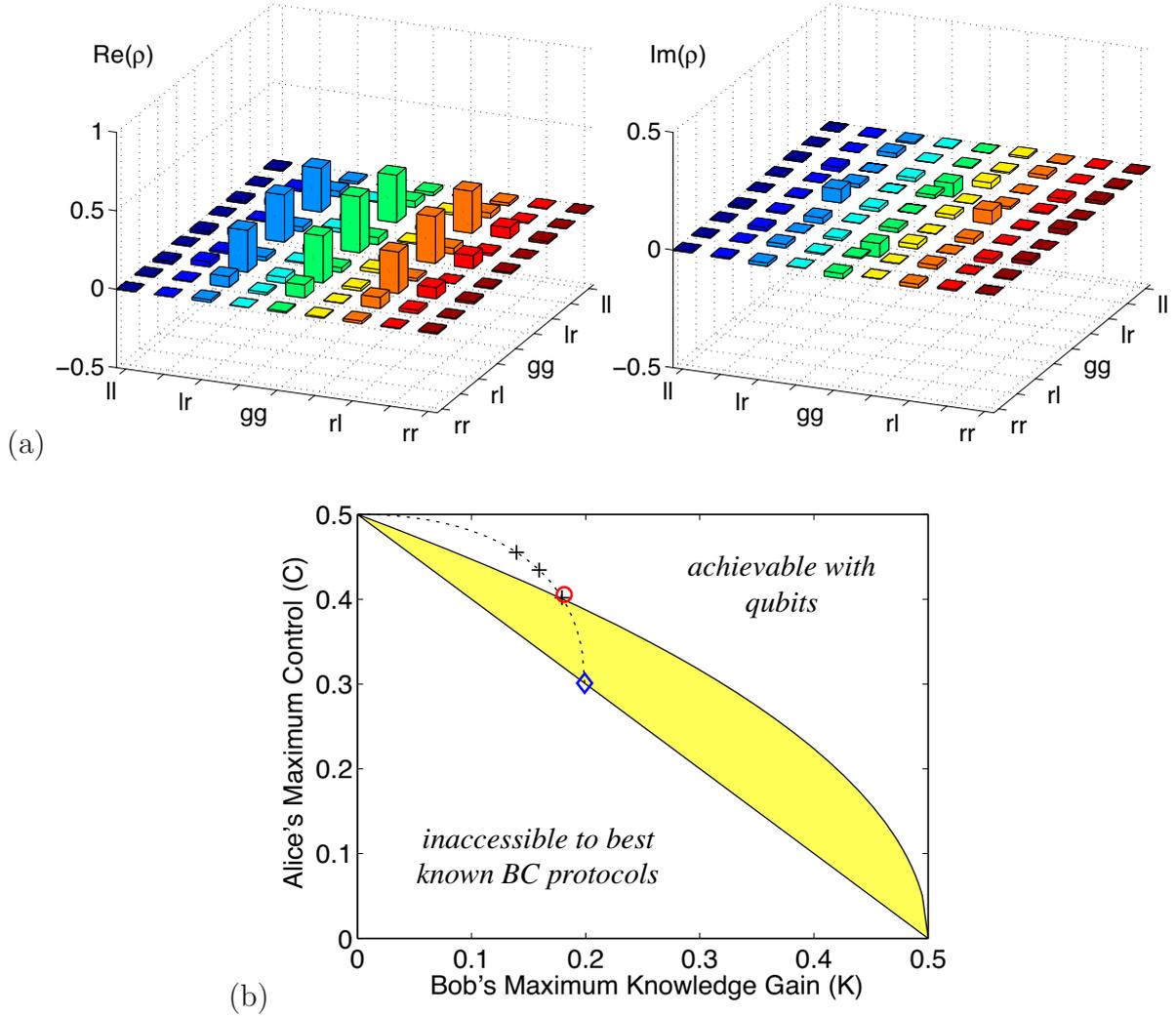


Figure 7.17: (a) Measured, two-photon density matrices within the nondegenerate qutrit subspace, reanalysed with compensation for non-optimal measurement settings. (b) Security of the purification BC protocol which uses the reanalysed state from (a) as the entanglement source: \circ (red): the measured protocol; \diamond (blue): the closest ideal protocol; dashed curve: imperfect purification protocol with token states of the form, $\rho_{0,1} = p/3 I + (1-p) \rho_{0,1}^{ideal}$, with $\lambda=0.40$. The positions for $p = 0.1, 0.2, 0.3$ are marked with $+$.

erate and nondegenerate spatial encodings, demonstrating strong entanglement in both situations. Following on from this and more recent theoretical results, I have also shown how to compensate for tomographic data taken with a non-optimal set of measurements. This work is the most comprehensive exploration of spatially encoded qubits and qutrits to date.

Finally, we also outlined a scheme for using this system to implement a purification bit commitment protocol which has the potential to achieve the best known levels of security. We have shown that the protocol would not reach maximal security using our measured state, but the results indicate what improvements are required. This analysis would have been impossible without access to the complete two-qutrit state, obtained using quantum tomography.

7.8 Chapter 7 References

- [1] N. K. Langford, R. B. Dalton, M. D. Harvey, J. L. O'Brien, G. J. Pryde, A. Gilchrist, S. D. Bartlett, and A. G. White. Measuring entangled qutrits and their use for quantum bit commitment. *Physical Review Letters*, 93(5):053601, 2004.
- [2] M. Fujiwara, M. Takeoka, J. Mizuno, and M. Sasaki. Exceeding classical capacity limit in quantum optical channel. *Physical Review Letters*, 90:167906, 2003.
- [3] R. W. Spekkens and T. Rudolph. Degrees of concealment and bindingness in quantum bit commitment protocols. *Physical Review A*, 65:012310, 2001.
- [4] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White. Measurement of qubits. *Physical Review A*, 64:052312, 2001.
- [5] R. T. Thew, K. Nemoto, A. G. White, and W. J. Munro. Qudit quantum-state tomography. *Physical Review A*, 66:012303, 2002.
- [6] R. T. Thew, A. Acin, H. Zbinden, and N. Gisin. Bell-type test of energy-time entangled qutrits. *Physical Review Letters*, 93:010503, 2004.
- [7] G. Molina-Terriza, J. P. Torres, and L. Torner. Orbital angular momentum of photons in noncollinear parametric downconversion. *Optics Communications*, 228:155, 2003.
- [8] A. Mair, A. Vaziri, G. Weihs, and A. Zeilinger. Entanglement of the orbital angular momentum states of photons. *Nature*, 412:313, 2001.
- [9] A. Vaziri, J.-W. Pan, T. Jennewein, G. Weihs, and A. Zeilinger. Concentration of higher dimensional entanglement: Qutrits of photon orbital angular momentum. *Physical Review Letters*, 91:227902, 2003.
- [10] W. J. Munro, K. Nemoto, and A. G. White. The Bell inequality: a measure of entanglement? *Journal of Modern Optics*, 48(7):1239, 2001.
- [11] A. Vaziri, G. Weihs, and A. Zeilinger. Experimental two-photon, three-dimensional entanglement for quantum communication. *Physical Review Letters*, 89(24):240401, 2002.
- [12] H. Kogelnik. Hologram efficiency and response. *Microwaves*, 6:68, November 1967.
- [13] M. D. Harvey. Production of computer generated diffractive optical elements by controlled laser damage. Honours thesis, University of Queensland, Brisbane, Australia, 1998.
- [14] M. H. Rubin. Transverse correlation in optical spontaneous parametric down-conversion. *Physical Review A*, 54:5349, 1996.
- [15] C. Kurtsiefer, M. Oberparleiter, and H. Weinfurter. High-efficiency entangled photon pair collection in type-II parametric fluorescence. *Physical Review A*, 64:023802, 2001.
- [16] M. J. Padgett and J. Courtial. Poincaré-sphere equivalent for light beams containing orbital angular momentum. *Optics Letters*, 24(7):430, 1999.

-
- [17] C. M. Caves and G. J. Milburn. Qutrit entanglement. *Optics Communications*, 179:439, 2000.
- [18] P. Rungta, V. Bužek, C. M. Caves, M. Hillery, and G. J. Milburn. Universal state inversion and concurrence in arbitrary dimensions. *Physical Review A*, 64:042315, 2001.
- [19] K. Audenaert, F. Verstraete, and B. De Moor. Variational characterizations of separability and entanglement of formation. *Physical Review A*, 64:052304, 2001.
- [20] P. Dita. Factorization of unitary matrices. *Journal of Physics A: Mathematical and General*, 36:2781, 2003.
- [21] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters. Mixed-state entanglement and quantum error correction. *Physical Review A*, 54:3824, 1996.
- [22] H.-K. Lo and H. F. Chau. Is quantum bit commitment really possible? *Physical Review Letters*, 78(17):3410, 1997.
- [23] D. Mayers. Unconditionally secure quantum bit commitment is impossible. *Physical Review Letters*, 78(17):3414, 1997.

*“Will you walk a little faster?” said a whiting to a snail.
“There’s a porpoise close behind us, and he’s treading on
my tail.”*

Alice’s Adventures in Wonderland, Lewis Carroll

Chapter 8

Hyperentangled photons

Acknowledgement

Most of the work in this chapter was first published in Ref. [1], and where appropriate, I have incorporated some of the text of that paper. The experiments were carried out in Paul Kwiat’s quantum information group at the University of Illinois at Urbana-Champaign in close collaboration with Julio Barreiro as part of his PhD research project. The results will also form a major part of his PhD thesis.

The work in Sec. 8.3.1 is a result of my own theoretical investigations, but I would like to acknowledge useful discussions with Julio Barreiro, Joe Altepeter and Paul Kwiat which helped to clarify some of my early ideas. With Julio Barreiro, I carried out preliminary experiments to support this theory [see Sec. 8.5] also while I was at Paul Kwiat’s group.

8.1 Multi-degree entanglement

The smallest quantum system that can exhibit entanglement is made up of two qubits. Its states lie in a two-particle, four-dimensional Hilbert space that includes the (maximally entangled) Bell states, which are the simplest and most commonly used entangled states. However, as demonstrated in the example of quantum bit commitment in the previous chapter, more complicated forms of entanglement can provide real advantages over the standard Bell-type entanglement which is available to two-qubit systems. Higher-order entanglement can be realised in two ways—between multiple particles (e.g. three or more qubits), and between two multi-dimensional particles (i.e. qudits), both of which have been demonstrated experimentally [2–7].

Many systems have more than one accessible degree of freedom (DOF). For example, a single photon can simultaneously store information in its polarisation [Ch. 4], its spatial-momentum profile [Ch. 5], and its time-frequency profile [see Sec. 8.3 below]. This produces a special type of qudit—a single particle with multiple (three for single photons),

independently accessible labels. In other words, although there is only one physical system, the different degrees of freedom can be manipulated and measured independently, identifying some underlying structure in a larger Hilbert space. Note, however, that for single photons, the most common forms of detection destroy the photon. In this case, the projection part of the measurement can be performed independently for the different degrees of freedom before the detection event.

For a system of two such *multi-degree* qudits, there are two quite different forms of entanglement available. For example, consider two single photons with information encoded in the polarisation ($\mathbf{0} \equiv H$ and $\mathbf{1} \equiv V$) and transverse spatial ($\mathbf{0} \equiv h$ and $\mathbf{1} \equiv v$) degrees of freedom. When the photons exhibit entanglement which is *simultaneously and independently* measurable in each degree of freedom, then they are *hyperentangled* [8]. The strongest form of such a state is one which consists of a Bell-type entangled state in each DOF combined in a tensor product, e.g.,

$$|\psi_{\text{hyper}}\rangle \sim (|HH\rangle + |VV\rangle) \otimes (|hh\rangle + |vv\rangle). \quad (8.1)$$

In an experiment, ignoring either DOF (i.e. tracing over it in mathematical terms) leaves the other in a maximally entangled state—the entanglement in the different DOFs is independently measurable. In the alternative picture described above, the polarisation and spatial labels are actually structure in a multi-dimensional qudit Hilbert space. The hyperentangled state is in fact also a maximally entangled state in the larger qudit basis ($\{\mathbf{0} \equiv Hh, \mathbf{1} \equiv Hv, \mathbf{2} \equiv Vh, \mathbf{3} \equiv Vv\}$):

$$|\psi_{\text{hyper}}\rangle \sim |Hh, Hh\rangle + |Hv, Hv\rangle + |Vh, Vh\rangle + |Vv, Vv\rangle = |\mathbf{00}\rangle + |\mathbf{11}\rangle + |\mathbf{22}\rangle + |\mathbf{33}\rangle. \quad (8.2)$$

The entropy of entanglement¹ for this maximally hyperentangled state is $E(\psi_{\text{hyper}}) = 2$. Note this is the maximum entanglement possible for a four-dimensional subsystem, and exactly *twice* the entanglement of a single Bell state. Hyperentangled systems enable the implementation of 100%-efficient complete Bell-state analysis with only linear elements [9] and techniques for state purification [10, 11].

The second form of multi-degree entanglement (also discussed in Ref. [12]), which I will call *hypoentanglement*, looks superficially like the GHZ entangled states for multi-particle systems, i.e.,

$$|\psi_{\text{hypo}}\rangle \sim |HHhh\rangle + |VVvv\rangle = |Hh, Hh\rangle + |Vv, Vv\rangle = |\mathbf{00}\rangle + |\mathbf{33}\rangle. \quad (8.3)$$

Unlike with hyperentanglement, ignoring either DOF of a *hypoentangled* state in an experiment will leave the other in a maximally *mixed* state with no measurable entanglement—the entanglement is only apparent if the entire qudit space is considered. In other words, hypoentangled photons exhibit simultaneous *but not independent* entanglement, displaying quite different characteristics from their hyperentangled cousins (this serves as a useful definition of hypoentanglement). Moreover, from the qudit form of this state, it is easy to see that it is not maximally entangled ($E(\psi_{\text{hypo}}) = 1$). Interestingly, this lower entanglement does not result from mixture or uneven weighting in the superposition in the computational basis, as is most common with non-maximally entangled states. It arises because the state does not “fill” the qudit basis for each photon. In other words, some

¹The entropy of entanglement is the appropriate measure of entanglement for bipartite pure states [Sec. 2.1.6].

information can be obtained about the state by measuring just one photon, because all results are not equally likely. Mathematically, the number of elements in the Schmidt decomposition ([13], p. 109), or Schmidt number, is fewer than the maximum possible value, though the elements are still balanced in terms of probability. The non-maximal qudit entanglement and the fact that the different DOFs exhibit no entanglement when viewed independently provide the motivations for the term, hypoentanglement², which I have proposed here. As an aside, note that with a single half-wave plate operation, the form of the hypoentangled state can be converted into

$$|\psi_{\text{hypo}}\rangle \sim |Hh, Vh\rangle + |Vv, Hv\rangle = |\mathbf{02}\rangle + |\mathbf{31}\rangle, \quad (8.4)$$

which, superficially, looks quite different in the qudit form, although the two states are clearly equivalent.

When the term, hyperentanglement, was originally introduced in Ref. [8], it was used to describe photons that were simultaneously entangled in *every* degree of freedom (i.e. polarisation, spatial-momentum *and* time-frequency), explicitly referring to states in a tensor-product form, as in Eq. (8.1). We also followed this convention in Ref. [1]. However, as discussed above, this is not the only way for two photons to be entangled in every degree of freedom. Therefore, in this thesis, I will distinguish between *hyper*entanglement and *hypo*entanglement as forms of entanglement across multiple degrees of freedom (i.e. more than one). For the states referred to in Ref. [8] which exhibit entanglement in every DOF, I will describe them as completely hyperentangled.

8.2 Hyperentanglement in down-conversion

The pairs of single photons produced by spontaneous parametric down-conversion can exhibit many different forms of entanglement. In general, the conditions for entanglement are produced by the conservation of energy and momentum which results from the phase-matching conditions [see Ch. 6]. In a *single* nonlinear crystal, entanglement in the spatial-momentum DOF results from conservation of momentum. This has been demonstrated using longitudinal, or “which path?” encoding [14], and transverse spatial modes [4, 15] [see also Ch. 7]. Similarly, with a continuous-wave (cw) pump laser, conservation of energy produces entanglement in the time-frequency DOF [see Sec. 8.3.1 below], and this has also been verified experimentally using time-bin encoding (see, e.g., Refs [16] and [17]). In contrast, down-conversion does not produce photons that are intrinsically entangled in polarisation. In our experiments, we engineered the polarisation entanglement using the Type-I crystal sandwich source introduced in Ref. [18] [see Sec. 8.4 for details].

Previous experiments have demonstrated two-degree hyperentanglement using polarisation and time-bin encoding [19], and polarisation and longitudinal spatial encoding [20, 21]. In this work, however, we produced *completely* hyperentangled pairs of single photons, i.e. they exhibited simultaneously and independently measurable entanglement in *every* degree of freedom—polarisation, spatial-momentum, and time-frequency. This was the first demonstration of complete hyperentanglement in any quantum system. We verified the entanglement by measuring Bell-inequality violations in each DOF. We also pro-

²I also acknowledge Andrew White for his part in coining this term.

duced maximally hyperentangled states and novel states simultaneously exhibiting both quantum and classical correlations. Finally, our full tomographic analysis of a $2 \times 2 \times 3 \times 3$ -dimensional system (i.e. $\dim=36$) is the largest entangled system of this size to be so characterised (making no prior assumptions about the form of the quantum state).

Using the two-crystal source described above, we generated a $2 \times 2 \times 3 \times 3 \times 2 \times 2$ -dimensional, completely hyperentangled state of the form:

$$|\psi_{\text{dc}}\rangle \sim \underbrace{(|HH\rangle + |VV\rangle)}_{\text{polarisation}} \otimes \underbrace{(|rl\rangle + \alpha|gg\rangle + |lr\rangle)}_{\text{spatial modes}} \otimes \underbrace{(|t', t'\rangle + |t, t\rangle)}_{\text{time-frequency}}. \quad (8.5)$$

Here the states are labelled as described in Chs 4 and 5, and Sec. 8.3.1: H (V) represents horizontal (vertical) polarisation; l , g and r , represent the Laguerre-Gauss vortex modes carrying $-\hbar$, 0 , and $+\hbar$ units of orbital angular momentum, respectively [22] (α describes the spatial mode balance prescribed by the source and selected via the mode-matching conditions [see Ch. 7]); and t' and t represent the relative early and late emission times, respectively, of the down-conversion photon pairs [23]. Note that by selecting only the degenerate, first-order spatial subspace (i.e. r and l), the down-conversion state in Eq. (8.5) becomes a tensor product of three Bell states $\Phi_{\text{poln}}^+ \otimes \Phi_{\text{spa}}^+ \otimes \Phi_{\text{t-f}}^+$.

8.3 Encoding in the time-frequency domain

In Chapters 4 and 5, I discussed techniques for storing information in, respectively, the polarisation and spatial mode of a single photon. The third and final degree of freedom of a single photon is the time-frequency domain. The frequency modes of a single photon are simply the eigenstates for the Hamiltonian of the electromagnetic field (i.e. the energy eigenstates). As with the spatial degree of freedom, no photon source is truly monochromatic (cf. plane waves), and its frequency distribution (its spectrum) is related to its temporal wave packet via a simple Fourier transform (cf. its spatial and momentum distributions). Consequently, information can be encoded in both the time domain (also “arrival time” and “time-bin” encoding) and the frequency domain (also “side-band” encoding). Although these methods are related and in some sense equivalent, the experimental details in the two situations are generally quite different.

The time-frequency domain has many similarities to the spatial-momentum domain. For example, like its spatial analogue, the Hilbert space for the time-frequency domain is infinite and continuous, and thus also offers the potential for encoding qudits. In fact, it would be an interesting and potentially fruitful research project to see how far the analogy would go—e.g. perhaps specially-engineered frequency filters could be used as “temporal holograms”. This however extends well beyond the scope of this thesis, so I merely mention it in passing.

In this section, I will focus on the “time-bin” method of encoding, where the information is stored in the arrival time of the single-photon at some optical element (e.g. a wave plate or a detector), which is grouped into discrete bins. This method has been widely used in recent times, and in particular, it underpins many of the efforts into practical realisations of quantum key distribution (see, e.g., [24, 25]). Obviously, these experiments will normally

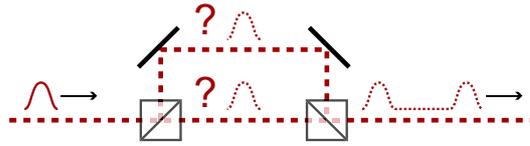


Figure 8.1: The Franson interferometer—a highly unbalanced Mach-Zehnder interferometer. This interferometer can be used both to create and measure photons in a superposition of arrival times (time-bin encoding).

require precision timing electronics and reference clocks to be able to distinguish between the different modes, which can be extremely difficult to control. Conveniently, however, in down-conversion experiments this precision is normally supplied by the simpler counting electronics required to perform coincidence detection.

A highly unbalanced interferometer³ (i.e. path length difference, $\Delta L \gg l_c$, the coherence length of the light) is perhaps the simplest piece of optical equipment that can manipulate the temporal information stored in a photon [see Fig. 8.1]. When a temporally localised pulse passes through this interferometer, it emerges as a pulse in a superposition of two different time bins. Of course, this idea can also be reversed. By placing the interferometer before a detector which can discriminate between pulses arriving at different times, it can be used to detect photons which *were initially* in a superposition of two different time bins.

8.3.1 Entanglement in the time-frequency domain

The EPR paradox as originally proposed by Einstein, Podolsky, and Rosen considered the correlations between two continuous, conjugate variables (e.g. position and momentum), which were predicted by quantum mechanics, and which they felt were inconsistent with reasonable notions of physical reality (including an implicit assumption of locality) [26]. Much later, when Bell showed quantitatively that the predictions of local realism could not match those of quantum mechanics [27], he considered an alternative version of the EPR paradox based on spin or polarisation variables as originally proposed by Bohm (p. 614, [28]), and this was the form used in the CHSH experiment to provide a realisable test of these limits [29]. Not surprisingly, all of the early experiments demonstrating a violation of Bell’s inequality were based on spin correlations.

In Ref. [23], however, Franson proposed a Bell violation experiment based on the continuous, conjugate time and frequency variables which was much closer to the original EPR thought experiment. The original Franson proposal, which involved photon pairs emitted from a single atom, can also be applied to the photon pairs emitted via spontaneous down-conversion, where the narrow bandwidth laser pump gives rise to comparatively broadband d.c. photon pairs. Because the pump has a long coherence time, the time-correlated photon pairs (they are emitted essentially simultaneously) are labelled by their emission time, t , and since the down-conversion process is coherent, so are the correlations—i.e. the photons are entangled. To observe these quantum correlations, the d.c. photons emit-

³In quantum information, this is sometimes referred to as a *Franson interferometer* [23].

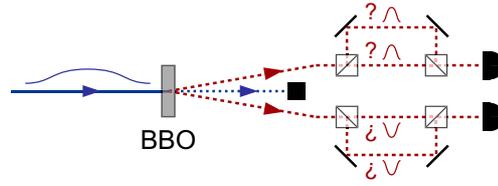


Figure 8.2: The Franson time-frequency Bell violation experiment in parametric down-conversion. Photons can either travel the fast (f) or slow (s) paths of the interferometer. Fast coincidence counting electronics can distinguish the f - s and s - f combinations, but cannot distinguish between photons emitted early and taking the s - s path and those emitted later taking the f - f path. If the time delay between the paths is longer than the d.c. photon coherence, but shorter than the pump coherence, then the s - s and f - f will interfere and the photons will be entangled in the time-frequency DOF.

ted from the nonlinear crystal are then collected by a detector assembly which consists of an unbalanced interferometer (normal beam splitters) and a single-photon counter. The path-length differences between the two arms of each interferometer (ΔL_j) are made greater than the short coherence length of the d.c. photons (L_{dc}), so that no classical interference effects are observed in the single-photon count rates.

For any given photon pair exiting the crystal, there are in fact four possible combined paths through the interferometers—the fast (f - f) and slow (s - s) paths, where ideally the photons arrive simultaneously at the detectors, and the s - f and f - s paths, which are distinguishable both from each other and the first two paths, because the photons arrive at significantly different times ($\Delta L/c$ apart). Therefore, provided the unbalanced interferometers are matched to within the d.c. coherence length ($|\Delta L_1 - \Delta L_2| \ll L_{dc}$) and the path-length differences are less than the pump coherence length ($\Delta L \ll L_p$), the detectors cannot distinguish between a photon pair (at t) taking the f - f path, and an earlier pair (at $t' = t - \Delta L/c$) taking the s - s path. Thus, the indistinguishable f - f and s - s paths interfere, and varying one of the ΔL_j will produce sinusoidal oscillations in the coincidence count rate, in principle with 100% fringe visibility. Coincidence detection is required to postselect the interfering paths and reject all non-simultaneous pair detection events. In early experiments (e.g. Ref. [30]), the counting electronics were not fast enough to achieve this, and the coincidence fringes (now including the non-interfering s - f -type terms) were limited to at most 50% visibility, but this was resolved by combining large path-length differences ($\Delta L \gtrsim 50\text{cm}$) with fast electronics (see, e.g., Refs [16] and [31]). In all of these papers, the experimenters measure the fringe visibility, which allows them to infer whether a Bell inequality can be violated.

It is important to note here that using this postselection raises questions about the conclusiveness of these experiments with regard to possible claims about local realism [32, 33]. Specifically, an extra assumption about the measured subensemble is required which is a particularly strong form of the *fair sampling* assumption that is invoked to account for detectors that are not perfectly efficient (i.e. all of them). This is sometimes referred to as the *postselection loophole*.

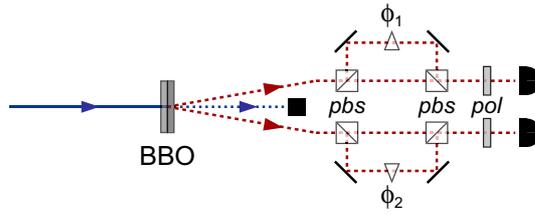


Figure 8.3: The modified Franson experiment uses a polarisation-entangled pair of photons (Φ^+ , say) and replaces the beam splitters in the Franson interferometers with polarising beam splitters (PBSs). The HH and VV components always travel the f - f and s - s paths, respectively, and time-based superpositions can be created without any non-interfering paths. The time-based interference is mapped onto the polarisation DOF by projecting into an unbiased measurement basis such as DD .

8.3.2 Polarisation-entanglement-assisted Bell violations

In Ref. [19], the authors demonstrate a postselection-free version of the Franson experiment which works if the photons are *also* entangled in the polarisation degree of freedom. I will present here a modified description of this experiment which simplifies its interpretation and leads naturally to a more general understanding of the power and flexibility of this technique, including a notion of state tomography in the time-frequency domain.

The standard Franson experiment is modified by replacing the normal beam splitters in the unbalanced interferometers with polarising beam splitters, and a polarisation analyser in front of each detector [Fig. 8.3]. The phase shifts are controlled by tuning the path-length differences in the unbalanced interferometers. If the two-photon state is entangled in polarisation (Φ^+ , say), then the HH and VV components always travel the f - f and s - s paths, respectively, and time-based superpositions can be created without any non-interfering paths that need to be postselected out.

In more detail, as the initial polarisation-entangled state traverses the unbalanced interferometers, it evolves as follows (ignoring normalisations):

$$|HH\rangle + |VV\rangle \longrightarrow |Hf, Hf\rangle + e^{i(\phi_1+\phi_2)}|Vs, Vs\rangle, \quad (8.6)$$

$$= (|Df\rangle + |Af\rangle) \otimes (|Df\rangle + |Af\rangle) + e^{i(\phi_1+\phi_2)} (|Ds\rangle - |As\rangle) \otimes (|Ds\rangle - |As\rangle) \quad (8.7)$$

$$= |DD\rangle [|f, f\rangle + e^{i(\phi_1+\phi_2)} |s, s\rangle] + |DA\rangle [|f, f\rangle - e^{i(\phi_1+\phi_2)} |s, s\rangle] + \dots \quad (8.8)$$

Thus, by setting the polarisers to analyse DD , the detectors then measure a superposition of the f - f and s - s paths. Once again, they cannot distinguish between a photon pair (at t) taking the f - f path, and an earlier pair (at $t' = t - \Delta L/c$) taking the s - s path, and by varying ΔL_j (i.e. ϕ_j) the interference of these two paths will result in sinusoidal fringes. Now, however, the fringes potentially reach 100% visibility without any need for precision time-based postselection. One could argue that there is still postselection performed by the polariser, but it can be seen above that all outputs in the $\{D, A\} \otimes \{D, A\}$ basis give the same full-visibility fringes, so there is no need for any form of fair-sampling

assumption⁴.

It is easy to generalise the above calculation to allow the polarisation-sensitive interferometers to redirect the incoming photons according to an arbitrary polarisation. By appropriate selection of this polarisation, the experiment can move between the original Franson and the postselection-free regimes.

8.3.3 Time-frequency tomography

Alternatively, the operation of this apparatus can be viewed in the following way. As the d.c. photons are transmitted through the polarisation-sensitive interferometer, information stored in the time-frequency domain is mapped onto the polarisation degree of freedom which can then be analysed independently. Detecting the photons then traces over the time-frequency degree of freedom because there is no access to any information about the emission times of the photon pairs. As a result, analysis of the polarisation actually measures some convolution of the information originally stored separately in both the polarisation and time-frequency domains. In other words, the fringe visibility in the above postselection-free Franson experiment will also be limited by the quality of the polarisation entanglement.

Using this interpretation, performing polarisation tomography in this context is actually equivalent to performing quantum state tomography in the time-frequency domain, subject to an initial characterisation of the polarisation entanglement. However, in the limit of perfect polarisation entanglement, the results are entirely determined by the time-frequency information. I will describe these ideas in detail below.

If the down-conversion apparatus is designed to produce polarisation-entangled photons, then the general state in the polarisation and time degrees of freedom is

$$|\psi_{\text{dc}}\rangle \sim (|HH\rangle + |VV\rangle) \otimes \sum_{\tau} |\tau, \tau\rangle, \quad (8.9)$$

where I have partitioned the continuous time degree of freedom into discrete time-bins which are separated by $\Delta L/c$. The evolution of the photon states through the unbalanced interferometers is described by:

$$|H, \tau\rangle \longrightarrow |H, \tau\rangle \quad \& \quad |V, \tau\rangle \longrightarrow e^{i\phi} |V, \tau+1\rangle, \quad (8.10)$$

where $\phi = \Delta L \omega_{\text{dc}}/c \pmod{2\pi}$, and I will assume that⁵ $\phi = 0 \pmod{2\pi}$. Therefore, after the interferometers and after tracing over the time degree of freedom, this becomes

$$|\psi_{\text{poln}}\rangle \sim \text{Tr}_{\tau} \left\{ \sum_{\tau} (|HH\rangle |\tau, \tau\rangle + |VV\rangle |\tau+1, \tau+1\rangle) \right\}, \quad (8.11)$$

$$= |HH\rangle + |VV\rangle. \quad (8.12)$$

⁴It is only necessary to ensure that polarisers analyse in a basis which is conjugate to the one used in the polarisation-sensitive interferometer so that the polarisation labelling of the entanglement is erased.

⁵This can be fixed using a bright alignment beam by adjusting ΔL so that incoming, diagonally polarised light is still diagonally polarised after the unbalanced interferometer.

This shows that, if the down-conversion state has the described form, then sending it through this apparatus does not change the polarisation entanglement. This, by itself, is perhaps somewhat surprising, because the HH and VV components of the state have been pulled apart by more than the coherence length of the d.c. photons by the “decohering interferometers”. What is not so obvious, however, is that measuring this polarisation state actually provides quite a lot of information about the time-frequency degree of freedom of the down-conversion state. To demonstrate this, I will consider two different key circumstances.

To begin with, suppose that the down-conversion photon pairs are not *entangled* in time and frequency, but only *correlated*. In other words, suppose that the state is

$$\rho_{\text{dc}} \sim |\Phi^+\rangle\langle\Phi^+| \otimes \sum_{\tau} |\tau, \tau\rangle\langle\tau, \tau|. \quad (8.13)$$

The state after the interferometers then becomes:

$$\begin{aligned} \rho \sim & \sum_{\tau} |H\tau, H\tau\rangle\langle H\tau, H\tau| + \sum_{\tau} |H\tau, H\tau\rangle\langle V(\tau+1), V(\tau+1)| + \\ & \sum_{\tau} |V(\tau+1), V(\tau+1)\rangle\langle H\tau, H\tau| + \sum_{\tau} |V(\tau+1), V(\tau+1)\rangle\langle V(\tau+1), V(\tau+1)|. \end{aligned} \quad (8.14)$$

So after tracing over the time information, the remaining polarisation state is

$$\rho_{\text{poln}} = \text{Tr}_{\tau} \{\rho\} = |HH\rangle\langle HH| + |VV\rangle\langle VV|, \quad (8.15)$$

because $\text{Tr} \{|\tau, \tau\rangle\langle\tau+1, \tau+1|\} = \sum_{\tau_1\tau_2} \langle\tau_1, \tau_2|\tau, \tau\rangle\langle\tau+1, \tau+1|\tau_1, \tau_2\rangle = 0$.

Now, suppose that the down-conversion photons *are* entangled, but in a form that allows for a relative phase shift between pairs in adjacent time bins, i.e.,

$$|\psi_{\text{dc}}\rangle \sim |\Phi^+\rangle \otimes \sum_{\tau} e^{i\varphi_{\tau}} |\tau, \tau\rangle. \quad (8.16)$$

After the decohering interferometers the state becomes

$$|\psi\rangle \sim \sum_{\tau} e^{i\varphi_{\tau}} |H\tau, H\tau\rangle + \sum_{\tau} e^{i\varphi_{\tau}} |V(\tau+1), V(\tau+1)\rangle \quad (8.17)$$

$$= \sum_{\tau} e^{i\varphi_{\tau}} (|HH\rangle + e^{i\Delta_{\tau}} |VV\rangle) |\tau, \tau\rangle, \quad (8.18)$$

where $\Delta_{\tau} = \varphi_{\tau-1} - \varphi_{\tau}$. After tracing over the time domain, this becomes

$$\rho_{\text{poln}} = \text{Tr}_{\tau} \{\psi\} \quad (8.19)$$

$$= \text{Tr}_{\tau} \left\{ \sum_{\tau_1\tau_2} e^{i\varphi_{\tau_1}} e^{-i\varphi_{\tau_2}} (|HH\rangle + e^{i\Delta_{\tau_1}} |VV\rangle) (\langle HH| + e^{-i\Delta_{\tau_2}} \langle VV|) \otimes |\tau_1, \tau_1\rangle\langle\tau_2, \tau_2| \right\} \quad (8.20)$$

$$= \sum_{\tau} (|HH\rangle + e^{i\Delta_{\tau}} |VV\rangle) (\langle HH| + e^{-i\Delta_{\tau}} \langle VV|) \quad (8.21)$$

$$= |HH\rangle\langle HH| + |VV\rangle\langle VV| + \left(\sum_{\tau} e^{-i\Delta_{\tau}} \right) |HH\rangle\langle VV| + \left(\sum_{\tau} e^{i\Delta_{\tau}} \right) |VV\rangle\langle HH| \quad (8.22)$$

To interpret this, I need to consider two different regimes: $\Delta L > L_p$ and $\Delta L \ll L_p$. In the first, the photon pairs from adjacent time bins will not be coherent with each other, so that Δ_τ will be a randomly varying phase. As a result, $\sum_\tau \exp(i\Delta_\tau) = 0$, and the state will become

$$|\psi_{\text{dc}}\rangle \longrightarrow \rho_{\text{poln}} = |HH\rangle\langle HH| + |VV\rangle\langle VV|. \quad (8.23)$$

If, on the other hand, $\Delta L \ll L_p$, then sequential photon pairs will be coherent with at most a linear phase shift⁶, $\varphi_\tau = \tau\varphi$. In this case, the down-conversion state evolves becomes

$$|\psi_{\text{dc}}\rangle \longrightarrow |\psi_{\text{poln}}\rangle = |HH\rangle + e^{-i\varphi}|VV\rangle. \quad (8.24)$$

So a random phase variation in the time domain shows up as mixture in the polarisation tomography, and a regular phase shift becomes a phase shift in the polarisation entanglement.

I note here that analysing this experiment in this way makes it a simple matter to incorporate imperfections into the initial states in both the polarisation and time domains. For example, this allows us to calculate the maximum entanglement expected in the output state for a given input polarisation state.

