Order Number 9430568

Nonclassical effects from spontaneous parametric down-conversion: Adventures in quantum wonderland

Kwiat, Paul Gregory, Ph.D.

University of California, Berkeley, 1993

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

Nonclassical Effects from Spontaneous Parametric Down-Conversion: Adventures in Quantum Wonderland

by

Paul Gregory Kwiat

S.B. (Massachusetts Institute of Technology) 1987
M.S. (University of California at Berkeley) 1989

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

GRADUATE DIVISION

of the

UNIVERSITY of CALIFORNIA at BERKELEY

Committee in charge:

Professor Raymond Y. Chiao, Chair
Professor Charles H. Townes
Professor T. Kenneth Gustafson

1993

The dissertation of Paul Gregory Kwiat is approved:

Raymond Chiao — August 31, 1993
Chair — Date

Chas. H. Tanner — Aug. 31, 1993
Date

T K Gustafson — Aug 31, 1993
Date

University of California at Berkeley

1993

Abstract

Nonclassical Effects from Spontaneous Parametric Down-Conversion: Adventures in Quantum Wonderland

by

Paul Gregory Kwiat

Doctor of Philosophy in Physics

University of California at Berkeley

Professor Raymond Y. Chiao, Chair

The present treatise describes research investigating the inherent nonlocality in quantum mechanics and the role of (in-)distinguishability in interference. The intent is to promote an intuitive understanding of the underlying principles, by means of fairly straightforward yet nonetheless striking experiments (and proposals for experiments). Central to all of these are the pairs of correlated photons produced when an ultraviolet photon interacts with a nonlinear optical medium, and via the process of spontaneous parametric down-conversion splits into two infrared photons. These pairs possess the useful property that one can produce essentially single-photon states by triggering on one of the photons, and counting the other in coincidence. The photons may also be prepared in several sorts of entangled states, which are intrinsically nonlocal.

Coincidence detection of the photon pairs, in conjunction with a Michelson interferometer in which one member of each pair acquired a geometrical phase due to a cycle in polarization states, has allowed the first observation of Berry's phase at the single photon level. We have verified that each interfering photon was essentially in an

n=1 Fock state, by means of a beam splitter following the interferometer, combined with a triple-coincidence technique. When the interferometer is set outside the white-light fringe regime, the results can be interpreted in terms of a nonlocal "collapse of the wavefunction".

We have also observed an effect known as a *quantum eraser*, in which an interfering system is first rendered incoherent by making the alternate Feynman paths which contribute to the overall process distinguishable; with our apparatus, polarization is used to label the path of a given photon. The quantum "eraser" removes this information from the state vector, *after* the output port of the interferometer, so that one may recover interference upon coincidence detection. In addition, we show how this procedure may do more than merely erase, in that the act of "pasting together" two previously distinguishable paths can introduce a new relative phase between them. Finally, we propose three new quantum eraser schemes, each of which satisfies all the criteria for an optimal demonstration. As the proposed schemes are all modifications or combinations of previously completed experiments, they are deemed to be feasible.

We also report on a two-photon interference experiment proposed by Franson, in which each of the down-converted photons is sent into an unbalanced interferometer. Employing large path-length differences and high time-resolution detectors, we were able to observe sinusoidal coincidence fringes with visibilities of up to 90%, while the classical limit for such fringes is only 50%. Moreover, we were able to violate an appropriate Bell's inequality (after several reasonable supplementary assumptions were made), which limits the correlations permitted by any theory constrained by "local realism". Here the elements of reality under consideration are the energy and time of emission of the photons. We infer that photons produced in spontaneous parametric down-conversion know neither their color nor their birthdays, and that Nature is inherently nonlocal, but only just barely so.

To perform an experiment of this sort that is incontrovertible is very difficult, and has not yet been done; one reason is that detection efficiencies need to be quite high. To this end, we have used our light source to measure the absolute quantum efficiencies (and other operating characteristics) of several single-photon detectors. Efficiencies as high as 76.4±2.3% were seen, with implications that efficiencies as high as 90% may be achievable. A suitable *source* for a loophole-free test of Bell's inequalities is also required. We propose such a source, relying on type-II collinear phase-matching in *two* down-conversion crystals, which should permit a violation of Bell's inequalities without the need for the usual supplementary assumptions.

Raymond Chiao 8/31/93

Table of Contents

List of Figures

List of Cartoons

Preface

Contained within these pages are the significant experimental and theoretical results of my last six years' research, consisting of four main experiments and two recent proposals for new experiments. While it was not always clear during this time exactly what form the experiments would take, the general desire to foster a better intuition about quantum mechanics has been an underlying theme. In particular, using nonclassical states of light, I have attempted to investigate several fundamental quantum-mechanical concepts involving interference and nonlocality, by performing experiments that hitherto existed only at the *gedanken* level, and by proposing new experiments also within the realm of current technologies.

It is not my claim that the experiments described here teach us anything fundamentally *new*. All of the results are predictable using the formalism of quantum optics, and from that vantage point are unsurprising. However, the intent of the experiments is to generate an intuitive understanding of the principles underlying all of our quantum mechanical machinery -- in essence, to shed more light on these principles. While there are those who maintain that one should just blindly apply the calculational tools which have proven to be so successful over the last 70 years, I feel one must strive for a deeper understanding. If any modifications to or extensions of existing quantum theory are to made (as perhaps they *must* be to smoothly join the quantum world of superpositions to the classical world of definite outcomes, and to account for gravity, for example), one must have a clear picture of the heart of the theory one is trying to change--what elements are essential?

The results that follow highlight those areas where physical intuition is most lacking, and classical descriptions lead to incorrect predictions. (Hence, the *first* part of the title of this dissertation.) The reason is two-fold. First, because insight into these phenomena is needed, it is sensible to attack the most obvious problems first. Second, exactly because they *are* at such odds with one's classical intuitions, these experiments have a profound sense of "coolness" about them (a high "gee whiz"-factor) and have therefore been extremely enjoyable to carry out. Although there is nothing inherently more strange about a single photon travelling two paths simultaneously than a baseball knowing which way and how fast to fall to earth (an effect mediated by gravitational fields, constructs which only seem "friendly" due to our constant association with them), somehow the wonder of the latter effect is reduced by familiarity. (Hence, the *second* part of the title.)

Nearly all of the results contained herein have been reported in other articles (cf. the List of Publications), with the noteworthy exceptions of some of the detailed calculations in the Appendices, the most recent data for the Franson experiment (Figs. 8.9b - 8.13), and the cartoons used to open four of the chapters.

Lastly, throughout we have attempted to follow "intelligent" rules of punctuation, so that quotation marks mostly fall *inside* other punctuation marks.

List of Publications

1. "A Study of the Photorefractive Effect in $LiNbO_3$:Ti Waveguides by Using Pulsed Illumination at 810 nm", M. Lowry, B. Jander, B. Kidd, P. G. Kwiat, R. Peterson, G. McWright, and F. Roeske, Proceedings SPIE **720**, 105 (1986).

2. "Waveguide Electrooptic Modulators of High Speed Diagnostics", G. McWright, D. Jander, J. Koo, P. G. Kwiat, M. Lowry, F. Roeske, and W. Slivinsky, Proceedings SPIE **704**, 227 (1986).

3. "Correlated Two-Photon Interference in a Dual-Beam Michelson Interferometer", P. G. Kwiat, W. A. Vareka, C. K. Hong, H. Nathel, and R. Y. Chiao, Phys. Rev. A **41**, 2910 (1990).

4. "Optical Manifestations of Berry's Topological Phase: Classical and Quantum Aspects", R. Y. Chiao, P. G. Kwiat, C. K. Hong, H. Nathel, and W. A. Vareka, in *Coherence and Quantum Optics VI*, J. H. Eberly *et al.* (eds.), Plenum Press, New York, 155 (1990).

5. "Observation of a Nonclassical Berry's Phase for the Photon", P. G. Kwiat and R. Y. Chiao, Phys. Rev. Lett. **66**, 588 (1991).

6. "Analogies Between Electron and Photon Tunneling: A Proposed Experiment to Measure Photon Tunneling Times", R. Y. Chiao, P. G. Kwiat, and A. M. Steinberg, Physica B **175**, 257 (1991).

7. "Dispersion Cancellation in a Measurement of the Single-Photon Propagation Velocity in Glass", A. M. Steinberg, P. G. Kwiat, and R. Y. Chiao, Phys. Rev. Lett. **68**, 2421 (1992).