Returning now to the original problem, if performing this experiment gives the Φ^+ Bell state as the result from the polarisation tomography, then this demonstrates that the down-conversion state is entangled in the time-frequency degree of freedom with a regular phase relationship between photon pairs in adjacent time bins with $\varphi = 0$. Indeed, this is what one might expect as a result of the phase-matching conditions in parametric down-conversion.

The interpretation

Can this really be called time-frequency tomography? Ideally, the goal of tomography is to completely characterise the state of a quantum system in some (generally limited) Hilbert space in an impartial way so that no initial assumptions are made about the unknown state. Is this goal fulfilled by the technique described above? It may seem that it only characterises a few restricted parameters of those that might describe the time-frequency quantum state. However, I claim that these are the only *meaningful* parameters when considering the time-domain tomography of a continuous source (in this case, the down-conversion produced by a cw pump laser). For example, with a cw pump, the probability of finding a photon pair in a particular time must be the same for all time bins, so it does not matter that the above technique retrieves no information about the relative sizes of the probability amplitudes.

The situation is of course quite different if the pump laser is a pulsed source. However, the clock signal of the pump pulses also provides a regular time reference, so that tracing over the time domain is no longer necessary, and tomography could probably be performed directly in the time domain.

⁶With a cw pump, the output must not be sensitive to the arbitrary setting of a “time origin”. With a linear phase relationship, changing the time origin only results in an irrelevant global phase.

For completeness, I wish to add one final note to the above discussion. One important advantage of the polarisation-assisted technique is that the adjacent time bins need not be resolved by the coincidence counting electronics. Since the standard electronics is normally limited to time-scales of at least 1ns, which corresponds to path-length imbalances of $\sim 30\text{cm}$, this is quite a considerable advantage. For example, in our experiments in this chapter (as well as the original experiment in Ref. [19]) the decohering interferometer is implemented using a long ($\sim 5\text{cm}$) birefringent quartz crystal oriented to delay one linear polarisation ($\Delta L \sim 100\mu\text{m}$). However, if longer path imbalances are used, and the polarisation-based technique is combined with the fast electronics required to implement the original Franson scheme, then this adds an extra degree of flexibility to the tomography.

Consider an incoming two-photon state of the form

$$|\psi\rangle \sim |DD\rangle \otimes \sum_{\tau} \alpha_{\tau} |\tau, \tau\rangle + \beta_{\tau} |\tau, \tau+1\rangle + \gamma_{\tau} |\tau+1, \tau\rangle. \quad (8.25)$$

After the interferometers the state becomes

$$\begin{aligned} \psi \sim & \sum_{\tau} \alpha_{\tau} (|H\tau, H\tau\rangle + |H\tau, V(\tau+1)\rangle + |V(\tau+1), H\tau\rangle + |V(\tau+1), V(\tau+1)\rangle) + \\ & \sum_{\tau} \beta_{\tau} (|H\tau, H(\tau+1)\rangle + |H\tau, V(\tau+2)\rangle + |V(\tau+1), H(\tau+1)\rangle + |V(\tau+1), V(\tau+2)\rangle) + \\ & \sum_{\tau} \gamma_{\tau} (|H(\tau+1), H\tau\rangle + |H(\tau+1), V(\tau+1)\rangle + |V(\tau+2), H\tau\rangle + |V(\tau+2), V(\tau+1)\rangle). \end{aligned} \quad (8.26)$$

If the counting electronics can now resolve adjacent time bins, then the effect of coincidence detections is a modified partial trace, $\text{Tr}_{\tau} \{\rho\} = \sum_{\tau} \langle \tau, \tau | \rho | \tau, \tau \rangle$, so the measured state is

$$\rho_{\text{poln}} = \begin{bmatrix} \sum_{\tau} |\alpha_{\tau}|^2 & \sum_{\tau} \alpha_{\tau} \gamma_{\tau-1}^* & \sum_{\tau} \alpha_{\tau} \beta_{\tau-1}^* & \sum_{\tau} \alpha_{\tau} \alpha_{\tau-1}^* \\ \sum_{\tau} \gamma_{\tau-1} \alpha_{\tau}^* & \sum_{\tau} |\gamma_{\tau}|^2 & \sum_{\tau} \gamma_{\tau} \beta_{\tau}^* & \sum_{\tau} \gamma_{\tau} \alpha_{\tau}^* \\ \sum_{\tau} \beta_{\tau-1} \alpha_{\tau}^* & \sum_{\tau} \beta_{\tau} \gamma_{\tau}^* & \sum_{\tau} |\beta_{\tau}|^2 & \sum_{\tau} \beta_{\tau} \alpha_{\tau}^* \\ \sum_{\tau} \alpha_{\tau-1} \alpha_{\tau}^* & \sum_{\tau} \alpha_{\tau} \gamma_{\tau}^* & \sum_{\tau} \alpha_{\tau} \beta_{\tau}^* & \sum_{\tau} |\alpha_{\tau}|^2 \end{bmatrix} \quad (8.27)$$

Of course, for a down-conversion source, we expect $\alpha_{\tau} = \alpha_{\tau-1} e^{i\varphi_{\alpha}}$ and $\beta_{\tau} = \gamma_{\tau} = 0$. However, using this technique one could *measure* the relative sizes of $\sum_{\tau} |\alpha_{\tau}|^2$, $\sum_{\tau} |\beta_{\tau}|^2$ and $\sum_{\tau} |\gamma_{\tau}|^2$, and also obtain information about the coherence of the $|\tau, \tau+1\rangle$ (β) and $|\tau+1, \tau\rangle$ (γ) populations (if any are nonzero).

8.4 Experimental design

A conceptual layout of the experiment is illustrated in Fig. 8.4. Broadly, the experiment consisted of two main stages: the source, photon pairs produced by spontaneous parametric down-conversion, and the analysis, consisting of three sequential (local) projective measurements, one for each DOF, which included coupling the down-conversion photons into single-mode fibres. Not surprisingly, setting up this source of hyperentangled photons

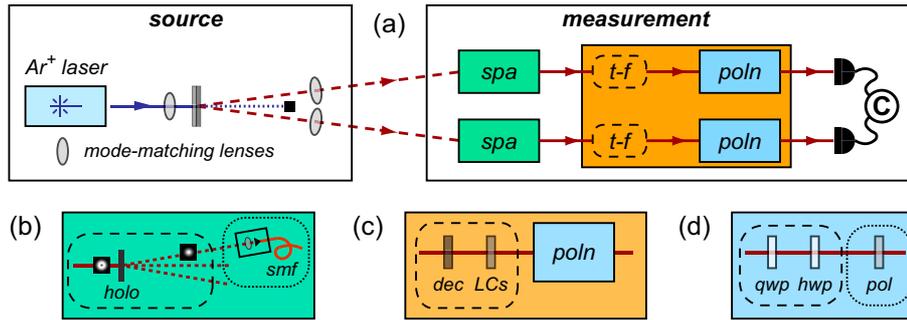


Figure 8.4: Conceptual experimental layout for the creation and analysis of hyperentangled photons. (a) The photons, produced using two sandwiched nonlinear crystals (BBO), pass through a state filtration process for each degree of freedom before coincidence detection. The measurement insets show the filtration processes as a transformation of the target state (dashed box) and a filtering step to discard the other components of the state (dotted box). (b) *Spatial Filtration* (spa): hologram (holo) and single-mode fiber (smf). (c) *Time-Frequency Transformation* (e-t): thick quartz decoherer (dec) and liquid crystal (LC). (d) *Polarisation Filtration* (poln): quarter-wave plate (qwp), half-wave plate (hwp) and polariser (pol).

involved dealing with many of the same issues as arose with setting up the source of spatially entangled photons [described in previous chapter; see Sec. 7.3]. The main difference is a mode-matching consideration related to generating polarisation entanglement.

Spatial and time-frequency entanglement result naturally from the energy and momentum conservation dictated by ideal Type-I phase-matching conditions [see Ch. 6]. To generate polarisation entanglement as well, we used the crystal sandwich configuration introduced by Kwiat et al. [18, 34] [Fig. 8.5]. Two contiguous β -barium borate (BBO) nonlinear crystals were aligned with their optic axes in perpendicular planes and pumped with a 120mW 351nm Ar^+ laser. Each 0.6mm-thick crystal was phase-matched to produce degenerate 702nm photon pairs into a cone of 3.0° half-opening angle (cut with optic axes at 33.9° to the propagation axis). The first (second) crystal was oriented to produce pairs of horizontally (vertically) polarised photons. When pumped by a diagonally polarised beam, two such down-conversion processes are coherent, provided the spatial modes emitted from each crystal are indistinguishable. With a pump focussed to a beam waist at the crystals, this constraint can be satisfied by using thin crystals and “large” beam waists (large relative to the mismatch in the overlap of the down-conversion cones from each crystal [18]). However, the spatial entanglement is maximised by balancing the relative populations of the low-order Gaussian modes [35]. As described in Sec. 7.3, this is optimised by imaging small spots at the crystal, in conjunction with a small pump beam waist, to collect a large area of the down-conversion cones. Here, we compromised by employing an intermediate waist size ($90\ \mu\text{m}$) at the crystal. We then used mode-matching lenses ($f=400\text{mm}$, $\approx 520\text{mm}$ from crystal) to optimise the coupling of the rapidly diverging down-conversion modes into single-mode fibres.

The measurement process was carried out by progressively filtering the down-conversion photons in each degree of freedom using the techniques described in Ch. 4, Ch. 5 and Sec. 8.3.1. For the spatial mode manipulation, we used binary, plane-wave phase holo-

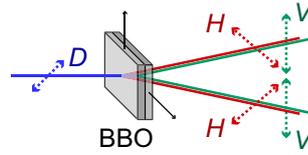


Figure 8.5: The Type-I crystal sandwich source of polarisation entanglement. Two non-linear BBO crystals with orthogonally oriented optic axes (H and V , say) are pumped by a diagonally-polarised pump beam (simultaneously producing VV and HH photon pairs, resp.). The output photons are then entangled in polarisation, provided the spatial modes emitted from the different crystals are indistinguishable.

grams (diffraction efficiency $\sim 40\%$) [see Sec. 5.4] mounted on precision, computer-controlled translation stages (this improved both time and accuracy of the spatial measurements). After coupling the photons into the single-mode fibres, it was important to maintain their polarisation, so the fibres were carefully coiled and taped to foam cushions for temperature stability and to prevent movement, both of which could cause changing geometrical effects. We compensated for the fixed polarisation rotation introduced by the fibres with “fibre wings” (in-fibre polarisation controllers) which we aligned to maintain the known polarisation of a diode laser⁷. We found that some realignment of the fibre wings was required each day or two to keep the nil polarisation rotation condition satisfied.

We were able to control the polarisation entanglement of the source in several ways [see Fig.8.6]. A half-wave plate directly before the crystal (the “epsilon” plate) controlled the linear polarisation of the pump beam, and consequently also what proportion of the power was pumping each crystal. Adjusting this allowed us to control the degree of entanglement and move smoothly between separable and maximally entangled polarisation states. The epsilon plate was mounted in a precision, computer-controlled rotation stage, so we could finely tune the balance of the HH and VV polarisation states to maximise the entanglement. The relative phase between these two components was controlled by a quarter-wave plate (the “phi” plate) also positioned before the crystal. With its optical axis aligned parallel to either H or V , this plate could be tilted around the vertical to apply a variable birefringent phase shift between the H and V components of the pump beam (by varying the optical path length travelled by the pump through the wave plate). Combining these two controls, we were able to produce arbitrary polarisation states of the form,

$$|\psi_{\text{poln}}\rangle = \frac{1}{\sqrt{1 + \varepsilon^2}} (|HH\rangle + \varepsilon e^{i\phi}|VV\rangle). \quad (8.28)$$

Once set almost to the desired value (generally to produce $|\Phi^+\rangle$), fine adjustments could be made to this wave plate without affecting the spatial alignment of the source (relative to the single-mode fibres).

We could also control to some degree the coherence of the polarisation state, by coupling to and tracing over the time degree of freedom [36]. For example, by separating the H and V

⁷The procedure, developed by Julio Barreiro, involved adjusting the fibre wings so that both an H - and D -polarised input emerged unchanged. This was made possible by using a Thorlabs polarimeter to observe the output polarisation real-time. The diode laser was temperature-controlled to run at around 693nm to match the conditions experienced by the down-conversion photons as closely as possible

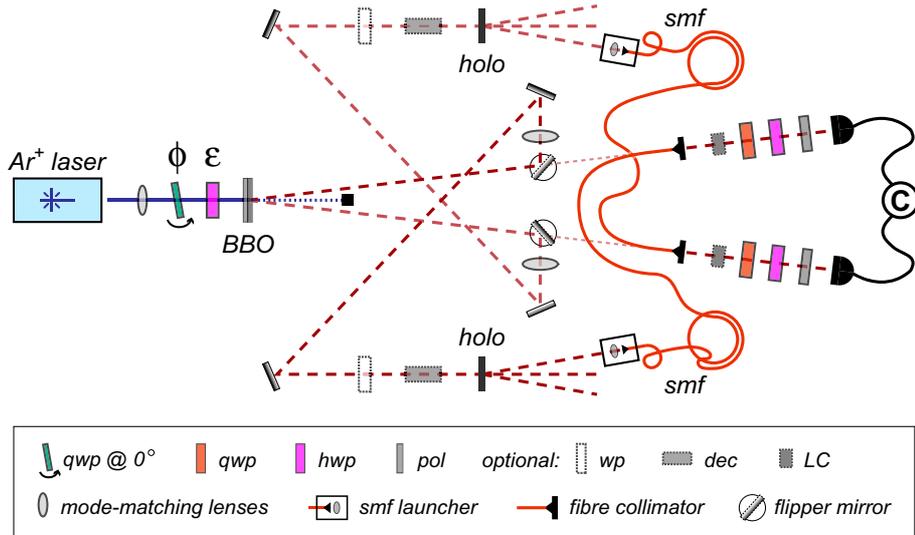


Figure 8.6: Detailed experimental layout for the creation and analysis of hyperentangled photons.

components of one d.c. photon in time, varying amounts of decoherence can be introduced between the corresponding subspaces depending on the size of time separation compared with the single-photon coherence length ($L_{dc} = \lambda^2 / \Delta\lambda \sim 70\lambda \sim 50\mu\text{m}$, with $\Delta\lambda = 10\text{nm}$ determined by the interference filters in the detection assembly). We introduced the delay between polarisations using long, birefringent crystals⁸ (quartz, $L \sim 11\text{mm}$) which acted as unbalanced, polarisation-sensitive Mach-Zender interferometers. This produced an effective path-length mismatch of $\Delta n_{\text{quartz}} L \sim 100\mu\text{m}$, which was significantly greater than L_{dc} , but much less than the pump coherence length of $L_p \sim 10\text{cm}$.

This procedure is obviously very similar to the method described in Sec. 8.3.1 for measuring the time-frequency correlations between the d.c. photons. Once again, we used the long quartz crystals to behave as polarisation-sensitive, unbalanced interferometers, and we carried out fine adjustment of the time shift using liquid crystals with a variable voltage supply (with their optic axes aligned parallel to the quartz rods).

The main stability problems we had with our experiments resulted from pointing instability of the Ar^+ pump laser, which was probably exacerbated by the propagation distance from the laser output to the crystal ($\sim 5\text{m}$). We were particularly sensitive to this because of the spatial precision required to align the single-mode fibres to the d.c. beams, and the holograms to the single-mode fibres. As a result, we had to carry out a reasonably thorough realignment of the source and fibre couplers every day or two. This caused some problems when we were running our large tomographies (1296 measurements, $\sim 10\text{hrs}$). We wanted to perform four long tomographies under the “same conditions” (i.e. without a major realignment), each requiring some preparation in between. This resulted in some deterioration in the quality of the source entanglement (particularly in the spatial DOF) over the course of the measurements.

⁸We had quartz crystals with $L \sim 6\text{mm}$ and 11mm , which corresponded to path-length differences of $\sim 80\lambda$ and 140λ , respectively. We were able to achieve different effects by using different combinations of crystals and altering their effective orientation (using wave plates).

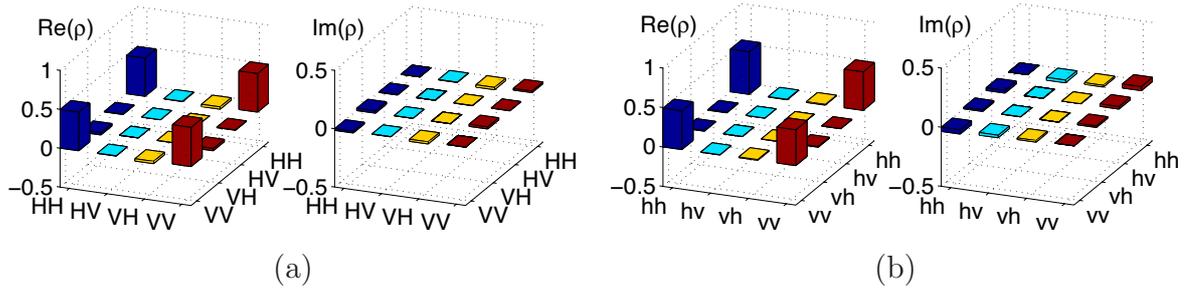


Figure 8.7: Typical two-qubit tomographies within the (a) polarisation and (b) spatial mode subspaces.

8.5 Results

Before beginning a discussion of our results, I would like to comment on the tomographic optimisations used in this chapter. In Ref. [1], we used a (fairly sophisticated) *local* optimisation algorithm to perform the tomographic reconstructions. In this chapter, I have reanalysed the results using our latest *global* optimisation routines [see Ch. 3]. I have reported results based on both “fixed weights” (FW) and standard maximum likelihood (ML) reconstructions defined in Sec. 3.8 [the latter in square brackets]. I will generally only report the latter case if they differ significantly from the former. However, for the large tomographies in Sec. 8.5.4, I have reported only the ML reconstructions, because the FW optimisations have failed to converge. Finally, note that I have also used slightly different definitional conventions from those in Ref. [1] (e.g. in defining R and L) to be consistent with the rest of the thesis. This produces some differences in the results (e.g. different signs in coherence phases).

As is the case throughout this thesis, all errors in physical quantities calculated from tomographic reconstructions were determined using Monte-Carlo simulations with 200 samples under the assumption that Poissonian fluctuations were the sole source of error [see Sec. 3.6.2 for details].

8.5.1 Individual subspace tomographies

As a first test of the hyperentanglement, we characterised the polarisation and spatial mode subspaces using tomographic reconstruction. We consistently measured high-quality states in both cases; typical examples are shown in Fig. 8.7 for (a) polarisation

$$(T=0.987\pm 0.002, N=0.991\pm 0.001, S_L=0.009\pm 0.001, F_{\Phi^+}=0.9939\pm 0.0005, Q=1.4\pm 0.2)$$

and (b) spatial mode

$$(T=0.961\pm 0.004, N=0.978\pm 0.002, S_L=0.023\pm 0.003, F_{\Phi^+}=0.987\pm 0.001, Q=1.4\pm 0.2)$$

subspaces. To our knowledge, these are the highest quality entangled states reported to date.

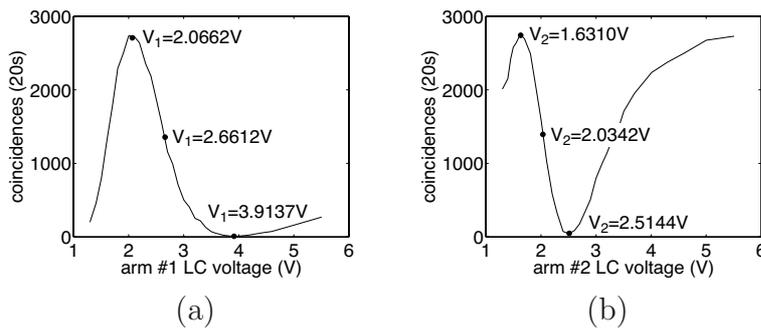


Figure 8.8: Liquid crystal voltage curves: (a) arm #1 curve taken with $V_2=1.6481\text{V}$; (b) arm #2 curve taken with $V_1=2.0657\text{V}$. The black dots indicate fitted voltage points where the liquid crystals apply birefringent phases separated by $\phi = \pi/2$. Note: the LC voltages, while precise, were difficult to set repeatably.

8.5.2 Time-frequency tomography

We followed the theory described in Sec. 8.3.3 to perform the first tomographic analysis of the time-frequency domain. We used long, birefringent quartz crystals to map the time-frequency information onto the polarisation DOF, and voltage-controlled liquid crystals to fine-tune the polarisation-sensitive time delay. Until my more recent theoretical investigations [Sec. 8.3.1], we did not perfectly understand the details of the time-frequency tomography technique. Consequently, our results were limited to a certain extent by the way we implemented the experiments, as I will briefly describe here.

The quartz crystal lengths⁹ were approximately matched in length, so that $|\Delta L_1 - \Delta L_2| \ll L_{\text{dc}}$. However, instead of setting the path-length differences ($\Delta L_{1,2}$) impartially so that $\phi_{1,2} = 0 \pmod{2\pi}$ [see Eq. (8.10)], we used the d.c. signal and set the LC voltages so that the LL polarisation component in the output was minimised. Unfortunately, our time-frequency tomography was then insensitive to the size of any relative phase in the time-entangled state, but we were still able to characterise the *nature* of the phase relationship between adjacent time-bins, i.e. whether it was regular or randomly varying.

It is now clear that the ideal time-frequency tomography involves setting the liquid crystals appropriately and then running standard polarisation tomography on the output (i.e. using the wave plates only). In our experiments, however, we also used the LCs to implement some of the wave plate transformations in the polarisation analysis, which achieved the same result but with slightly less accurate measurements. The liquid crystal birefringence responds nonlinearly with the applied voltage, so we determined the measurement positions via a data fit with the measured fringes [Fig. 8.8].

Before inserting the quartz crystals and liquid crystals into the apparatus, we reconstructed the initial polarisation state ($T=0.984\pm 0.003$, $N=0.991\pm 0.001$, $S_L=0.009\pm 0.002$, $F_{\text{opt}}=0.9957\pm 0.0006$, $Q=0.9\pm 0.2$) [Fig. 8.9(a)], which is essentially the input state to the time-to-polarisation mapping process used to make the time-frequency measurements.

⁹These were obviously fixed at manufacture, but the *effective* lengths were also affected by the alignment of the crystals, i.e. the tilt. This effect should have been reasonably small, however, because the crystals were back-reflected in the apparatus.

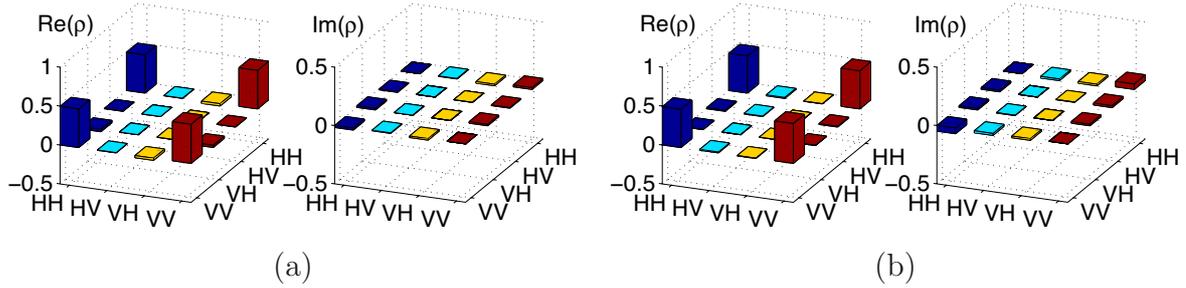


Figure 8.9: Time-frequency tomography: (a) the initial polarisation state; (b) the post-mapping polarisation state—also contains time-frequency information.

This state defines the best possible entanglement and mixture levels to be expected from the final time-frequency measurements. Figure 8.9(b) shows the reconstructed density matrix for the final polarisation state which contains the time-frequency tomography information ($T=0.964\pm 0.005$, $N=0.981\pm 0.002$, $S_L=0.02\pm 0.02$, $F_{\text{opt}}=0.991\pm 0.001$, $Q=1.3\pm 0.2$). This state shows a slight decrease in entanglement (98% of T) and purity (99% of P) over the initial state, indicating the high quality of the time-entangled state, especially since these small decreases would also have been partly caused by the non-ideal polarisation measurements.

It would be very interesting to investigate this concept of time-frequency tomography more thoroughly. First, I would like to repeat the measurements with the ideal method to give the best possible results. However, I would also like to consider different experimental conditions, for example by varying the size of the birefringent delay in comparison with the pump coherence length.

Finally, the main aim of this work was to demonstrate and characterise hyperentanglement. In this spirit, these experiments can also be interpreted in some sense as a simultaneous tomographic characterisation of the polarisation and time-frequency degrees of freedom, since high-quality entanglement in both is necessary to give such results. Moreover, the information contained in the reconstructed matrix (e.g. in the phase) is some sort of convolution of the properties of both polarisation- and time-entangled states.

8.5.3 CHSH Bell violations

To verify quantum mechanical correlations, we tested every degree of freedom against a Clauser-Horne-Shimony-Holt (CHSH) Bell inequality [29] [see Sec. 2.1.6]. To measure the strongest violation for the polarisation and spatial-mode DOFs, we determined the optimal measurement settings by first tomographically reconstructing the two-qubit subspace of interest.

Table 8.1 shows the Bell parameters measured for the polarisation, spatial mode, and energy-time subspaces, with various projections in the complementary degrees of freedom. For every entangled subspace, the Bell parameter exceeded the classical limit of $S=2$ by more than 20 standard deviations (σ). This verifies conclusively that the down-conversion photon pairs were completely hyperentangled—simultaneously entangled in *every* degree

DOF	Spatial mode projected subspaces				
	$ gg\rangle\langle gg $	$ rl\rangle\langle rl $	$ lr\rangle\langle lr $	$ hh\rangle\langle hh $	$ vv\rangle\langle vv $
Φ_{poln}^+	2.76[76 σ]	2.78[46 σ]	2.75[44 σ]	2.81[40 σ]	2.75[33 σ]
$\Phi_{\text{t-e}}^+$	2.78[77 σ]	2.80[40 σ]	2.80[40 σ]	2.72[30 σ]	2.74[29 σ]
DOF	Polarisation projected subspaces				
	no polarisers	$ HH\rangle\langle HH $	$ VV\rangle\langle VV $		
Φ_{spa}^+	2.78[78 σ]	2.80[36 σ]	2.79[37 σ]		
$\alpha gg\rangle + rl\rangle$	2.33[55 σ]	2.30[25 σ]	2.38[30 σ]		
$\alpha gg\rangle + lr\rangle$	2.28[47 σ]	2.26[20 σ]	2.31[26 σ]		

Table 8.1: Bell parameter S showing CHSH-Bell inequality violations in every degree of freedom. The local realistic limit ($S \leq 2$) is violated by the number of standard deviations shown in brackets, determined by counting statistics.

of freedom.

For both the polarisation and spatial-mode measurements, we traced over the energy-time DOF by not projecting in this subspace. We measured the polarisation correlations while projecting the spatial modes into the various states: l , g , r , h , and v . The measured Bell parameters agreed (within $\sim 2\sigma$) with predictions from tomographic reconstruction and violated the inequality by more than 30σ . In the spatial mode DOF, the correlations for the state Φ_{spa}^+ were close to maximal ($S=2\sqrt{2}\approx 2.88$), also in agreement with predictions from the measured state density matrix. We also tested Bell inequalities for the non-maximally entangled spatial states: $\alpha|gg\rangle + |rl\rangle$ and $\alpha|gg\rangle + |lr\rangle$; the measured Bell parameters in this case were slightly smaller than predicted from the tomography (e.g. $S_{\text{exp}}=2.28\pm 0.01 < S_{\text{pred}}=2.35$), yet still 20σ above the classical limit. There were two main possible causes for this discrepancy—the measurements were extremely sensitive to any inaccuracy in locating the position of the holographic phase singularity, and the correlations would also have been affected by the errors involved in making nondegenerate superposition measurements [see Sec. 5.6.6]. Finally, our measured Bell violation for the time-frequency DOF using particular phase settings is in good agreement with the prediction ($S=2\sqrt{2}V$) from the measured two-photon interference visibility, $V=0.985\pm 0.002$.

8.5.4 Multi-degree quantum state tomography

To investigate the hyperentanglement in more detail, we fully characterised the joint polarisation and spatial-mode state via quantum state tomography¹⁰. First, we performed the minimum set of $4^2 \times 9^2 = 1296$ measurements required for tomographic reconstruction in the $(2 \times 3) \times (2 \times 3)$ -dimensional Hilbert space consisting of two polarisation and three spatial modes for each photon¹¹. The measured state [Fig. 8.10] has a lin-

¹⁰All the reconstructions in this section result from standard maximum likelihood optimisations [see Sec. 3.8 for details]. So far, the optimisations based on the fixed-weight likelihood function have been unsuccessful, which seems to be due to limitations in the convex optimisation solvers which are currently accessible. To remain consistent with the rest of the thesis, I have continued to display all results based on the standard maximum likelihood optimisations in square brackets.

¹¹This data was collected for 40s per projection with ~ 600 detected photon pairs/s.

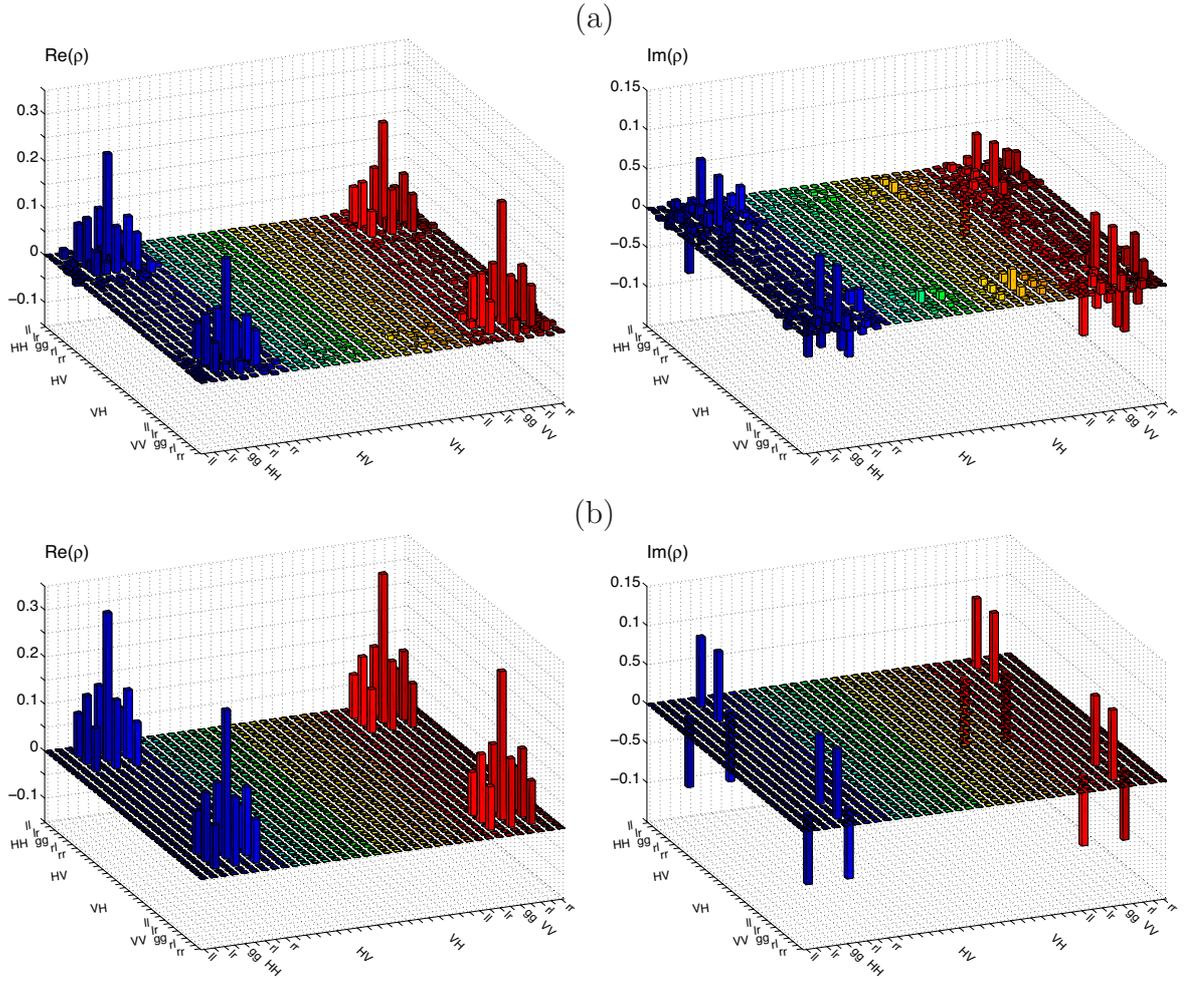


Figure 8.10: (a) Measured and (b) close pure (ψ_p) density matrices of a $2 \times 2 \times 3 \times 3$ -dimensional state of two-photon polarisation and spatial mode [$F(\varrho, \psi_p) = 0.738 \pm 0.002$].

ear entropy of [$S_L = 0.393 \pm 0.003$] and a fidelity of [$F(\varrho, \psi_p) = 0.738 \pm 0.002$] with the pure state $|\psi_p\rangle = |\Phi_{\text{poln}}^+\rangle \otimes (|lr\rangle + \alpha|gg\rangle + |rl\rangle)$, for [$\alpha = 1.85e^{-i0.175\pi}$]. Treating the photon pairs as a system of two six-level particles, we can quantify the entanglement using the negativity [see Sec. 2.1.6], which, in this 6×6 -dimensional Hilbert space, reaches a value of 5 for the maximally entangled state (i.e. $\alpha=1$). For our measured, somewhat mixed state, [$N(\varrho) = 3.23 \pm 0.01$] compared with [$N(\psi_p) = 4.468$]. The spatial-mode subspace alone has [$N(\varrho_{\text{spa}}) = 1.200 \pm 0.007$], greater than the maximum negativity ($N=1$) of any two-qubit system. In the polarisation subspace, [$T(\varrho_{\text{poln}}) = 0.646 \pm 0.004$] and [$N(\varrho_{\text{poln}}) = 0.790 \pm 0.003$]. Thus, our large state possesses both two-qubit and two-qutrit entanglement.¹² The fit quality parameter for the maximum likelihood reconstruction was [$Q(\varrho) = 4.18 \pm 0.03$].

We also selected a $(2 \times 2) \times (2 \times 2)$ -dimensional state that was maximally entangled in both polarisation and spatial mode¹³ (by neglecting the $|gg\rangle$ spatial component), with $F = 0.9748 \pm 0.0009$ between the measured state ρ [Fig. 8.11(a)] and the target $\Phi_{\text{poln}}^+ \otimes \Phi_{\text{spa}}^+$.

¹²With the original local optimisation routines used in Ref. [1]: $S_L = 0.46$; $F(\rho, \psi_p) = 0.69 \pm 0.01$ with $\alpha = 1.88e^{i0.16\pi}$; $N(\varrho) = 3.0$ and $N(\psi_p) = 4.4$; and $N(\varrho_{\text{spa}}) = 1.14$. The different sign in the phase of α arises from the different definitional conventions used here compared with Ref. [1] (e.g. in defining R and L).

¹³The data for these tomographies were collected for 20s per projection with ~ 100 detected pairs/s.

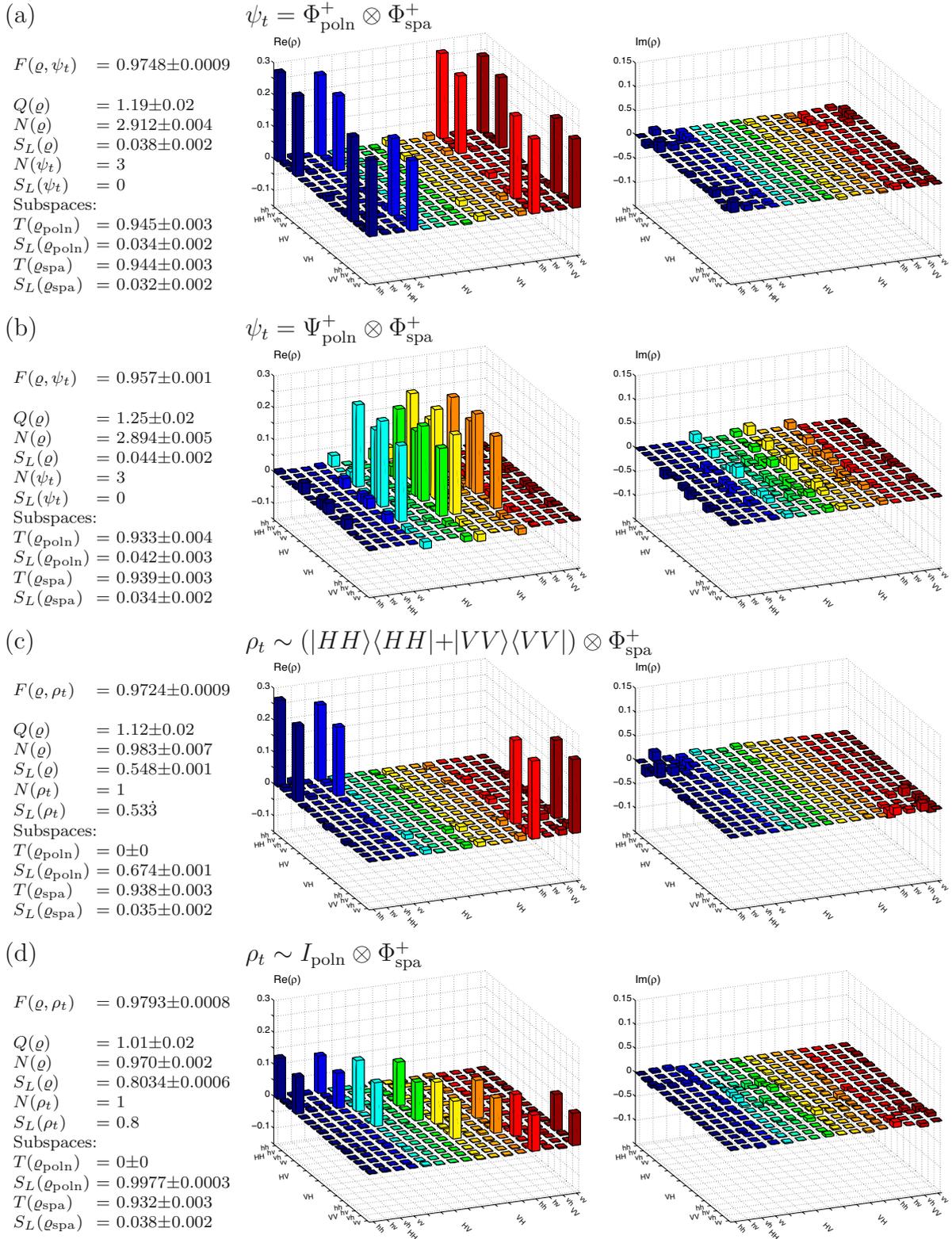


Figure 8.11: Measured density matrices (real parts) for the $2 \times 2 \times 2 \times 2$ -dimensional two-photon states in the polarisation and degenerate spatial qubit subspaces. For each state, we list: the target state ρ_t , the fidelity $F(\rho, \rho_t)$ of the measured state ρ with the target ρ_t , their negativities and linear entropies, and the tangle and linear entropy for each subspace. All results are based on standard maximum likelihood reconstructions [Sec. 3.8].

By tracing over either the polarisation or spatial mode, we look at the measured state in the remaining subspace. The reduced states in both DOFs are pure ($S_L < 0.04$) and highly entangled ($T > 0.94$).

With this precise source of hyperentanglement, we have the flexibility to prepare nearly arbitrary polarisation states [36], and to select arbitrary spatial-mode encodings. For example, we also generated a different maximally entangled state: $\Psi_{\text{poln}}^+ \otimes \Phi_{\text{spa}}^+$ [Fig. 8.11(b)]. As mentioned earlier, we were also able to introduce controlled levels of mixture into the polarisation state by coupling to and tracing over the time-frequency DOF using quartz decoherers. Doing this allowed us to prepare a previously unrealised state that simultaneously displays *classical correlations* in polarisation and maximal *quantum correlations* between spatial modes ($F(\rho, \rho_t) = 0.9724 \pm 0.0009$) [Fig. 8.11(c)]: $\rho_t \sim (|HH\rangle\langle HH| + |VV\rangle\langle VV|) \otimes |\Phi_{\text{spa}}^+\rangle\langle\Phi_{\text{spa}}^+|$. With high fidelity, we also prepared the state $\rho_t \sim I_{\text{poln}} \otimes |\Phi_{\text{spa}}^+\rangle\langle\Phi_{\text{spa}}^+|$, with no polarisation correlations at all (i.e. completely mixed or unpolarised), while still maintaining maximal entanglement in the spatial DOF ($F(\rho, \rho_t) = 0.979 \pm 0.001$) [Fig. 8.11(d)].

8.6 Conclusion

In this chapter, I have explored ways to produce and characterise multi-degree entanglement between pairs of single photons. In particular, I have focussed on hyperentanglement which can be used to enable 100%-efficient Bell-state analysis [9], an important component of a variety of quantum information protocols [37, 38].

In the experiments, we achieved the first realisation of complete hyperentanglement in pairs of single photons. To demonstrate this, we characterised the entanglement within each degree of freedom via tomographic analysis (producing the highest quality entangled states reported to date for both polarisation and spatial modes), and verified that the photon pairs were completely hyperentangled through a series of complementary violations of CHSH-Bell inequalities (each by greater than 20σ). We then fully characterised the polarisation-spatial hyperentanglement using simultaneous tomography of both degrees of freedom, reconstructing a $2 \times 2 \times 3 \times 3$ -dimensional state, which is the largest quantum system to date analysed via a full “black-box” tomography (i.e. making no prior assumptions about the state). Finally, we prepared a range of target states in a restricted $2 \times 2 \times 2 \times 2$ -dimensional subspace with unprecedented fidelities for quantum systems of this size. These included novel states with a controllable degree of correlation in the polarisation subspace.

In Sec. 8.3.3, I introduced a technique for performing quantum state tomography in the time-frequency DOF based on ideas from Ref. [19], and I have also reported some preliminary experimental results demonstrating the technique. In some sense, however, this is also a simultaneous tomography of polarisation and time-frequency DOFs, since it provides information about the quality of entanglement in both. In the same vein, I would note that it is also possible to perform a tomographic analysis of the complete hyperentangled system—polarisation, time-frequency *and* spatial-momentum—by combining this technique with the standard spatial tomography. This would be no more difficult than the combined polarisation-spatial tomographies reported above.

Finally, it is worth mentioning again in this context [recall Ch. 6] that the pairwise mechanism of the $\chi^{(2)}$ -down-conversion process inherently produces entanglement not only of single photons, but also in photon number [39].

8.7 Chapter 8 References

- [1] J. T. Barreiro, N. K. Langford, N. A. Peters, and P. G. Kwiat. Generation of hyperentangled photon pairs. *Physical Review Letters*, 95(26):260501, 2005.
- [2] Z. Zhao, Y.-A. Chen, A.-N. Zhang, T. Yang, H. J. Briegel, and J.-W. Pan. Experimental demonstration of five-photon entanglement and open-destination teleportation. *Nature*, 430:54, 2004.
- [3] P. Walther, K. J. Resch, T. Rudolph, E. Schenck, H. Weinfurter, V. Vedral, M. Aspelmeyer, and A. Zeilinger. Experimental one-way quantum computing. *Nature*, 434:169, 2005.
- [4] A. Vaziri, G. Weihs, and A. Zeilinger. Experimental two-photon, three-dimensional entanglement for quantum communication. *Physical Review Letters*, 89(24):240401, 2002.
- [5] R. T. Thew, A. Acin, H. Zbinden, and N. Gisin. Bell-type test of energy-time entangled qutrits. *Physical Review Letters*, 93:010503, 2004.
- [6] M. N. O’Sullivan-Hale, I. A. Khan, R. W. Boyd, and J. C. Howell. Pixel entanglement: Experimental realization of optically entangled $d=3$ and $d=6$ qudits. *Physical Review Letters*, 94:220501, 2005.
- [7] S. S. R. Oemrawsingh, X. Ma, D. Voigt, A. Aiello, E. R. Eliel, G. W. ’t Hooft, and J. P. Woerdman. Experimental demonstration of fractional orbital angular momentum entanglement of two photons. *Physical Review Letters*, 95:240501, 2005.
- [8] P. G. Kwiat. Hyper-entangled states. *Journal of Modern Optics*, 44:2173, 1997.
- [9] P. G. Kwiat and H. Weinfurter. Embedded Bell-state analysis. *Physical Review A*, 58:R2623, 1998.
- [10] P. Walther, K. J. Resch, Č. Brukner, A. M. Steinberg, J.-W. Pan, and A. Zeilinger. Quantum nonlocality obtained from local states by entanglement purification. *Physical Review Letters*, 94:040504, 2005.
- [11] C. Simon and J.-W. Pan. Polarization entanglement purification using spatial entanglement. *Physical Review Letters*, 89:257901, 2002.
- [12] X.-F. Ren, G.-P. Guo, J. Li, C.-F. Li, and G.-C. Guo. Engineering of multi-dimensional entangled states of photon pairs using hyper-entanglement. *Chinese Physics Letters*, 23:552, 2006.
- [13] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, UK, 2000.
- [14] J. G. Rarity and P. R. Tapster. Experimental violation of Bell’s inequality based on phase and momentum. *Physical Review Letters*, 64:2495, 1990.
- [15] N. K. Langford, R. B. Dalton, M. D. Harvey, J. L. O’Brien, G. J. Pryde, A. Gilchrist, S. D. Bartlett, and A. G. White. Measuring entangled qutrits and their use for quantum bit commitment. *Physical Review Letters*, 93(5):053601, 2004.

- [16] P. G. Kwiat, A. M. Steinberg, and R. Y. Chiao. High-visibility interference in a Bell-inequality experiment for energy and time. *Physical Review A*, 47:R2472, 1993.
- [17] R. T. Thew, A. Acín, H. Zbinden, and N. Gisin. Experimental realization of entangled qutrits for quantum communication. *Quantum Information and Computation*, 4(2):93, 2004.
- [18] P. G. Kwiat, E. Waks, A. G. White, I. Appelbaum, and P. H. Eberhard. Ultrabright source of polarization-entangled photons. *Physical Review A*, 60:R773, 1999.
- [19] D. V. Strekalov, T. B. Pittman, A. V. Sergienko, Y. H. Shih, and P. G. Kwiat. Postselection-free energy-time entanglement. *Physical Review A*, 54:R1, 1996.
- [20] C. Cinelli, M. Barbieri, R. Perris, P. Mataloni, and F. De Martini. All-versus-nothing nonlocality test of quantum mechanics by two-photon hyperentanglement. *Physical Review Letters*, 95:240405, 2005.
- [21] T. Yang, Q. Zhang, J. Zhang, J. Yin, Z. Zhao, M. Żukowski, Z.-B. Chen, and J.-W. Pan. All-versus-nothing violation of local realism by two-photon, four-dimensional entanglement. *Physical Review Letters*, 95:240406, 2005.
- [22] L. Allen, M. W. Beijersbergen, R. J. C. Spreeuw, and J. P. Woerdman. Orbital angular momentum of light and the transformation of Laguerre-Gaussian laser modes. *Physical Review A*, 45(11):8185, 1992.
- [23] J. D. Franson. Bell inequality for position and time. *Physical Review Letters*, 62:2205, 1989.
- [24] D. Stucki, N. Gisin, O. Guinnard, G. Ribordy, and H. Zbinden. Quantum key distribution over 67km with a plug&play system. *New Journal of Physics*, 4:41.1, 2002.
- [25] G. Ribordy, J.-D. Gautier, N. Gisin, O. Guinnard, and H. Zbinden. Automated ‘plug & play’ quantum key distribution. *Electronics Letters*, 34:2116, 1998.
- [26] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, 47:777, 1935.
- [27] J. S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195, 1964. As reprinted in *Speakable and Unsayable in Quantum Mechanics: Collected papers on quantum philosophy*, Cambridge University Press (1987).
- [28] D. Bohm. *Quantum Theory*. Prentice-Hall, Inc., Eaglewood Cliffs, New Jersey, USA, 1951.
- [29] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23:880, 1969.
- [30] J. G. Rarity, P. R. Tapster, E. Jakeman, T. Larchuk, R. A. Campos, M. C. Teich, and B. E. A. Saleh. Two-photon interference in a Mach-Zehnder interferometer. *Physical Review Letters*, 65:1348, 1990.
- [31] P. R. Tapster, J. G. Rarity, and P. C. M. Owens. Violation of Bell’s inequality over 4 km of optical fiber. *Physical Review Letters*, 73:1923, 1994.

-
- [32] E. Santos. Does quantum mechanics violate the Bell inequalities? *Physical Review Letters*, 66:1388, 1991.
- [33] L. De Caro and A. Garuccio. Reliability of Bell-inequality measurements using polarization correlations in parametric-down-conversion photon sources. *Physical Review A*, 50:R2803, 1994.
- [34] A. G. White, D. F. V. James, P. H. Eberhard, and P. G. Kwiat. Nonmaximally entangled states: Production, characterization, and utilization. *Physical Review Letters*, 83(16):3103, 1999.
- [35] G. Molina-Terriza, J. P. Torres, and L. Torner. Management of the angular momentum of light: Preparation of photons in multidimensional vector states of angular momentum. *Physical Review Letters*, 88:013601, 2002.
- [36] A. G. White, D. F. V. James, W. J. Munro, and P. G. Kwiat. Exploring Hilbert space: Accurate characterization of quantum information. *Physical Review A*, 65:012301, 2001.
- [37] C. H. Bennett and S. J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Physical Review Letters*, 69:2881, 1992.
- [38] C. Wang, F.-G. Deng, Y.-S. Li, X.-S. Liu, and G. L. Long. Quantum secure direct communication with high-dimension quantum superdense coding. *Physical Review A*, 71:044305, 2005.
- [39] H. S. Eisenberg, G. Khoury, G. A. Durkin, C. Simon, and D. Bouwmeester. Quantum entanglement of a large number of photons. *Physical Review Letters*, 93:193901, 2004.

*Eight eights are sixty-four,
Multiply by seven.
When it's done,
Carry one
And take away eleven.
Nine nines are eighty-one
Multiply by three.
If it's more,
Carry four,
And then it's time for tea.*

The Emperor's Rhyme
from *Now we are Six*, A. A. Milne

Chapter 9

Distance measures for comparing quantum processes

Acknowledgement

The work in this chapter was first published in Ref. [1], and the main body of the chapter largely incorporates the text of that paper. The work was carried out in close collaboration with Alexei Gilchrist and Michael Nielsen. Furthermore, I would particularly like to acknowledge that my co-workers were chiefly responsible for the work in Sec. 9.4, which I have included because it is a vital part of the overall scope of this work.

9.1 Introduction

In any experimental implementation of a quantum information processing task there are many real-world imperfections that may arise, either in the creation or measurement of a quantum state, or in the manipulation of the state via some quantum process. As an experimentalist, it is important to be able to quantitatively measure and characterise these imperfections in a way that is theoretically meaningful and experimentally practical.

How can this be done? It is well known that quantum states can be completely characterised using quantum state tomography [2, 3], and compared using a variety of common measures [4]. However, although quantum processes can be measured using an analogous procedure called quantum process tomography [4–6], the problem of developing quantitative measures to compare real and idealised quantum processes has not been comprehensively addressed.

Ideally there would be a single good measure, a “gold standard”¹, which would allow the sensible comparison of different experimental implementations of quantum information processing, and this would be agreed upon by experimentalists and theorists alike. In this chapter, I will refer to candidates for such a gold standard as “distance measures” for quantum processes, or as “error measures”, when stressing the comparison of real and idealised processes.

Having such an error measure would be extremely useful when comparing experiments with the theoretical ideal, as well as when wishing to compare different experiments that are designed to perform the same task. To date, many quantum information processing implementations have been assessed in a very *ad hoc* manner. For example, when demonstrating quantum logic gates, experimentalists have often showed only that they act in the correct way on computational basis states (i.e. verifying the truth table of the gate), and a few superposition states. While such demonstrations are important, a figure of merit that is standardised, theoretically well motivated and experimentally practical would be a considerable step forward. Moreover, such a measure would also be extremely valuable in connecting real experiments directly to results such as the fault-tolerance threshold for quantum computation².

The goal of the work in this chapter is to comprehensively address the problem of developing such error measures. Although there is already a sizeable literature on this subject, we believe that there has been a consistent gap between work motivated primarily by theoretical considerations, and work constrained by experimental realities. Here, we have aimed to address both theoretical and experimental desiderata.

The criteria—what makes a good distance measure?

The key to our work is to introduce a list of six simple, physically motivated criteria which we claim that any good measure of distance between quantum processes should satisfy. These criteria enable us to eliminate many ways to define an error measure that *a priori* appear highly plausible.

Suppose Δ is a candidate measure of the distance between two quantum processes. Such processes are described by maps between input and output quantum states, e.g. $\rho_{\text{out}} = \mathcal{E}(\rho_{\text{in}})$, known as *quantum operations* [see Sec. 2.1.3]. Physically, $\Delta(\mathcal{E}, \mathcal{F})$ may be thought of in two ways—either as the error in quantum information processing when one implements \mathcal{E} instead of the ideal (target) process \mathcal{F} , or as the distinguishability between the two processes \mathcal{E} and \mathcal{F} . We believe that any such measure must satisfy the following six properties, motivated by both physical and mathematical concerns:

1. *Metric*: Δ should be a metric. This requires three properties: (i) $\Delta(\mathcal{E}, \mathcal{F}) \geq 0$ with $\Delta(\mathcal{E}, \mathcal{F}) = 0$ if and only if $\mathcal{E} = \mathcal{F}$; (ii) symmetry: $\Delta(\mathcal{E}, \mathcal{F}) = \Delta(\mathcal{F}, \mathcal{E})$; and (iii) the triangle inequality $\Delta(\mathcal{E}, \mathcal{G}) \leq \Delta(\mathcal{E}, \mathcal{F}) + \Delta(\mathcal{F}, \mathcal{G})$.
2. *Easy to calculate*: it should be possible to evaluate Δ in a direct manner.

¹I gratefully acknowledge Michael Nielsen for suggesting this term based on some work by Carlton Caves.

²See Ch. 10 of Ref. [4], and references therein.

3. *Easy to measure*: there should be a clear and achievable experimental procedure for determining the value of Δ .
4. *Physical Interpretation*: Δ should have a well-motivated physical interpretation.
5. *Stability* [7]: $\Delta(\mathcal{I} \otimes \mathcal{E}, \mathcal{I} \otimes \mathcal{F}) = \Delta(\mathcal{E}, \mathcal{F})$, where \mathcal{I} represents the identity operation on an additional quantum system. Physically, this means that unrelated ancillary quantum systems do not affect the value of Δ .
6. *Chaining*: $\Delta(\mathcal{E}_2 \circ \mathcal{E}_1, \mathcal{F}_2 \circ \mathcal{F}_1) \leq \Delta(\mathcal{E}_1, \mathcal{F}_1) + \Delta(\mathcal{E}_2, \mathcal{F}_2)$. Thus, for a process composed of many smaller steps, the total error will be bounded above by the sum of the errors in the individual steps.

The chaining and stability criteria are key properties for estimating the error in a complex quantum information processing task. Because quantum information processing tasks typically consist of a sequence of simpler component operations, a conservative bound on the total error can be found by instead analyzing the individual components. This is critical for applications such as quantum computation, since full process tomography on an n -qubit computation requires exponentially many measurements, and is thus infeasible. Chaining and stability make it sufficient to benchmark the constituent processes involved in the computation to be able to infer that the entire computation is robust.

Many other properties follow from these six criteria. For example, from the metric and chaining criteria it follows that $\Delta(\mathcal{R} \circ \mathcal{E}, \mathcal{R} \circ \mathcal{F}) \leq \Delta(\mathcal{E}, \mathcal{F})$, where \mathcal{R} is any quantum operation. This corresponds to the requirement that post-processing by \mathcal{R} cannot *increase* the distinguishability of two processes \mathcal{E} and \mathcal{F} . Another elementary consequence of the metric and chaining criteria is *unitary invariance*, i.e. $\Delta(\mathcal{U} \circ \mathcal{E} \circ \mathcal{V}, \mathcal{U} \circ \mathcal{F} \circ \mathcal{V}) = \Delta(\mathcal{E}, \mathcal{F})$, where \mathcal{U} and \mathcal{V} are unitary operations.

Diagnostic measures

For both theoreticians and experimentalists, there are strong motivations to find a gold standard satisfying the above criteria—the need for a physically sensible way of evaluating the performance of a quantum process, and the need to compare the success of a theoretical model to the operation of a real, experimental system. For the experimentalist, however, there is also another important consideration—the need for *diagnostic measures* which can be used to build insight into the source of imperfections in experimental implementations. Diagnostic measures may not necessarily be good candidates for our sought-after gold standard (they may fail to satisfy one or more of our criteria), but they still may be extremely useful in the experimental context. Thus, some measures that we have discarded as unsuitable for a gold standard may still be useful as diagnostic measures. Furthermore, it is not difficult to construct many examples of useful diagnostic measures that are not considered here³. A detailed examination of such diagnostic measures is, however, beyond the scope of the work in this chapter.

³For example, I will introduce an example of such a measure, the *process purity*, in Ch. 10.

Prior work

The main contribution of this investigation is to comprehensively evaluate many plausible error measures for quantum information processing, within the broad framework of the criteria we have identified. We are not aware of any other work that has considered such a wide range of error measures and such a broad array of theoretical and experimental concerns.

Error measures for quantum teleportation have received particular attention in the prior literature, perhaps spurred by controversy over which experiments should be regarded as definitively demonstrating the teleportation effect [8]. Examples of this include Refs [9–14], and references therein. Except for Ref. [14], this work differs from ours by focussing primarily on the problem of teleportation. Reference [14] has a more general focus, but is primarily concerned with the question of when quantum information processing can be modelled classically, rather than the development of error measures.

There have also been more mathematical investigations of error measures, especially in the context of quantum communication and fault-tolerant quantum computation, e.g. [7, 15–23], and references therein. This work (which is often embedded in a larger investigation) typically focusses only on measures of specific interest for the problem at hand. Unlike our work, the authors do not attempt a comprehensive survey of possible error measures against some set of abstract criteria; nor do they typically address experimental criteria such as ease of measurement. Nonetheless, while this work is different in character from ours, it greatly informed our point of view and will be acknowledged in the appropriate places. Of particular relevance is Ref. [7], which introduced a key measure, the stabilised process distance, or S distance (referred to as the diamond norm in Ref. [7]), and emphasised some of its important properties.

Structure of the chapter

I begin with a summary of relevant background information about distance measures for quantum states [Sec. 9.2], while quantum processes are described in some detail in Sec. 2.1.3.

Section 9.3 contains the core of the work in the form of a comprehensive survey of possible approaches to the definition of error measures. Using our list of criteria, we are able to reject certain seemingly plausible candidate measures. This process of elimination has considerable benefit, by providing confidence that the few measures we identify as promising should be preferred over all other measures. Indeed, we eliminate all but four of the measures we define: the *Jamiolkowski process fidelity* (J fidelity), the *Jamiolkowski process distance* (J distance), the *stabilised process fidelity* (S fidelity), and the *stabilised process distance* (S distance). In several instances, we also show that error measures proposed previously in the literature (in one case, by M.A. Nielsen, one of the authors of our paper on this work) should be rejected as inadequate.

In Sec. 9.4, we apply the four promising measures identified in Sec. 9.3 to the concrete problem of quantum computation, showing that each measure has a useful operational

interpretation in terms of the success or failure of a quantum computation.

I conclude the chapter with a summary of our results, and a discussion and comparison of the properties of the measures that are the most attractive candidates for use as a gold standard in quantum information processing. We do not make a final recommendation as to which of these measures should be used, but we do make definite recommendations regarding the reporting of quantum information processing experiments. Finally, we sketch future research directions which may make it possible to definitively choose a single measure as a gold standard.

Before launching into the main body of the chapter, I will make a brief comment on our notational conventions. We often use notation like ψ to denote either a pure state $|\psi\rangle$ or the corresponding density matrix $|\psi\rangle\langle\psi|$, with the meaning to be determined from context. Thus, for example, we may write $\psi = \alpha|0\rangle + \beta|1\rangle$ to indicate a pure state of a single qubit, while also writing $\mathcal{E}(\psi)$ to indicate a quantum operation \mathcal{E} acting on the density matrix corresponding to that pure state.

9.2 Distance measures for quantum states

A natural starting place for defining distance measures for quantum processes is to look at distance measures for quantum states. The quantum information science community has identified the *trace distance* and the *fidelity* as particularly important approaches in the case of states, and these two measures will serve as the basis for our definitions for quantum operations. In keeping with the aims of the work, we do not make a choice between the trace distance and the fidelity at the outset. The following sections provide a brief review of the basic properties of the trace distance and the fidelity⁴.

The trace distance

The *trace distance* between density matrices ρ and σ is defined by $D(\rho, \sigma) \equiv \frac{1}{2} \text{Tr} \{|\rho - \sigma|\}$, where $|X| \equiv \sqrt{X^\dagger X}$. From this definition it follows that the trace distance is a genuine metric on quantum states, with $0 \leq D \leq 1$ ([4], p. 406). The trace distance also has many other attractive properties as a measure of distance between quantum states.

First, the trace distance has a compelling physical interpretation in terms of state distinguishability. Suppose Alice prepares a quantum system in the state ρ with probability $\frac{1}{2}$, and in the state σ with probability $\frac{1}{2}$. She gives the system to Bob, who performs a POVM measurement [4] to distinguish the two states. It can be shown that Bob's probability of correctly identifying which state Alice prepared is $1/2 + D(\rho, \sigma)/2$. That is, $D(\rho, \sigma)$ can be interpreted, up to the factor $1/2$, as the optimal *bias* in favour of Bob correctly determining which of the two states was prepared. This physical interpretation follows from the identity⁵ $D(\rho, \sigma) = \max_{E \leq I} \text{Tr} \{E(\rho - \sigma)\}$, where the maximum is over all positive operators E satisfying $E \leq I$.

⁴See also Ch. 9 of Ref. [4].

⁵Eq. (9.22) in Ref. [4].

The second feature of the trace distance is the *contractivity* property [24] which describes the fact that $D(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \leq D(\rho, \sigma)$ whenever \mathcal{E} is a trace-preserving quantum operation. This statement expresses the physical fact that a quantum process acting on two quantum states cannot increase their distinguishability. Contractivity follows from the physical interpretation of $D(\rho, \sigma)$ described above.

Third, the trace distance is *doubly convex*, i.e. if p_j are probabilities then

$$D\left(\sum_j p_j \rho_j, \sum_j p_j \sigma_j\right) \leq \sum_j p_j D(\rho_j, \sigma_j). \quad (9.1)$$

This inequality can be physically interpreted to mean that the distinguishability between the states $\sum_j p_j \rho_j$ and $\sum_j p_j \sigma_j$, where j is not known, can never be greater than the average distinguishability when j is known, but has been chosen at random according to the distribution p_j .

Fidelity

The *fidelity* between density matrices ρ and σ is defined by

$$F(\rho, \sigma) \equiv \text{Tr} \left\{ \sqrt{\sqrt{\rho} \sigma \sqrt{\rho}} \right\}^2. \quad (9.2)$$

When $\rho = |\psi\rangle\langle\psi|$ is a pure state, this reduces to $F(\psi, \sigma) = \langle\psi|\sigma|\psi\rangle$, the overlap between ψ and σ .

The fidelity also has many attractive properties. It can be shown that $0 \leq F(\rho, \sigma) \leq 1$, with equality in the second inequality if and only if $\rho = \sigma$. The fidelity is thus not a metric as such, but serves rather as a generalised measure of the overlap between two quantum states. The fidelity is also symmetric in its inputs, $F(\rho, \sigma) = F(\sigma, \rho)$, a fact that is not obvious from the definition we have given, but which follows from other equivalent definitions.

There is an ambiguity in the literature in the definition of fidelity that is worth commenting on here [see also Sec. 2.1.4]. Both the quantity defined above and its square root have been referred to as the fidelity, and both have many appealing properties⁶.

Nevertheless, we strongly advocate using the definition of Eq. (9.2), despite the other definition being used in references such as [4]. In Sec. 9.4, we show that adopting the definition of Eq. (9.2) gives rise to a measure of distance between quantum processes with a physically compelling interpretation in terms of the *probability of success* of a quantum computation [cf. Sec. 2.1.4]. Adopting the other definition of fidelity would make about as much sense as reporting the square root of the probability that the quantum computation succeeded.

Although not in itself a metric, the fidelity can easily be converted into one. Two common derived metrics are the *Bures metric*, defined by $B(\rho, \sigma) \equiv \sqrt{2 - 2\sqrt{F(\rho, \sigma)}}$, and the

⁶It is noteworthy that using the square root of the definition in Eq. (9.2) results in “better” values for the fidelity, and this may account for some of the attractiveness of that quantity. Needless to say, this is not a compelling reason for using it.

angle, defined by $A(\rho, \sigma) \equiv \arccos \sqrt{F(\rho, \sigma)}$. The origin of these metrics can be seen intuitively by considering both ρ and σ to be pure states. The Bures metric is just the Euclidean distance between the two pure states, with respect to the usual norm on state space⁷, while the angle is, as the name suggests, just the angle between the two states, with respect to the usual inner product on state space.

In addition to the *angle* and the Bures metric, it is useful to introduce a third example, $C(\rho, \sigma) \equiv \sqrt{1 - F(\rho, \sigma)}$. We have not found this metric in previous literature, but it arises naturally later in the chapter in the context of quantum computation. The only difficult step in verifying $C(\rho, \sigma)$ to be a metric is to prove it satisfies the triangle inequality⁸.

The discussion in later sections will sometimes focus on the fidelity, and sometimes on metrics derived from the fidelity. We define a metric $\Delta^F(\rho, \sigma)$ on state space to be a *fidelity-based* metric if it is a monotonically decreasing function of the fidelity $F(\rho, \sigma)$. Obviously the angle, the Bures metric and $C(\cdot, \cdot)$ are all fidelity-based metrics. It is often the case that the specific details of the metric used are not important, and whenever possible our results will be stated directly in terms of the fidelity to provide a single unifying concept. Sometimes, however, it will be necessary to use the fidelity-based metrics directly. For example, unlike the fidelity, they satisfy the triangle inequality, which will be useful in proving the chaining criterion (property 6).

Like the trace distance, the fidelity and its derived metrics have many other nice properties. It can be shown [26] that $F(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \geq F(\rho, \sigma)$ for any trace-preserving quantum operation \mathcal{E} . We call this the *monotonicity* property of the fidelity. It follows that any fidelity-based metric satisfies a contractivity property analogous to that satisfied by the trace distance.

The fidelity also satisfies a property analogous to the double convexity of the trace distance. Specifically, the square root of the fidelity is *doubly concave*, i.e.,

$$F\left(\sum_j p_j \rho_j, \sum_j p_j \sigma_j\right)^{1/2} \geq \sum_j p_j F(\rho_j, \sigma_j)^{1/2}. \quad (9.3)$$

This double concavity can be used to prove that certain fidelity-based metrics are doubly convex⁹. In particular, supposing Δ^F is a fidelity-based metric which is convex in the square root of the fidelity (the angle, the Bures metric and $C(\cdot, \cdot)$ all satisfy this), then it is easy to verify that Δ^F is doubly convex.

One drawback of the fidelity is that it is difficult to find a compelling physical interpretation. When ρ and σ are mixed states, no completely satisfactory interpretation

⁷Strictly, this statement is only true in projective state space, in which states which are the same up to an overall phase factor are regarded as identical.

⁸To prove the triangle inequality for $C(\rho, \sigma)$: according to Uhlmann's theorem (see Ref. [25], or Theorem 9.4 on p410 of Ref. [4]), given any three states ρ, σ and τ , there exist corresponding purifications $|r\rangle, |s\rangle$ and $|t\rangle$ such that $F(\rho, \sigma) = |\langle r|s\rangle|^2$, $F(\rho, \tau) = |\langle r|t\rangle|^2$, and $F(\sigma, \tau) = |\langle s|t\rangle|^2$. Let θ_{rs} be an angle in the range 0 to $\pi/2$ so that $\cos(\theta_{rs}) = |\langle r|s\rangle|$. Define θ_{rt} and θ_{st} similarly. With these definitions it is clear that $C(\rho, \sigma) = \sin(\theta_{rs})$, with similar relations holding for the other pairs of density matrices, and so the triangle inequality reduces to proving $\sin(\theta_{rt}) \leq \sin(\theta_{rs}) + \sin(\theta_{st})$. But elementary geometry implies that $\theta_{rt} \leq \theta_{rs} + \theta_{st}$, and simple algebra can be used to deduce the triangle inequality from that.

⁹The fidelity-based metrics are "inverted" with respect to the fidelity itself, in that for two identical states, $\rho = \sigma$, $F(\rho, \sigma)=1$ and $\Delta^F(\rho, \sigma)=0$.

of the fidelity is known (but cf. Refs [27, 28]). When $\rho = \psi$ is a pure state, then $F(\psi, \sigma) = \langle \psi | \sigma | \psi \rangle$, the overlap between ψ and σ . Physically, one could imagine σ is an attempt to prepare the pure state ψ , and the fidelity is then the probability that a perfect measurement of the state will find the result ψ . We use this property in Sec. 9.4 to connect our fidelity-based error measures for quantum processes to the probability of success of a quantum computation.

General comments

The fidelity is, at present, perhaps somewhat more widely used in the quantum information science community than is the trace distance. However, the trace distance and the fidelity have complementary advantages as a basis for developing distance measures for quantum operations, and so it is useful to investigate both (this will be illustrated at various points of this chapter). In any case, the two measures are, as one might expect, quite closely related. In particular, it is possible to verify the inequalities [29]:

$$1 - \sqrt{F(\rho, \sigma)} \leq D(\rho, \sigma) \leq \sqrt{1 - F(\rho, \sigma)}, \quad (9.4)$$

and it is not difficult to construct examples of saturation for both inequalities. Note that the second inequality is always saturated for pure states, i.e. $D(\psi, \phi) = \sqrt{1 - F(\psi, \phi)}$ for pure ψ and ϕ .

9.3 Error measures for quantum processes

The main goal of this work is to recommend a single error measure enabling researchers to compare the performance of quantum information processing experiments against the theoretical ideal. To provide the basis for such a recommendation, we have comprehensively surveyed possible definitions of such error measures, assessing each against the criteria introduced above.

We considered three fundamentally different approaches to defining an error measure for processes. Section 9.3.1 considers measures based on the process matrix, $\rho_{\mathcal{E}}$, and is followed by a description of the *average* behaviour of a process [Sec. 9.3.2] and the *worst-case* behaviour of a process [Sec. 9.3.3]. In each case, we appraised measures based on both the trace distance and the fidelity, describing connections between the various measures, and identifying four of particular merit. Their properties will be discussed in more detail in the next section.

Nomenclature

In the following treatment, the unadorned symbol Δ will refer to a metric between states. In our approach we have used state-based metrics to form metrics between processes, and these will also be represented by Δ but with a subscript denoting the method used, e.g. Δ_{ave} is a process metric based on the average over input states. Where we have specialised

to a specific state-metric (e.g. A , B , C , and D from section 9.2), we have represented the related process metric by the symbol for the state metric with a subscript for the method, e.g. D_{ave} is the *process* metric based on the average trace distance. In contrast, since the fidelity is not itself a metric, Δ^F will refer to any *metric* derived from the fidelity (e.g. A , B , and C), and the symbol F with a subscript will represent a process measure based on fidelity, e.g. F_{ave} is the average fidelity.

9.3.1 Error measures based on the process matrix

Suppose $\Delta(\rho, \sigma)$ is any metric on the space of quantum states. A natural approach to defining a measure Δ_{pro} of the distance between two quantum processes is to calculate the distance between the two process matrices using a state-based metric:

$$\Delta_{\text{pro}}(\mathcal{E}, \mathcal{F}) \equiv \Delta(\rho_{\mathcal{E}}, \rho_{\mathcal{F}}). \quad (9.5)$$

Defining Δ_{pro} in this way automatically gives Δ_{pro} the metric property. Provided $\Delta(\cdot, \cdot)$ is easy to calculate, Δ_{pro} is also easy to calculate. Furthermore, since \mathcal{E} can be experimentally determined using quantum process tomography, it follows that Δ_{pro} can also be measured, at least in principle.