8. "Dispersion Cancellation and High-Resolution Time Measurements in a Fourth-Order Optical Interferometer", A. M. Steinberg, P. G. Kwiat, and R. Y. Chiao, Phys. Rev. **A 45**, 6659 (1992).

9. "Observation of a 'Quantum Eraser': A Revival of Coherence in a Two-Photon Interference Experiment", P. G. Kwiat, A. M. Steinberg and R. Y. Chiao, Phys. Rev. A **45**, 7729 (1992).

10. "Experimental Observation of a Quantum Eraser: How Much is Nature Willing to Forget?", P. G. Kwiat, A. M. Steinberg and R. Y. Chiao, in *Proceedings of the Second Conference on the Foundations of Quantum Mechanics*, T. D. Black *et al.* (eds.), World Scientific, Singapore, 193 (1992).

11. "The Energy-Time Uncertainty Principle and the EPR Paradox: Experiments Involving Correlated Two-Photon Emission in Parametric Down-Conversion", R. Y. Chiao, P. G. Kwiat, and A. M. Steinberg, from *Workshop on Squeezed States and Uncertainty Relations*, D. Han *et al.* (eds.), NASA CP 3135, Washington, DC, 61 (1992).

12. "Observation of a Nonclassical Berry's Phase in Quantum Optics", R. Y. Chiao, P. G. Kwiat, I. H. Deutsch, and A. M. Steinberg, *Recent Developments in Quantum Optics*, R. Inguva, ed., Plenum Press, New York, 145 (1993).

13. "Measurement of the Single-Photon Tunneling Time", A. M. Steinberg, P. G. Kwiat, and R. Y. Chiao, Phys. Rev. Lett. **71**, 708 (1993).

14. "High-Visibility Interference in a Bell-Inequality Experiment for Energy and Time", P. G. Kwiat, A. M. Steinberg, and R. Y. Chiao, Phys. Rev. A **47**, R2472 (1993).

15. "Faster than Light?", R. Y. Chiao, P. G. Kwiat, and A. M. Steinberg, Sci. Am. **269**, 52 (1993).

16. "The Single-Photon Tunneling Time", A. M. Steinberg, P. G. Kwiat and R. Y. Chiao, in *XXVIIIe Rencontre de Moriond*, J. T. T. Van *et al.*(eds.) Editions Frontières, France (1993).

17. "High Efficiency Single-Photon Detectors", P. G. Kwiat, A. M. Steinberg, R. Y. Chiao, P. Eberhard and M. Petroff, Phys. Rev. A **48**, R867 (1993).

Accepted for publication

18. "Absolute Efficiency and Time-Response Measurement of Single-Photon Detectors", P. G. Kwiat, A. M. Steinberg, R. Y. Chiao, P. Eberhard, and M. Petroff, Appl. Opt. (1993).

19. "Comment on 'Einstein-Podolsky-Rosen state for space-time variables in a two-photon interference experiment'", P. G. Kwiat, A. M. Steinberg, and R. Y. Chiao, Phys. Rev. A (1993).

20. "Three Proposed 'Quantum Erasers'", P. G. Kwiat, A. M. Steinberg and R. Y. Chiao, Phys. Rev. A (1993).

Submitted for publication

21. "Proposal for a loophole-free Bell's inequality experiment", P. G. Kwiat, P. H. Eberhard, A. M. Steinberg and R. Y. Chiao, Phys. Rev. A, submitted (1993).

Acknowledgements

Ack.I. Introduction

There are so many people who are in some sense responsible for the present work that it is hard to know where to begin. I suppose in a very general sense I might "play it safe" and acknowledge everyone within my backward light cone (more correctly, every-one whose backward light cone intersected with mine). And if our notions concerning the impossibility of superluminal communication turn out to be false, then perhaps I should also be thanking people from my future light cone. But this seems extreme.

Ack.II. Scientific Support

Without a doubt the person who has had the biggest influence on me over the last six years has been my advisor and friend, Professor Raymond Y. Chiao. He constantly demonstrated what it means to be a Principle Investigator, to look beyond formal mathematics to the underlying physical issues. His encouragement to "be creative" and willingness to listen to (and occasionally generate) "fringe" ideas has been central to my growth as a physicist. Moreover, his support as a mentor and a friend, and his continual sense of humor have made the last six years seem like only five and a half. I would also like to recognize Prof. Chiao's daughters as responsible for filling me with a constant sense of urgency throughout my graduate career, so that I might graduate before at least *one* of them.

I should in addition thank Dr. John Garrison for many interesting discussions about some of our experiments, in particular the issues surrounding Berry's phases in a GRIN lens. Dr. John Clauser has also been very helpful, sharing his insights on tests of Bell's inequalities, the "quantum" nature of the photon, and the personalities that participated in the early days of these topics.

I would also like to acknowledge my collaborator Dr. Philippe Eberhard for being always supportive and understanding in times of experimental hardship, for knowing when *not* to come into the lab, and for reducing the required detection efficiency for a loophole-free test of Bell's inequalities to only 67%. Without Philippe's political savvy and organizational skills, the tests on the high-efficiency detectors, and the efforts toward a final Bell-type experiment would never have happened (and this thesis would be considerably lighter!). Along the same lines, Dr. Michael Petroff of Rockwell International Corp. is also to be acknowledged, for supplying the detectors (and supporting equipment), and his own personal expertise in these matters.

Ack.III. Sponsors

I am extremely grateful to the Office of Navel Research for its long-term support of my graduate studies (under Contract N00014-88-K-0126, Grant N00014-90-J-1259, and the AASERT Graduate Funding Program), and also to the Department of Education for a two-year fellowship. Without these funds none of the work described herein would have been possible. I should like at this time to thank in addition my other corporate sponsors, PepsiCo, for contributing millions[1] to my undergraduate education at M. I. T., and the makers of all caffeinated beverages, for keeping me awake the past decade.

Ack.IV. Family

It will not go without saying that my family played a crucial role in the above work. Besides the obvious "begatting", my parents have always been supportive of and encouraging me in my endeavors. In particular, I am most grateful to my father, who knows all about the SPICE business, and always asks the most important question: "But are you having any fun?" With her numerous "care-packages" of chocolate-chip

[1] pennies

cookies, my sister Donna has carried my through many a long night. And my brother Glenn essentially forced me to get my doctorate, in order to keep up with the men of the family. Lastly, I thank my wonderful inlaws, Nancy and Dean, without whom I would not have such a superb collection of nieces and nephew to coax into the world of physics.

Ack.V. Friends and/or Roommates

I am especially grateful to the many friends I have had over the past years, whose companionship has eased a difficult journey. Particularly I would like to recognize Tim Dalby, who is nothing if not my friend, supportive to the end; and Gail Tillman, as kind a woman as I've ever met.

I would also like to thank my roommates, past and present--James Schaeffer, Peter Garafolo, Alaine Gold, and Dallas Middaugh--who never complained about my graduate student hours or the massive entropy experiment being conducted in our living room.

I must also extend my gratitude to my many sparring partners at the Berkeley Aikikai, and of course to Shibata Sensei, for helping me stay centered, and for not permanently damaging me. Our practice was invaluable to me.

To the many women who subtly encouraged me to spend long nights and weekends in the lab, by refusing their company; and to the few who truly offered solace and distraction in times of experimental hardship--I thank you.

And lastly (in this section), I would like to thank Trevor, Zuke, Mal-Zyliq-Bravos, Pectro-Kismet, Durinna-Gwen, Evero, Odessia, and especially Targa, who taught me to struggle forth against all odds, in the face of great adversity (partially self-inflicted!)

Ack.VI. Partners

I would first like to acknowledge the invaluable contributions of the annual lab assistants: Steve Wilkinson, Ken Honer, Rick Bentley, Rob Larisch, Bruce Johnson,

and Grant McKinney. Much like the "roadies" of a touring rock band, it is these gentlemen who make sure that all the reprints are sent out, all the viewgraphs made, and all the soda cans crushed. Thanks for your help.

Next, I must offer my deep thanks to Ivan Deutsch, theorist extraordinaire, who patiently tolerated my constant questioning as to exactly what it was he did, and amazed me with his tremendous knowledge of and insight into Brady Bunch trivia, and who was always willing to lend an empathetic ear. I am not the least bit bitter that he *stole* my NRC fellowship at NIST, thereby ruining my chances for the good life!

Lastly, I come to the people who are most directly responsible for the present opus, as well as for my present frame of mind, to my partners Bill Vareka and Aephraim Steinberg, who in the deep of night trustingly shared their deepest, darkest secrets with me[2] via walkie-talkie as we collected data. It was Bill's great experimental know-how which enabled us to set up the laboratory from scratch, Bill's negotiating expertise which gained us an extra three months coverage on the laser (so that it would die a mere *one* month out of warrantee), and Bill's wife who ensured that we did not spend *every* night up to the wee hours (thanks, Sue). Aephraim has throughout been a comrade beyond compare. Although occasionally I despised him for making me *think* (the nerve!), in the end I was always the better for it. His Pod-liness, P, is only barely surpassed by his ability to somehow not fail to avoid annoying people by it. If we label this ability A, then we may express the relation as

$$P \leq A \, . \qquad \text{(Ack.1)}$$

It is a shame that he and I will have to compete for the same positions down the line. Presumably, whichever one of us discovers a mechanism for sending superluminal signals will have the head start.

---

[2] After completion of the present work, I plan to publish a steamy best-seller based on these sordid details.

Chapter 1: Entanglement, Interference, and an Overview

"What do you wish from me?" the master asked.

"I wish to be your student and become the finest ... in the land," the boy replied. "How long must I study?"

"Ten years at least," the master answered.

"Ten years is a long time," said the boy. "What if I studied twice as hard as all your other students?"

"Twenty years," replied the master.

"Twenty years! What if I study day and night with all my effort?"

"Thirty years," was the master's reply.

"How is it that each time I say I will work harder, you tell me that it will take longer?" the boy asked.

"The answer is clear. When one eye is fixed upon your destination, there is only one eye left with which to find the Way."

-Anonymous; taken from Zen in the Martial Arts, by Joe Hyams

1.I. Introduction

The present treatise describes research focussing on the nonlocality inherent in quantum mechanics, as manifested by entangled states of two or more particles, and the role of (in-)distinguishability in interference. Central to all of the experiments and proposals included herein are the pairs of correlated photons produced when an ultraviolet photon interacts with a nonlinear optical medium, and via the process of spontaneous parametric down-conversion splits into two infrared photons. These pairs possess the useful property that one can produce essentially single-photon states by triggering on one of the photons, and counting the other in coincidence. The photons

may also be prepared in various types of entangled states. In the next section we shall briefly discuss these states, why they are preferred over others, and how the process of parametric down-conversion is an ideal source of them. Sect. 1.III reviews the Feynman approach to understanding interference, which we shall employ throughout, while Sect. 1.IV gives a preview of the remainder of this work.

1.II. Entangled States

It has been said that "entangled states lead to interesting effects"[1]. Nearly all of our experiments and proposals rely on them. A quantum-mechanical state $|\psi\rangle_{1,2}$ of two particles (or more generally, two systems) is said to be an "entangled" state if there is no basis in which the state can be factorized into a single product state of the form $|\chi\rangle_1 \otimes |\Phi\rangle_2$. The canonical example is the spin (here, polarization) singlet state:

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|H_1, V_2\rangle - |V_1, H_2\rangle\right), \tag{1.1}$$

which describes a pair of particles--photons, say--whose polarizations are strongly correlated, even though the polarization of either photon individually is not well-defined. If we measure the polarization of one of them to be horizontal (H), then the other is certain to be vertically (V) polarized, and vice versa. Moreover, the state described by Eq. (1.1) is symmetric, so no matter what polarization-component we measure for the first particle, the polarization of the other one will be orthogonal. This rather strange state of affairs constitutes an undeniable *nonlocality* in quantum mechanics. Although one can concoct theories which are completely local to explain some of the correlations, *no* such theory can completely mimic all the predictions of quantum mechanics, as proved by Bell in 1964 [Bell, 1964], and discussed in Chaps. 7, 8, and 10. In

[1] Although admittedly this is a quote of the author several years back, and thus of dubious origin, it should be pointed out that a great many people nodded when he spoke it at a conference.

particular, simple classical-field models are in contradiction with quantum mechanics, and more importantly, with experimental results.

Using the correlated photons produced in spontaneous parametric down-conversion, it is possible to create polarization-entangled states like that of Eq. (1.1) (see, for instance, Chaps. 5 and 10). However, it is a remarkable feature of the down-conversion process that the photons are automatically produced in an *energy*-entangled state, as discussed in the next chapter:

$$|\psi\rangle = \int dE\, A(E)\, |E\rangle_1\, |E_0 - E\rangle_2 \,. \tag{1.2}$$

According to the standard Copenhagen interpretation, the meaning of this entangled state is that when a measurement of the energy of one photon results in a sharp value E', there is a sudden "collapse" of the wavefunction such that the other photon, no matter how remote, instantly also possesses a sharp value of energy E_0 - E'.

Just as one can investigate nonlocality with Eq. (1.1) by using variable polarization analyzers, one can similarly investigate the nonlocality associated with Eq. (1.2) by using variable *frequency* analyzers. We have done this, using both narrow-band filters (Chap. 4), and unbalanced interferometers for this purpose (Chap. 8). The interpretation of our results is that Nature is nonlocal, and contrary to what one might naively think, the energy and time of emission of the down-converted photons have no definite values until we make a precise measurement, at which point both the photon we measured and the other one acquire well-defined properties.

The states described by Eqs. (1.1) and (1.2) involve entangling similar types of quantum systems, i.e., entangling the *polarization (energy)* of one photon with the *polarization (energy)* of a second photon. We shall see in the discussion of quantum

erasers (Chaps. 5 and 6) that it is also possible to entangle different types of systems. For example, the trajectory of a photon may be linked to its *own* polarization (Chap. 5), or to the energy of a different photon (Chap. 6). In fact, not only is it *possible* for this to occur--it is crucial, from the standpoint of measurements. The quantum-mechanical description of a measurement (as given by von Neumann [von Neumann, 1983], and modified by others [Peres, 1980; Peres, 1986 ; Wheeler, 1983; Zurek, 1986; Zurek, 1991]) is that the system which one is measuring becomes entangled with the states of the measuring apparatus. The measuring apparatus similarly becomes entangled with the states of environment, and an effective irreversibility is achieved, somehow destroying the subtle quantum coherences (of course, the whole problem in measurement theory is that a linear theory such as standard QM cannot actually lose the information). Clearly, we need to have a better understanding of the role entanglement plays in the measurement process.

As alluded to above, our source of entangled states are the correlated photon pairs from spontaneous down-conversion. We shall describe their characteristics in much greater detail in the next chapter; it suffices here to stress that due to the strong time, energy, and momentum correlations of the twin photons, they are ideal for investigating many fundamental quantum-mechanical phenomena.

I.III. Interference á la Feynman

Most of the experiments discussed in this manuscript involve interference, or *lack* of it, in some form. While the rigorous formalism of quantum optics is certainly capable of explaining the results, we find that Feynman's procedure for calculating final probabilities is much more useful (and definitely quicker!) for gaining an intuitive, comprehensive picture regarding the nature of interference. There are two basic rules to calculate the probability of a certain outcome, given a specific initial state [Feynman, 1965]:

1. If the processes leading to the particular outcome are indistinguishable, then one must add the probability amplitudes of these processes, and then take the absolute square to obtain the probability.

2. If, on the other hand, the contributing processes are distinguishable, then their probabilities (i.e., the absolute square of the amplitude of each process) should be added.

It is remarkable that these two simple rules, correctly applied, can account for all interference effects. In particular, it is the indistinguishable processes which lead to interference, for the *relative* phase information is not lost in the taking of the absolute square; in contrast, the distinguishable processes contribute to a non-interfering background, because any coherence between them is lost when the absolute square is taken. The key point is that one must thoroughly understand what is meant by *distinguishable* and *indistinguishable*. As used in Feynman's rules, these constructs concern what is *in-principle* knowable. It matters not whether one makes an effort to actually distinguish two "distinguishable" processes in practice. The mere possibility that one *could* distinguish means that rule 2 applies.

Of course, the physical content of the Feynman rules is contained within the standard quantum formalism, according to which one determines the total state of a system (e.g., an interfering particle and any "measuring apparatus" along the way), projected onto a given outcome, and then takes the absolute square. If the various processes leading to the outcome of interest are distinguishable, then the component final states will be orthogonal, and no cross-terms will survive. Alternately, for indistinguishable processes the final states will not be orthogonal, and therefore the cross-terms containing relative phase information will remain, yielding interference. Thus, we see that by "distinguishable" processes, we mean those that lead to orthogonal final states.