What about the other properties? The properties of stability and chaining can be obtained by making the following natural assumptions about the state metric Δ . Suppose first that the metric Δ is *stable* in the sense that $\Delta(\rho \otimes \tau, \sigma \otimes \tau) = \Delta(\rho, \sigma)$, which, for example, can easily be verified for the trace distance and for any fidelity-based metric. The stability property for the process metric, Δ_{pro} , then follows immediately:

$$\Delta_{\text{pro}}(I \otimes \mathcal{E}, I \otimes \mathcal{F}) = \Delta(\rho_I \otimes \rho_{\mathcal{E}}, \rho_I \otimes \rho_{\mathcal{F}}) = \Delta(\rho_{\mathcal{E}}, \rho_{\mathcal{F}}) = \Delta_{\text{pro}}(\mathcal{E}, \mathcal{F}). \quad (9.6)$$

The chaining property can be proved, with some caveats to be described below, by assuming that $\Delta(\cdot, \cdot)$ is *contractive*, i.e. $\Delta(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \leq \Delta(\rho, \sigma)$, for trace-preserving operations \mathcal{E} . As already mentioned in Sec. 9.2, this is a natural, physical assumption which is satisfied by the trace distance and any fidelity-based metric.

Recall that the chaining criterion requires that

$$\Delta(\mathcal{E}_2 \circ \mathcal{E}_1, \mathcal{F}_2 \circ \mathcal{F}_1) \leq \Delta(\mathcal{E}_1, \mathcal{F}_1) + \Delta(\mathcal{E}_2, \mathcal{F}_2). \quad (9.7)$$

To prove that this is satisfied by Δ_{pro} , we need to assume that \mathcal{F}_1 is *doubly stochastic*, which means that \mathcal{F}_1 is trace-preserving and satisfies $\mathcal{F}_1(I) = I$. This may seem like a significant assumption, since physical processes such as relaxation to a finite temperature are not doubly stochastic. However, in most typical quantum information processing tasks, \mathcal{F}_1 and \mathcal{F}_2 are ideal unitary processes, and the process metric, Δ_{pro} , compares the composition of these two ideal processes to the experimentally realised process $\mathcal{E}_2 \circ \mathcal{E}_1$. Since unitary processes are automatically doubly stochastic, the assumption is normally justified.

The proof of chaining begins by applying the triangle inequality to obtain

$$\Delta_{\text{pro}}(\mathcal{E}_2 \circ \mathcal{E}_1, \mathcal{F}_2 \circ \mathcal{F}_1) = \Delta(\rho_{\mathcal{E}_2 \circ \mathcal{E}_1}, \rho_{\mathcal{F}_2 \circ \mathcal{F}_1}) \quad (9.8)$$

$$\leq \Delta(\rho_{\mathcal{E}_2 \circ \mathcal{E}_1}, \rho_{\mathcal{E}_2 \circ \mathcal{F}_1}) + \Delta(\rho_{\mathcal{E}_2 \circ \mathcal{F}_1}, \rho_{\mathcal{F}_2 \circ \mathcal{F}_1}). \quad (9.9)$$

Then note the easily verified identity

$$\rho_{\mathcal{E} \circ \mathcal{F}} = [\mathcal{F} \circ \mathcal{E}](\Phi) = [\mathcal{F}^T \otimes \mathcal{E}](\Phi), \quad (9.10)$$

where $|\Phi\rangle = \frac{1}{\sqrt{d}} \sum |\mathbf{j}\rangle_a |\mathbf{j}\rangle_b$ is the maximally entangled state introduced in Sec. 2.1.3, and $\mathcal{F}^T(\rho) \equiv \sum_j F_j^T \rho F_j^*$, if F_j are the operation elements for \mathcal{F} [cf. Eq. (2.22)]. Applying this identity to both density matrices in the second term on the right-hand side of Eq. (9.9) gives

$$\Delta_{\text{pro}}(\mathcal{E}_2 \circ \mathcal{E}_1, \mathcal{F}_2 \circ \mathcal{F}_1) \leq \Delta(\rho_{\mathcal{E}_2 \circ \mathcal{E}_1}, \rho_{\mathcal{E}_2 \circ \mathcal{F}_1}) + \Delta([\mathcal{F}_1^T \otimes \mathcal{E}_2](\Phi), [\mathcal{F}_1^T \otimes \mathcal{F}_2](\Phi)). \quad (9.11)$$

The double stochasticity of \mathcal{F}_1 implies that \mathcal{F}_1^T is a trace-preserving quantum operation. Finally, the desired result is achieved by applying contractivity to both the first and the second terms on the right-hand side of Eq. (9.11).

The only remaining question about Δ_{pro} is whether it has a good physical interpretation. In fact, we have shown that it does [see Sec. 9.4] by relating both D_{pro} and F_{pro} in a natural way to the average probability of success for a quantum computation.

As mentioned above, the process metric, Δ_{pro} , is easy to calculate provided the corresponding state metric $\Delta(\cdot, \cdot)$ is easy to calculate, which is certainly true for both the trace distance and fidelity-based metrics. However, it seems that Δ_{pro}^F may be even simpler to determine than the trace distance metric, D_{pro} . To the best of our knowledge, measuring D_{pro} requires full process tomography, involving the estimation of at least $d^4 - d^2$ observable averages for a d -dimensional quantum system [see Sec. 3.5]. In contrast, when the target operation is a unitary, U , we have shown that the fidelity $F_{\text{pro}}(\mathcal{E}, U)$ (and related error measures) can be determined by estimating at most $2d^2$ observable averages, in general, and d^2 observable averages for qubits, in particular. This makes $F_{\text{pro}}(\mathcal{E}, U)$ and its related error measures substantially easier to calculate experimentally than D_{pro} . The key to the proof (given below) is the observation [16],

$$F_{\text{pro}}(\mathcal{E}, U) = \frac{1}{d^3} \sum_j \text{Tr} \left\{ U U_j^\dagger U^\dagger \mathcal{E}(U_j) \right\}, \quad (9.12)$$

where the $\{U_j\}$ are a basis of unitary operators which are orthogonal under the Hilbert-Schmidt inner product, i.e. satisfying $\text{Tr}\{U_j^\dagger U_k\} = d\delta_{jk}$. Up to a scaling factor, an example of such a set is the n -qubit tensor products of the Pauli matrices, $\{\sigma_j \otimes \dots \otimes \sigma_k\}$ [see Sec. 2.1.2].

Start by expanding the U_j in terms of a set of d^2 *input states*, ρ_k : $U_j = \sum_k a_{jk} \rho_k$, which span the entire operator space (an explicit example is given below for two qubits), and expand the $U U_j U^\dagger$ in terms of a complete set of *observables*, σ_l : $U U_j U^\dagger = \sum_l b_{jl} \sigma_l$. Substituting into Eq. (9.12) gives

$$F_{\text{pro}}(\mathcal{E}, U) = \frac{1}{d^3} \sum_{kl} M_{kl} \text{Tr} \{ \sigma_l \mathcal{E}(\rho_k) \}, \quad (9.13)$$

where $M_{kl} \equiv \sum_j b_{jl} a_{jk}$. This equation provides a general method for evaluating F_{pro} . Choose a spanning set of d^2 input states, ρ_k , which can be prepared experimentally, and a set of observables, σ_l , which can be reliably measured. Then determine the matrix

$M = (M_{kl})$, which depends only on known quantities (ρ_k, σ_l , and the idealised operation U), and not on the unknown \mathcal{E} . The non-zero matrix elements in M determine which observable averages must be estimated in order to calculate $F_{\text{pro}}(\mathcal{E}, U)$.

In general, d^4 observable averages will be required, which is clearly no improvement on quantum process tomography. However, by choosing some fixed set of ρ_k , and defining¹⁰ $\sigma_l \equiv \sum_k a_{kl} U U_k U^\dagger$, then Eq. (9.13) simplifies to:

$$F_{\text{pro}}(\mathcal{E}, U) = \frac{1}{d^3} \sum_k \text{Tr} \{ \sigma_k \mathcal{E}(\rho_k) \}, \quad (9.14)$$

which only requires between d^2 and $2d^2$ measurements. The drawback to this method is that there is no freedom to choose the σ_l ; they are determined by U and the ρ_k .

In practical situations, however, some input states and measurements are more accessible than others. Therefore, we envisage that an experimentalist will choose the most convenient set of input states and measurements and use the prescription above to determine which combinations are necessary. In general, this will be less than the number required to perform full process tomography. As an added bonus, it is much easier to estimate the experimental error in F_{pro} with this direct method than with full tomographic analysis.

To illustrate the procedure, consider an n -qubit process, U . Suppose the U_j are just the n -fold tensor products of Pauli matrices, as described above. Furthermore, for each qubit, let the set of input states be $\{I, I + X, I + Y, I + Z\}$ (where X, Y, Z are the usual Pauli operators), so that the ρ_k are just the set of all possible tensor products of the single-qubit input states. These definitions ensure that the a_{kl} will always be real, and since the U_k are Hermitian, choosing $\sigma_l \equiv \sum_k a_{kl} U U_k U^\dagger$ ensures that the σ_l are also Hermitian—i.e. they are themselves observables. Thus Eq. (9.14) shows that only d^2 measurements of observables are required to evaluate $F_{\text{pro}}(\mathcal{E}, U)$ for *any* U . This is much fewer ($d^2 - 1$ times fewer) than the $d^4 - d^2$ observable averages which must be estimated in full n -qubit process tomography.

The only drawback in the above example is that some of the required measurements are likely to be entangled. Therefore, an interesting problem deserving further exploration is to find the minimum number of measurements required to estimate F_{pro} when there are constraints on what input states and observables are available. For instance, it would be useful to know the optimal number in an experiment where one is restricted to separable inputs and tensor-product observables, i.e. inputs and observables that can be given direct local implementations.

9.3.2 Error measures based on the average case

Another natural way to define error measures for quantum operations is to compare output states using quantum state metrics [Sec. 9.2] and average over all input states, i.e.,

$$\Delta_{\text{ave}}(\mathcal{E}, \mathcal{F}) \equiv \int d\psi \Delta(\mathcal{E}(\psi), \mathcal{F}(\psi)), \quad (9.15)$$

¹⁰Note that the σ_l defined in this way are not, in general, Hermitian, and thus may not be observables. However, they are easily split up into $\sigma_l = A_l + iB_l$, where A_l and B_l are Hermitian. Thus measuring $\text{Tr} \{ \sigma_l \rho \}$ can be achieved by simply measuring two separate observables.

where the integral is over the uniform (Haar) measure on state space.

While this approach seems sensible intuitively, the resulting measures satisfy few of the desired criteria. It appears that the only two properties they satisfy in general (for an arbitrary state metric Δ), are the metric and chaining criteria, both of which follow immediately from the metric property of Δ . With the other criteria, they are less successful. Even when $\Delta(\mathcal{E}(\psi), \mathcal{F}(\psi))$ is easy to calculate, it is not obvious that the full integral in Eq. (9.15) will have the simple form which would make the calculation of Δ_{ave} similarly easy. Moreover, this also means that Δ_{ave} may not be easy to determine experimentally. To date, we do not know of any simple expressions for Δ_{ave} for any of the state metrics we have considered.

Not surprisingly, the physical interpretations of these metrics rely heavily on the possible interpretations of the corresponding state metrics, which were discussed Sec. 9.2. The earlier discussion of the trace distance, for example, extends to give a meaning for D_{ave} . Suppose one needs to distinguish between $\mathcal{E}(\psi)$ and $\mathcal{F}(\psi)$ for some ψ which is known, but has been chosen uniformly at random. On average, the optimal probability of successfully distinguishing the two processes will be $1/2 + D_{\text{ave}}(\mathcal{E}, \mathcal{F})/2$. Thus, $D_{\text{ave}}(\mathcal{E}, \mathcal{F})$ may be interpreted as a measure of the average bias in favour of correctly distinguishing which process was applied to a state ψ . However, for the fidelity-based average metrics, Δ_{ave}^F , there does not appear to be any clear physical interpretation because there is no clear meaning for the corresponding state metrics.

Finally, completing the checklist of criteria, numerical analysis shows that Δ_{ave} is not naturally stable for any of the four candidate state metrics we have investigated, but there is a method for “stabilising” measures which are not stable, described in detail in Sec. 9.3.3. For now, the basic idea is to introduce an ancillary system A , and define the *stabilised* quantity $\Delta_{\text{stab-ave}}(\mathcal{E}, \mathcal{F}) \equiv \lim \Delta_{\text{ave}}(\mathcal{I} \otimes \mathcal{E}, \mathcal{I} \otimes \mathcal{F})$ which is calculated in the limit of large ancilla dimension. Using the result that a randomly chosen state of a composite system AQ ($\dim A \gg \dim Q$) has very close to maximal entanglement [30, 31], it follows that $\Delta_{\text{stab-ave}}(\mathcal{E}, \mathcal{F}) = \Delta_{\text{pro}}(\mathcal{E}, \mathcal{F})$, i.e. the stabilised average distance reduces to the process distance considered earlier.

An alternative approach is to create a measure based on the average fidelity:

$$F_{\text{ave}}(\mathcal{E}, \mathcal{F}) \equiv \int d\psi F(\mathcal{E}(\psi), \mathcal{F}(\psi)). \quad (9.16)$$

This is a genuinely different approach, because the fidelity-based metrics are nonlinear functions of the fidelity. When \mathcal{F} is a unitary operation, U , the average fidelity has a physical interpretation that is at least plausible, as the average overlap between $U|\psi\rangle$ and $\mathcal{E}(\psi)$. In this case, it was shown in Ref. [32] (see also Ref. [16]) that F_{ave} and F_{pro} are related by the equation

$$F_{\text{ave}}(\mathcal{E}, U) = \frac{F_{\text{pro}}(\mathcal{E}, U)d + 1}{d + 1}, \quad (9.17)$$

where d is the dimension of the quantum system. This relationship makes $F_{\text{ave}}(\mathcal{E}, U)$ easy to calculate [16, 17] and also easy to measure experimentally, using the techniques described for $F_{\text{pro}}(\mathcal{E}, U)$.

Nevertheless, although F_{ave} has several advantages (ease of calculation, ease of measurement, and a physical interpretation), the outlook for the other criteria is not so good. Not only is F_{ave} not a metric, but it is not stable either, a fact that follows from Eq. (9.17) and the knowledge that F_{pro} is stable. By the same argument, measures analogous to A , B , and C based on F_{ave} are also not stable. We do not know of any stable metrics that may be derived as a function of F_{ave} , and since Eq. (9.17) renders any such metrics equivalent in content to functions based on F_{pro} , they do not provide any advantages over the Δ_{pro}^F .

In summary, despite the fact that the above approaches are fairly obvious candidates for a plausible error measure, none of the average-case error measures are particularly attractive. It was, however, important to scrutinise them carefully before choosing to reject them.

9.3.3 Error measures based on the worst case

Our final approach to defining process error measures is based on the worst case distance between the output states, $\mathcal{E}(\psi)$ and $\mathcal{F}(\psi)$,

$$\Delta_{\text{max}}(\mathcal{E}, \mathcal{F}) \equiv \max_{\psi} \Delta(\mathcal{E}(\psi), \mathcal{F}(\psi)), \quad (9.18)$$

where the maximum is over all pure input states, ψ , and Δ is a metric on quantum states.

When $\Delta = \Delta^F$ is a fidelity-based metric, then Δ_{max}^F is a function of the *minimal fidelity*,

$$F_{\text{min}}(\mathcal{E}, \mathcal{F}) \equiv \min_{\psi} F(\mathcal{E}(\psi), \mathcal{F}(\psi)). \quad (9.19)$$

Before proceeding further, is it sufficient to optimise over all *pure* input states? In fact, it is fairly simple to show that this gives the same value achieved by optimising over *all* physical inputs, including mixed states. It is therefore unnecessary to consider mixed states, which would make the task significantly harder¹¹. Suppose Δ is a *doubly convex* metric, as are all the metrics discussed in this chapter (cf. Sec. 9.2). If the maximum is achieved at some mixed state, ρ , then $\Delta_{\text{max}} = \Delta(\mathcal{E}(\rho), \mathcal{F}(\rho))$. Expanding $\rho = \sum_j p_j \psi_j$ as a mixture of pure states, and applying double convexity gives

$$\Delta_{\text{max}} \leq \sum_j p_j \Delta(\mathcal{E}(\psi_j), \mathcal{F}(\psi_j)) \leq \max_j \Delta(\mathcal{E}(\psi_j), \mathcal{F}(\psi_j)) \sum_j p_j = \Delta(\mathcal{E}(\psi_j^{\text{max}}), \mathcal{F}(\psi_j^{\text{max}})). \quad (9.20)$$

This shows that the maximum must also be attained at some pure state ψ_j^{max} . A similar argument holds for F_{min} , based on the double concavity of the fidelity.

The next task is to assess these measures against our list of criteria. It has already been shown that the worst-case trace distance, D_{max} , is not stable in general [7], and similar arguments show that this also applies to the fidelity-based measures. In Ref. [7], Aharonov *et al.* resolve this difficulty by constructing a variant of D_{max} which is stable,

¹¹Obviously, maximising over pure states will be numerically easier, as it involves only $2d - 2$ real parameters for a d -dimensional state space, instead of the $d^2 - 1$ required for normalized mixed states.

and otherwise has extremely similar properties to D_{\max} . In a similar way, a stable version of Δ_{\max} can be defined for an arbitrary state metric Δ .

Suppose \mathcal{E} and \mathcal{F} act on the d -dimensional system, Q (subscripts will indicate the system on which a process operates, e.g. $\mathcal{E} = \mathcal{E}_Q, \mathcal{F} = \mathcal{F}_Q$). By introducing a fictitious d -dimensional ancillary system A , which is left alone (i.e. acted on by the identity operation, \mathcal{I}_A), a stabilised quantity can be defined,¹²

$$\Delta_{\text{stab}}(\mathcal{E}_Q, \mathcal{F}_Q) \equiv \Delta_{\max}(\mathcal{I}_A \otimes \mathcal{E}_Q, \mathcal{I}_A \otimes \mathcal{F}_Q). \quad (9.21)$$

The proof that Δ_{stab} is stable under the addition of ancillary systems is simple and has been included in Appendix 9.A.1. Using a similar definition (and proof of stability), there is also a stable form of the minimum fidelity, $F_{\text{stab}}(\mathcal{E}_Q, \mathcal{F}_Q) \equiv F_{\min}(\mathcal{I}_A \otimes \mathcal{E}_Q, \mathcal{I}_A \otimes \mathcal{F}_Q)$. Note that the stabilised fidelity-based metrics Δ_{stab}^F are functions of F_{stab} in the obvious way (e.g. we define as usual $A_{\text{stab}}, B_{\text{stab}}$ and C_{stab}).

Which of the other criteria does Δ_{stab} satisfy? It is straightforward to show that Δ_{stab} satisfies the metric and chaining criteria. Furthermore, the stabilised trace-distance D_{stab} has an appealing physical interpretation—it is the worst-case bias in the probability of being able to distinguish $[\mathcal{I} \otimes \mathcal{E}](\psi)$ from $[\mathcal{I} \otimes \mathcal{F}](\psi)$, with an ancilla of arbitrary size. The physical interpretation of the fidelity-based measures will be discussed in the next section, where, along with D_{stab} , we have shown they have elegant interpretations in the context of quantum computation.

What of the remaining criteria, ease of calculation and ease of measurement? Unfortunately, no powerful general formulae for calculating Δ_{stab} are known. Reference [7] gives a general formula for the distance D_{stab} between two unitary operations, but the more interesting case of the distance between an idealised unitary operation and a noisy quantum process has not been solved, even for single-qubit operations.

Fortunately, however, D_{stab} and F_{stab} (and thus $A_{\text{stab}}, B_{\text{stab}}$ and C_{stab}) are easy to calculate numerically, because they can all be reduced to *convex optimisation* problems [33] [see App. 9.A.2 for the proof for F_{stab} (the proof for D_{stab} is similar)]. For this special class of problems, the task is to minimise a convex function defined on a convex set, and extremely efficient numerical solutions can be implemented. Among many other nice properties, it is possible to show that a local minimum of a convex optimisation problem is always a global minimum, and thus techniques such as gradient descent typically converge extremely rapidly, with no danger of finding false minima.

Since D_{stab} and F_{stab} can be calculated numerically with relative ease, this enables a two-step procedure for measuring either quantity experimentally—process tomography, followed by a numerical optimisation. Of course, finding general formulae along the lines of $F_{\text{pro}}(\mathcal{E}, U)$ or D_{pro} is still a highly desirable goal. Aside from the intrinsic benefit, finding general formulae would also simplify the estimation of experimental errors in D_{stab} and F_{stab} , and perhaps obviate the need for a full process tomography, as Eq. (9.12) did for $F_{\text{pro}}(\mathcal{E}, U)$.

¹²When Δ is the trace distance, Aharonov *et al.* [7] refer to this as the diamond norm.

9.4 Application to quantum computing

Is there a good, generally applicable physical interpretation for any of the error measures that we have identified? In this section we have considered interpretations that arise within the context of quantum computation and shown that four of the error measures discussed have particularly outstanding properties: D_{pro} , F_{pro} , D_{stab} and F_{stab} . (In the case of the fidelity, it will be more convenient to state the results in terms of the equivalent measures, C_{pro} and C_{stab} .)

Based on the criteria described in the introduction, these four measures already stand out as superior to all the other measures we have investigated. That each also arises naturally in the context of quantum computation, strongly indicates that these four measures should be the main candidates for error measures in quantum information processing. The conclusion [Sec. 9.5] then discusses which of these may be the best.

There are a variety of ways of describing quantum computations, and each of the four error measures arises naturally in different contexts. We have considered two broad divisions of quantum computation, *function computation* and *sampling computation*, and both worst-case and average-case performance in each.

Most algorithms on classical computers are framed as function computations, and the four main error measures can be given particularly compelling interpretations relating to the probability of error in a function computation. However, in the context of simulating quantum systems it is often more natural to consider sampling computations, where the goal is to reproduce the statistics obtained from a measurement of the system in some specified configuration. Again, all four error measures can be given good interpretations in this context, if somewhat more complex than for function computation.

The reason for treating the two types of computation separately is at least partially practical, since both types of computation arise naturally in the context of quantum computation. More fundamentally, however, it is not yet known how to reduce sampling computation to function computation. Rather remarkably, even when there is an efficient way of *computing* a probability distribution, it does not seem possible, in general, to convert that into an efficient way of *sampling* from that distribution.

9.4.1 Function computation

In function computation, the goal of the quantum computation is to *compute a function*, f , exactly or with high probability of success. More precisely, the goal is to take as input an instance, x , of the problem, and to produce a final state ρ_x of the computer that is either equal to $|f(x)\rangle$, or sufficiently close that when a measurement in the computational basis is performed, the outcome is $f(x)$ with high probability¹³. Grover's algorithm [34] is usually cast in this way, where the target is to determine the identity of the state marked by the oracle.

¹³Both x and $f(x)$ are elements of the input and output computational space.

Function computation in the worst case

Suppose that one wishes to perform a quantum computation represented by an ideal operation \mathcal{F} that acts on an input $|x\rangle$, where x represents the instance of the problem to be solved, e.g. a number to be factored¹⁴. This process succeeds in computing $f(x)$ with an error probability of at most p_e^{id} , where ‘id’ indicates that this is the *ideal* worst-case error probability. Of course, in reality some non-ideal operation, \mathcal{E} , is performed. A good measure of error in the real computation is the *actual* probability p_e that the measured output of the computation is not equal to $f(x)$. Appendix 9.B.1 shows that

$$p_e \leq p_e^{\text{id}} + D_{\text{stab}}(\mathcal{E}, \mathcal{F}) \quad (9.22)$$

$$p_e \leq \left[\sqrt{p_e^{\text{id}}} + C_{\text{stab}}(\mathcal{E}, \mathcal{F}) \right]^2. \quad (9.23)$$

Which of these gives the tighter bound depends upon the exact circumstances. For example, when $p_e^{\text{id}} = 0$, it depends upon whether $D_{\text{stab}}(\mathcal{E}, \mathcal{F})$ is larger or smaller than $C_{\text{stab}}(\mathcal{E}, \mathcal{F})^2$, and with Eq. (9.4) in mind, it is not difficult to see that both cases are possible.

Function computation in the average case

Once again, the goal is to compute a function $f(x)$ using an approximation \mathcal{E} to some ideal operation \mathcal{F} . Now, however, consider the average-case error probability, \bar{p}_e , that the measured output of $\mathcal{E}(x)$ is not equal to $f(x)$, where the average is taken with respect to a uniform distribution over instances x . Correspondingly, \bar{p}_e^{id} is the average case error probability for the idealised operation \mathcal{F} . Appendix 9.B.2 shows that:

$$\bar{p}_e \leq \bar{p}_e^{\text{id}} + D_{\text{pro}}(\mathcal{E}, \mathcal{F}). \quad (9.24)$$

Unfortunately, we have been unable to develop a full natural analogue of Eq. (9.23) based on the fidelity. However, we have proved a partial analogue when the ideal computation succeeds with probability one (i.e. $\bar{p}_e^{\text{id}} = 0$). In this case,

$$\bar{p}_e \leq C_{\text{pro}}(\mathcal{E}, \mathcal{F})^2 = 1 - F_{\text{pro}}(\mathcal{E}, \mathcal{F}). \quad (9.25)$$

The proof (omitted) uses very similar techniques to those which establish Eqs (9.24) and (9.23).

9.4.2 Sampling computation

In sampling quantum computation, the goal is to *sample* from some ideal distribution $\{p_x(y)\} \equiv p_x$ of measurement outcomes y , with x representing input data for the problem. For instance, x might represent the coupling strengths and temperature of some spin glass model, with the goal being to sample from the thermal distribution of configurations y

¹⁴One can also consider a model in which the starting state of the computer is a standard state, $|s\rangle$, and the operation to be performed, \mathcal{F}_x , itself depends on the problem instance. Exactly analogous results can be proved in this alternative context.

for that spin glass. This type of computation is particularly useful for simulating the dynamics of another quantum system.

Unlike Grover’s algorithm, Shor’s algorithm [35] is usually described as a sampling computation. The goal is not to directly produce a factor or list of factors, but rather to produce a distribution over measurement outcomes. By sampling from this distribution and doing some classical post-processing it is possible to extract factors of some number x . Of course, as noted in Ref. [36], it is possible to modify Shor’s algorithm to be a function computation, taking an instance x and producing a list of all the factors of x .

The desired result in sampling computation is that the measurement outcomes y are distributed according to the *ideal* probabilities $p_x(y)$, for a given problem instance x . Suppose, however, that they are instead distributed according to some nonideal set of *real* probabilities $q_x(y)$. How should these two distributions be compared? There are two widely used classical measures for comparing probability distributions p and q . The first is the *Kolmogorov* or l_1 distance, defined by $D(p, q) \equiv \sum_y |p(y) - q(y)|/2$. The second is the *Bhattacharya overlap*, defined by $F(p, q) \equiv \sum_y \sqrt{p(y)q(y)}$. Since these measures are in fact commutative (“classical”) analogues of the trace distance and fidelity, respectively, they are represented here by the same symbols as their quantum counterpart (D and F). As with the trace distance, the Kolmogorov distance can be given an appealing interpretation as the bias in probability when trying to distinguish the distributions p and q . No similarly simple interpretation for the Bhattacharya overlap seems to be known, although it is related to the Kolmogorov distance through inequalities analogous to Eq. (9.4).

The Kolmogorov distance and Bhattacharya overlap, together with the quantum error measures we have introduced, can be used to relate ideal and real probability distributions obtained as the result of a quantum computation.

Sampling computation in the worst case

Suppose again that one wishes to perform a quantum computation represented by an ideal operation \mathcal{F} that acts on an input $|x\rangle$, where x represents the instance of the problem to be solved. The goal is to produce a final state $\mathcal{F}(x)$ which, when measured in the computational basis, gives rise to an ideal distribution p_x . Instead, one performs the operation \mathcal{E} , giving rise to a distribution q_x on measurement outcomes. In Appendix 9.B.3, we prove that:

$$\max_x D(q_x, p_x) \leq D_{\text{stab}}(\mathcal{E}, \mathcal{F}) \quad (9.26)$$

$$\max_x [1 - F(q_x, p_x)] \leq C_{\text{stab}}(\mathcal{E}, \mathcal{F})^2. \quad (9.27)$$

As before, which inequality is better depends upon the details of the situation under study.

Sampling computation in the average case

Given the same situation as for the worst case, assume now that problem instances are chosen uniformly at random. Therefore, the Kolmogorov distance and Bhattacharya

overlap between the *joint* distributions $\{p(x, y)\} \equiv p$ and $\{q(x, y)\} \equiv q$ must be used to measure how well \mathcal{E} has approximated \mathcal{F} . Arguments analogous to those used in the worst case establish:

$$D(q, p) \leq D_{\text{pro}}(\mathcal{E}, \mathcal{F}) \quad (9.28)$$

$$1 - F(q, p) \leq C_{\text{pro}}(\mathcal{E}, \mathcal{F})^2. \quad (9.29)$$

9.5 Summary, recommendations, and conclusion

We have formulated a list of criteria that must be satisfied by a good measure of error in quantum information processing. These criteria provide a broad framework that can be used to assess candidate measures, incorporating both theoretical and experimental desiderata.

We have used this framework to comprehensively survey possible approaches to the definition of an error measure, rejecting many seemingly plausible error measures that fail to satisfy many of our criteria. Although many of these rejected error measures are of some interest as diagnostic measures, none are suitable for use as a *primary* measure of the error in a quantum information processing task.

Four error measures were identified to have particular merit, each of which satisfies most or all of the criteria that we introduced. To give them useful names, we have called them: the *J distance* (Jamiolkowski process distance, D_{pro}) and *J fidelity* (Jamiolkowski process fidelity, F_{pro}), which are based on the process matrix (i.e. the Jamiolkowski isomorphism); and the *S distance* (stabilised process distance, D_{stab}) and *S fidelity* (stabilised process fidelity, F_{stab}), which are based on the worst-case performance of the processes in a stabilised form.

All four measures either are metrics (in the case of the process distances) or give rise to a variety of associated metrics (for the process fidelities). Moreover, all of the metrics can be shown to satisfy the stability and chaining properties which greatly simplify the analysis of multistage quantum information processing tasks, as described in the introduction. The main differences arise in the criteria of easy calculation, measurement and sensible physical interpretation. The list below briefly summarises these remaining properties for the four measures. Throughout this section, assume that the goal in each case is to compare a quantum operation \mathcal{E} to an ideal unitary operation U ; the results vary somewhat when \mathcal{E} is being compared to an arbitrary process \mathcal{F} .

- **J distance:** There is a straightforward formula enabling D_{pro} to be calculated directly from the process matrix, thus also allowing it to be experimentally determined using quantum process tomography. The *J distance* can be given an operational interpretation as a bound on the average probability of error \bar{p}_e experienced during quantum computation of a function, or as a bound on the distance between the real and ideal joint distributions of the computer in a sampling computation:

$$\bar{p}_e - \bar{p}_e^{\text{id}} \leq D_{\text{pro}}(\mathcal{E}, U) \quad (9.30)$$

$$D(q, p) \leq D_{\text{pro}}(\mathcal{E}, U). \quad (9.31)$$

In the first expression \bar{p}_e^{id} is the average probability of error in the *ideal* computation, represented by U . In the second expression, $D(q, p)$ is the Kolmogorov distance between the real joint probability distribution $\{p(x, y)\} \equiv p$ over problem instances x and measurement outcomes y and the ideal joint distribution $\{q(x, y)\} \equiv q$, for a uniform distribution of problem instances.

- **J fidelity:** Once again, the J fidelity can be calculated directly from the process matrix. However, there is also a simpler formula for F_{pro} , Eq. (9.13), allowing easy calculation and measurement, without the need for full process tomography. This is much more straightforward than the calculation for the J distance, and is likely to simplify the determination of experimental errors. As for the J distance, the J fidelity can be given an operational interpretation related to average error probabilities:

$$\bar{p}_e \leq 1 - F_{\text{pro}}(\mathcal{E}, U). \quad (9.32)$$

$$F(q, p) \geq F_{\text{pro}}(\mathcal{E}, U). \quad (9.33)$$

In this case, the first expression holds only for ideal computations U which succeed perfectly, i.e. $\bar{p}_e^{\text{id}} = 0$. In the second expression, $F(q, p)$ is the Bhattacharya overlap between the real and ideal joint probability distributions, p and q , again for a uniform distribution of problem instances.

- **S distance:** There is no known elementary formula for D_{stab} , but we have proved that calculating the S distance is equivalent to a convex optimisation problem, which can be efficiently solved numerically, given knowledge of the process. This, in turn, enables D_{stab} to be measured experimentally, by performing full quantum process tomography. The S distance can be simply interpreted as a bound on the worst-case error probability p_e for a function computation, and as a bound on the maximum distance between the real and ideal output distributions of a sampling computation:

$$p_e - p_e^{\text{id}} \leq D_{\text{stab}}(\mathcal{E}, U). \quad (9.34)$$

$$\max_x D(q_x, p_x) \leq D_{\text{stab}}(\mathcal{E}, U). \quad (9.35)$$

In the first expression p_e^{id} is the worst-case error probability in the ideal computation, U . In the second expression $D(q_x, p_x)$ is the Kolmogorov distance between the real and ideal output probability distributions $\{q_x(y)\} \equiv q_x$ and p_x , taking the worst case over all problem instances x .

- **S fidelity:** Once again, no elementary formula for the S fidelity is known, but we have proved that the determination of F_{stab} can be formulated as a convex optimisation problem, and thus F_{stab} can be efficiently determined numerically. As a result, F_{stab} can again be determined experimentally, using process tomography. As with the S distance, F_{pro} has an operational interpretation related to worst-case error probabilities:

$$p_e \leq \left(\sqrt{p_e^{\text{id}}} + C_{\text{stab}}(\mathcal{E}, U) \right)^2. \quad (9.36)$$

$$\min_x F(q_x, p_x) \geq F_{\text{stab}}(\mathcal{E}, U). \quad (9.37)$$

The notation here is the same as above, with the definition $C_{\text{stab}} \equiv \sqrt{1 - F_{\text{stab}}}$.

As yet, it is difficult to choose between these four error measures, because we do not yet have a complete understanding of their properties. In particular, the discovery of simpler formulae for calculating them or simpler procedures for measuring them experimentally would have an important impact on which measures may be superior.

Since they all obey the stability and chaining criteria, it is, in all cases, only necessary to characterise component processes in order to bound the total error in a complex quantum information processing task. This makes it conceivable to use these measures for assessing processes in large-scale systems.

One important difference between the measures is that the S distance and S fidelity bound worst-case error probabilities, whereas the J distance and J fidelity bound the average-case errors. Traditionally, worst-case errors are usually of more interest than the average case, and this would seem to be a significant advantage for the S distance and S fidelity. On the other hand, given the linear nature of quantum mechanics, it seems likely that the average errors could be used to find relatively tight bounds for the worst-case errors. Moreover, the average errors and worst-case errors in fact contain complementary information about the process, since knowing both provides information about the *distribution* of errors in the sample space, as well as the size of the errors.

The measure which is simplest to calculate is the J fidelity, which has a simple formula, and is relatively easy to determine experimentally compared with the other measures. Unfortunately, this measure also has the weakest operational interpretation of the four, since we have not found a full form of Eq. (9.32) which holds true for function computations where the ideal case suffers an intrinsic error. Therefore, while the other measures may be more generally applicable, this measure will be convenient for early, proof-of-principle experimental demonstrations. It may also be particularly useful when trying to optimise the performance of a given experimental component, where many iterations of the characterisation may be required in a fine-tuning process.

The J distance has different strengths and weaknesses from the J fidelity. For example, it does allow the analysis of function computations with intrinsic errors in the ideal case. However, it requires a full process tomography to be determined experimentally and is not as easy to calculate.

From a traditional theoretical perspective, the S distance and S fidelity have the most attractive operational interpretations, since they relate to worst-case error probabilities. Unfortunately, they also require full process tomography to determine them experimentally, making them more difficult to measure than the J fidelity. Moreover, unlike both the J distance and J fidelity, no elementary formula for either is known. They are easy to calculate numerically, at least in principle, but as demonstrated by the difficulties we encountered with the convergence of some tomographies in Ch. 8, implementing a convex optimisation in practice is not a completely trivial task, and the optimisations themselves may still take a substantial time to find solutions for larger systems.

Finally, is it possible to justify a definite preference for either the fidelity or trace-distance measures? At the moment, we know of no convincing argument to choose one over the other. For instance, one can find examples of different processes where either the trace-distance or the fidelity measures give the better bounds for the results in Sec. 9.4. In

fact, it may ultimately turn out that the fidelity and trace distance measures provide complementary and somewhat distinct information, and that both measures must be calculated to completely characterise a process. For example, one might draw an analogy with a vector in cartesian coordinates, which is not completely characterised by its length alone. Further work is required before any definitive choice can be made.

At the present time we believe that *both* trace-distance and fidelity measures should be reported in experiments. For example, in Ch. 10, I report both the J distance and J fidelity. Once process tomography has been carried out, there is little extra difficulty in calculating the two measures over just one.

There is still much work that remains to be done. Some tasks of obvious importance include: (a) obtaining closed-form formulae and simple experimental measurement procedures for the S distance and S fidelity; (b) finding procedures which can be used to calculate experimental error bars for the S distance and S fidelity; (c) expressing the threshold condition for fault-tolerant quantum computation and communication using the error measures we have identified; and (d) extending our work so that it applies to quantum operations which are not trace-preserving, such as arise naturally in certain optical proposals for quantum computation [37, 38], where measurements and post-selection are critical elements.

Broadening the scope of this work, it would also be useful to develop additional diagnostic measures, which could be used experimentally to understand and improve specific aspects of a process' operation. For example, the *process purity*, $\text{Tr}\{\rho_{\mathcal{E}}^2\}$, can be regarded as a measure of the extent to which a quantum operation \mathcal{E} maintains the purity of the quantum state [see Ch. 10]. Although this measure is easily seen to be deficient in terms of the criteria developed in the introduction, and thus is not suitable as a general-purpose measure, it can provide useful information about one specific aspect of \mathcal{E} 's performance.

9.A Appendix: Worst case proofs

9.A.1 Proof of worst-case stabilisation

Let \mathcal{E}_Q and \mathcal{F}_Q be trace-preserving quantum operations acting on a d -dimensional system Q . We will show, following Ref. [7], that $\Delta_{\text{stab}}(\mathcal{E}_Q, \mathcal{F}_Q)$ is stable under the addition of an arbitrary d' -dimensional system Q' , i.e. $\Delta_{\text{stab}}(\mathcal{E}_Q, \mathcal{F}_Q) = \Delta_{\text{stab}}(\mathcal{I}_{Q'} \otimes \mathcal{E}_Q, \mathcal{I}_{Q'} \otimes \mathcal{F}_Q)$.

The definition of $\Delta_{\text{stab}}(\mathcal{E}_Q, \mathcal{F}_Q)$ states that $\Delta_{\text{stab}}(\mathcal{E}_Q, \mathcal{F}_Q) \equiv \Delta_{\text{max}}(\mathcal{I}_A \otimes \mathcal{E}_Q, \mathcal{I}_A \otimes \mathcal{F}_Q)$, where we have introduced a fictitious d -dimensional ancillary system A that is acted upon by the identity operation \mathcal{I}_A . Similarly, it follows that $\Delta_{\text{stab}}(\mathcal{I}_{Q'} \otimes \mathcal{E}_Q, \mathcal{I}_{Q'} \otimes \mathcal{F}_Q)$ is equal to $\Delta_{\text{max}}(\mathcal{I}_B \otimes \mathcal{I}_{Q'} \otimes \mathcal{E}_Q, \mathcal{I}_B \otimes \mathcal{I}_{Q'} \otimes \mathcal{F}_Q)$, where \mathcal{I}_B acts as the identity on a $d \times d'$ -dimensional ancilla B . Thus, to prove stability it suffices to show that the quantity $\Delta_{\text{max}}(\mathcal{I}_S \otimes \mathcal{E}_Q, \mathcal{I}_S \otimes \mathcal{F}_Q)$ is independent of the dimension of the system S that \mathcal{I}_S acts on, provided S is at least d -dimensional.

To see this, let ψ be a state achieving the maximum in $\Delta_{\text{max}}(\mathcal{I}_S \otimes \mathcal{E}_Q, \mathcal{I}_S \otimes \mathcal{F}_Q)$, with a Schmidt decomposition $\psi = \sum_j \psi_j |e_j\rangle |f_j\rangle$, where $|e_j\rangle$ are orthonormal states of S , and

$|f_j\rangle$ is an orthonormal basis set for Q . Since Q is d -dimensional, the state ψ has at most d Schmidt coefficients, and it is sufficient to consider only that d -dimensional subspace of S spanned by the states $|e_j\rangle$ with nonzero Schmidt coefficients. Thus, the maximum can be obtained working only in this subspace, and this concludes the proof.

9.A.2 Proof of convex optimisation property for F_{stab}

Our goal is to show that the problem of computing F_{stab} can be reduced to the minimisation of a convex function defined on a convex set. To show this we introduce a new function, denoted $F(\rho_Q, \mathcal{E}_Q, \mathcal{F}_Q)$, where subscripts indicate the system on which the variable is defined. The value of $F(\rho_Q, \mathcal{E}_Q, \mathcal{F}_Q)$ is defined to be the state fidelity $F([\mathcal{I}_A \otimes \mathcal{E}_Q](\psi), [\mathcal{I}_A \otimes \mathcal{F}_Q](\psi))$, where A is an ancilla of at least the same dimension as Q , and ψ is any purification of ρ_Q to AQ . It is easily verified that this definition is independent of which purification ψ of ρ_Q is used.

From this definition, it can be seen that the problem of computing $F_{\text{stab}}(\mathcal{E}_Q, \mathcal{F}_Q)$ is equivalent to minimising $F(\rho_Q, \mathcal{E}_Q, \mathcal{F}_Q)$ over all density matrices ρ_Q of system Q . Therefore, to prove that finding F_{stab} is a convex optimisation problem, we only need to show that $F(\rho_Q, \mathcal{E}_Q, \mathcal{F}_Q)$ is a convex function of ρ_Q , which takes values in a convex set.

To do this, let p_j be probabilities, and let ρ_Q^j be corresponding states of the system Q , with purifications ψ_j to a system AQ . It is helpful to introduce another ancillary system A' with an orthonormal basis $|j\rangle$ in one-to-one correspondence with the index on the states ρ_Q^j , and we define a state $|\psi\rangle \equiv \sum_j \sqrt{p_j} |j\rangle |\psi_j\rangle$ of the joint system $A'AQ$. By observing that $|\psi\rangle$ is a purification of $\sum_j p_j \rho_Q^j$, it follows that

$$F\left(\sum_j p_j \rho_Q^j, \mathcal{E}_Q, \mathcal{F}_Q\right) = F([\mathcal{I}_{A'A} \otimes \mathcal{E}_Q](\psi), [\mathcal{I}_{A'A} \otimes \mathcal{F}_Q](\psi)). \quad (9.38)$$

We then apply the monotonicity of the fidelity (cf. Sec. 9.2) under decoherence in the $|j\rangle$ basis, giving

$$F\left(\sum_j p_j \rho_Q^j, \mathcal{E}_Q, \mathcal{F}_Q\right) \leq F\left(\sum_j p_j |j\rangle\langle j| \otimes [\mathcal{I}_A \otimes \mathcal{E}_Q](\psi_j), \sum_j p_j |j\rangle\langle j| \otimes [\mathcal{I}_A \otimes \mathcal{F}_Q](\psi_j)\right). \quad (9.39)$$

Finally, by expanding the sums explicitly and applying some elementary algebra to simplify the right-hand side, we obtain

$$F\left(\sum_j p_j \rho_Q^j, \mathcal{E}_Q, \mathcal{F}_Q\right) \leq \sum_j p_j F(\rho_Q^j, \mathcal{E}_Q, \mathcal{F}_Q),$$

which implies that $F(\rho_Q, \mathcal{E}_Q, \mathcal{F}_Q)$ is convex in ρ_Q , as desired.

A similar construction shows that the computation of D_{stab} is equivalent to the maximisation of a concave function over a convex set, and thus is also a convex optimisation problem, with concomitant numerical benefits. The construction is sufficiently similar that we have omitted the details.

9.B Appendix: Application to quantum computing

9.B.1 Function computation in the worst case

Suppose \mathcal{E} and \mathcal{F} are real and ideal quantum operations, respectively, that act on an input $|x\rangle$, where x represents a problem instance. \mathcal{E} succeeds in computing the desired function $f(x)$ with an error probability of at most p_e , whereas \mathcal{F} succeeds with an (ideal) error probability of at most p_e^{id} .

We wish to show:

$$p_e \leq p_e^{\text{id}} + D_{\text{stab}}(\mathcal{E}, \mathcal{F}) \quad (9.40)$$

$$p_e \leq \left(\sqrt{p_e^{\text{id}}} + C_{\text{stab}}(\mathcal{E}, \mathcal{F}) \right)^2. \quad (9.41)$$

To prove the first inequality [Eq. (9.40)], we introduce a quantum operation \mathcal{M} representing the process of measurement, $\mathcal{M}(\rho) = \sum_y |y\rangle\langle y|\rho|y\rangle\langle y|$, where the sum is over all possible measurement outcomes y . Now observe that

$$p_e = D([\mathcal{M} \circ \mathcal{E}](x), f(x)) \quad (9.42)$$

$$\leq D([\mathcal{M} \circ \mathcal{E}](x), [\mathcal{M} \circ \mathcal{F}](x)) + D([\mathcal{M} \circ \mathcal{F}](x), f(x)) \quad (9.43)$$

$$\leq D(\mathcal{E}(x), \mathcal{F}(x)) + p_e^{\text{id}}, \quad (9.44)$$

where we used simple algebra in the first line, the triangle inequality in the second line, and contractivity of trace distance and some simple algebra in the third line. The desired result, Eq. (9.40), now follows from the definition of D_{stab} .

To prove the second inequality, Eq. (9.41), note that

$$p_e = 1 - F(\mathcal{E}(x), f(x)) \quad (9.45)$$

$$= C(\mathcal{E}(x), f(x))^2 \quad (9.46)$$

$$\leq [C(\mathcal{E}(x), \mathcal{F}(x)) + C(\mathcal{F}(x), f(x))]^2, \quad (9.47)$$

where the first line follows from the definition of p_e and the state fidelity, the second line follows from the definition of the metric $C(\cdot, \cdot)$, and the third line follows from the triangle inequality for $C(\cdot, \cdot)$. The proof of Eq. (9.41) is completed by noting that $C(\mathcal{E}(x), \mathcal{F}(x)) \leq C_{\text{stab}}(\mathcal{E}, \mathcal{F})$ and $C(\mathcal{F}(x), f(x)) \leq \sqrt{p_e^{\text{id}}}$.

9.B.2 Function computation in the average case

As in the worst case, \mathcal{E} and \mathcal{F} are real and ideal quantum operations that act on an input $|x\rangle$ to compute a desired function $f(x)$. \mathcal{E} succeeds with an average error probability \bar{p}_e , whereas \mathcal{F} succeeds with an average error probability \bar{p}_e^{id} .

The first steps in the proof of Eq. (9.24) are directly analogous to the proof of Eq. (9.22), resulting in the inequality

$$\bar{p}_e \leq \bar{p}_e^{\text{id}} + \frac{1}{d} \sum_x D(\mathcal{E}(x), \mathcal{F}(x)), \quad (9.48)$$

where d is the total number of possible inputs x . Recall that

$$D_{\text{pro}}(\mathcal{E}, \mathcal{F}) = D([\mathcal{I} \otimes \mathcal{E}](\Phi), [\mathcal{I} \otimes \mathcal{F}](\Phi)), \quad (9.49)$$

where \mathcal{I} acts on an ancilla which is a copy of the system \mathcal{E} and \mathcal{F} act on, and $|\Phi\rangle = \sum_x |x\rangle|x\rangle/\sqrt{d}$ is a maximally entangled state of the two systems. Now let \mathcal{M} be a quantum operation representing measurement on the ancilla system, defined similarly to the definition of \mathcal{M} just above. By contractivity of the trace distance,

$$D_{\text{pro}}(\mathcal{E}, \mathcal{F}) \geq D([\mathcal{M} \otimes \mathcal{E}](\Phi), [\mathcal{M} \otimes \mathcal{F}](\Phi)). \quad (9.50)$$

Elementary algebra gives

$$D([\mathcal{M} \otimes \mathcal{E}](\Phi), [\mathcal{M} \otimes \mathcal{F}](\Phi)) = \frac{1}{d} \sum_x D(\mathcal{E}(x), \mathcal{F}(x)). \quad (9.51)$$

Combining these results, we obtain Eq. (9.24).

As already remarked we have not found a natural average-case analogue of Eq. (9.23). However, if $\bar{p}_e^{\text{id}} = 0$, i.e. our computation succeeds with probability one, then it is possible to prove an average-case analogue. The result is

$$\bar{p}_e \leq C_{\text{pro}}(\mathcal{E}, \mathcal{F})^2 = 1 - F(\mathcal{E}, \mathcal{F}). \quad (9.52)$$

The proof uses very similar techniques to those used to establish Eqs (9.24) and (9.23), and is therefore omitted.

9.B.3 Sampling computation in the worst case

The quantum operation \mathcal{E} is an imperfect attempt to reproduce the statistics of the ideal operation \mathcal{F} which acts on an input $|x\rangle$. Measured in the computational basis, \mathcal{F} gives rise to a distribution $\{p_x(y)\} \equiv p_x$, whereas \mathcal{E} gives a distribution $\{q_x(y)\} \equiv q_x$.

The inequalities Eqs (9.26) and (9.27) that we want to prove may be stated as follows:

$$\max_x D(q_x, p_x) \leq D_{\text{stab}}(\mathcal{E}, \mathcal{F}) \quad (9.53)$$

$$\min_x F(q_x, p_x) \geq F_{\text{stab}}(\mathcal{E}, \mathcal{F}). \quad (9.54)$$

To prove the first inequality [Eq. (9.53)], let \mathcal{M} again be a quantum operation representing measurement in the computational basis. Note that for all x

$$D(q_x, p_x) = D([\mathcal{M} \circ \mathcal{E}](x), [\mathcal{M} \circ \mathcal{F}](x)) \quad (9.55)$$

$$\leq D(\mathcal{E}(x), \mathcal{F}(x)) \quad (9.56)$$

$$\leq D_{\text{stab}}(\mathcal{E}, \mathcal{F}), \quad (9.57)$$

where we used simple algebra in the first line, contractivity in the second line, and the definition of D_{stab} in the third line. An analogous argument can be used to establish the second inequality, (9.54).

9.3 Chapter 9 References

- [1] A. Gilchrist, N. K. Langford, and M. A. Nielsen. Distance measures to compare real and ideal quantum processes. *Physical Review A*, 71(6):062310, 2005.
- [2] K. R. W. Jones. Principles of quantum inference. *Annals of Physics*, 207:140, 1991.
- [3] U. Leonhardt. Discrete Wigner function and quantum-state tomography. *Physical Review A*, 53:2998, 1996.
- [4] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, UK, 2000.
- [5] I. L. Chuang and M. A. Nielsen. Prescription for experimental determination of the dynamics of a quantum black box. *Journal of Modern Optics*, 44:2455, 1997.
- [6] J. F. Poyatos, J. I. Cirac, and P. Zoller. Complete characterization of a quantum process: The two-bit quantum gate. *Physical Review Letters*, 78:390, 1997.
- [7] D. Aharonov, A. Kitaev, and N. Nisan. Quantum circuits with mixed states. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computation (STOC)*, page 20, 1997.
- [8] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 70:1895, 1993.
- [9] C. M. Caves and K. Wódkiewicz. Classical phase-space descriptions of continuous-variable teleportation. *Physical Review Letters*, 93:040506, 2004.
- [10] S. L. Braunstein, C. A. Fuchs, H. J. Kimble, and P. van Loock. Quantum versus classical domains for teleportation with continuous variables. *Physical Review A*, 64:022321, 2001.
- [11] T. Rudolph and B. C. Sanders. Requirement of optical coherence for continuous-variable quantum teleportation. *Physical Review Letters*, 87:077903, 2001.
- [12] F. Grosshans and P. Grangier. Quantum cloning and teleportation criteria for continuous quantum variables. *Physical Review A*, 64:010301(R), 2001.
- [13] T. C. Ralph, P. K. Lam, and R. E. S. Polkinghorne. Characterizing teleportation in optics. *Journal of Optics B: Quantum and Semiclassical Optics*, 1:483, 1999.
- [14] R. Schack and C. M. Caves. Classical model for bulk-ensemble NMR quantum computation. *Physical Review A*, 60:4354, 1999.
- [15] B. M. Terhal and G. Burkard. Fault-tolerant quantum computation for local non-Markovian noise. *e-print arXiv*, quant-ph/0402104, 2004.
- [16] M. A. Nielsen. A simple formula for the average gate fidelity of a quantum dynamical operation. *Physics Letters A*, 303:249, 2002.

- [17] M. D. Bowdrey, D. K. L. Oi, A. J. Short, K. Banaszek, and J. A. Jones. Fidelity of single-qubit maps. *Physics Letters A*, 294:258, 2002.
- [18] M. Raginsky. A fidelity measure for quantum channels. *Physics Letters A*, 290:11, 2001.
- [19] A. M. Childs, I. L. Chuang, and D. W. Leung. Realization of quantum process tomography in NMR. *Physical Review A*, 64:012314, 2001.
- [20] A. M. Childs, J. Preskill, and J. Renes. Quantum information and precision measurement. *Journal of Modern Optics*, 47:155, 2000.
- [21] B. Schumacher. Sending entanglement through noisy quantum channels. *Physical Review A*, 54:2614, 1996.
- [22] B. Schumacher. Quantum coding. *Physical Review A*, 51:2738, 1995.
- [23] C. H. Bennett, E. Bernstein, G. Brassard, and U. Vazirani. Strengths and weaknesses of quantum computing. *SIAM Journal on Computing*, 26(5):1510, 1997. arXiv:quant-ph/9701001.
- [24] M. B. Ruskai. Beyond strong subadditivity: improved bounds on the contraction of generalized relative entropy. *Reviews in Mathematical Physics*, 6:1147, 1994.
- [25] A. Uhlmann. The ‘transition probability’ in the state space of a $*$ -algebra. *Reports on Mathematical Physics*, 9:273, 1976.
- [26] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. Schumacher. Noncommuting mixed states cannot be broadcast. *Physical Review Letters*, 76:2818, 1996.
- [27] J. Dodd and M. A. Nielsen. Simple operational interpretation of the fidelity of mixed states. *Physical Review A*, 66:044301, 2002.
- [28] C. A. Fuchs. *Distinguishability and Accessible Information in Quantum Theory*. PhD thesis, The University of New Mexico, Albuquerque, NM, 1996. arXiv:quant-ph/9601020.
- [29] C. A. Fuchs and J. van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Transactions on Information Theory*, 45:1216, 1999.
- [30] D. N. Page. Average entropy of a subsystem. *Physical Review Letters*, 71:1291, 1993.
- [31] S. Sen. Average entropy of a quantum subsystem. *Physical Review Letters*, 77:1, 1996.
- [32] M. Horodecki, P. Horodecki, and R. Horodecki. General teleportation channel, singlet fraction, and quasidistillation. *Physical Review A*, 60:1888, 1999.
- [33] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2004.
- [34] L. K. Grover. Quantum mechanics helps in searching for a needle in a haystack. *Physical Review Letters*, 79:325, 1997.

-
- [35] P. W. Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In *Proceedings of the 35th Annual Symposium on the Foundations of Computer Science*, page 124, 1994.
 - [36] N. A. Gershenfeld and I. L. Chuang. Bulk spin-resonance quantum computation. *Science*, 275:350, 1997.
 - [37] E. Knill, R. Laflamme, and G. J. Milburn. A scheme for efficient quantum computation with linear optics. *Nature*, 409:46, 2001.
 - [38] M. A. Nielsen. Optical quantum computation using cluster states. *Physical Review Letters*, 93:040503, 2004.

Whenever the horse stopped (which it did very often), he fell off in front; and, whenever it went on again (which it generally did rather suddenly), he fell off behind. Otherwise he kept on pretty well, except that he had a habit of now and then falling off sideways;

Through the Looking-Glass (and what Alice found there),
Lewis Carroll

Here is Edward Bear, coming downstairs now, bump, bump, bump, on the back of his head, behind Christopher Robin. It is, as far as he knows, the only way of coming downstairs, but sometimes he feels that there really is another way, if only he could stop bumping for a moment and think of it.

Winnie-the-Pooh, A. A. Milne

Chapter 10

A simple entangling optical gate and its use in Bell-state analysis

Acknowledgement

The work reported in this chapter was first published in Ref. [1], back-to-back with Refs [2] and [3]. Where appropriate, I have incorporated some of the text of Ref. [1]. I performed the experiments with the continuous-wave (cw) down-conversion source, and I will focus mainly on these results in this thesis. However, complementary experiments using a pulsed down-conversion source (implementing the gate only), carried out by Till Weinhold, Robert Prevedel and Kevin Resch, were also reported in Ref. [1] and formed part of Robert Prevedel's Diploma thesis submitted to the University of Vienna [4]. I will make occasional reference to these other experiments where appropriate.

I also carried out the original theoretical investigations of real-world effects in the two-photon CZ gate described in Sec. 10.2.2 [5].

10.1 Introduction

Throughout this thesis, I have repeatedly discussed the importance of entanglement in quantum information. In particular, I have investigated down-conversion as a natural

source of various types of entanglement, as well as different ways to measure and characterise entanglement. However, perhaps the most flexible and powerful way to create or manipulate entanglement is to use unitary quantum gates—this is a key component of all implementations of quantum computing. In the standard circuit model, any entangling gate can be used to construct a universal set of operations [6]. Moreover, when combined with measurements, disentangling gates (normally the CNOT gate) are used in quantum teleportation, for analysing in the Bell-state basis, and error correction protocols, for performing parity measurements¹. In the cluster-state model, entangling gates (normally the CZ gate) can be used to build the cluster-state which encodes the computation.

Conceptually, one of the simplest entangling two-qubit gates is the controlled-Z (CZ) gate, which in the logical basis produces a π phase shift on the $|11\rangle$ term, (i.e. $|00\rangle \rightarrow |00\rangle$; $|01\rangle \rightarrow |01\rangle$; $|10\rangle \rightarrow |10\rangle$; $|11\rangle \rightarrow -|11\rangle$). As a maximally entangling gate, it is universal for quantum computing when coupled with single qubit rotations². However, in optical systems, where interactions between photons are generally very weak, it is extremely difficult to produce such a highly nonlinear phase shift, which must be triggered by the presence of a single photon in the control arm. In the KLM scheme for deterministic linear optics quantum computing [7], the key for circumventing this problem is the recognition that the nonlinear interaction can be realised non-deterministically using measurement (itself a highly nonlinear process). The non-deterministic gates can then be converted into deterministic gates using teleportation [8]. Since then, there have been a number of demonstrations of quantum logic gates derived from this concept [9–13], and further theoretical development of linear optics schemes [5, 14–17]. Moreover, there is also a recent suggestion to use non-deterministic CZ gates to construct cluster-states [18].

In this chapter, I discuss our experiment that demonstrated a non-deterministic linear-optics CZ gate, and its application as a Bell-state analyser. With a new, totally polarisation-based architecture, this CZ gate is the simplest entangling (or dis-entangling) linear optics gate realised to date, requiring only three partially polarising beam splitters (PPBSs), two half-wave plates, no classical interferometers, and no ancilla photons. It is non-deterministic and success is heralded by detection of a single photon in each of the outputs. We fully characterised the operation of this gate using quantum process tomography, and we also demonstrated its use for fully resolving Bell measurements. This simple entangling optical gate is particularly promising for micro-optics or guided optics implementations where extremely good non-classical interference is realisable [19].

10.1.1 A note on pulsed experiments

In my experiments, I used a continuous-wave (cw) Ar⁺-ion pump laser ($\lambda_p=351.1\text{nm}$) to produce the down-conversion photon pairs ($\lambda_{dc}=702.2\text{nm}$, $\Delta\lambda=0.36\text{nm}$). At the same time, Till Weinhold and others carried out a complementary experimental project to implement a PPBS CZ gate with a pulsed, frequency-doubled Ti:Sa laser ($\lambda_p=410\text{nm}$)

¹Parity measurements determine whether two qubits are in the same or different logical states, without providing any information about the states themselves. The standard way to implement a non-destructive parity measurement uses an ancilla qubit and two CNOT gates.

²The more familiar CNOT gate is formed by applying a Hadamard gate H to the target qubit before and after a CZ gate.

pumping the down-conversion crystal ($\lambda_{dc}=820\text{nm}$, $\Delta\lambda=3.0\text{nm}$) (also in Ref. [1]). Both experiments used BBO (β -barium borate) crystals pumped under Type-I phase-matching conditions. Not surprisingly, these experiments met with different challenges which I will touch upon below. Overall, however, we found that the gate operation was slightly better in the pulsed case, from which we can conclude that temporal mode mismatch was not a significant factor in the gate's performance.

10.2 The two-photon CZ gate

The two-photon CZ gate has provided an important intermediate step in developing optical quantum computing, allowing the demonstration of the quantum mechanical effects without resorting to the complexity of the fully scalable, four-photon schemes [7]. In this section, I will provide a theoretical description of, first, the original two-photon circuit design, and then our new simplified architecture.

10.2.1 Non-classical interference

The key element in current implementations of linear optical quantum computing is the effect of non-classical interference between two photons which arrive in separate ports of a beam splitter [recall Fig. 2.3]. This is quite different from its classical counterpart which describes the interference between two alternative paths of a single photon or bright coherent state.

When two photons ($|1_a, 1_b\rangle_1$) arrive at a beam splitter (reflectance, γ), there are four possible outcomes: both are reflected ($-\gamma|1_a, 1_b\rangle_2$), both are transmitted ($(1-\gamma)|1_a, 1_b\rangle_2$), or one each is reflected and transmitted (two ways). Of these four possibilities, the first two both result in photons at each output port of the beam splitter, and since different photons can be indistinguishable in quantum mechanics, these two alternatives can interfere (i.e. the amplitudes combine in the output state). If, however, the photons are distinguishable in any way, then the two possibilities will not interfere, but instead combine incoherently (i.e. the probabilities will add in the output, not the amplitudes). It is important to note that this condition is decided not by whether the photons have been distinguished, but rather by whether they can be distinguished *in principle*. The photons must be indistinguishable in every way, e.g. in arrival time, frequency spectrum, spatial mode (mode shape and longitudinal beam position) and polarisation. For example, if the photons arrive at the beam splitter separated by more than their coherence time, then they will be distinguishable and the interference will be destroyed.

Assume that detectors are placed at the two output ports of the beam splitter. Under the two different interference conditions, the probabilities that *both* detectors see an arriving photon are:

$$\begin{aligned} \text{indistinguishable:} & \quad p_{\text{indist}} = ((-\gamma) + (1-\gamma))^2 = (1 - 2\gamma)^2, \\ \text{distinguishable:} & \quad p_{\text{dist}} = \gamma^2 + (1-\gamma)^2. \end{aligned} \tag{10.1}$$

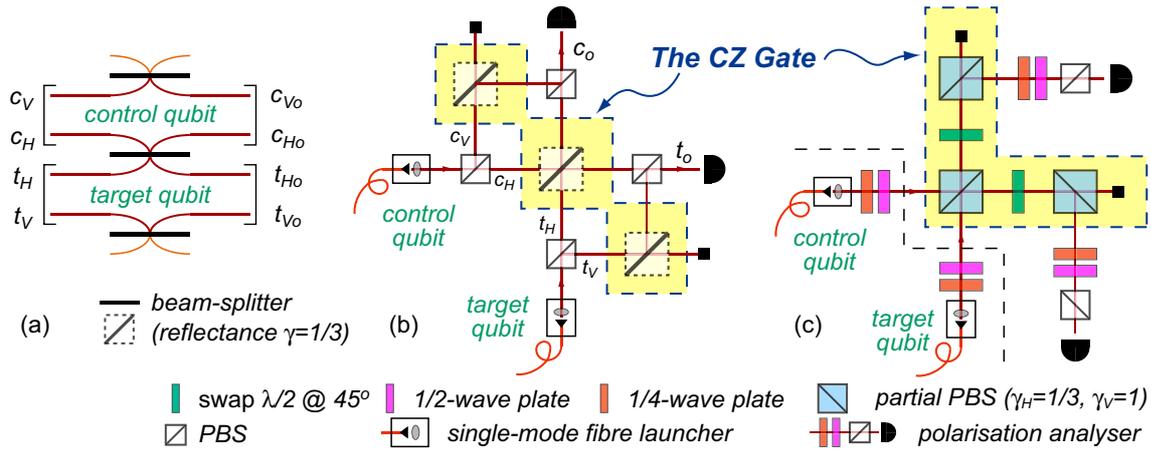


Figure 10.1: The interferometric, optical CZ gate introduced in Refs. [5, 17] (a) in the dual-rail representation, and (b) in the form of a more complete optical circuit. In practice, the dual-rail encoding is implemented by transforming each qubit from polarisation to spatial encoding, and back again. This requires high interferometric stability and spatio-temporal mode-matching for correct operation. (c) Partially polarising beam splitter (PPBS) gate. The qubits can remain polarisation-encoded, since the vertically polarised modes are completely reflected by the first PPBS, and do not interact. Non-classical interference occurs between the horizontally polarised modes, with $\gamma_H=1/3$. The subsequent PPBSs balance the circuit by providing the required losses in the c_V and t_V modes as shown in (a) and (b).

The non-classical interference visibility is therefore defined to be $(p_{\text{dist}} - p_{\text{indist}})/p_{\text{dist}}$. For a standard 50:50 beam splitter, this results in completely destructive interference when the photons are indistinguishable (100% visibility). This is sometimes referred to as *photon bunching*, since the two photons always exit from the same port of the beam splitter. In an experiment, the non-classical interference visibility is normally measured by delaying the arrival time of one of the photons by more than the coherence length, and then scanning it through the coincident timing condition. If aligned correctly, this produces a measurable dip in the coincidence counts, often called a Hong-Ou-Mandel (HOM) dip [20]. In the following experiments there is non-classical interference at a beam splitter with 1/3 reflectance ($\gamma=1/3$). In this case, the best achievable interference visibility is 80%.

10.2.2 The original design

Figure 10.1(a) illustrates the original two-photon, entangling CZ gate³ introduced in Refs [5, 17] using the so-called “dual-rail” representation [see also Sec. 2.2.3], where a logical qubit (marked in the figure by square brackets) is represented by two evolution paths (*rails*), one for each basis mode, and the logical basis states are defined to be $|\mathbf{0}\rangle \equiv |H\rangle = |1_H, 0_V\rangle$ and $|\mathbf{1}\rangle \equiv |V\rangle = |0_H, 1_V\rangle$. The original KLM linear optics scheme [7] relied upon a conditional entangling gate which required two ancilla photons and perfect

³Invented by Tim Ralph [5, 21] and independently discovered by Holger Hofmann and Shigeki Takeuchi [17].

photon detectors capable of reliably discriminating zero and one photon detection events. Conversely, the simplified gate in Fig. 10.1(a) uses only two photons, those encoding the logical qubits. It is *non-deterministic*, with a heralded success probability of 1/9 flagged by detecting a single photon in each logical qubit output (i.e. coincidence measurement), and *post-selective*, in that it only produces the correct gate behaviour in this coincidence subspace.

In linear optics quantum information, the behaviour of each logical basis mode is characterised by the field (annihilation) operator for that mode (e.g. c_H or t_V) [see Sec. 2.2]. Consider then an input state of the form:

$$|\psi_{\text{in}}\rangle = \alpha_{HH}|HH\rangle + \alpha_{HV}|HV\rangle + \alpha_{VH}|VH\rangle + \alpha_{VV}|VV\rangle, \quad (10.2)$$

$$= \left\{ \alpha_{HH}c_H^\dagger t_H^\dagger + \alpha_{HV}c_H^\dagger t_V^\dagger + \alpha_{VH}c_V^\dagger t_H^\dagger + \alpha_{VV}c_V^\dagger t_V^\dagger \right\} |0\rangle, \quad (10.3)$$

and for convenience, assume that the beam splitters are *symmetric* beam splitters, i.e. $\varphi = \pi/2$ in Eq. (2.109), giving:

$$a_o = \frac{1}{\sqrt{2}} (i\sqrt{\gamma}a + \sqrt{\tau}b) \quad \& \quad b_o = \frac{1}{\sqrt{2}} (\sqrt{\tau}a + i\sqrt{\gamma}b), \quad (10.4)$$

where γ and $\tau=1-\gamma$ are the reflectance and transmittance, respectively. Using the technique summarised in Sec. 2.2.2, it is then fairly straightforward to calculate the output state:

$$\begin{aligned} |\psi_{\text{out}}\rangle = & (1 - 2\gamma)\alpha_{HH}|HH\rangle - \gamma\alpha_{HV}|HV\rangle - \gamma\alpha_{VH}|VH\rangle - \gamma\alpha_{VV}|VV\rangle \\ & + 2i\sqrt{2\gamma\tau}\alpha_{HH}|2000\rangle|00\rangle + 2i\sqrt{2\gamma\tau}\alpha_{HH}|0020\rangle|00\rangle \\ & + i\sqrt{\gamma\tau}\alpha_{HV}|0011\rangle|00\rangle + i\sqrt{\gamma\tau}\alpha_{VH}|1100\rangle|00\rangle \\ & + \{\tau\alpha_{VH}|1000\rangle + i\sqrt{\gamma\tau}\alpha_{VH}|0010\rangle + i\sqrt{\gamma\tau}\alpha_{VV}|0001\rangle\}|10\rangle \\ & + \{\tau\alpha_{HV}|0010\rangle + i\sqrt{\gamma\tau}\alpha_{HV}|1000\rangle + i\sqrt{\gamma\tau}\alpha_{VV}|0100\rangle\}|01\rangle + \tau\alpha_{VV}|0000\rangle|11\rangle, \end{aligned} \quad (10.5)$$

where the states outside the computational basis are written $|n_{c_H}n_{c_V}n_{t_H}n_{t_V}\rangle|n_{d_c}n_{d_t}\rangle$, and $d_{c,t}$ are dump modes for the control and target qubits⁴. After measuring in coincidence, the only remaining contributions are the computational basis states, and the gate operation is described by the operator matrix,

$$A_{\text{CZ}} = \begin{bmatrix} 1-2\gamma & 0 & 0 & 0 \\ 0 & -\gamma & 0 & 0 \\ 0 & 0 & -\gamma & 0 \\ 0 & 0 & 0 & -\gamma \end{bmatrix}, \quad (10.6)$$

where $|\psi_{\text{out}}\rangle = A_{\text{CZ}}|\psi_{\text{in}}\rangle$. In the ideal case, where $\gamma = 1/3$, this is the same as the standard CZ gate (barring an irrelevant global phase and bit flips on each qubit which are equivalent to a simple redefinition of the logical states).

There are two distinct types of error states. In principle, the states with one or more photons in the dump ports (the last two lines) can be rejected by detecting those modes—this does not affect the computational basis states. However, the other states, which

⁴These are required to balance the probabilities of the different computational states in the coincidence subspace.

have two photons in one qubit and none in the other modes (e.g. $|2000\rangle$ or $|1100\rangle$), can only be removed by detecting photons in the actual qubit modes. With current detector technology, this makes the gate *destructive* as well as post-selective. In principle, this two-photon gate could be made scalable if photon number QND measurements were made on each output qubit.

10.2.3 The all-polarisation approach

The dual-rail representation is very convenient for analysing optical circuits because it exactly matches the field operators which describe the circuit behaviour. Its weakness, perhaps, is that it obscures an important practical complication. Moving from a polarisation encoding to an explicitly dual-rail encoding requires polarising beam splitters (PBSs) (so that one can apply different optical elements to different modes). A more complete physical description of the CZ circuit is illustrated in Fig. 10.1(b). Including the PBSs in the circuit diagram makes it obvious that the gate actually contains two coupled classical interferometers, as well as the non-classical interference that is required at the central beam splitter. Although the best performing entangling gate implementations to date have been based upon this design, it is a very difficult optical experiment to align. Moreover, since the phase relationship between the logical modes must be maintained during the gate operation, interferometric stability is required between the control and target modes. Previously, this has been achieved by using inherently stable interferometers [10, 11], but these may not be suitable for scaling to large numbers in micro- or integrated-optical realisations.

Figure 10.1(c) shows an alternative CZ design⁵ which operates directly on the polarisation degree of freedom without requiring the classical interferometers necessary to an explicitly dual-rail approach. The key elements are partially polarising beam splitters (PPBSs), which are designed to perfectly reflect vertically polarised light and couple the horizontal inputs (here, $\gamma_V=1$ and $\gamma_H=1/3$), thereby allowing direct manipulation of the individual polarisation modes without any spatial separation. Experimentally, this is much easier to align, because the gate performance only depends on one mode-matching condition—the non-classical interference at the central PPBS. Moreover, while classical interferometers require path lengths to be stable to less than a wavelength, non-classical interference is only sensitive to path-length shifts on the order of the (much longer) coherence length ($L_c=\lambda^2/\Delta\lambda \gg \lambda$).