Given the preceding discussion, one is led to ask, just how distinguishable is distinguishable enough (to remove the possibility of interference)? Of course, there is no hard cutoff--the more information one has about which way a particle goes in a two-slit experiment, the lower the visibility of the interference fringes[2], until finally they are indistinguishable (in an experimental sense) from random fluctuations. In Fig. 1.1 we show a plot of the visibility of fringes from a Mach-Zehnder interferometer as a function of attenuation in one of the arms. Remarkably, the attenuation must be quite high (implying a high degree of certainty about which way a given photon took) before the visibility is significantly affected.

Although we shall use the Feynman rules throughout, nowhere are they as conspicuous as in the quantum eraser experiments, discussed in Chaps. 5 and 6. In

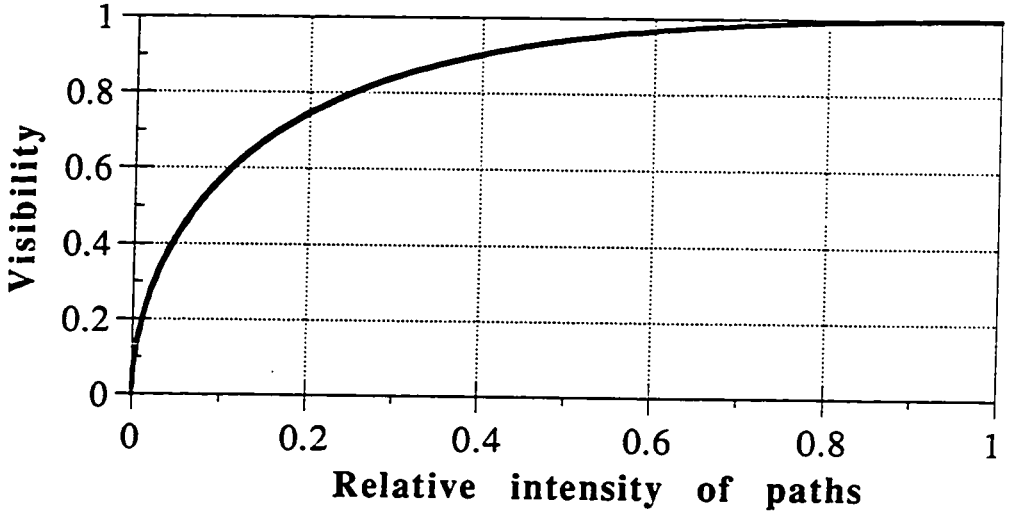


Figure 1.1 A plot of the visibility of interference fringes in a two-path interferometer, in which one of the paths is attenuated. Unless the attenuation is rather severe, implying a strong predilection for which path a given interfering particle might take, the interference is seen to persist.

[2] This has been put in more quantitative terms by Ya'sin and Greenberger [Ya'sin and Greenberger, 1986] and by Mandel [Mandel, 1991].

some sense, these are the archetypical demonstrations of the rules "at work". As is standard knowledge, if one attempts to determine through which slit a particle went in the canonical double-slit experiment, then interference will be lost. This is just the principle of complementarity. Typically some form of Heisenberg's uncertainty relation is cited as the underlying "mechanism" by which phase information is lost. However, in terms of Feynman's rules, it is the fact that the final states are distinguishable which is responsible. If one somehow manages to *erase* this distinguishability, then the interference can be recovered. That is the basic idea of the quantum eraser. Its interpretation relies wholly on the notions of distinguishability and indistinguishability leading to which-path information and interference, respectively.

I.IV. What Lies Ahead

Each chapter of this dissertation is intended to be somewhat self-contained, although all rely on the basic results of Chap. 2, in which we discuss the nature of the source of our correlated photons (viz., spontaneous parametric down-conversion), both in an abstract sense, and for our particular nonlinear crystal. In the following two chapters we present the results of two similar arrangements that involve sending one photon through an interferometer and the other straight to a detector. Chap. 3 focusses on the fact that the interference observed arises from a Berry's phase; by violating an inequality regarding the statistics of any classical light source, we shall conclude that the topological phase is essentially quantum mechanical. In Chap. 4 the above interferometer is set outside the white-light fringe regime, so that no fringes exist in the rate of singles detection at the output port of the interferometer. Nevertheless, by placing a narrow-band filter in front of the detector collecting the conjugate photons, one can observe the interference in the rate of *coincidences* between the detectors. The results are interpreted in terms of the "collapse of the wavefunction".

Quantum erasers are the subject of Chaps. 5 and 6. In the former, the results of the first quantum eraser are presented, after a general discussion of the phenomenon. Chap. 6 discusses several proposals for "new and improved" quantum erasers, which are pedagogically superior to the experiment of the previous chapter. The central element of these experiments is the entangled state of system plus measuring apparatus. In Chap. 7 we discuss entangled states from the viewpoint of the EPR paradox and Bell's inequalities. No substantively new results are reported in this review of the issues involved, although it may be useful to see in one place two rather different derivations of Bell's inequalities. We also discuss in some detail the various technical loopholes still to be overcome in any "true test" (i.e., completely incontrovertible) of local realism, as well as the leading interpretations and implications of violations of Bell's inequalities. Finally, we briefly point out how the nonlocal correlations may used to send un-eavesdroppable messages via quantum cryptographic schemes.

Having laid the groundwork in Chap. 7, in Chap. 8 we turn to a Bell-type test relying on energy and time correlations, instead of the typical spin (or equivalently, polarization) correlations. Based on a two-photon interference effect (predicted first by Franson [Franson, 1989]) which is interesting in its own right, the experiment allows us to challenge the locality of nature in a never-before tested realm. We shall discuss the relation to the original EPR proposal, as well as the explicit non-classicality of the results.

In Chap. 9 we describe an experiment to determine the absolute detection efficiency of single-photon detectors, relying on the strong momentum correlations of the photons from our source, and the fact that they are always produced in pairs. We believe our measured efficiencies of ~75% to be the highest reported to date, with indications that they may actually be improvable to 90% or better. Such high efficiencies are absolutely crucial if one desires to close the technical loopholes that have plagued

every Bell's-inequality until now. A suitable *source* is also required to close these loopholes. In Chapter 10 we propose such a source, and investigate in detail its advantages over previous schemes, both actual and proposed. Specifically, we show how it should permit a loophole-free test of Bell's inequalities. Lastly, in Chap. 11 we summarize and discuss some of the most important results, emphasizing the relationships between the various experiments, and concluding with some general remarks concerning the various sorts of non-classicality revealed in our experiments.

A number of Appendices have been included also, to admit detailed calculations not appropriate for the general text, and also specifications of our various detectors. In Appendix A an inequality for the statistics of a classical light field, which is violated by the photon pairs from our light source, is derived. Appendices B1 and B2 present formal calculations of two quantum eraser effects. Two approaches to the error analysis in our efficiency measurements are given in Appendix C. Various aspects of the Bell's inequality-proposal scheme of Chap. 10 are detailed in Appendix D: D1 contains the derivation of type-II vector phase matching in a negative uniaxial crystal, such as BBO; D2 considers the effect of finite iris sizes on collection efficiency; and D3 presents a general calculation for our proposal, allowing for non-ideal optical elements. In Appendix E we compile specification sheets for the three types of detectors used for our experiments.

Chapter 2: Spontaneous Parametric Down-Conversion in Potassium Dihydrogen Phosphate

I am never forget the day I am given first original paper to write. It was on analytic and algebraic topology of local Euclidean metrization of infinitely differentiable Riemannian manifold. *Bozhe moi*! This I know from nothing.

--Tom Lehrer, singing "Lobachevsky"

Man: Hello my boy. And what is your dog's name?
Boy: I don't know. I call him Rover.

--Anonymous

2.I. Introduction

Because all of the experiments described herein use the correlated photons emitted in spontaneous down-conversion, it is important to understand the relevant characteristics of this source. Basically, contrary to the somewhat daunting name, the salient features are simply described: Due to a nonlinear interaction in a crystal (KDP, for our experiments), an ultraviolet "parent" photon may spontaneously split into two infrared photons, historically called the "signal" and "idler". Due to the nature of the production process, they are strongly correlated in time (they are emitted within 100 femtoseconds of each other [Hong et al., 1987; Steinberg et al., 1992], as determined by the inverse-bandwidth of each), in energy (the sum of their energies is essentially a constant, to within the extremely narrow pump bandwidth; see Chaps. 4 and 8), and in momentum (given the direction of one photon, the direction of the twin is determined to within less than a milliradian).