We showed already in Ref. [5] that the two-photon CZ is reasonably forgiving to beam splitters with incorrect reflectivities, but much more sensitive to mode mismatch in the non-classical interference at the central beam splitter. Will the new architecture described in Fig. 10.1(c) behave any differently? Obviously, misaligned modes at the central PPBS will still reduce the visibility of the non-classical interference, and the resulting distinguishability will manifest itself as mixture instead of entanglement in output states. Indeed, one of the main purposes of the new architecture is to make this alignment process simpler in the lab. In this section, I will model the gate operation under other effects, such as using real-world PPBSs and wave plates. To do this, I first need to introduce the appropriate tools.

⁵Developed independently by Geoff Pryde and the groups in Refs [2] and [3].

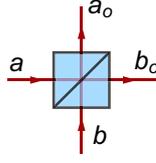


Figure 10.2: A partially polarising beam splitter (PPBS) is designed to perfectly reflect vertically polarised light and couple the horizontal inputs (e.g. in our experiments, $\gamma_V=1$ and $\gamma_H=1/3$).

The PPBS is a cube beam splitter with a multilayer dielectric stack at the interface specially manufactured to give the desired reflectivities. Not surprisingly, in reality these beam splitters also induce polarisation-sensitive phase shifts in the reflected beams. Fortunately, in this circuit, these have no effect on the entangling quality of the gate, only applying correctable local polarisation rotations to each qubit. With modes defined as in Fig. 10.2, the field operator evolution equations for the PPBS can be written (in a symmetric form):

$$\begin{aligned}
 a_{Ho} &= i\sqrt{\gamma_H}a_H + \sqrt{\tau_H}b_H, \\
 a_{Vo} &= e^{i\phi_a}a_V, \\
 b_{Ho} &= \sqrt{\tau_H}a_H + i\sqrt{\gamma_H}b_H, \\
 b_{Vo} &= e^{i\phi_b}b_V.
 \end{aligned} \tag{10.7}$$

This can be summarised in matrix form:

$$\begin{bmatrix} a_{Ho} \\ a_{Vo} \\ b_{Ho} \\ b_{Vo} \end{bmatrix} = \begin{bmatrix} i\sqrt{\gamma_H} & 0 & \sqrt{\tau_H} & 0 \\ 0 & e^{i\phi_a} & 0 & 0 \\ \sqrt{\tau_H} & 0 & i\sqrt{\gamma_H} & 0 \\ 0 & 0 & 0 & e^{i\phi_b} \end{bmatrix} \begin{bmatrix} a_H \\ a_V \\ b_H \\ b_V \end{bmatrix} = U_{\text{ppbs}} \begin{bmatrix} a_H \\ a_V \\ b_H \\ b_V \end{bmatrix}. \tag{10.8}$$

This matrix describes the evolution of the field operators through the central PPBS with photons arriving from both sides. However, consider a situation where photons arrive at one side only, and only the reflected output is measured [cf. the two dump PPBSs in Fig. 10.1(c)]. Then the evolution matrix is⁶:

$$\begin{bmatrix} a_{Ho} \\ a_{Vo} \end{bmatrix} = \begin{bmatrix} i\sqrt{\gamma_H} & 0 \\ 0 & e^{i\phi_a} \end{bmatrix} \begin{bmatrix} a_H \\ a_V \end{bmatrix} = i \begin{bmatrix} \sqrt{\gamma_H} & 0 \\ 0 & e^{i\phi'_a} \end{bmatrix} \begin{bmatrix} a_H \\ a_V \end{bmatrix} = U_{\text{ppbs}}^{(1)} \begin{bmatrix} a_H \\ a_V \end{bmatrix}, \tag{10.9}$$

where $\phi'_a = \phi_a - \pi/2$ is the phase that would be measured in such a situation—without interferometers, only relative phases are accessible to measurement. The final element required to analyse the circuit in Fig. 10.1(c) is the wave plate evolution matrix. For an arbitrary wave plate with retardation phase θ , and optical axis at angle φ , the matrix is [Sec. 4.4]:

$$\begin{bmatrix} a_{Ho} \\ a_{Vo} \end{bmatrix} = \begin{bmatrix} e^{i\theta/2} \cos^2 \varphi + e^{-i\theta/2} \sin^2 \varphi & (e^{i\theta/2} - e^{-i\theta/2}) \sin \varphi \cos \varphi \\ (e^{i\theta/2} - e^{-i\theta/2}) \sin \varphi \cos \varphi & e^{-i\theta/2} \cos^2 \varphi + e^{i\theta/2} \sin^2 \varphi \end{bmatrix} \begin{bmatrix} a_H \\ a_V \end{bmatrix} = M_{\text{wp}}(\theta, \varphi) \begin{bmatrix} a_H \\ a_V \end{bmatrix}. \tag{10.10}$$

⁶Note that this sub-matrix is no longer unitary, because I have not included all the vacuum inputs and dump outputs. I can only use this form because we are restricted to the coincidence subspace, and the non-unitarity indicates that energy is being lost from this subspace.

To model the circuit in Fig. 10.1(c) realistically, I will allow the PPBSs to have arbitrary phases, but assume their (horizontal) reflectances are equal ($\equiv \gamma$). Also, although I will allow the possibility that the swap wave plates do not induce a perfect half-wave phase shift, I will assume there is no error in the alignment of their optical axes (i.e. $\varphi = 45^\circ$), since computer-controlled rotation mounts allow these to be measured and positioned accurately and precisely. Then, using the reverse-time formalism [see Sec. 2.2.2],

$$\begin{aligned}
[\text{inputs}] &= (\text{central PPBS}) \times (\text{swap HWPs @ } 45) \times (\text{dump PPBSs}) \times [\text{outputs}] \\
\begin{bmatrix} c_H \\ c_V \\ t_H \\ t_V \end{bmatrix} &= (-i) \begin{bmatrix} \sqrt{\gamma} & 0 & i\sqrt{\tau} & 0 \\ 0 & e^{-i\phi'_{cc}} & 0 & 0 \\ i\sqrt{\tau} & 0 & \sqrt{\gamma} & 0 \\ 0 & 0 & 0 & e^{-i\phi'_{ct}} \end{bmatrix} \cdot \begin{bmatrix} \cos \theta_c/2 & \sin \theta_c/2 & 0 & 0 \\ \sin \theta_c/2 & \cos \theta_c/2 & 0 & 0 \\ 0 & 0 & \cos \theta_t/2 & \sin \theta_t/2 \\ 0 & 0 & \sin \theta_t/2 & \cos \theta_t/2 \end{bmatrix} \\
&\times (-i) \begin{bmatrix} \sqrt{\gamma} & 0 & 0 & 0 \\ 0 & e^{-i\phi'_{dc}} & 0 & 0 \\ 0 & 0 & \sqrt{\gamma} & 0 \\ 0 & 0 & 0 & e^{-i\phi'_{dt}} \end{bmatrix} \begin{bmatrix} c_{Ho} \\ c_{Vo} \\ t_{Ho} \\ t_{Vo} \end{bmatrix} = M_{\text{exp}} \begin{bmatrix} c_{Ho} \\ c_{Vo} \\ t_{Ho} \\ t_{Vo} \end{bmatrix}, \tag{10.11}
\end{aligned}$$

where $\phi_{cc,ct}$ are the reflection phases for the central PPBS from the control and target sides, respectively, and $\phi_{dc,dt}$ are those for the dump PPBSs in the control and target arms. In the case where $\theta=\pi$ (perfect swap wave plates), this reduces to:

$$M_{\text{exp}} = - \begin{bmatrix} 0 & e^{-i\phi'_{dc}}\sqrt{\gamma} & 0 & ie^{-i\phi'_{dt}}\sqrt{\tau} \\ e^{-i\phi'_{cc}}\sqrt{\gamma} & 0 & 0 & 0 \\ 0 & ie^{-i\phi'_{dc}}\sqrt{\tau} & 0 & e^{-i\phi'_{dt}}\sqrt{\gamma} \\ 0 & 0 & e^{-i\phi'_{ct}}\sqrt{\gamma} & 0 \end{bmatrix}. \tag{10.12}$$

If a general two-qubit, linear optical circuit transforms the field operators according to the reverse-time equations summarised by the matrix $M = (m_{jk})$, then it is not too difficult to show that the state (in the coincidence basis) is transformed according to $|\psi_{\text{out}}\rangle = A|\psi_{\text{in}}\rangle$, where

$$A = \begin{bmatrix} m_{13}^*m_{31}^* + m_{11}^*m_{33}^* & m_{13}^*m_{41}^* + m_{11}^*m_{43}^* & m_{23}^*m_{31}^* + m_{21}^*m_{33}^* & m_{23}^*m_{41}^* + m_{21}^*m_{43}^* \\ m_{14}^*m_{31}^* + m_{11}^*m_{34}^* & m_{14}^*m_{41}^* + m_{11}^*m_{44}^* & m_{24}^*m_{31}^* + m_{21}^*m_{34}^* & m_{24}^*m_{41}^* + m_{21}^*m_{44}^* \\ m_{13}^*m_{32}^* + m_{12}^*m_{33}^* & m_{13}^*m_{42}^* + m_{12}^*m_{43}^* & m_{23}^*m_{32}^* + m_{22}^*m_{33}^* & m_{23}^*m_{42}^* + m_{22}^*m_{43}^* \\ m_{14}^*m_{32}^* + m_{12}^*m_{34}^* & m_{14}^*m_{42}^* + m_{12}^*m_{44}^* & m_{24}^*m_{32}^* + m_{22}^*m_{34}^* & m_{24}^*m_{42}^* + m_{22}^*m_{44}^* \end{bmatrix}. \tag{10.13}$$

For example, with perfect wave plates,

$$A_{\text{exp}} = \begin{bmatrix} 0 & 0 & 0 & e^{i(\phi'_{cc}+\phi'_{ct})\gamma} \\ 0 & 0 & e^{i(\phi'_{cc}+\phi'_{dt})\gamma} & 0 \\ 0 & e^{i(\phi'_{dc}+\phi'_{ct})\gamma} & 0 & 0 \\ e^{i(\phi'_{dc}+\phi'_{dt})}(-1+2\gamma) & 0 & 0 & 0 \end{bmatrix}. \tag{10.14}$$

This result shows that the polarisation-sensitive phase shifts have no effect on the entangling quality of the gate. In fact, they can be undone by the local phase rotations,

$$\begin{bmatrix} e^{-i\phi'_{cc}} & 0 \\ 0 & e^{-i\phi'_{dc}} \end{bmatrix} \otimes \begin{bmatrix} e^{-i\phi'_{ct}} & 0 \\ 0 & e^{-i\phi'_{dt}} \end{bmatrix}, \tag{10.15}$$

leaving an entangling CZ-type operation of the form: $A_{\text{exp}} \sim (X \otimes X) A_{\text{CZ}}$. Therefore, I will ignore these phase shifts in the analysis below. The extra X (bit-flip) gates can also

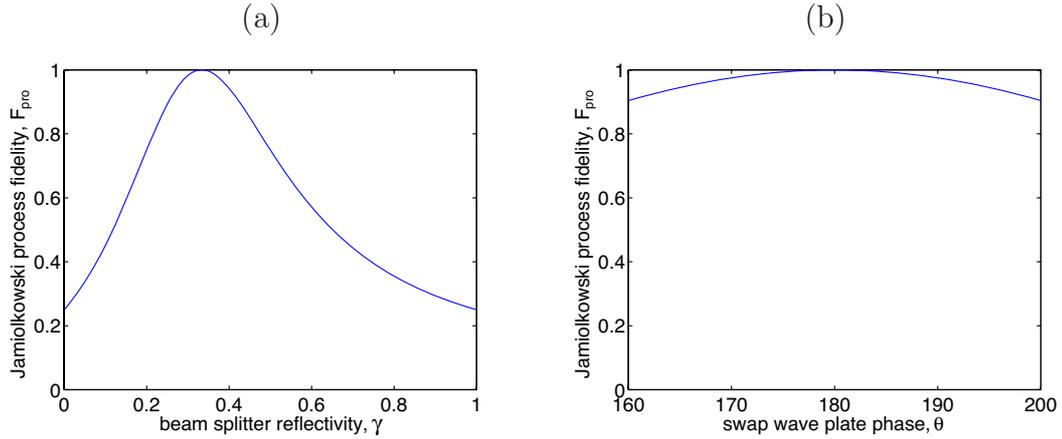


Figure 10.3: Real-world effects in the PPBS CZ gate due to (a) imperfect PPBSs and (b) imperfect swap wave plates.

be corrected with wave plates or by relabelling the logical states of the outputs—in these experiments, we simply chose to relabel.

To quantify these real-world effects, I will use the Jamiolkowski process fidelity (J fidelity) defined in Sec. 9.3.1. The target process matrix is $\rho_U = [I \otimes U_t](\Phi)$, where Φ is the d^2 -dimensional maximally entangled state, and $U_t = (X \otimes X)U_{\text{CZ}}$. The above real-world gate model also results in a “pure” process, $\rho_A = [I \otimes A](\Phi)$, the difference being that A is not necessarily unitary, so that the output state may not be normalised, i.e. the probability of the output state lying completely in the coincidence basis may be less than one, indicating that the gate is non-deterministic. However, because we always measure in the coincidence basis, we do not “see” this conditional operation and we will measure the *normalised* process, $\tilde{\rho}_A = \rho_A / \text{Tr}\{\rho_A\}$, where it is easy to show that $\text{Tr}\{\rho_A\} = \text{Tr}\{A^\dagger A\} / d$. For any such process with an operator-sum representation containing only one term, it follows that $\tilde{\rho}_A^2 = \tilde{\rho}_A$. Moreover, it can also be shown that the J fidelity,

$$F_{\text{pro}} = \text{Tr}\{\rho_U, \tilde{\rho}_A\} = \frac{|\text{Tr}\{U^\dagger A\}|^2}{d \text{Tr}\{A^\dagger A\}}. \quad (10.16)$$

For theoretical calculations, this is a far easier quantity to calculate than the full process fidelity definition⁷.

Figure 10.3 shows how the J fidelity depends on the horizontal reflectivity of the PPBSs, γ (a), and the retardation phase of the swap half-wave plates, $\theta = \theta_c = \theta_t$ (b). In both cases, the analysis indicates that these errors should not be a major problem for realistic values for the experimental parameters—e.g. with $\gamma = 0.33 \pm 0.05$, the fidelity remains above 95%, while with $\theta = 180^\circ \pm 10^\circ$, the fidelity is greater than 97.5%. These new results for the quality of the gate performance with incorrect beam splitter reflectivities support the conclusions that we reached in Ref. [5], although the original analysis techniques were somewhat more *ad hoc*.

⁷Because one often knows the operator and not the process matrix, using the full definition requires finding the corresponding process matrix for each calculation, which is conceptually simple, but in practice, is far more effort (and numerically intensive) than using the simple operator definition.

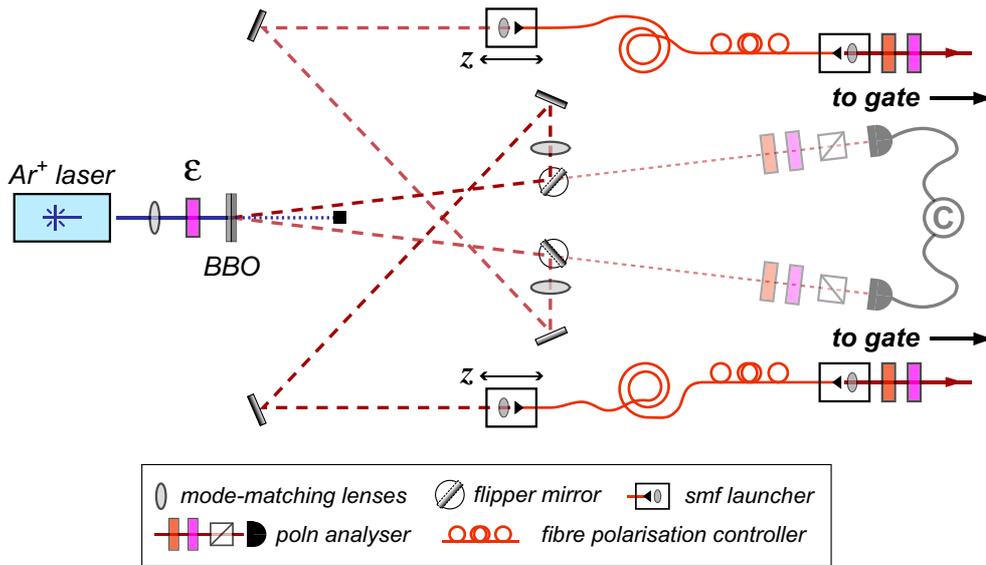


Figure 10.4: A detailed schematic of the entangled source apparatus. The polarisation-entangled state is produced by Type-I down-conversion in a crystal-sandwich configuration, and directed to single-mode fibres by two flipper mirrors. The fibre launchers are mounted on z translation stages to allow optimisation of the non-classical interference between the photons once they are injected into the gate apparatus.

10.3 Experimental design

In my experiment, we aimed not only to demonstrate and characterise the new CZ architecture, but also to use the gate to perform fully discriminating Bell-state analysis, i.e. a single-shot measurement which (non-deterministically) projects a two-qubit state into the Bell basis. The basic apparatus can therefore be broken up into two main stages—the entangled source and the entangling quantum gate, or state preparation and state discrimination.

10.3.1 State preparation—the entangled source

To create polarisation-entangled photon pairs, we used a fibre-coupled version of the Type-I, crystal-sandwich scheme introduced in Ref. [22], very similar in many respects to the fibre-coupled sources I described in Chs 7 and 8. We used a pair of 0.6mm-thick BBO (β -barium borate crystals which were cut for Type-I phase matching (optical axis at 33.9° ; half-opening angle $\sim 3^\circ$) and pumped by a cw Ar^+ -ion UV laser ($\lambda_p=351.1\text{nm}$). We then coupled the down-conversion photons into single-mode fibres to improve their spatial mode, and thereby also the spatial mode overlap at the central PPBS of the CZ gate, which is required for good non-classical interference. A detailed schematic of the source is illustrated in Fig. 10.4.

Not surprisingly, the important considerations for mode-matching in this experiment were also similar to those in the earlier experiments [see Secs 7.3 and 8.4 for details]. Like in the hyperentanglement experiment, the polarisation entanglement was maximised by using a

relatively loose beam waist in the pump beam to limit the distinguishability of the two perpendicularly oriented BBO crystals. However, this did not really matter here, because we were not interested in the spatial degree of freedom, and so were not constrained to couple a large section of the d.c. cone anyway. Moreover, this also simplified the mode-matching because we did not have to align a hologram to the beam, and could therefore use shorter arm lengths. To balance the various considerations, we chose to image modes with a $40\mu\text{m}$ (w_0) beam waist at the crystal, with the pump focussed⁸. The d.c. mode-matching lenses ($f=400\text{mm}$) were placed $\sim 620\text{mm}$ from the crystal, with the fibre-coupling lenses $\sim 970\text{mm}$ beyond them.

The “input” fibre couplers were mounted on z -axis translation stages (i.e. along the propagation axis), because to demonstrate the non-classical interference required in the CZ gate, the beam path lengths had to be matched to less than the coherence length of the d.c. photons⁹ ($L_{\text{dc}}=\lambda^2/\Delta\lambda\sim 1.4\text{mm}$). For fine-tuning the path-length difference, one of these was a high-precision, computer-controlled translation stage. By using the input fibre couplers, the path lengths could be tuned without affecting the optical alignment of the gate, where spatial mode-matching is so crucial. The disadvantage is that coupling efficiency can now be affected. This effect was negligible with the manual translation stage, but because the computer-controlled stage was more difficult to mount, reasonable coupling was only possible over $\sim 15\text{mm}$ of its motion. We used the manual translation stage to ensure that the lengths were approximately matched when the computer-controlled stage was centred, and we aligned the fibres to give optimal coupling efficiency in these positions.

The polarisation state of the d.c. photons at the input fibre couplers was of the form,

$$|\psi\rangle \sim |HH\rangle + \varepsilon e^{i\phi}|VV\rangle. \quad (10.17)$$

We used the coarse adjustment of the fibre polarisation controllers (“fibre wings”) to ensure that H and V at the fibre input were still H and V , respectively, when they emerged from the fibres. So the two-photon polarisation state exiting the fibres had approximately the same form as the input state (barring a potentially different value for ϕ). The wave plates at the fibre outputs were then used to prepare the desired entangled state. We were able to control the phase, ϕ , by using the result obtained in Sec. 4.6.6, Eq. (4.35). Ultimately, because the fibre wings only allowed a coarse level of control, we actually used the wave plate settings from Eq. (4.35) as a starting point in a numerical search for the optimal settings based on a tomographic reconstruction of the state at the output of the fibres.

⁸We used a $f=1\text{m}$ lens, 1.125m before the crystal. The pump waist estimate was based on a rough calculation with a slightly expanding beam, and because of this divergence, the focus was slightly after the crystal. More accuracy was not necessary, since the effective pump spot is in fact determined by the imaged mode, but some mismatch may have reduced the efficiency of the fibre-coupling.

⁹For most of the experiments described in this chapter, the data was recorded using interference filters of 4nm in one arm and 0.36nm in the other (FWHM, centred at 702.2nm). Using coincidence detection, this limited the bandwidth of *both* d.c. photons to 0.36nm to maximise the visibility of the non-classical interference in the CZ gate, while keeping count rates as high as possible (broader filters generally have higher peak efficiencies). Some of the early data (not involving the CZ gate) was collected using two 4nm filters—this will be noted in the appropriate places.

10.3.2 State discrimination—the entangling (disentangling) gate

A detailed schematic of the experimental apparatus for the entangling CZ gate has already been included as Fig. 10.1(c). To optimise the spatial mode matching at the central PPBS, we matched the incident beams by using identical output fibre couplers positioned exactly the same distance from the central PPBS. In more recent experiments, this has been further improved by collecting the gate output into single-mode fibres.

We also coupled beams from a 633nm HeNe laser into the d.c. collection fibres to align the gate. In particular, we aligned the position and tilt of the central PPBS by optimising the classical interference of the visible beams in the near *and* far field¹⁰. We achieved extremely good visibility, but the slowly pulsing, random fluctuations of the intensity made this too difficult to quantify¹¹.

Since the d.c. wavelength differed quite substantially from the HeNe at 633nm, some further alignment was needed when we switched to photon counting. Unfortunately, the PPBS position was too sensitive to adjust based on small coincidence count rates (and there was no repeatability in the mount, because the actuators were too coarse and had no scale markings). Therefore, we tweaked the alignment using the tilt and translation of one of the input beams, which we could control with the output fibre couplers. We adjusted the position of the beam with precision translation stages, and the tilt of the beam by displacing the coupling lens relative to the end of the fibre, although this too proved to be quite difficult (discussed more in the next section). The divergence of the beam could be optimised by fine-tuning the separation of the fibre from the coupling lens.

Using flipper mirrors¹², we aligned an alternative beam path from the fibres to the detectors [not shown in Fig. 10.1(c)] which did not pass through the CZ gate [the yellow box in the figure]. This allowed us to switch between looking at the input and the output of the gate as necessary. Since the alternative paths were reflected off the flipper mirrors, switching between paths had no effect on the gate alignment.

In the photon counting, we used a coincidence window of 5.0ns and made no correction for accidental counts¹³.

¹⁰This was only possible because the two coherent alignment beams came from the same HeNe laser, split using a half-wave plate and a polarising beam splitter. Because the PPBS was not a 50:50 beam splitter, the intensity of the two beams were adjusted so they made equal power contributions to the output port.

¹¹These slowly pulsing fluctuations were actually extremely useful. The interferometer in alignment laser beams was not stable, since it consisted of several mirrors and two separate optical fibres in each arm. However, we needed the random path-length fluctuations to see the visibility when aligning the PPBS. Because the CZ gate is *not* sensitive to wavelength-scale fluctuations, these did not affect the gate operation.

¹²NewFocus 1" Flipper Mount, Model 9891

¹³The rate of accidental coincidences is affected by the singles rates, and is therefore different in every situation. If desired, it can be estimated using a simple formula: $A = S_1 S_2 \times \text{coincidence window}$.

power (mW)	pump poln	detector config.		singles (s ⁻¹)		coinc. (s ⁻¹)	eff. η	brightness factor
		det #1	det #2	det #1	det #2			
100	<i>H</i>	FS [‡]	FS [‡]	192k	192k	7800	0.041	33
100	<i>V</i>	FS [‡]	FS [‡]	163k	166k	5900	0.036	25
50	<i>V</i>	FS	FS	116k	102k	5000	0.049	42
50	<i>V</i>	FS	MMF	115k	64k	2400	0.038	20
450	<i>V</i>	FS	MMF	850k	475k	11.2k	0.025	10
460	<i>V</i>	FS	SMF	810k	25k	1000	0.04	0.9
460	<i>V</i>	SMF	SMF	28k	25k	1000	0.04	0.9

Table 10.1: Typical count rates during the alignment process. This data was recorded using interference filters with 4nm bandwidth (FWHM, centred at 702.2nm).

10.4 Results

Throughout this section, I have normally reported tomographic reconstructions based on the “fixed weights” (FW) penalty function [defined in Sec. 3.8]. Where appropriate, I have also reported results based on standard maximum likelihood (ML) reconstructions [in square brackets].

10.4.1 Characterising the source

We had some significant difficulties in setting up our Type-I down-conversion source. It did not perform as well as the similar sources which we aligned for the experiments in either Ch. 7 or Ch. 8, particularly in terms of brightness. Ultimately, it seems that most of our problems were caused by poor quality nonlinear (BBO) crystals (from EKSMA Co.), although this is very difficult to confirm without buying new crystals. Some evidence, however, is provided by looking at typical coincidence count rates with the source in different coupling configurations [Tab. 10.1]. The brightness factor (B.F.) is a scaled indication of intensity designed to allow comparison between different down-conversion sources. It is defined:

$$\text{B.F.} = \frac{[\text{coincidence count rate}]}{[\text{pump power}] \times [\text{d.c. bandwidth}] \times [\text{crystal thickness}]}, \quad (10.18)$$

and is generally reported in units of (mW.nm.mm.s)⁻¹. All of the measurements in Tab. 10.1 were taken using 4nm interference filters (FWHM, centred at 702nm) in front of free space detectors. The detector configurations marked “FS[‡]” refer to free space detectors with no polarisation analysis and no irises in front of the collection lenses. All other configurations involve detectors with polarisation analysers set to measure the appropriate state (i.e. maximum counts), with 12.5mm irises for those marked “FS”. In the MMF and SMF configurations, the power was coupled out of the fibres using a collimator and directed at the free space detectors.

In both brightness and efficiency, this source was considerably weaker than the blue diode source described in Ch. 7, which produced comparable count rates with 1/20 the power

(albeit with slightly broader filters) [cf. Tab. 7.1; with FS[†]/FS[†] config., B.F.~133; with SMF/SMF config., B.F.~21]. This effect was compounded, because we needed to use even narrower filtering of the d.c. photons (0.36nm FWHM) to maximise the stability of the CZ gate. We had to rely on the high power of the Ar⁺ ion pump laser to produce workable count rates, which had to be as high as possible, since the non-deterministic CZ gate reduced the output counts by a factor of 9 in principle¹⁴. Furthermore, under identical conditions, the “identical” crystals (pumped by orthogonal polarisations) gave consistently different down-conversion efficiencies [see first two lines in the table]. This continued throughout the experiment and we had to compensate by tuning either the pump polarisation or the fibre couplers.

We also noticed that we were not able to remove all distinguishability between the two cones, although this was only observable with the high spatial sensitivity of the single-mode fibre couplers. We first pumped one crystal and optimised the fibres to both arms of the down-conversion cone. Then we tilted the crystals in the appropriate direction to tune the opening angle of the second d.c. cone, and we observed that the singles rates peaked at slightly different angles. If we set the opening angle of the second cone to optimise the coupling into one coupler, then different horizontal positions of the second coupler would be required to optimise the coincidences for the two cones. Normally these two positions were separated by around $\Delta x \sim 0.1\text{mm}$, and with one cone optimised, the count rate for the other cone was normally slightly less than half its maximum value. We were unable to find the cause for this effect, but one possible suggestion is that the crystals were slightly wedged so that the pump travelled at a slightly different direction through the two crystals. At a practical level, we positioned the fibre between these maximum values at the point which balanced the count rates when pumping the different cones. I note here that we also observed this effect to some degree in the hyperentanglement experiment [Ch. 8].

Figure 10.5(a) shows the down-conversion polarisation state measured with free-space detectors (FS/FS config.) with irises set to around 2.8mm diameter. As expected, it is a (nearly) maximally entangled state ($T=0.901\pm 0.005$, $N=0.949\pm 0.002$, $S_L=0.065\pm 0.003$) of the form: $|HH\rangle + \exp(i\phi)|VV\rangle$. In contrast, when coupled through single-mode fibres (SMF/SMF config.), the polarisation state has been scrambled [Fig. 10.5(b)], while the entanglement and purity have been maintained ($T=0.876\pm 0.005$, $N=0.934\pm 0.003$, $S_L=0.085\pm 0.004$).

After the source was coupled into the CZ gate apparatus, we used fibre wings to ensure that the H and V polarisations were roughly maintained by the fibres. In this way we removed most of the unwanted HV and VH components. A typical density matrix for the state directly after the fibres is shown in Fig. 10.5(c) ($T=0.929\pm 0.015$, $N=0.960\pm 0.008$, $S_L=0.05\pm 0.01$). Although this state has an optimum fidelity of $F_{\text{opt}}=0.980\pm 0.004$ with a maximally entangled state, its fidelity with the standard Φ^+ Bell state is only $F_{\Phi^+}\sim 0.25$, and the difference is almost entirely due to the phase on the main coherence terms. Using Eq. (4.35), however, we were able to mostly correct this phase and prepare the state shown in Fig. 10.5(d) ($T=0.965\pm 0.010$, $N=0.981\pm 0.005$, $F_{\text{opt}}=0.990\pm 0.003$, $S_L=0.024\pm 0.007$), which has a fidelity with Φ^+ of $F_{\Phi^+}\sim 0.96$. The small remaining imaginary components are probably due to the inaccuracies associated with characterising the effect of the alternative

¹⁴In practice, this is probably further reduced by unavoidable losses such as reflection losses at each of the optical elements.

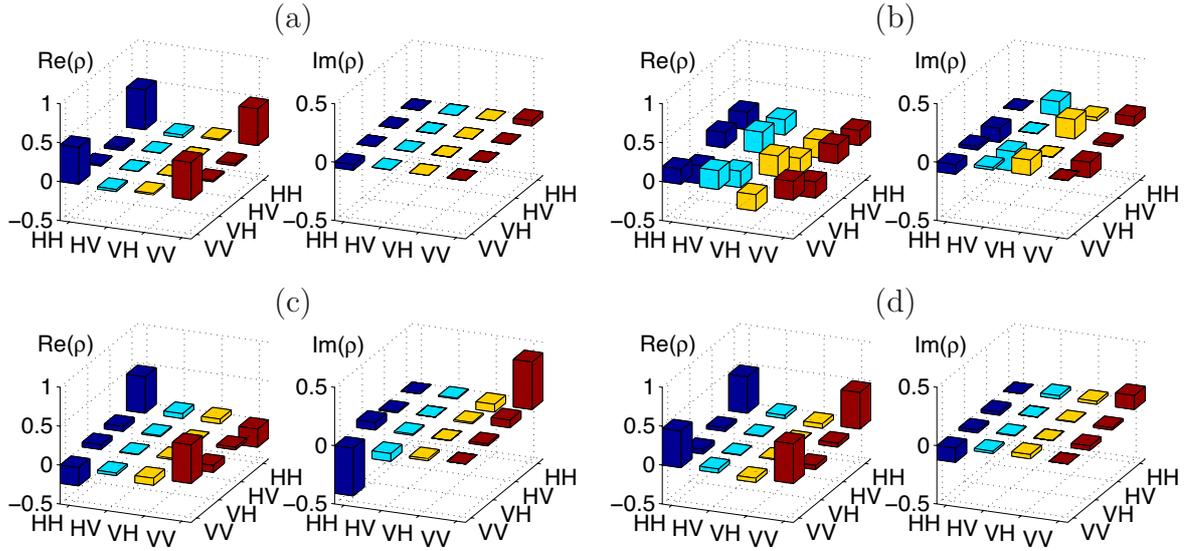


Figure 10.5: Characterising the crystal-sandwich entangled source—near-maximally entangled states: (a) measured with free-space detectors; (b) coupled through SMFs; (c) coupled through SMFs with polarisation controllers (fibre wings); and (d) using wave plates to correct the phase at the fibre outputs.

beam paths (see below).

One property of a fibre-coupled photon source is that it can be very sensitive to the spatial alignment of the down-conversion apparatus. In our experiments, this had some unexpected but profound implications. We observed strong, cyclic oscillations in our count rates (period ~ 30 minutes) which correlated with the temperature oscillations induced by air-conditioning in the laboratory (stability $\sim 0.5\text{--}2^\circ\text{C}$ ¹⁵). The peak-to-peak amplitude of these count rate oscillations could be up to 30% of the mean rate which would obviously have a significant impact on results. To minimise the effect of these fluctuations, where possible, we took all measurements in complete POVM sets [e.g. for tomographies, we used the measurement order described in Tab. 3.1]. By ensuring that the integration time for a complete POVM set was much less than the air-conditioner cycle and normalising within each one, the counts could be converted into probabilities with much greater accuracy. Of course, this requires the assumption that the main effect of the fluctuations is just to cause a simple scaling in the overall count rate.

To check the success of this technique, I performed eight successive tomographies of the same entangled state emerging from the single-mode fibres. In Figure 10.6, I have displayed the total count rate as quantified by the POVM normalisations as a function of time (the number of tomographies). Each tomography took about 14 minutes to run (10s integration + ~ 14 s moving wave plates per measurement), with a short break between some of the measurement sets. The plot shows oscillations with a peak-to-peak amplitude $\sim 22\%$ of mean value, but successive normalisations generally differ by less than

¹⁵Originally, the air-conditioner specifications were designed and verified to keep the lab at a stable temperature to an accuracy of $\pm 0.5\text{--}1^\circ\text{C}$ (something over three years ago). Recently, however, the stability has been remeasured to be around $\pm 1.5\text{--}2^\circ\text{C}$. Our experiments were carried out approximately half-way between these two times, so we do not have data specifically relevant to that time period.

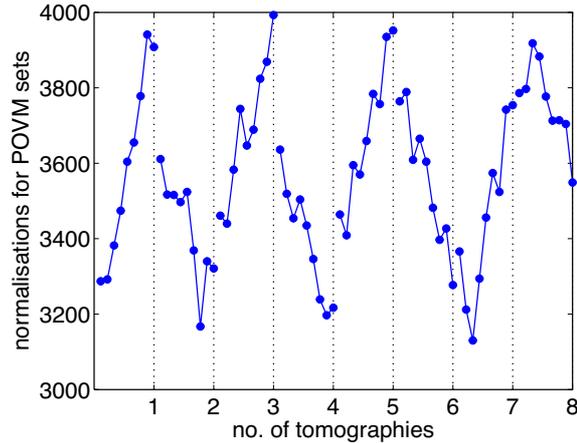


Figure 10.6: The effect of the laboratory air-conditioner cycle on tomographic counts.

$\sim 3\%$ showing that the probabilities would be significantly more accurate when normalised with the POVM sets. The oscillation period agrees well with the air-conditioning cycle.

The tomographic reconstructions demonstrated a high degree of repeatability with

$T=0.96\pm 0.01$, $N=0.977\pm 0.007$, $F_{\text{opt}}=0.989\pm 0.004$, $S_L=0.028\pm 0.009$, $Q=0.70\pm 0.17$, and the average overlap between the different reconstructed states was 0.993, where the errors are the standard deviations in the sample set of eight values. For comparison, the errors determined by a Monte-Carlo simulation of one of the eight example data sets were: $\Delta T=0.01$, $\Delta N=0.007$, $\Delta F_{\text{opt}}=0.004$, $\Delta S_L=0.009$, $\Delta Q=0.14$, and the average overlap between different reconstructions was 0.994. Thus, the repeatability of the measured tomography agreed very well with what could be expected as a result of Poisson-distributed noise.

This technique works best with the smallest integration time per POVM set. In later measurements, we used integration times of up to 50s (+ wave plate motion), so the air-conditioner cycle could be expected to have a greater effect on these results. In some cases we were able to replace a single, long POVM integration with a series of repeated, shorter measurements, at the cost of more “wasted” time spent moving wave plates¹⁶.