For our source, we employ type-I vector phase-matching, so that the identically-polarized signal and idler photons are emitted on opposite sides of the "down-conversion cone"--see Fig. 2.1. The opening angle of the cone depends on the wavelength of the light, for although the *sum* of the frequencies is well-defined, each photon individually possesses a broad bandwidth. Selecting out one photon with a narrow-bandwidth filter and small iris immediately constrains the energy and direction of the conjugate photon.

In Sect. 2.II we describe theoretically the state of light produced in down-conversion, while the calculation of type-II vector phase-matching is reviewed in Sect. 2.III. The particulars of our source are described in Sect. 2.IV. Finally, in Sect. 2.V, we discuss briefly three "practical" applications of the down-converted pairs.

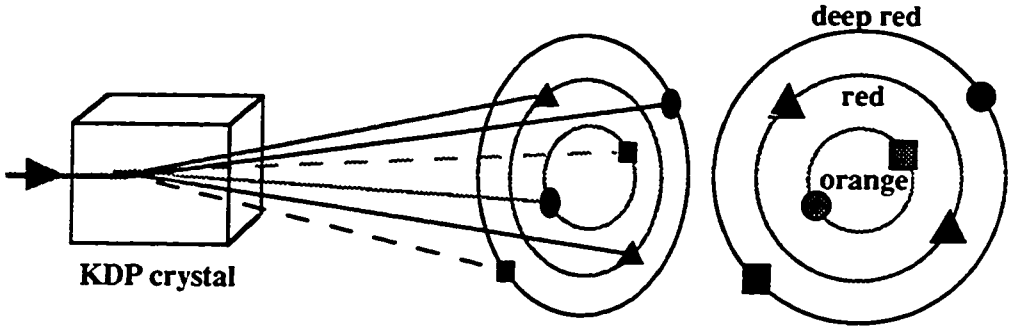


Figure 2.1 A sketch of the down-converted light cones emitted from a nonlinear crystal, pumped from the left. Some of the pump photons are split into two lower-frequency photons--three such pairs are shown, i.e., the two squares represent a conjugate pair, as do the two circles and the two triangles. The opening angle of the cone on which a given photon exits depends on the color of the photon. In the diagram the triangles represent the degenerate case, where both signal and idler have exactly half the energy of the pump.

2.II. General Theory

A full treatment of parametric down-conversion is beyond the scope of this work, and superfluous in any event, as the theory is described in many standard books on nonlinear optics [Shen, 1984; Yariv, 1988; Dmitriev et al., 1991]. We will focus on those aspects not generally covered in the standard texts, which are relevant for our experiments. Following Hong [Hong, 1988; Hong and Mandel, 1985], we can write the pump input state to the crystal as a superposition of single-photon states $|\mathbf{k_p}\rangle$ (since the pump photons are assumed not to interact with one another):

$$|\psi\rangle_{in} = \int d^3k_p\, \phi_p(\mathbf{k_p})\, |\mathbf{k_p}\rangle\ , \tag{2.1}$$

where $\phi_p(\mathbf{k_p})$ is assumed to describe a nearly-monochromatic (at ω_p), highly-directional pump. After interacting[1] with the nonlinear medium, the state is

$$|\psi\rangle_{out} = \alpha\,|\psi\rangle_{in} + \int d^3k_p\, \phi_p(\mathbf{k_p}) \int d^3k_s \int d^3k_i\, \tilde{\chi}^{(2)}[\omega_p, \omega_s, \omega_i]$$

$$\times\, \delta(\omega_p - \omega_s - \omega_i - \Delta\omega) \prod_1^3 \mathrm{sinc}\left[(\kappa_p - \kappa_s - \kappa_i)_m L_m/2\right]\, |\mathbf{k_s}, \mathbf{k_i}\rangle\ , \tag{2.2}$$

where the first term represents the unconverted pump photons, and the second term describes the signal and idler photons, with frequencies ω_s and ω_i, respectively. $\tilde{\chi}^{(2)}$ is related to the second-order susceptibility, but has different dimensions, and is assumed

[1] A much-simplified interaction Hamiltonian for down-conversion is $H = \chi^{(2)} a_s^{\dagger} a_i^{\dagger} a_p + \text{h. a.}$, where $a_{s(i)}^{\dagger}$ denotes the creation operator of the signal (idler) photon, and a_p the annihilation operator for the pump photon. See [Hong, 1988; Hong, 1985] for the complete interaction Hamiltonian that leads to (2.2).

to be a slowly-varying function of the frequencies. The delta-function describes energy-conservation; the uncertainty $\Delta\omega$ is essentially the reciprocal of the effective interaction time, effectively infinite for a cw pump. (Note: There has been a bit of confusion about this in the literature, so we shall discuss it in greater detail presently.) The sinc-function [$\equiv (\sin x)/x$] demonstrates the need for phase matching: the photon momentum vectors *inside* the material (the κ's, related to the free-space k-vectors by $\kappa_j = n(\omega_j)\, k_j$, where n is the index of refraction) must satisfy momentum conservation to within $\sim 1/L_m$ (where L_m is the relevant dimension of the nonlinear medium) to have efficient down-conversion. In practice, the transverse L's are much larger than the transverse extent of the pump beam [as governed by $\phi_p(\mathbf{k}_p)$], so it is the divergence of the pump that leads to a range of efficient phase-matching. Also, although the crystal's finite longitudinal length will lead to a finite bandwidth of down-converted light (along any specific direction), this does not affect the frequency-correlation imposed by the delta-function of (2.2).

Eq. (2.2) has contained within it the tight time, energy, and momentum correlations of the photons. However, it will be easier to discuss the relevant features using the less cumbersome state:

$$|\psi\rangle = \int d\omega_s\, A(\omega_s)\, |\omega_s\rangle_s\, |\omega_p - \omega_s\rangle_i\ , \qquad (2.3)$$

where $A(\omega_s) = A(\omega_p - \omega_s)$ is the complex probability amplitude for finding one signal photon with a frequency ω_s (i.e., in the n=1 Fock state $|\omega_s\rangle_s$) and one idler photon with a frequency $\omega_p - \omega_s$ (i.e., in the n=1 Fock state $|\omega_p - \omega_s\rangle_i$). In going from (2.2) to (2.3) we have dropped the uninteresting term corresponding to non-conversion of the pump photon, and have assumed a plane-wave pump and a crystal very large compared to a wavelength (i.e., we have assumed that the phase-matching constraints are

satisfied[2]), allowing us to focus on the energy-correlations of the photons. We have also assumed a cw, monochromatic pump, so that the delta-function enforcing energy-conservation ($\omega_s + \omega_i = \omega_p$) is exact.

Note that energy conservation does not forbid the signal photon from having a broad spread in energy, however, and the idler photon from having a conjugately broad spread in energy. In principle, the spread in signal and idler frequencies is limited only by phase-matching considerations (see Sect. 2.III), but in practice, it is usually limited by the bandwidth of the filters and size of irises placed in front of the detectors (see Sect. 2.IV). Nevertheless, the energy-correlation is exact (to within the bandwidth of the pump). We now elaborate on this. It has been suggested [Rubin and Shih, 1992; Shih et al., 1993] that the idealized delta function in the frequencies of the two photons, $\delta(\omega_s + \omega_i - \omega_p)$, is in practice "replaced by functions with nonzero widths, giving . . . $\omega_s + \omega_i \pm \Delta\omega$", where the uncertainty is due to the "finite size of the crystal and the finite interaction time of the down-conversion". These same authors acknowledge that "one can easily arrange a narrow enough spectral bandwidth of the pump field by means of a single mode laser". Thus, the implication is that the energy uncertainty is primarily due to the *finite length* of the crystal. This is not so, because although the finite crystal size does lead to an uncertainty $\hbar\Delta k$ in total momentum, it is not correct to simply multiply this by c to obtain the uncertainty in total energy. Energy conservation imposes a much stricter requirement, due to the fact that the pump photon originated in a continuous-wave, single-mode laser. The effective interaction time of the down-conversion process is determined in this case by the coherence length of the pump laser, *not* by the much-

[2] The assumption of phase-matching also fixes the polarizations of the down-converted photons--for type-I phase-matching in a crystal such as KDP, they will be ordinary-polarized, while the pump is extraordinary-polarized. In Chap. 10 and in Appendix D1 we will examine type-II phase-matching, in which only one of the down-converted photons is ordinary-polarized, and the other is extraordinary-polarized (in a crystal such as BBO, the pump is also extraordinary-polarized).

shorter crystal length. Although a finite crystal size gives an energy-uncertainty, a finite bandwidth, to each of the photons individually, the *sum* of their energies remains extremely well defined (to within the inverse of the pump coherence time). This is the central feature of the entangled state (2.3). The claim that a finite interaction time can limit the two-photon coherence length is correct, in that the two-photon coherence length can never be greater than the *pump coherence length* (cf. the total loss of interference for path lengths greater than the pump coherence length, shown in Fig. 8.2).

The tight time-correlation implied by Eq. (2.3) can be seen by calculating the joint probability of detecting a signal photon at time t_s and its conjugate idler at time t_i. According to the standard Glauber theory of photodetection [Glauber, 1963], this is given by

$$G^{(2)}(t_s,t_i;\, t_i,t_s) = \langle\psi \,\big|\, \hat{E}_s^{(-)}(t_s)\, \hat{E}_i^{(-)}(t_i)\, \hat{E}_i^{(+)}(t_i)\, \hat{E}_s^{(+)}(t_s) \,\big|\, \psi\rangle$$

$$= \left| \hat{E}_i^{(+)}(t_i)\, \hat{E}_s^{(+)}(t_s) \,|\, \psi\rangle \right|^2, \qquad (2.4)$$

where $\hat{E}_{s(i)}^{(-)}(t_{s(i)})$ and $\hat{E}_{s(i)}^{(+)}(t_{s(i)})$, the negative- and positive-frequency parts of the electric field for the signal (idler) mode, may be Fourier expanded in terms of frequency-dependent detection efficiencies $\eta_s(\omega_s)$ and $\eta_i(\omega_i)$, and creation and destruction operators $\hat{a}^\dagger(\omega)$ and $\hat{a}(\omega)$:

$$\hat{E}_s^{(+)}(t_s) = \int d\omega_s\, \eta_s^*(\omega_s)\, \hat{a}_s(\omega_s)\, e^{-i\omega_s t_s}\ , \qquad (2.5a)$$

$$\hat{E}_s^{(-)}(t_s) = \left(\hat{E}_s^{(+)}(t_s)\right)^\dagger, \qquad (2.5b)$$

$$\hat{E}_i^{(+)}(t_i) = \int d\omega_i\, \eta_i^*(\omega_i)\, \hat{a}_i(\omega_i)\, e^{-i\omega_i t_i}\ , \qquad (2.5c)$$

and

$$\hat{E}_i^{(-)}(t_i) = \left(\hat{E}_i^{(+)}(t_i)\right)^{\dagger} . \qquad (2.5d)$$

Substituting these and Eq. (2.3) into Eq. (2.4) we have

$$G^{(2)}(t_s, t_i;\ t_i, t_s) =$$

$$\left| \iiint d\omega_s\, d\omega_i\, d\omega\, \eta_s^*(\omega_s)\, \eta_i^*(\omega_i)\, A(\omega)\, e^{-i\omega_s t_s}\, e^{-i\omega_i t_i}\ \hat{a}_s(\omega_s)\, \hat{a}_i(\omega_i)\ |\,\omega\rangle_s\ |\,\omega_p - \omega\rangle_i \right|^2$$

$$= \left| \int d\omega\, \eta_s^*(\omega)\, \eta_i^*(\omega_p - \omega)\, A(\omega)\, e^{-i\omega t_s}\, e^{-i(\omega_p - \omega) t_i} \right|^2 , \qquad (2.6)$$

where we have used the canonical commutation relations to set $\omega_s = \omega$ and $\omega_i = \omega_p - \omega$. We are interested in the ultimate limit of the time correlation between members of a signal-idler pair, so infinitely fast detectors are assumed, with frequency-independent efficiencies. Then Eq. (2.6) becomes

$$G^{(2)}(t_s, t_i;\ t_i, t_s) = \left| \int d\omega\, A(\omega)\, e^{i\omega (t_i - t_s)} \right|^2 , \qquad (2.7)$$

which clearly vanishes unless the time difference t_i - t_s is less than frequency spread described by $A(\omega)$. Ultimately the spread is limited by the phase-matching constraints (see below), but in practice it is determined by the size of collection irises and by filters in front of the detectors. This behavior has now been verified in many experiments [Hong et al., 1987; Kwiat and Chiao, 1991; Steinberg et al., 1993].

2.III. Phase Matching in KDP

KDP is a negative uniaxial crystal ($n_o > n_e$). The indices of refraction are given [Dmitriev et al., 1991] by

$$n_o^2 = 2.259276 + \frac{0.01008956}{\lambda^2 - 0.012942625} + \frac{13.00522\,\lambda^2}{\lambda^2 - 400}\ , \tag{2.8a}$$

and

$$n_e^2 = 2.132668 + \frac{0.008637494}{\lambda^2 - 0.012281043} + \frac{3.2279924\,\lambda^2}{\lambda^2 - 400}\ . \tag{2.8b}$$

The effective nonlinearity for type-I phase matching (extraordinary-polarized pump, ordinary-polarized down-converted light) is $d_{eff} = d_{36} \sin\theta \sin 2\phi$. Here θ and ϕ are polar coordinates referring to z (the optic axis) and x, respectively. The nonlinear coefficient d_{36} for conversion of 351-nm light is ~5 x 10^{-13} m/V. The linear absorption coefficient in the range 350-530 nm is less than 0.005 cm^{-1}, while by 780 nm it has climbed to 0.024 cm^{-1}.

We now present the calculation for type-I vector phase-matching, which is basically just a statement of momentum conservation inside the crystal. From the schematic in Fig. 2.2 we can immediately write down the longitudinal- and transverse-momentum conservation relations:

$$\kappa_p = \kappa_s \cos\theta_s + \kappa_i \cos\theta_i\ , \tag{2.9a}$$

and

$$\kappa_s \sin\theta_s = -\kappa_i \sin\theta_i\ , \tag{2.9b}$$

where the κ's are the momenta *inside* the crystal, given by $\kappa_p = \omega_p\, n_e(\omega_p, \theta_{pm})$ $\kappa_s = \omega_s\, n_o(\omega_s)$, and $\kappa_i = \omega_i\, n_o(\omega_i)$. Here θ_{pm} is the angle between the pump and the

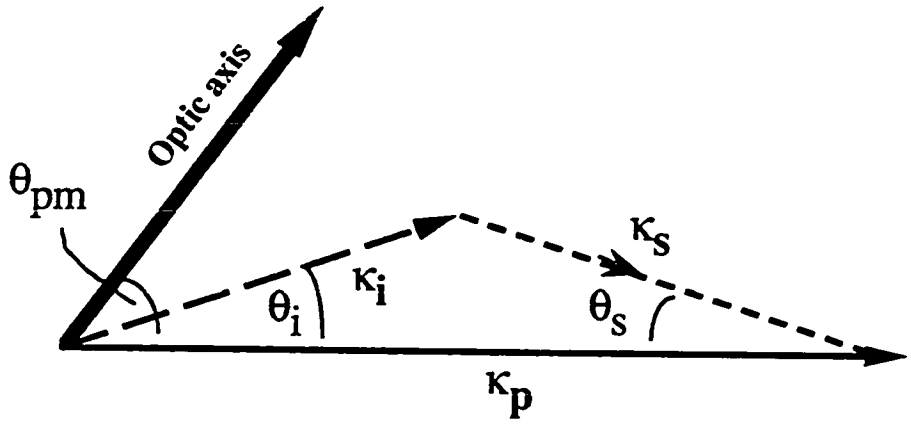


Figure 2.2. Schematic representation of type-I vector phase-matching (angles are exaggerated for clarity).

optic axis. Squaring Eq. (2.9b) and using the trigonometric identity, $\sin^2\theta = 1 - \cos^2\theta$, we have

$$\kappa_s^2 \cos^2\theta_s = \kappa_i^2 \left(\cos^2\theta_i - 1\right) + \kappa_s^2 \,, \tag{2.10}$$

Moving $\kappa_i \cos\theta_i$ to the left-hand side of Eq. (2.9a), and squaring and substituting (2.10) for $\kappa_s^2 \cos^2\theta_s$ gives

$$\kappa_{p,\theta_{pm}}^2 + \kappa_i^2 \cos^2\theta_i - 2\kappa_{p,\theta_{pm}} \kappa_i \cos\theta_i = \kappa_i^2 \cos^2\theta_i + \kappa_s^2 - \kappa_i^2 \,, \tag{2.11a}$$

or

$$2\kappa_{p,\theta_{pm}} \kappa_i \cos\theta_i = \kappa_{p,\theta_{pm}}^2 - \kappa_s^2 + \kappa_i^2 \,, \tag{2.11b}$$

so that

$$\theta_i = \cos^{-1}\left[\frac{\kappa_{p,\theta_{pm}}^2 - \kappa_s^2 + \kappa_i^2}{2\kappa_{p,\theta_{pm}} \kappa_i}\right] . \tag{2.11c}$$

Lastly, using Snell's law to obtain the laboratory angles θ'_s and θ'_i:

$$\sin\theta'_s = n_o(\omega_s)\sin\theta_s \quad \text{and} \quad \sin\theta'_i = n_o(\omega_i)\sin\theta_i \,, \tag{2.12}$$

where it has been assumed that the pump is incident normal to the crystal face. Using the formulas (2.8) for the indices of refraction, one can now generate plots of the output angles of the down-converted light (with respect to the pump beam), for various values of θ_{pm}, as shown in Fig. 2.3.