In Reference [5], we showed that the operation of the two-photon gate was far more sensitive to the quality of the mode overlap in the central non-classical interference, than to other real-world effects such as incorrect beam-splitter reflectivities [see also Sec. 10.2], and this was supported by later experiments [10, 11]. Because we used narrow bandwidth interference filters to select the frequency of the d.c. beams, and high-precision translation stages to match the path lengths traversed by the photon pairs, the most significant source of mode-mismatch at the central PPBS in our experiment was spatial misalignment of the incident beams. In the all-polarisation gate design, this is the only interference condition and is the most difficult part of the gate to align.

As described earlier, we performed the first stage of the gate alignment using bright, 633nm HeNe lasers, and we would typically produce non-classical visibilities of $\sim 60\text{--}65\%$

¹⁶Our rotation stages were limited to move at around $4^\circ/\text{s}$, so this was an important consideration, but much faster rotation stages are also available if necessary.

when we switched to the d.c. beams ($\sim 75\text{--}81\%$ of the maximum expected visibility of 80% [Sec. 10.2.1]). We then optimised the interference by translating the coupler and adjusting the tilt of the beam emerging from the fibre coupler. Unfortunately, the tilt of the beam proved to be extremely difficult to align repeatably because of the low count rates and the poor quality of the actuators controlling the lens position of the Thorlabs KT110/M fibre launchers. Ultimately, we chose to optimise the d.c. interference using only the translation of the coupler and the beam divergence (by fine-tuning the distance between fibre and coupling lens—also extremely sensitive). During our experiments, optimised visibilities ranged between $\sim 65\%$ and 73% ($\sim 81\text{--}91\%$ of max.). Because this is the only interference condition in our CZ gate, this non-optimal overlap of spatial modes at the central PPBS was one of the main sources of error in our experiments. By creating distinguishability between the d.c. beams, the mode mismatch introduced mixture into the output states produced by superposition input states.

In recent experiments by my colleagues, the output modes from the CZ gate have been coupled again into single-mode fibres. This improves the quality of the non-classical interference by ensuring that all detected photons originate from the same spatial mode at the central PPBS. The alignment procedure is also much simpler, being reduced to a series of optimisations of coupling efficiencies. Typical visibilities achieved using this arrangement¹⁷ are $\sim 95\%$ of the maximum value (80%).

10.4.2 Characterising the alternative beam paths

As I mentioned in the previous section, we used the alternative optical paths around the CZ gate to look at the state emerging from the fibres. Each path comprised three extra reflections of steering mirrors, and thus in fact implemented a general polarisation-sensitive rotation of the form:

$$U_{\text{reflections}} = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{bmatrix}. \quad (10.19)$$

Therefore, to accurately characterise the polarisation state, we needed to measure and account for these phases for each path. To do this, we prepared the DD polarisation state (with polarisers and the preparation wave plates) and tomographically reconstructed the state passing through the alternative beam paths. We then calculated the phases using the average of several tomographies. Typical results for the two arms were 1.58 ± 0.02 and 0.11 ± 0.06 , respectively. With these measurements, we were able to “undo” the effects of the extra reflection and the preparation wave plates and calculate the state exiting the fibres.

10.4.3 The CZ as an entangling gate

One of the most obvious uses of a CZ gate is to generate entanglement between systems in a separable state. In fact, looking at how much entanglement is generated by a quantum gate is one standard way of gauging how well the gate is performing (see, e.g., Ref. [23]

¹⁷Private communication from Rohan Dalton.

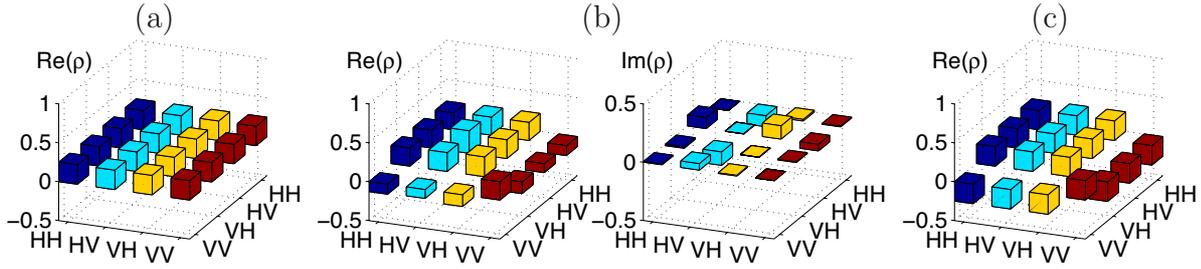


Figure 10.7: An example entangled state created by our CZ gate: (a) the input state, $|DD\rangle$; (b) the measured output state; (c) the expected output state, $|HD\rangle+|VA\rangle$. (The imaginary parts for (a) and (c) were zero and therefore omitted.)

input state	tangle (T)	negativity (N)	linear entropy (S_L)	fit quality (Q)
DD	0.57 ± 0.03	0.67 ± 0.02	0.35 ± 0.02	0.74 ± 0.16
DR	0.57 ± 0.03	0.71 ± 0.02	0.32 ± 0.02	0.90 ± 0.16
RD	0.58 ± 0.03	0.61 ± 0.01	0.42 ± 0.01	0.97 ± 0.17
RR	0.53 ± 0.02	0.66 ± 0.01	0.37 ± 0.01	0.65 ± 0.15
input state	target fidelity (F_t)	optimum fidelity (F_{opt})	model #1 fidelity (F_1)	model #2 fidelity (F_2)
DD	0.797 ± 0.008	0.831 ± 0.009	0.947 ± 0.005	0.971 ± 0.004
DR	0.791 ± 0.008	0.854 ± 0.009	0.922 ± 0.006	0.967 ± 0.004
RD	0.789 ± 0.009	0.790 ± 0.008	0.980 ± 0.003	0.972 ± 0.004
RR	0.811 ± 0.008	0.811 ± 0.007	0.969 ± 0.005	0.984 ± 0.004

Table 10.2: Typical entangled states created by the PPBS CZ gate.

and references therein). Figure 10.7(b) shows an example entangled state created by sending DD [Fig. 10.7(a)] into the PPBS CZ gate (the expected output state is $HD+VA$ [Fig. 10.7(c)]). This state has a tangle of $T=0.57\pm 0.03$, showing that it is quite strongly entangled; a linear entropy of $S_L=0.35\pm 0.02$, indicating that it is also quite mixed; and a fidelity of $F_t=0.797\pm 0.008$ with the target state in (c). Table 10.2 summarises the properties of this state and three other typical entangled states that we created¹⁸. At a preliminary level, these results already show that our gate operated quite effectively as an entangling gate, comparing well in quality with the results of Ref. [10].

There are two interesting points to note about the measured state in Fig. 10.7(b). First, the non-zero imaginary components confirm that the photons experienced polarisation-sensitive phase shifts as they reflected off the optical elements in the gate, as predicted in Sec. 10.2.3. Their size, however, indicates that the polarisation rotations were small. Supporting this observation, the state's optimum fidelity with a maximally entangled state is $F_{\text{opt}}=0.831\pm 0.009$, which is larger than the target state fidelity, but only by a small amount. The second point to note is that most of the coherences (off-diagonal terms) are strong (similar in size to the corresponding populations), except for the coherences

¹⁸These four data sets were in fact taken as part of the full process tomography described in the next section (tomography set #1).

with the VV population (the negative terms), which seem to be about half the size of the others. While lower coherences are expected in states that are mixed, this also gives an indication of the type of mixture present. For example, if the source of the mixture were a depolarising process, then all of the coherences would be affected, which is not the case here. In fact, a state like Fig. 10.7(b) could be produced by an incoherent (classical) mixture of the target output state (the main contribution) [(c)] and the unchanged input state (a smaller amount) [(a)], and this provides a clue to understanding the gate operation.

As I have already discussed, the main source of error in our gate was the imperfect spatial mode matching between the input qubits at the central PPBS. For example, before taking the above data, the non-classical interference visibility was $\sim 70\%$ ($\sim 87\%$ of max.). The effect of this spatial mode mismatch can be understood by recognising that the interfering qubit modes can be broken up into distinguishable and indistinguishable spatial modes. Most of the time, the two photons are indistinguishable, so the non-classical interference works perfectly, and the CZ gate implements the correct transformation, $A_{\text{exp}} \sim (X \otimes X) A_{\text{CZ}}$. However, with some error probability, p , the photons are distinguishable and the incident qubit modes do not interfere. Now, the operation components involving the reflected photons must be separated from those involving the transmitted photons and then combined incoherently, i.e.,

$$\begin{aligned} \rho_{\text{out}} &= A_{\gamma} \rho_{\text{in}} A_{\gamma}^{\dagger} + A_{\tau} \rho_{\text{in}} A_{\tau}^{\dagger}, & (10.20) \\ \text{where } A_{\gamma} &= \gamma X \otimes X \\ \text{and } A_{\tau} &= -\tau X \otimes X |HH\rangle\langle HH| = -\tau |VV\rangle\langle HH|. \end{aligned}$$

A_{γ} and A_{τ} are the operators involving the photon pairs which are either both reflected or both transmitted, respectively¹⁹, n.b. $A_{\text{exp}} \sim A_{\gamma} + A_{\tau}$. Therefore, the CZ gate with imperfect mode matching can be modelled by the following quantum operation,

$$\mathcal{E}_1(\rho_{\text{in}}) = p A_{\text{exp}} \rho_{\text{in}} A_{\text{exp}}^{\dagger} + (1-p) (A_{\gamma} \rho_{\text{in}} A_{\gamma}^{\dagger} + A_{\tau} \rho_{\text{in}} A_{\tau}^{\dagger}). \quad (10.21)$$

This model can then be extended by allowing the process to implement polarisation-sensitive phase shifts in each arm of the gate, i.e.,

$$\begin{aligned} \mathcal{E}_2(\rho_{\text{in}}) &= U \mathcal{E}_1(\rho_{\text{in}}) U^{\dagger} & (10.22) \\ \text{where } U &= U_1 \otimes U_2 = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\phi_1} \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 0 & e^{i\phi_2} \end{bmatrix}. \end{aligned}$$

These two models show good agreement with the four measured output states from Tab. 10.2. The fidelities between the measured states and those predicted by the two models are also recorded in the table (F_1 and F_2). The first, simpler model is closest when $p=0.128$, and the fidelities (F_1) show a marked increase over the target fidelities (F_t). The second model, however, provides the best agreement with the observed states. It is closest when $p=0.124$, $\phi_1 = -0.158$ and $\phi_2 = 0.247$, giving an average fidelity over the four entangled states of $\overline{F_2} = 0.974$. Using this error probability, it is possible to predict the non-classical interference visibility by calculating the expected minimum and

¹⁹Recall that $\gamma \equiv \gamma_H$ and $\tau \equiv \tau_H$ are the reflectance and transmittance, respectively, for the horizontal polarisations arriving at the PPBS.

maximum coincidence probabilities for a scan through the HOM dip: $p_{\max} = p_{\text{dist}}$ and $p_{\min} = p \cdot p_{\text{dist}} + (1-p) \cdot p_{\text{indist}}$. The latter is determined by noting that the non-interfering error photons will produce a flat background to the dip. Therefore, for an error probability of $p=0.124$, and a beam splitter with $1/3$ reflectance ($p_{\text{dist}}=5/9$, $p_{\text{indist}}=1/9$), the expected visibility is 70.1%, which agrees extremely well with the measured visibility of $\sim 70\%$.

10.4.4 Process tomography

The most comprehensive way to characterise the operation of a quantum gate is to use quantum process tomography [see Sec. 3.5]. In our experiment, we used 16 separable states as inputs to the CZ gate and analysed the output for each with the over-complete set of 36 measurements described in Sec. 3.7. An example measured process matrix is shown in Fig. 10.8(a), where we used 30 second integration times for each measurement.

Following the conclusions of Ch. 9 (also Ref. [24]), we quantify the performance of our CZ gate by calculating the two complementary measures²⁰: the J process fidelity, $F_{\text{pro}}(\mathcal{E}, U_{\text{CZ}}) = F(\rho_{\mathcal{E}}, \rho_{\text{CZ}})$, and the J process distance, $D_{\text{pro}}(\mathcal{E}, U_{\text{CZ}}) = D(\rho_{\mathcal{E}}, \rho_{\text{CZ}})$ [see Sec. 9.2 for definitions]. Furthermore, we can use the J fidelity to calculate the average fidelity between expected and actual output states: $F_{\text{ave}}(\mathcal{E}(\psi), U_{\text{CZ}}(\psi)) = (4F_{\text{pro}}(\mathcal{E}, U_{\text{CZ}}) + 1)/5$ [see Sec 9.3.2, Eq. (9.17)]. For the process in Fig. 10.8(a), the process fidelity with the ideal is $F_{\text{pro}} = 0.759 \pm 0.003$ [0.740 ± 0.003], the process distance is $D_{\text{pro}} = 0.309 \pm 0.004$ [0.323 ± 0.004], and the average fidelity is $F_{\text{ave}} = 0.807 \pm 0.003$ [0.792 ± 0.003].

As I have already discussed, the PPBS CZ gate produces systematic, fixed rotations of the input and output polarisations. In practice, these have no effect on gate quality and, if necessary, could be corrected with appropriate wave plates. To demonstrate this, we modelled their effect numerically, identifying single qubit unitary corrections [for Fig. 10.8(a)] which increased the process fidelity to 0.778 ± 0.003 [0.758 ± 0.004] and average gate fidelity to 0.822 ± 0.003 [0.806 ± 0.003], and decreased the process distance to 0.258 ± 0.003 [0.277 ± 0.004]. These optimal quantities provide information about the maximum entangling power of the gate. In some sense, they are the process analogue of the state-based optimum fidelity with a maximally entangled state (F_{opt}).

Given the results of the previous section, it is not surprising that removing the polarisation rotations only makes a small difference to the values of the process measures. I have already shown that the major imperfections in the gate operation result from mismatched spatial modes at the central PPBS which reduce the visibility of the non-classical interference. This is further supported by the form of the process matrix in Fig. 10.8(a) which deviates from the ideal process matrix [Fig. 10.8(b)] mainly in the II population, which is larger than expected (0.37 instead of 0.25). As discussed above, with some probability, the control and target qubits do not interfere and the circuit implements a transformation similar to the identity operation, and this process is incoherent with the correct operation of the gate. Most strongly affected will be the input states which involve superpositions of HH with other polarisation components (e.g. those which give entangled output states).

²⁰We use the J process measures (based on the process matrix) rather than the stabilised S process measures (worst-case) because they are significantly easier to calculate.

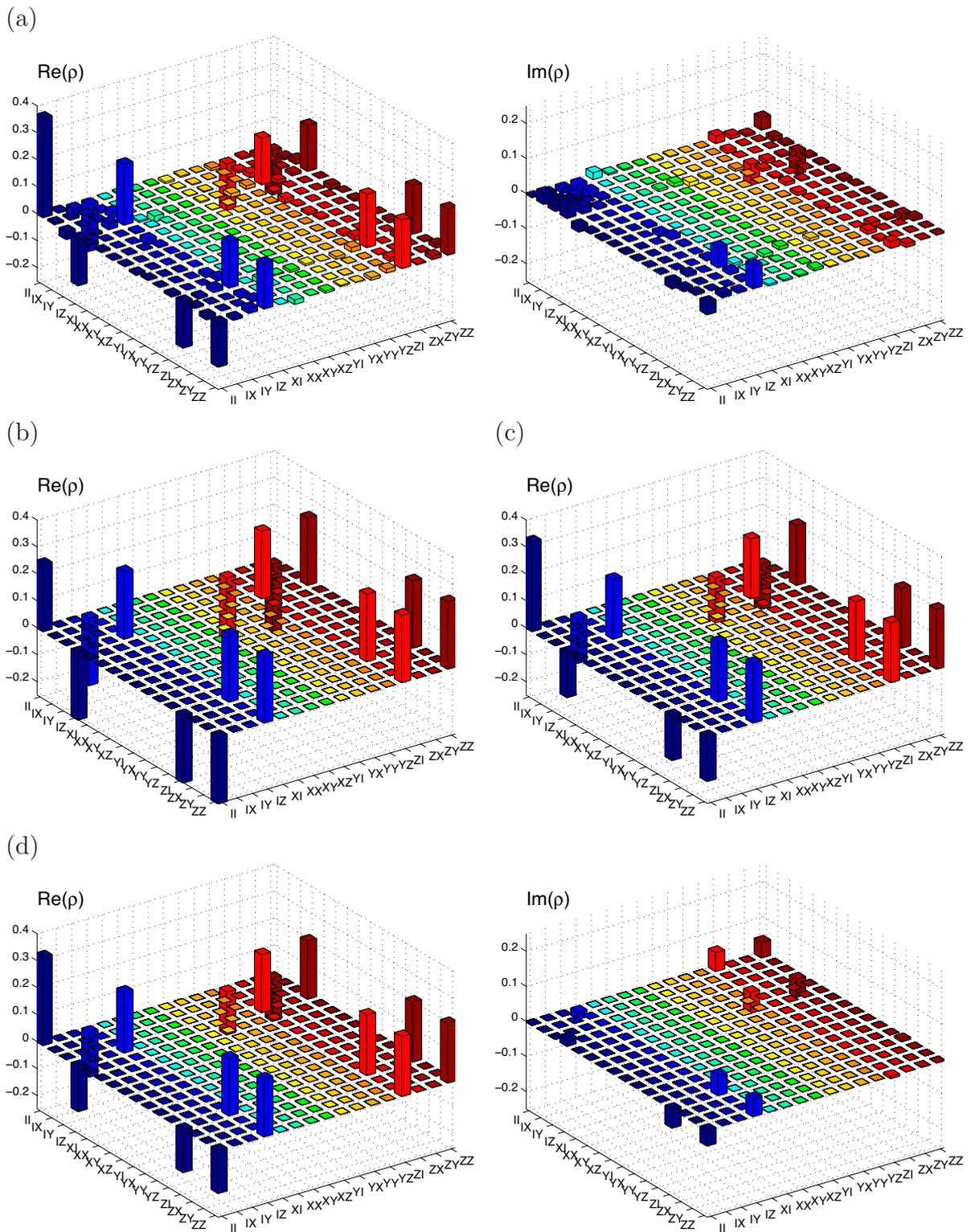


Figure 10.8: Quantum process tomography of the cw CZ gate: (a) measured process (a few elements of the imaginary part of the process matrix are on the order of 0.05, the average is ~ 0.005); (b) ideal CZ gate ($\text{Im}\{\rho\}=0$); (c) numerical fit for model #1 ($\text{Im}\{\rho\}=0$); (d) numerical fit for model #2. For ease of visualisation, the process matrices have been rotated to correct for the extra $X \otimes X$ produced by the PPBS implementation.

tomo set	integration time	tomo duration	process fidelity (F_{pro})	process distance (D_{pro})	average fidelity (F_{ave})	fit quality (Q)
#1	30s	~7h	0.759 ± 0.003 [0.740 ± 0.003]	0.309 ± 0.004 [0.323 ± 0.004]	0.807 ± 0.003 [0.792 ± 0.003]	1.64 ± 0.04 [1.50 ± 0.04]
#2	50s	>10h	0.758 ± 0.003 [0.745 ± 0.003]	0.312 ± 0.003 [0.322 ± 0.003]	0.807 ± 0.002 [0.796 ± 0.002]	1.95 ± 0.05 [1.82 ± 0.04]
#3	30s	~7h	0.758 ± 0.004 [0.741 ± 0.003]	0.313 ± 0.004 [0.325 ± 0.003]	0.806 ± 0.003 [0.793 ± 0.003]	1.55 ± 0.05 [1.42 ± 0.04]

Table 10.3: Consistency results for the PPBS CZ gate—three successive process tomographies.

To examine this more quantitatively, I have compared the measured process matrix with the models introduced above.

Using the first model, which only includes the effect of imperfect mode matching, provides substantially better agreement with the measured process. Based on a numerical fit, the model is closest when $p=0.125$ [Fig. 10.8(c)] giving a process fidelity of $F_{\text{pro}}=0.910 \pm 0.002$ and a process distance $D_{\text{pro}}=0.221 \pm 0.003$. This agreement is again substantially improved. The second model, which also includes the polarisation phase shifts, provides some further improvement, but only a small amount. The closest fit is obtained when $p=0.124$, $\phi_1 = -0.035$ and $\phi_2=0.271$ [Fig. 10.8(d)], giving a process fidelity of $F_{\text{pro}}=0.923 \pm 0.002$ and a process distance $D_{\text{pro}}=0.185 \pm 0.003$. Note that the rotations produced by this second model explain all of the more pronounced components in the imaginary part of the measured process matrix. The fit parameters for both models agree very well with the measured non-classical interference visibility of $\sim 70\%$ and with the values found for the entangled output states in the previous section. These results show again that spatial mode mismatch, which can be improved by further spatial filtering, is the main source of error in the experiment, but the fidelity of 0.923 for the second model suggests that it may not be the only one.

Throughout our experiments, one of the main difficulties was the low count rates which led to longer data runs. For example, the total duration of the process tomography shown in Fig. 10.8(a) was about 7 hours (using 30 second counts²¹). Consequently, long-term drift in the alignment or operating conditions of the experiment was a significant concern and was perhaps responsible for further reduction in the performance quality of the gate. Although the effects of short term temperature cycling driven by the air conditioning were significantly reduced by measuring the data in POVM sets, we suspect there was a significant difference between the experimental operation under day-time and night-time ambient temperatures, not wholly removed by the air conditioning control.

To gauge the effects of any long-term drift on the results of the process tomography, we performed a series of three successive tomography measurements and the results for the different reconstructions are summarised in Tab. 10.3 (Fig. 10.8(a) shows tomography #1). They are, perhaps surprisingly, very consistent—the fidelities between the different reconstructed process matrices are $F_{\text{pro}}(\#1, \#2)=0.995$, $F_{\text{pro}}(\#1, \#3)=0.997$ and

²¹Moving the wave plates in between measurements took about 15 seconds, on average.

$F_{\text{pro}}(\#2, \#3)=0.995$. This suggests that while the drift *during* a tomography may be detrimental, the drift *between* different tomographies is not substantial.

The reconstruction fit qualities also provide some interesting information about the process tomographies. The values in Tab. 10.3 are significantly larger than the average fit quality for states reconstructed from the constituent data sets for the 16 different input states: $\overline{Q}=0.80$ [0.78] for tomography #1, $\overline{Q}=0.92$ [0.91] for the #2 tomography, and $\overline{Q}=0.66$ [0.65] for tomography #3²². In Section 3.6.1, I showed that, when tomographic data are limited by Poissonian statistics, the fit quality should have a mean of $\overline{Q}=1$ and a standard deviation related to the number of tomographic measurements (M): $\Delta Q=1/\sqrt{2M}$. Not surprisingly then, these latter results for the state reconstructions agree quite well with this prediction²³ ($\Delta Q\sim 0.12$ for $M=36$). For the process tomographies, however, while the fit qualities are still reasonably close to 1, they lie well outside the expected range determined by $\Delta Q\sim 0.03$ for $M=576$. These observations suggest that the experiment was affected by drift on a medium time-scale which slightly altered the operating conditions of the gate between data sets.

This idea is further supported by analysing the predictive power of the process tomographies via a comparison of the measured output state reconstructions with the predicted output states for each input state. For example, for tomography #1, the average prediction fidelity and trace distance were $\overline{F}_{\text{pred}}=0.98\pm 0.01$ and $\overline{D}_{\text{pred}}=0.07\pm 0.03$, indicating that the reconstructed process matrix provides a good description of the observed gate operation. However, the predicted states were generally more mixed than the measured states (on average, the linear entropies were large by 0.03), which seems to be reflected more using trace distance than the fidelity. In other words, there was some process which did not affect the individual output state tomographies but which has shown up as extra decoherence in the reconstructed quantum process. A medium time-scale drift in the gate's operating conditions is a plausible explanation for these results.

The process purity

To quantify the effect of decoherence in quantum processes, in Ref. [11] we introduced the *process purity*: $P_{\text{pro}}(\mathcal{E}) = \text{Tr}\{\rho_{\mathcal{E}}^2\}$. In the terminology of Ch. 9 [24], the process purity is a *diagnostic measure*, i.e. it is *not* an appropriate candidate for the canonical measure of distance between two processes, but it has some useful properties and provides valuable information about an experiment. For example, by analogy with its state-based equivalent, it is immediately clear that $0 \leq P_{\text{pro}}(\mathcal{E}) \leq 1$. However, its most useful property

²²Note: for both state and process reconstructions, the fit qualities are poorer for the data taken with longer integration times (50s for #2), which is a result of the count rate oscillations induced by the air-conditioner cycle. The compensation provided by measuring and normalising in POVM sets [see Fig. 10.6 and related text] was less effective when using longer integration times.

²³In fact, when Poissonian fluctuations are the only noise source, the mean seems to be consistently less than 1. This is because the calculations in Sec. 3.6.1 actually determine the expected distance between the “true” quantum state and the measured data. If there were no noise, then the reconstructed state would indeed be the “true” quantum state. However, the noise “pulls” the results away from the true and towards the actual data, thus reducing the value of the fit quality. This effect is likely to be smaller in larger systems and more measurements.

provides its interpretation:

$$P_{\text{ave}}(\mathcal{E}) = \frac{P_{\text{pro}}(\mathcal{E})d + 1}{d + 1}, \quad (10.23)$$

where $P_{\text{ave}} = \overline{\text{Tr}\{\mathcal{E}(\psi)^2\}} = \overline{P(\mathcal{E}(\psi)^2)}$ [25] is the purity of the output state averaged over all pure input states. This can be shown as follows²⁴.

$$P_{\text{ave}}(\mathcal{E}) = \overline{\text{Tr}\{\mathcal{E}(\psi)^2\}} = \int d\psi \text{Tr} \left\{ \sum_{jk} E_j |\psi\rangle\langle\psi| E_j^\dagger E_k |\psi\rangle\langle\psi| E_k^\dagger \right\} \quad (10.24)$$

$$= \int d\psi \text{Tr} \left\{ |\psi\rangle\langle\psi| \sum_{jk} E_j^\dagger E_k |\psi\rangle\langle\psi| E_k^\dagger E_j \right\} \quad (10.25)$$

$$= \int d\psi F(\psi, \mathcal{E}^\dagger \circ \mathcal{E}(\psi)) \quad (10.26)$$

$$= F_{\text{ave}}(\mathcal{I}, \mathcal{E}^\dagger \circ \mathcal{E}) \quad (10.27)$$

$$= \frac{F_{\text{pro}}(\mathcal{I}, \mathcal{E}^\dagger \circ \mathcal{E}) d + 1}{d + 1} \quad (10.28)$$

In the last step, we use Eq. (9.17). Using similar steps to the first few lines, it can also be shown that $F_{\text{pro}}(\mathcal{I}, \mathcal{E}^\dagger \circ \mathcal{E}) = \text{Tr}\{\rho_{\mathcal{E}}^2\}$, which completes the proof. A simple, intuitive explanation of Eq. (10.23) derives from the key step of recognising that $P_{\text{ave}}(\mathcal{E}) = F_{\text{ave}}(\mathcal{E}^\dagger \circ \mathcal{E})$, where \mathcal{E}^\dagger is the time-reversed process. In other words, the average purity of the output state, and in turn the process purity, is a measure of the average reversibility of the process as determined by comparing the “undone” process with the identity operation. It is not surprising that $P_{\text{pro}}(U) = P_{\text{ave}}(U) = 1$ for all unitary processes. Note also that $P_{\text{pro}}(U \circ \mathcal{E}) = P_{\text{pro}}(\mathcal{E} \circ U) = P_{\text{pro}}(\mathcal{E})$.

The process purities for the three tomographies from Tab. 10.3 are $P_{\text{pro}}=0.659\pm 0.004$ [0.630 ± 0.004], $P_{\text{pro}}=0.659\pm 0.004$ [0.640 ± 0.004] and $P_{\text{pro}}=0.659\pm 0.005$ [0.635 ± 0.004], which is once again quite consistent between the three data sets. By comparison, the process purity of the numerically fitted models above are $P_{\text{pro}}=0.722$ and $P_{\text{pro}}=0.724$, respectively. Thus the main source of decoherence is the spatial mode mismatch, but the process reconstructions predict some extra mixture which is probably the effect of experimental drift during the long tomographies. Since the longer tomography (#2) does not have a lower purity, this again suggests that the problematic drift occurs on a medium time-scale (longer than a state tomography, shorter than a process tomography).

Further comment on PPBS architecture

A potential drawback of the PPBS architecture is that the beam splitter reflectances are determined in the manufacturing process, in contrast to schemes where a half-wave plate setting controls the beam splitters [10, 11]. We measured the PPBS reflectances in the cw experiment (Asahi Spectra; optimised for 702.2nm) to be within ± 0.01 of the required value ($\gamma=1/3$, normalised to output power), but for the pulsed experiment

²⁴Note that $[\mathcal{E} \circ \mathcal{F}](\rho) = \mathcal{E}(\mathcal{F}(\rho))$ is the *composition* of \mathcal{E} and \mathcal{F} —the two processes are implemented successively, first \mathcal{F} and then \mathcal{E} .

computation basis truth table		Bell measurement truth table	
input	output	input	output
$ HH\rangle$	$ HH\rangle$	$ \Phi'^+\rangle = HD\rangle + VA\rangle$	$ AA\rangle$
$ HV\rangle$	$- HV\rangle$	$ \Phi'^-\rangle = HD\rangle - VA\rangle$	$ DA\rangle$
$ VH\rangle$	$- VH\rangle$	$ \Psi'^+\rangle = HA\rangle + VD\rangle$	$ AD\rangle$
$ VV\rangle$	$- VV\rangle$	$ \Psi'^-\rangle = HA\rangle - VD\rangle$	$ DD\rangle$

Table 10.4: Ideal CZ gate truth tables [from Eq. (10.6)].

(Special Optics; optimised for 820nm), we found the reflectances were $\sim 0.28 \pm 0.01$. From the model in Sec. 10.2.3, a gate using such beam splitters gives a maximum process fidelity of $F_{\text{pro}}(\gamma=0.28) = 96\%$, which is still close to ideal. As we showed in Ref. [5], the CZ gate is quite forgiving of the exact splitting ratios, making it an eminently suitable gate to be realised with a PPBS architecture. In fact, the entangling quality of this gate is limited almost exclusively by spatial mode matching. Thus it is also a promising candidate for micro- or integrated-optical implementations, where non-classical mode matching in excess of 99% can be expected [19].

10.4.5 The CZ as a disentangling gate for Bell-state analysis

To complement the earlier experiments, as a final test of the PPBS CZ gate, we investigated its use in a disentangling capacity. In particular, we combined the gate with the cw entangled source described in Sec. 10.3.1 to test its ability to perform Bell-state analysis (measurement in an entangled basis) which plays an important role in many quantum information protocols such as teleportation. The theoretical operation of the CZ gate [from Eq. (10.6)] as a Bell analyser is summarised by the truth tables in Tab. 10.4 (cf. the computational basis truth table). When the appropriate maximally entangled states are injected into the gate, they are transformed into separable states which can then be distinguished by separable measurements. The entangled states which are discriminated by the CZ gate are just the usual four Bell states, with the second qubit rotated by a Hadamard.

Using the techniques described in Secs 10.3.1 and 10.4.1, we produced and characterised all of the four maximally entangled states from Tab. 10.4 as input states for the gate. The reconstructed density matrices [Fig. 10.9(c)] displayed an average target state fidelity of $\overline{F}_t = 0.965 \pm 0.006$, tangle of $\overline{T} = 0.958 \pm 0.008$, and linear entropy of $\overline{S}_L = 0.026 \pm 0.007$. For ease of visualisation, we have numerically rotated these states into the more familiar form by applying a Hadamard gate to the second qubit [Fig. 10.9(b)]. The top half of Tab. 10.5 summarises the basic properties of the entangled input states.

For each of the four input states in Fig. 10.9(c), we measured and characterised the output states of the gate [shown in Fig. 10.9(d)]. Because the PPBS gate applies additional local, polarisation-sensitive phase shifts to each qubit, the output states were not simply the D/A basis states from Tab. 10.4. In fact, they were close to the four orthogonal separable states $(|H\rangle \pm e^{i\varphi_1}|V\rangle) \otimes (|H\rangle \pm e^{i\varphi_2}|V\rangle)$, with $\varphi_1 = 0.07$ and $\varphi_2 = 0.32$ determined by a

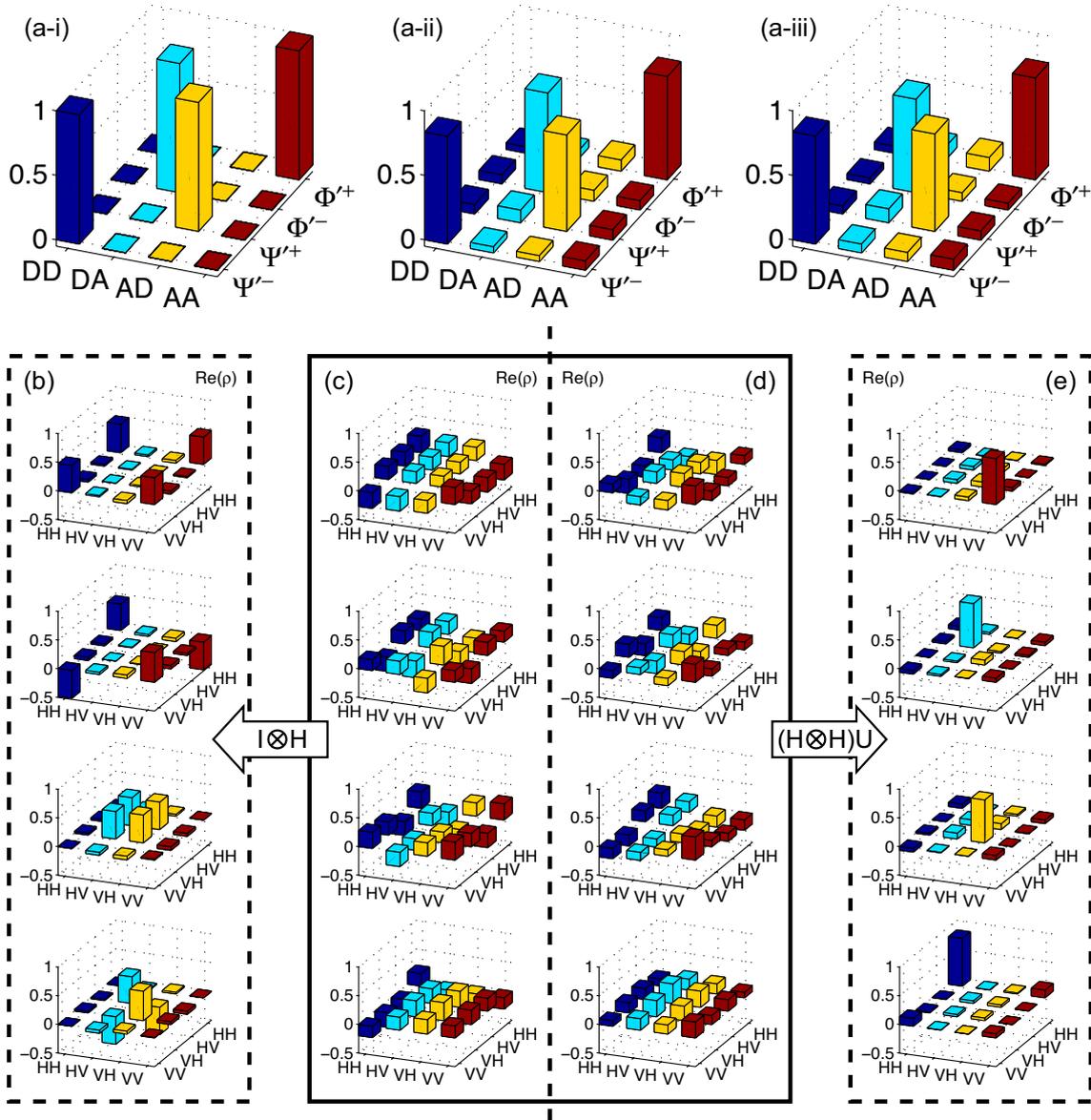


Figure 10.9: The CZ gate operating as a Bell-state analyser. (a) The Bell analysis truth tables: (i) the ideal case [see Tab. 10.4]; (ii) calculated from the measured input and output density matrices ($\bar{p}_{\text{success}} = 0.79 \pm 0.03$); (iii) measured directly ($\bar{p}_{\text{success}} = 0.78 \pm 0.05$). (b-e) The measured density matrices showing the transformation of near-maximally entangled states to near-separable states. (b) The input Bell-states determined from (c) the measured input states, with the second qubit rotated by a Hadamard. (d) The measured output states, (e) transformed by applying local rotations to each qubit (see text). This is the basis used in the measured truth tables in (a-ii) and (a-iii).

target input state	tangle (T)	linear entropy (S_L)	fit quality (Q)	target fidelity (F_t)
$HD+VA$	0.965 ± 0.006	0.019 ± 0.004	1.05 ± 0.16	0.971 ± 0.002
$HD-VA$	0.947 ± 0.008	0.035 ± 0.005	1.29 ± 0.17	0.958 ± 0.002
$HA+VD$	0.959 ± 0.007	0.022 ± 0.005	1.48 ± 0.16	0.967 ± 0.002
$HA-VD$	0.962 ± 0.007	0.028 ± 0.005	1.16 ± 0.15	0.964 ± 0.002
expected output state	tangle (T)	linear entropy (S_L)	fit quality (Q)	model #2 fidelity (F_2)
AA	0.020 ± 0.004	0.468 ± 0.007	0.75 ± 0.16	0.983 ± 0.002
DA	0.005 ± 0.002	0.413 ± 0.006	0.42 ± 0.13	0.982 ± 0.003
AD	0.11 ± 0.01	0.315 ± 0.007	0.44 ± 0.15	0.984 ± 0.003
DD	0.006 ± 0.002	0.488 ± 0.008	0.73 ± 0.16	0.977 ± 0.006

Table 10.5: Key properties of measured input and output states for Bell measurements.

best fit. For ease of visualisation, we have rotated these states into the logical basis²⁵ in Fig. 10.9(e). The average of the fidelities between all combinations of the measured output states was $24\pm 6\%$ (ideally zero), demonstrating that the states are close to orthogonal. Their average tangle, $\bar{T}=0.03\pm 0.05$, and linear entropy, $\bar{S}_L=0.42\pm 0.08$, indicates that they were unentangled, albeit somewhat mixed [see Tab. 10.5 for details]. This circuit worked quite well as a Bell-state analyser.