2.IV. *Our* Source

Our KDP crystal (custom-manufactured by Cleveland Crystals) was 10-cm long, with a 2.54-cm square cross-section[3]. It was housed in a sealed aluminum container, with index matching fluid (n=1.29) and fused silica end-windows. The crystal "input" was anti-reflection coated [single-layer of MgF_2, 1/4 wave at 351 nm] to reduce losses of the pump beam. The crystal "output" was broadband anti-reflection (BBAR) coated [multi-layer coating] to reduce output losses to less than 0.5% over the range 400-750 nm[4]. Our pump was a Coherent Innova-200 Argon-ion laser, operated at 351.1 nm[5]; the polarization was vertical, with a purity of 100:1. A "black-glass" filter was used to supress the blue-fluorescence from the laser. Typically, we found it necessary to further attenuate the pump (with a variable neutral density filter wheel), so that typical power into the crystal was only ~150 mW. After passing through the crystal, the remaining, unscattered uv beam was directed by a small mirror to a beam dump. A uv-

---

[3] In principle, since our pump-beam diameter was only 2 mm, this was "overkill"--in practice, it was useful on several occasions to work with a different part of the crystal, due to some sort of optical damage from the ultraviolet pump.

[4] Note: These directions were accidentally reversed in all experiments presented here, except for a few of the detector-efficiency measurements (Chap. 9).

[5] With the exception of the Berry's phase experiments and the dual-beam Michelson experiment (Section 8.III), all the experiments had the pump operating in single-line mode (from a temperature-stabilized intracavity etalon), with a nominal bandwidth of 50 MHz, corresponding to a coherence length of about 6 m.

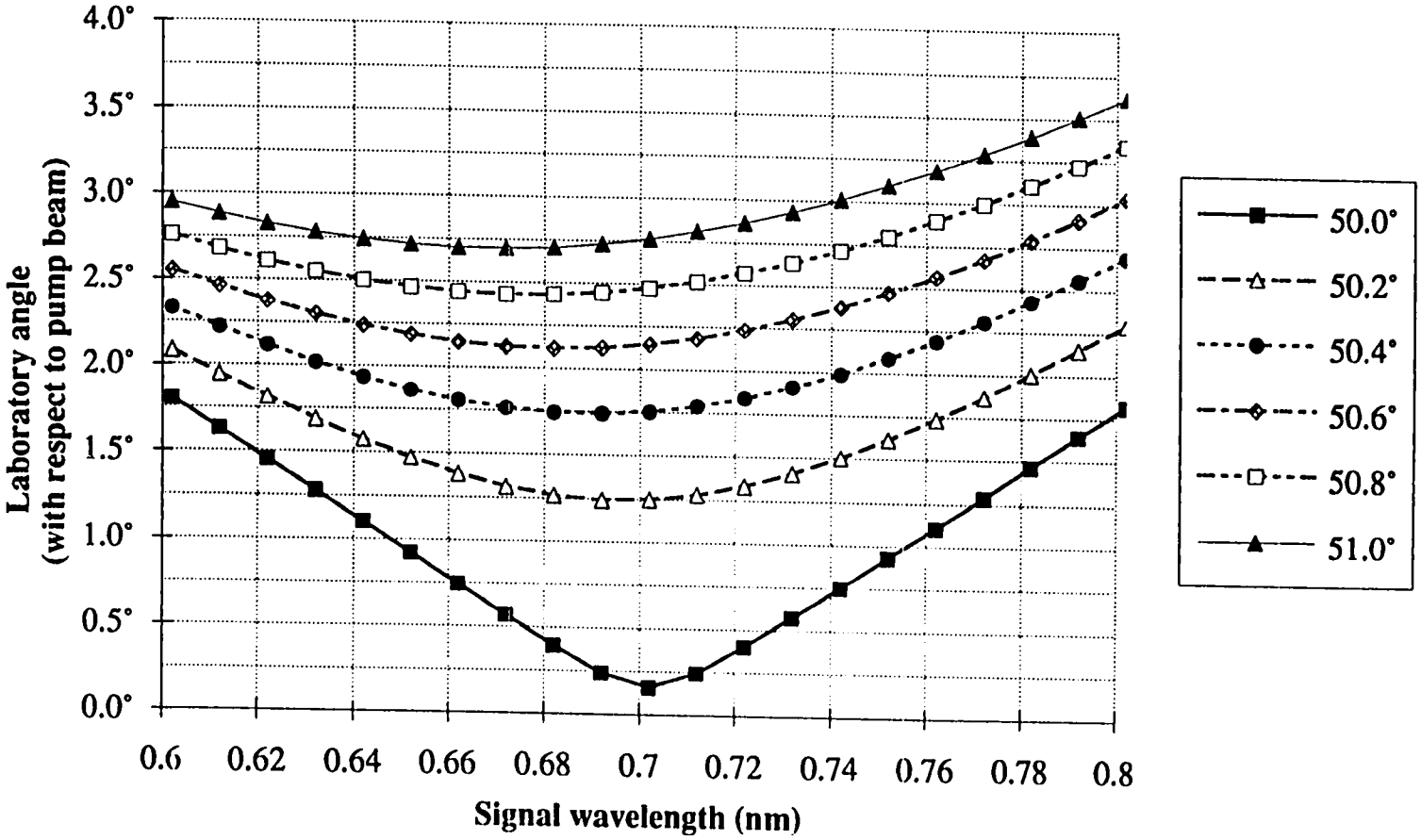


Figure 2.3 Calculation of the cone half-angles, with respect to the pump beam, versus wavelength in KDP. The various curves correspond to values of the angle between the pump beam (at 351.1 nm) and the crystal optic axis.

cutoff filter [Schott GG-475] was used to remove any scattered uv and laser fluorescence from the down-converted beamlines.

One can see from the curves of Fig. 2.3, that the bandwidth (over all angles) of the detected output light can be quite broad[6]. In practice, it is usually determined by filters and irises before the detectors. In most of the experiments described herein, we select out the nearly-degenerate signal and idler pairs at 702.2 nm. However, we did have occasion to map out parts of the angular spectrum, by translating a detector preceded by an interference filter transversly to the beam (see Fig. 2.4). The transverse profiles from several interference filters are shown in Fig. 2.5a. From the positions of the peaks of the various colors, a curve similar to Fig. 2.3 can be obtained (see Fig. 2.5b). We find that the angle (*in* the crystal) between the pump beam and the crystal optic axis is 50.5°; the degenerate (λ = 702 nm) signal and idler beams emerge on opposite sides of a cone whose half opening angle is 2.1°.[7]