The average fidelity of the measured output states with the above separable states is $F=79\pm 3\%$: i.e. if we analyzed the output of the circuit in this rotated basis, we would correctly identify the Bell-state with a probability of 79%. This can be calculated from the average overlap of the calculated and ideal truth tables [Fig. 10.9(a-ii) and (a-i), respectively]. More directly, we also measured each of these separable states for the different Bell-state inputs by explicitly analysing in the rotated basis—this gave the directly measured truth table for the Bell-state analysis operation [Fig. 10.9(a-iii)], and the average probability of success was $78\pm 5\%$, in agreement with the tomography results.

What are the major limitations on the quality of Bell measurements with our measured gate? The most obvious source of error is the input states themselves—they are close to the target states, but not perfect, with an average target state fidelity of $\bar{F}_t=0.965\pm 0.006$. In fact, it is not too difficult to see that this would also be the maximum achievable average success probability for the Bell analysis, even if one had access to a perfect CZ gate. The other, more major source of error is the imperfect mode matching at the central PPBS. Before these measurements we measured the non-classical interference visibility to be²⁶ $\sim 69\%$. Note that for the measured output states, the average of the largest eigenvalues is 0.81 ± 0.04 . Since this agrees with the average truth table fidelities above, this suggests (not surprisingly) that mixture in the output states, as opposed to partly overlapping output states²⁷, was the main cause of failed measurement events.

²⁵In the figure, U is the appropriate separable operator of the form $U_1\otimes U_2$ to undo the gate-induced phase shifts.

²⁶This set of measurements was taken at a different time to the data reported in the previous sections.

²⁷Overlapping in the sense that H and D are partly overlapping, not in the sense that two mixed states will have some non-zero fidelity.

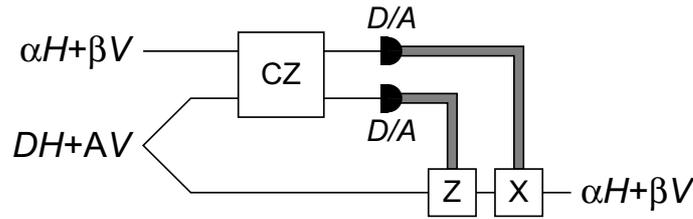


Figure 10.10: A simple, simulated teleportation protocol using our CZ gate to perform the Bell measurement step. The thick grey path indicates a classical channel. The use of the correction operations is conditioned upon the measurement outcomes from the Bell-state analysis using feed-forward.

Using the measured input and output states for the Bell analysis, I have compared the gate operation with the model developed in Sec. 10.4.3 (model #2). The results are closest to the model when $p=0.139$, $\phi_1=-0.039$ and $\phi_2=0.101$ (fitted numerically), with an average fidelity of $\overline{F_2}=0.981\pm 0.003$ over the four output states. This error probability corresponds to a non-classical interference visibility of 68.9%, which matches the measured visibility, and given the input states, the model predicts a Bell measurement success probability of $80\pm 2\%$, agreeing with the above values within error. This means that the limitations of our CZ gate as a Bell analyser are entirely explained by the imperfect input states and spatial mode mismatch.

After performing the Bell analysis measurements, we also ran a full process tomography without altering the operating conditions (100s counts, >18h tomography duration). The resulting process reconstruction can also be applied to the measured input states to simulate the Bell measurement process, and this predicts an average success probability of $71\pm 5\%$, which is somewhat lower than the measured and simulated results above. The difference is probably due to the long data run and the related effects of drift on the decoherence properties of the reconstructed process matrix. Because the tomography was substantially longer than those reported in the previous section, it is not surprising that the process purity for this measurement set was also significantly lower ($P_{\text{pro}}\sim 61\%$).

It is interesting to note that whenever a post-selected event occurs, the Bell measurement has effectively discriminated one of four input wave plate settings applied to a single input qubit, i.e. two bits of classical information have been encoded into a single qubit. This demonstrates the basic effect of quantum dense coding [26–28], although because the Bell measurement is non-deterministic, a protocol using this gate would be less efficient than ordinary classical communication. Nevertheless, this still illustrates the power of entanglement for dense coding given a deterministic Bell analyser, such as can be constructed in principle using measurement-induced nonlinearity.

Using our CZ gate in a teleportation protocol

To this point, I have assessed the quality our CZ Bell analyser by calculating the truth table fidelity for the measurement process which represents the average probability of successfully discriminating between the Bell-state inputs. But what are the implications of this for using our gate in a real application? To investigate this, I have simulated

a simple teleportation protocol [shown in Fig. 10.10] using the truth table information obtained above. The protocol is described below.

We wish to teleport an unknown single qubit state, $|\psi\rangle = \alpha|H\rangle + \beta|V\rangle$, using a perfect entangled resource of the form, $|\phi\rangle = (|DH\rangle + |AV\rangle)/\sqrt{2}$, and our real CZ gate to perform the Bell measurement. The following simple algebra describes the action of the protocol.

$$|\psi\rangle \otimes |\phi\rangle = (\alpha|H\rangle + \beta|V\rangle) \otimes \frac{1}{\sqrt{2}} (|DH\rangle + |AV\rangle) \quad (10.29)$$

$$= \frac{1}{\sqrt{2}} (\alpha|HDH\rangle + \alpha|HAV\rangle + \beta|VDH\rangle + \beta|VAV\rangle) \quad (10.30)$$

$$\begin{aligned} &= \frac{1}{2} (|HD\rangle + |VA\rangle) \otimes \alpha|H\rangle + \frac{1}{2} (|HD\rangle - |VA\rangle) \otimes \alpha|H\rangle \\ &\quad + \frac{1}{2} (|HA\rangle + |VD\rangle) \otimes \alpha|V\rangle + \frac{1}{2} (|HA\rangle - |VD\rangle) \otimes \alpha|V\rangle \\ &\quad + \frac{1}{2} (|HA\rangle + |VD\rangle) \otimes \beta|H\rangle - \frac{1}{2} (|HA\rangle - |VD\rangle) \otimes \beta|H\rangle \\ &\quad + \frac{1}{2} (|HD\rangle + |VA\rangle) \otimes \beta|V\rangle - \frac{1}{2} (|HD\rangle - |VA\rangle) \otimes \beta|V\rangle \end{aligned} \quad (10.31)$$

$$\begin{aligned} &= \frac{1}{2} (|HD\rangle + |VA\rangle) \otimes (\alpha|H\rangle + \beta|V\rangle) + \frac{1}{2} (|HD\rangle - |VA\rangle) \otimes (\alpha|H\rangle - \beta|V\rangle) \\ &\quad + \frac{1}{2} (|HA\rangle + |VD\rangle) \otimes (\alpha|V\rangle + \beta|H\rangle) + \frac{1}{2} (|HA\rangle - |VD\rangle) \otimes (\alpha|V\rangle - \beta|H\rangle) \end{aligned} \quad (10.32)$$

Now that the three-qubit state has been rewritten with the first two qubits in the appropriate basis to be discriminated by the CZ Bell analyser, it is easy to see that applying an ideal, unitary CZ gate of the form implemented by the PPBS CZ gate, $U_{\text{exp}} \propto A_{\text{exp}}$ [Eq. (10.14)], will create the following separable state,

$$\begin{aligned} [U_{\text{exp}} \otimes I] (|\psi\rangle \otimes |\phi\rangle) &= \frac{1}{2} |AA\rangle \otimes -(\alpha|H\rangle + \beta|V\rangle) + \frac{1}{2} |DA\rangle \otimes (\alpha|H\rangle - \beta|V\rangle) \\ &\quad + \frac{1}{2} |AD\rangle \otimes (\alpha|V\rangle + \beta|H\rangle) + \frac{1}{2} |DD\rangle \otimes -(\alpha|V\rangle - \beta|H\rangle). \end{aligned} \quad (10.33)$$

The final step of the protocol is to measure the first two qubits in the D/A basis, and then apply a correction operation to the third qubit based on the measurement outcome. No correction is needed for the first term (AA), but for the DA , AD and DD outcomes, the appropriate corrections are Z , X and XZ , respectively. With a perfect Bell measurement, the third qubit will end up in the correct state, regardless of the measurement outcome, which is the classic signature of quantum teleportation. However, with an imperfect CZ gate, sometimes the Bell state will be misdiagnosed, and the wrong correction operation will be applied, leaving the output qubit in the wrong state. In other words, quantifying the success of this teleportation protocol provides an indicator of the performance of our CZ gate in a practical context.

Unfortunately, the truth table for the gate does not provide enough information to calculate any useful measure of this success, for the same reason that the standard truth table cannot be used as a characterisation of a quantum operation—it does not differentiate between mixtures and superpositions of states, and provides no phase information in superpositions. The basic problem is that when we make an entangled measurement with an imperfect gate, the state of the output qubit is some combination of the four *non-orthogonal* possibilities, and the required superposition information cannot be obtained from the truth table. For example, the output qubit could be mixed, or it could be in a pure, but rotated state. Therefore, instead of using the truth tables, I have used the results of the tomographic reconstruction techniques described above which do contain the extra phase information.

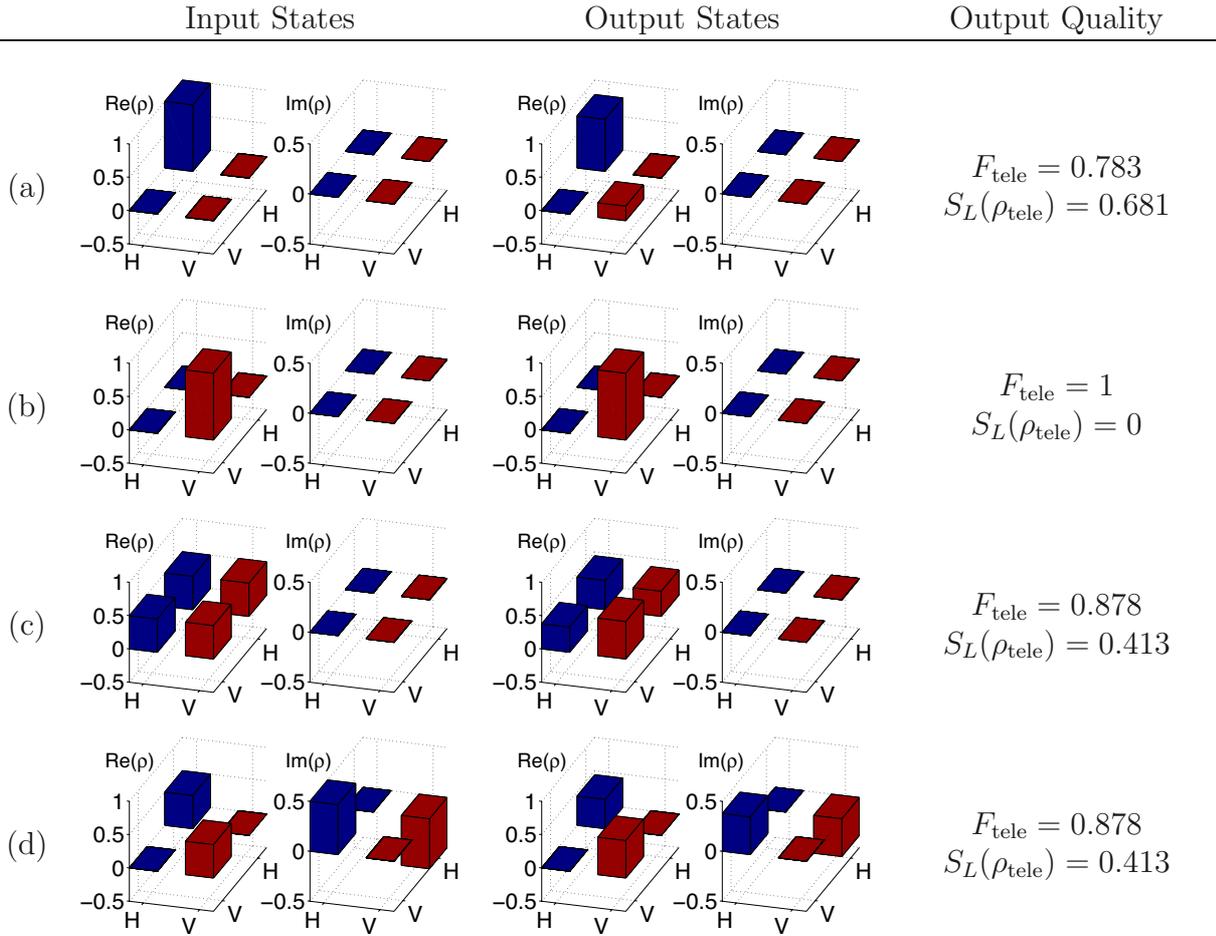


Figure 10.11: Example input states for the simple teleportation protocol: (a) $\psi_{\text{in}} = H$; (b) $\psi_{\text{in}} = V$; (c) $\psi_{\text{in}} = D$; (d) $\psi_{\text{in}} = R$.

In the teleportation protocol, I have directly simulated the behaviour of our CZ gate using the modelled CZ process found above by numerical fit ($p=0.139$, $\phi_1 = -0.039$ and $\phi_2=0.101$), except I have used the simpler form of model #1, ignoring the phase shifts produced by the PPBS gate since they can be controllably reversed at the gate's output. As a first, qualitative test of the teleportation protocol, I considered the input states H , V , D and R [Fig. 10.11]. These show that the teleportation protocol fails only when there is an H component in the input state, a result of the imperfect interference of the HH component at the central PPBS of the CZ gate (the only imperfection included in model #1), and that the error creates mixture in the output. To test its performance more comprehensively, however, I generated a set of 1000 randomly selected pure states²⁸ for the unknown input qubit. The average teleportation fidelity (between the teleported state and the input state) was²⁹ $\overline{F_{\text{tele}}}=0.877\pm 0.002$ and the minimum (worst case) fidelity was $F_{\text{tele},\text{min}}=0.784$ for a state which was very close to H . The average linear entropy of the output states was $\overline{S_L}=0.391\pm 0.006$, indicating once again that failure of the protocol relates mostly to the effects of mixture. Note that the equal superposition states, D and

²⁸Uniformly selected according to the Haar measure.

²⁹The errors reported here are errors in the mean of the distribution over all pure states ($\sigma/\sqrt{\text{no. samples}}$), not the width of the distribution itself (σ).

R , are very close to the average.

For comparison purposes, there are two important limits for the teleportation fidelity which are useful to consider: the best average fidelity of a classical teleportation protocol (i.e. with no shared entanglement) is $2/3$ (worst case fidelity of $1/2$); and the minimum average fidelity required to guarantee that the recipient has the best available copy of the initial state, as prescribed by the No Cloning theorem, is $5/6$ [29]. Thus, using our modelled CZ gate, which predicts an average fidelity of 0.877 ± 0.002 , gives a protocol which surpasses both limits, even making allowances for the slight difference between the modelled and measured Bell-state analysis ($\sim 80\%$ instead of $\sim 78\%$). The minimum fidelity for our predicted protocol was also far above the best classical protocol.

I have also simulated the teleportation protocol using the tomographically reconstructed process matrix³⁰, although the earlier results indicated that it did not fit the Bell measurement input and output state tomographies as well as the modelled process. Not surprisingly, therefore, the average teleportation fidelity (between the teleported state and the input state) was $\overline{F_{\text{tele}}} = 0.817 \pm 0.002$, the minimum (worst case) fidelity was $F_{\text{tele},\text{min}} = 0.688$, and the average linear entropy of the output states was $\overline{S_L} = 0.534 \pm 0.005$. Once again, these results surpass the mark set by best possible classical protocol.

10.5 Conclusions

In this chapter, I have investigated a new optical architecture for the entangling gates which play such an important role in many quantum information applications. Like previous optical implementations it uses non-classical interference and measurement to realise the large optical non-linearity required to generate entanglement, but its key advantage is its simplicity. By using partially polarising beam splitters (PPBSs), the two-photon CZ gate operation is performed directly in a polarisation-based encoding and therefore requires only one non-classical mode-matching condition and no classical interferometers. This makes this a particularly promising possibility for micro-optics and integrated-optics implementations, where extremely good mode matching can be achieved.

I have developed detailed theoretical models of the gate which include all of the main real-world effects, such as imperfect PPBSs and wave plates [Sec. 10.2.3] and misalignment in the spatial mode overlap for the non-classical interference [Sec. 10.4.3], and which agree well with a broad array of experimental results. I have showed that the most significant factor affecting the performance of the gate was spatial mode mismatch which reduced the visibility of the non-classical interference. In recent experiments this has been greatly improved by using further spatial filtering at the gate output.

We experimentally investigated many aspects of the gate's operation, demonstrating its ability to both create entanglement and undo existing entanglement. In a disentangling configuration, we used our CZ gate in conjunction with a Type-I, crystal-sandwich source of entangled photon pairs to perform Bell-state analysis, showing that it could discriminate quite effectively between the four orthogonal, maximally entangled Bell states. This Bell

³⁰I used the phase shifts for the modelled process to undo the polarisation effects of the gate implemented by the process matrix.

analyser has a higher success probability than alternative recent demonstrations [13, 30] and requires no ancilla photons. Using these results, I have simulated the gate's behaviour in a simple teleportation protocol and showed that it performs better than the best possible classical protocol, and on average, guarantees that the recipient has the best copy of the initial state. We also used quantum process tomography to completely characterise the operation of the gate in the two-photon coincidence subspace, and compared the predictions of the process matrix reconstruction with the results of both the entangling and disentangling experiments.

As an aside, we have also introduced the process purity, calculated from the process matrix, which is a measure of the decoherence properties of a quantum process. It is related in a simple way to the average purity of the output states of the process.

10.6 Chapter 10 References

- [1] N. K. Langford, T. J. Weinhold, R. Prevedel, K. J. Resch, A. Gilchrist, J. L. O'Brien, G. J. Pryde, and A. G. White. Demonstration of a simple entangling optical gate and its use in Bell-state analysis. *Physical Review Letters*, 95(21):210504, 2005.
- [2] N. Kiesel, C. Schmid, U. Weber, R. Ursin, and H. Weinfurter. Linear optics controlled-phase gate made simple. *Physical Review Letters*, 95:210505, 2005.
- [3] R. Okamoto, H. F. Hofmann, S. Takeuchi, and K. Sasaki. Demonstration of an optical quantum controlled-NOT gate without path interference. *Physical Review Letters*, 95:210506, 2005.
- [4] R. Prevedel. Experimental realization of a simple entangling optical gate for quantum computation. Diplomarbeit zur erlangung des grades eines "magister der naturwissenschaften", University of Vienna, Wien, Austria, 2005.
- [5] T. C. Ralph, N. K. Langford, T. B. Bell, and A. G. White. Linear optical controlled-NOT gate in the coincidence basis. *Physical Review A*, 65(6):062324, 2002.
- [6] M. J. Bremner, C. M. Dawson, J. L. Dodd, A. Gilchrist, A. W. Harrow, D. Mortimer, M. A. Nielsen, and T. J. Osbourne. Practical scheme for quantum computation with any two-qubit entangling gate. *Physical Review Letters*, 89:247902, 2002.
- [7] E. Knill, R. Laflamme, and G. J. Milburn. A scheme for efficient quantum computation with linear optics. *Nature*, 409:46, 2001.
- [8] D. Gottesman and I. L. Chuang. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature*, 402:390, 1999.
- [9] T. B. Pittman, M. J. Fitch, B. C. Jacobs, and J. D. Franson. Experimental controlled-NOT logic gate for single photons in the coincidence basis. *Physical Review A*, 68:032316, 2003.
- [10] J. L. O'Brien, G. J. Pryde, A. G. White, T. C. Ralph, and D. Branning. Demonstration of an all-optical quantum controlled-NOT gate. *Nature*, 426:264, 2003.
- [11] J. L. O'Brien, G. J. Pryde, A. Gilchrist, D. F. V. James, N. K. Langford, T. C. Ralph, and A. G. White. Quantum process tomography of a controlled-NOT gate. *Physical Review Letters*, 93(8):080502, 2004.
- [12] S. Gasparoni, J.-W. Pan, P. Walther, T. Rudolph, and A. Zeilinger. Realization of a photonic controlled-NOT gate sufficient for quantum computation. *Physical Review Letters*, 93:020504, 2004.
- [13] Z. Zhao, A.-N. Zhang, Y.-A. Chen, H. Zhang, J.-F. Du, T. Yang, and J.-W. Pan. Experimental demonstration of a nondestructive controlled-NOT quantum gate for two independent photon qubits. *Physical Review Letters*, 94:030501, 2005.
- [14] T. C. Ralph, A. G. White, W. J. Munro, and G. J. Milburn. Simple scheme for efficient linear optics quantum gates. *Physical Review A*, 65:012314, 2001.

- [15] E. Knill. Quantum gates using linear optics and postselection. *Physical Review A*, 66:052306, 2002.
- [16] T. B. Pittman, B. C. Jacobs, and J. D. Franson. Demonstration of nondeterministic quantum logic operations using linear optical elements. *Physical Review Letters*, 88(25):257902, 2002.
- [17] H. F. Hofmann and S. Takeuchi. Quantum phase gate for photonic qubits using only beam splitters and postselection. *Physical Review A*, 66:024308, 2002.
- [18] M. A. Nielsen. Optical quantum computation using cluster states. *Physical Review Letters*, 93:040503, 2004.
- [19] T. B. Pittman and J. D. Franson. Violation of Bell's inequality with photons from independent sources. *Physical Review Letters*, 90:240401, 2003.
- [20] C. K. Hong, Z. Y. Ou, and L. Mandel. Measurement of subpicosecond time intervals between two photons by interference. *Physical Review Letters*, 59:2044, 1987.
- [21] T. C. Ralph. Quantum optical cnot gate. Provisional Patent No. 2002328651, Granted in UK and US (No. WO2003/019468), August 2001.
- [22] P. G. Kwiat, E. Waks, A. G. White, I. Appelbaum, and P. H. Eberhard. Ultrabright source of polarization-entangled photons. *Physical Review A*, 60:R773, 1999.
- [23] A. G. White, A. Gilchrist, G. J. Pryde, J. L. O'Brien, M. J. Bremner, and N. K. Langford. Measuring controlled-not and two-qubit gate operation. *quant-ph*, page 0308115, 2003.
- [24] A. Gilchrist, N. K. Langford, and M. A. Nielsen. Distance measures to compare real and ideal quantum processes. *Physical Review A*, 71(6):062310, 2005.
- [25] J. F. Poyatos, J. I. Cirac, and P. Zoller. Complete characterization of a quantum process: The two-bit quantum gate. *Physical Review Letters*, 78:390, 1997.
- [26] C. H. Bennett and S. J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Physical Review Letters*, 69:2881, 1992.
- [27] T. Schaetz, M. D. Barrett, D. Liebfried, J. Chiaverini, J. Britton, W. M. Itano, J. D. Jost, C. Langer, and D. J. Wineland. Quantum dense coding with atomic qubits. *Physical Review Letters*, 93:040505, 2004.
- [28] K. Mattle, H. Weinfurter, P. G. Kwiat, and A. Zeilinger. Dense coding in experimental quantum communication. *Physical Review Letters*, 76:4656, 1996.
- [29] H.-A. Bachor and T. C. Ralph. *A Guide to Experiments in Quantum Optics*. WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, Deutschland, 2004.
- [30] P. Walther and A. Zeilinger. Experimental realization of a photonic bell-state analyzer. *Physical Review A*, 72:010302(R), 2005.

*Thus grew the tale of Wonderland:
Thus slowly, one by one,
Its quaint events were hammered out
And now the tale is done,
And home we steer, a merry crew,
Beneath the setting sun.*

Alice's Adventures in Wonderland,
Lewis Carroll

Chapter 11

Discussion and conclusions

In this thesis, I have described work that spans a range of contexts within quantum optics and quantum information, and is linked by the desire to realise experimentally the use of quantum systems for encoding and processing information. In particular, I have studied different methods for encoding information in optical systems, for measuring that information and for creating entanglement in those systems. This diversity is perhaps the most exciting feature of quantum information, but it also brings its own challenges of communication and understanding. I have tried here to synthesise the ideas from the different areas and present them in a consistent and compatible way.

The first main goal of my thesis was *to study the techniques for characterising the two main components of any quantum information processing task*—namely, the states and processes—and assessing the extracted data. Complete characterisation is only possible using tomographic reconstruction procedures, which are well known and becoming more standard. Exploring these in more detail, however, I have shown that they are not yet well understood, particularly with regard to how errors propagate through the reconstruction process, what measurements are ideal and what method should be used to find the optimal reconstruction. My investigations are not intended to be comprehensive, so there is much work here that remains to be done. For example, how do the above results vary in systems where Poissonian fluctuations are not the main source of error? Also, can significant improvements be achieved using adaptive tomographic schemes, and which are the most effective?

Once a state or process has been characterised, the next step is to assess the outcome, perhaps by comparing it to some expected ideal. While appropriate measures are both well known and well understood for comparing states, the issue has not been comprehensively addressed in relation to quantum processes. In Chapter 9, we developed a set of experimentally and theoretically motivated criteria describing the desirable properties of a measure of process comparison. We used these criteria to assess a broad range of possible measures, and isolated a small number which have particularly promising attributes. A key open problem in this area is to derive quantum computing error correction thresholds

directly using these promising measures. We also defined a secondary, *diagnostic* measure called the process purity, which cannot be used to compare quantum processes, but does provide information about the decoherence properties of a process [Sec. 10.4.4].

The second major aim of my work has been *to investigate different ways to encode, measure and manipulate information in single photons*, using the three degrees of freedom of polarisation, spatial-momentum and time-frequency distributions. Polarisation is still the most well developed and common of the three possibilities—it has the longest history of being actively studied and the largest array of tools available for its precise manipulation—but there are issues which need further study. For example, I have shown that implementing arbitrary unitary operations with wave plates is more complicated than commonly thought.

The key feature of polarisation is that in normal circumstances it forms a natural qubit (discrete and two-dimensional), but this is also its main limitation. Although higher-dimensional systems do not provide an exponential speed-up over qubit-based quantum computing, they may provide improvements in computing power in practical situations, where systems are subject to physical constraints. They also have significant advantages over qubits in many quantum communication protocols, such as increased channel capacity and higher levels of security. Both spatial-momentum and time-frequency degrees of freedom have infinite-dimensional Hilbert spaces and allow access to qudit encoding schemes.

I have developed a simple procedure for performing tomography of the transverse spatial quantum state, and through modelling and experiments, I have studied in detail the properties and effects of the real-world holograms that we use to transform the spatial modes. In particular, plane-wave holograms can be used to implement reliable spatial tomography by making over-complete sets of measurements. It is also important to realise there is a fundamental distinction between two different types of transverse spatial encoding according to whether they exhibit either degenerate or nondegenerate superpositions.

The main limitations of transverse spatial encodings of information are that, with current technology, there are no deterministic beam splitters and unitary rotations are difficult. By combining holograms with spatial filtering, and perhaps also standard beam splitters, it is possible to produce probabilistic spatially dependent beam splitters, but these have drastic effects on count rates and also limit the scalability of spatially encoded quantum computing schemes. There are natural difficulties associated with developing beam splitters for an infinite-dimensional state space, but with the further advantage of fibre technologies, one might envisage a special optical fibre with a series of output ports each picking off one particular spatial component of the incident beam with the final output containing whatever is left. With regard to unitary rotations, the first step would be to develop a simple way of converting one arbitrary transverse spatial mode into another. Techniques such as using pairs of cylindrical lenses to convert between vortex modes and Hermite-Gauss modes are only able to perform certain transformations and are not flexible. Arbitrary spatial transformations are possible using thick holograms, but these are extremely difficult to make, let alone with any precision. Once again, with the appropriate developments in technology, one could imagine using a three-dimensional spatial light modulator to controllably and flexibly manipulate spatial quantum states. Incidentally, I note that these issues are related to an important open problem: we have

a good technique for creating single higher-order Gaussian modes with precision, but not for measuring them, because the technique is probabilistic and therefore non-reversible.

I have also developed a technique for performing tomographic analysis of the time-frequency domain which is an extension of the ideas introduced in Ref. [1] and uses existing polarisation entanglement as a resource [Sec. 8.3.1]. Time-frequency encodings have similar fundamental properties and limitations to the spatially-encoded systems. Diffraction gratings may provide a good beam splitter for frequency-encoded states, but there are no satisfactory beam splitters for time-bin encoding—either they work in the correct way with only a limited probability, or they rely on classical, incoherent processes such as fast optical switching. It may be that useful ideas can be obtained by investigating more deeply the related properties of the time-frequency and spatial-momentum domains. For example, can one define time-encoded quantum states using the temporal distribution of the wave packet instead of separate arrival time-bins (cf. Hermite-Gauss spatial modes)? If so, can engineered frequency filters be used to produce similar effects as holograms?

My final major objective in this thesis was *to demonstrate and characterise different methods for producing entanglement between single photons*. In Chapter 7 we performed the first full characterisation of entangled, spatially encoded quantum states, and the first complete measurement of an entangled, two-qutrit state in *any* encoding. To obtain a meaningful measure of the quality of the two-qutrit entanglement, we simulated a quantum bit commitment protocol using the measured state as a resource. We found that despite its strong entanglement, it would need to be significantly more pure in order to provide better security than is achievable with qubit-based schemes. Finally, I have also shown that the tomographic reconstructions of nondegenerate systems can be greatly improved using a better understanding of the plane-wave hologram behaviour.

A range of research activity by other groups has followed on from these experiments, including some detailed theoretical analysis of entanglement between Hermite-Gauss modes in down-conversion [2, 3]. From this point, there are several interesting directions that future work could take. Replacing the holographic plates in this experiment with spatial light modulators would allow a more complete characterisation of the spatial quantum state of the down-conversion source, and these results could be compared with the theoretical predictions (which would require some modification to fit realistic experimental conditions). Another worthwhile experiment would be to implement the bit commitment protocol described above—with appropriate optimisation of the entangled state, one could demonstrate a milestone with qutrits which could not be achieved with qubit-based systems. On a more fundamental level, further exploration of the differences between degenerate and nondegenerate encoding schemes is required. For example, in higher dimensions, one could compare the nondegenerate qutrit entanglement described above with the entanglement in a degenerate second-order qutrit.

In Chapter 8, we demonstrated for the first time the production of completely hyperentangled photons using both Bell-violation and tomographic techniques. Using combined spatial and polarisation tomography, we reconstructed a 36-dimensional quantum state—the largest such characterisation to date. We produced a variety of 16-dimensional spatial-polarisation states with unprecedented fidelities for such large systems, and along the way, we also demonstrated the best reported entanglement in both polarisation and spatial sys-

tems, and the first tomographic analysis of the time-frequency domain. More recently, similar hyperentangled states have been used to perform complete Bell-state analysis [4]. In future work, it would be interesting to find and explore operations which produce direct interactions between the different degrees of freedom, such as would probably be necessary to produce hypoentanglement, the other form of multi-degree entanglement. For example, it may be possible to produce hypoentangled photons by slightly modifying the bright source of polarisation-entangled photons introduced in Ref. [5].

In Chapter 10, I modelled and demonstrated a new optical architecture for an entangling, two-qubit CZ gate which uses partially polarising beam splitters to implement the operation directly in the polarisation domain (see also Refs [6] and [7]). The key feature of this new gate is its stability, because it requires no classical interferometers to move between polarisation and longitudinal spatial encodings. As a result, the most significant factor limiting the performance of the gate was the spatial mode overlap for the non-classical interference condition. I also combined the gate with a high-quality source of polarisation-entangled photons to demonstrate that it operated effectively as a Bell-state analyser. Indeed, by modelling a simple teleportation protocol which used the measured CZ gate as its only source of imperfection, I was able to show that the protocol would perform better than any classical teleportation, with a predicted teleportation fidelity of $\sim 88\%$. Moreover, this fidelity indicates that the recipient of the teleported state would be guaranteed by the limits of the No Cloning theorem to receive the best available copy of the original state.

In more recent experiments, the non-classical interference visibility has been greatly improved by further spatial filtering at the output of the gate, and it would be valuable to see how far the limits could be pushed by this method and by micro- or integrated-optics techniques. If the precision of this gate could be improved sufficiently, it would open the door for demonstrations of many interesting small-scale quantum computing experiments, e.g. the creation of certain types of cluster states for use in one-way quantum computing. Finally, with such a gate, it would also be possible to controllably introduce mixture into the entangled output states by precisely adjusting the spatial mode overlap at the beam splitter.

Quantum information science is a field where the cutting edge of both theoretical and experimental research examines the truly fundamental characteristics of quantum mechanics. In this thesis, I have studied ways to encode, manipulate, measure, characterise and quantify information in quantum systems, particularly within the context of single-photon optics. This work spans both experimental and theoretical endeavours and is the result of a continued effort to tackle them cooperatively rather than separately. Perspectives developed through experience in practical research have provided direction for my theoretical investigations, and these have in turn supported and stimulated my experiments. Through this combined approach, my work has developed tools which will help to explore the possibilities of quantum information processing, and which will be required in order to successfully build and understand intrinsically quantum computing machines.

11.1 Chapter 11 References

- [1] D. V. Strekalov, T. B. Pittman, A. V. Sergienko, Y. H. Shih, and P. G. Kwiat. Postselection-free energy-time entanglement. *Physical Review A*, 54:R1, 1996.
- [2] S. P. Walborn, S. Pádua, and C. H. Monken. Conservation and entanglement of Hermite-Gaussian modes in parametric down-conversion. *Physical Review A*, 71:053812, 2005.
- [3] X.-F. Ren, G.-P. Guo, J. Li, and G.-C. Guo. Entanglement of the Hermite-Gaussian modes states of photons. *Physics Letters A*, 341:81, 2005.
- [4] C. Schuck, G. Huber, C. Kurtsiefer, and H. Weinfurter. Complete deterministic linear optics bell state analysis. *Physical Review Letters*, 96:190501, 2006.
- [5] T. Kim, M. Fiorentino, and F. N. C. Wong. Phase-stable source of polarisation-entangled photons using a polarisation Sagnac interferometer. *Physical Review A*, 73:012316, 2006.
- [6] N. Kiesel, C. Schmid, U. Weber, R. Ursin, and H. Weinfurter. Linear optics controlled-phase gate made simple. *Physical Review Letters*, 95:210505, 2005.
- [7] R. Okamoto, H. F. Hofmann, S. Takeuchi, and K. Sasaki. Demonstration of an optical quantum controlled-NOT gate without path interference. *Physical Review Letters*, 95:210506, 2005.

*'Twas brillig, and the slithy toves
Did gyre and gimble in the wabe;
All mimsy were the borogroves,
And the mome raths outgrabe.*

*Jabberwocky from *Through the Looking-Glass*
(and what Alice found there), Lewis Carroll*