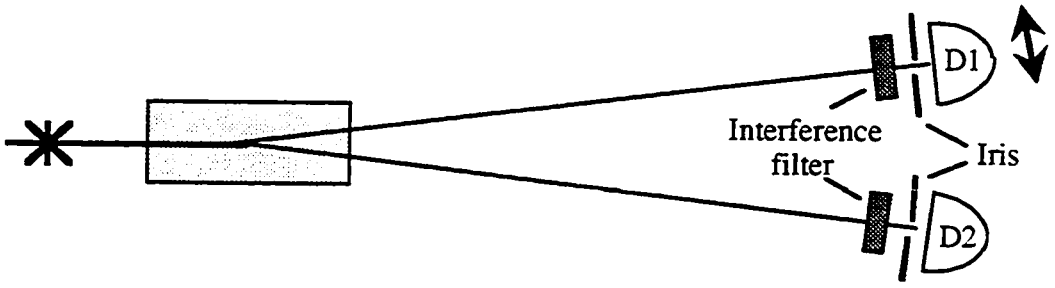
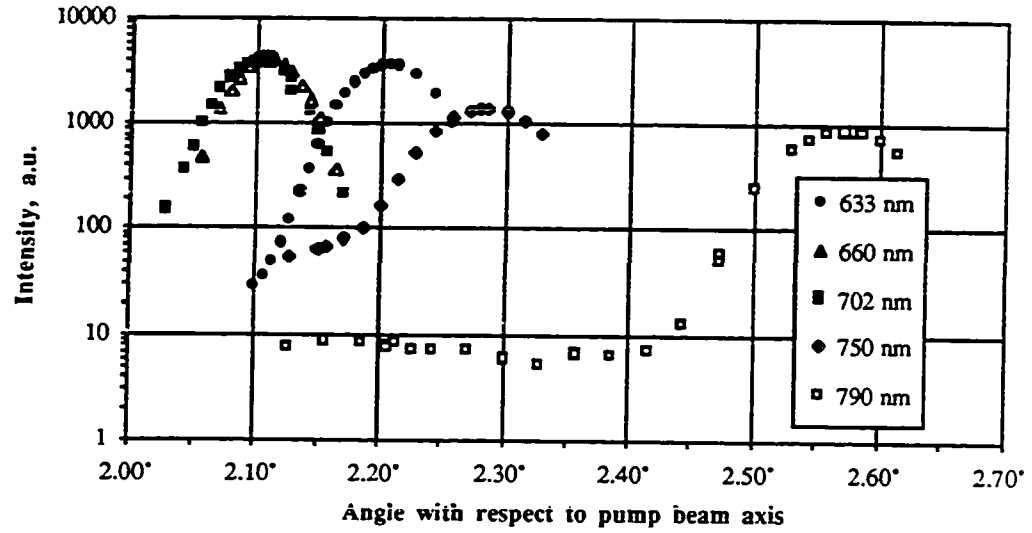


Figure 2.4 Simple setup to measure spectral profiles, either in singles or coincidence. The effective size of the iris, which must be deconvolved from all profiles, was typically 0.5 mm.

[6] As stressed earlier, though, the correlation between photons in a given signal-idler pair is very tight.

[7] The crystal optic axis was nominally cut at 50.3° to the normal to the crystal input face, but due to pump-beam wander, the pump was not always perfectly normal to this face.



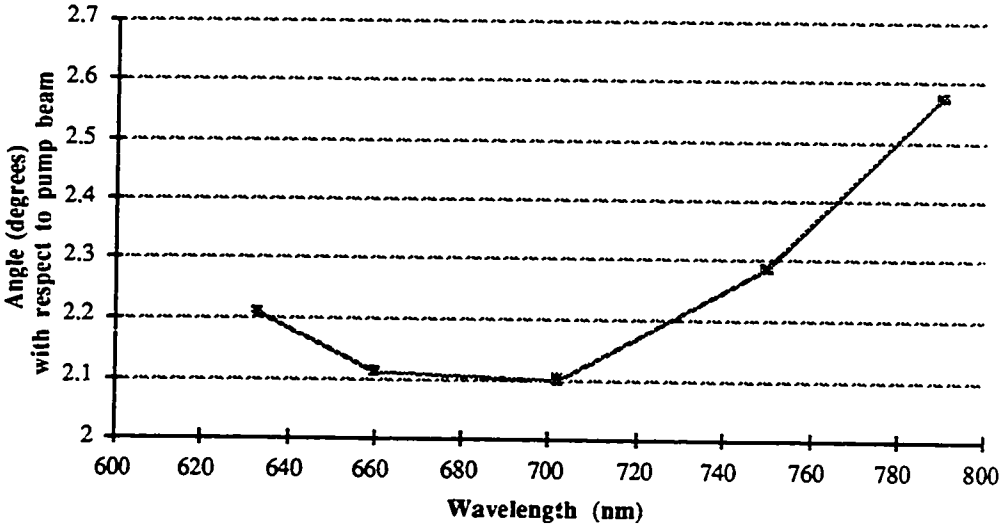


Figure 2.5 Data from transverse profiles of detector (see Fig. 2.4). a) Color filters used had different bandwidths; curves have been rescaled to a nominal 10-nm filter width, for an approximation of true relative intensity. b) The location of the peaks in a) is plotted against wavelength; cf. Fig. 2.3.

By looking at coincidence rates between the signal and idler detectors, as one of them is translated, one may investigate the conditional angular-profile; i.e., given that a photon of a well-defined color (determined by a narrow bandwidth filter) is detected at the fixed detector, what is the angular spread of its conjugate? We performed many such profiles. For a plane-wave monochromatic pump, in the limit of perfect phase-matching (i.e., infinitely long crystal), there will be no spread. In practice, the angular spread of a given color in coincidence arises from the pump-beam divergence.

There is an assymetry in the width for different parts of a given cone, however, and different directions at a given part. In particular, the angular correlation width for photons whose k-vectors lie in the plane determined by the pump and the optic axis (which we shall call the POA plane) should be roughly twice the width for photons whose k-vectors form a plane perpendicular to the POA plane. The reason can be seen from the curves in Fig. 2.3, where a 1° difference in the angle between the pump beam and optic axis corresponds to a ~2° shift in the opening angle of a given color. In other words, if one had a plane-wave pump, angling it slightly transversely to the POA plane would have almost no effect on the opening angles (with respect to the pump beam) of the down-converted cone. In contrast, angling the pump *in* the POA plane would actually change the opening angles, by roughly twice the angular tilt of the pump. All of the experiments described herein selected out signal and idler photons whose k-vectors form a plane perpendicular to the POA plane; i.e, the conjugate k-vectors lay on either side of the crystal optic axis.

Our pump was specified to have a full-angle beam divergence of 0.4 mrad, and a beam diameter (at the crystal) of 2 mm. Spatial profiles performed in coincidence showed that 1.5 m from the crystal, the signal beam conjugate to an idler beam selected by a 0.4-mm pinhole and a 0.86-nm FWHM filter (centered at 702.2 nm) had a horizontal size of 1.1 mm and a vertical size of 2.5 mm. Using near and far irises, or placing the iris in the focal-plane of a lens, we measured a horizontal (vertical)

divergence angle of ~1 mrad (~2 mrad).

We used three types of photon detectors in the experimental work described here. Specifications are reproduced in Appendix E. Initially, we employed photomultiplier tubes (Burle #C31034A-02), selected for high-sensitivity (typical efficiencies were 5% at 702 nm) and low dark count rate (~10 cps). These were cooled to approximately -30° C with Pacific Instruments #3470 Thermoelectric PMT Housings, and required biasing of about 2000 V. The active area of the PMT was 4 x 10 mm. Next we switched to silicon avalanche photodiodes (EG&G C30902S) operated in the Geiger mode, requiring a bias voltage of approximately –200 V. When cooled to -18° C (temperature controller circuit is listed in Appendix E, along with the passive-quench circuit), dark count rates were typically 500 cps. Detection efficiencies of 40% were measured at 702 nm. The active diameter of these devices is about 0.7 mm. Finally, we switched to single-photon counting modules (EG&G SPCM-200-PQ), which also employ silicon avalanche photodiodes, but with a much better performance than typical APDs. (See Sect. 9.IV for more explanation of *why* they are better.) Specifically, we have measured efficiencies of 75%, and dark counts of only about 60 cps. Moreover, these devices require only standard low-voltage power supplies, and produce a TTL-like pulse for each detection event. The main drawback is the small active area (0.1 mm-diameter), which makes alignment difficult.

2.V. Applications

Aside from the fundamental sorts of experiments which make up the bulk of this work, there are several more "practical" applications of the spontaneously-generated photons, which we will discuss briefly here. The first is communication. One can view the tight time correlations as leading to correlated intensity (noise) fluctuations between the signal and idler pairs[8]. Several researchers have proposed utilizing these

correlations in an optical binary communication channel [Mandel, 1984; Hong et al., 1985]. By modulating both the signal and idler beams, and using two separate receivers to detect their arrivals in coincidence, one can in principle achieve tremendous discrimination against background and noise. Other suggestions and demonstrations have been made, using the photons in various quantum cryptographic schemes [Ekert et al., 1992; Ekert, 1991], in which the nonlocal correlations arising from an entangled state allow two collaborators to share a secret, random key. This is discussed in more detail in Sect. 7.VII.

A second application of the down-converted pairs is in the measurement of propagation times, with very high resolution. In the simplest case, one would simply use one photon as a timing trigger for the other. The arrival times with and without the presence of some sample could be directly compared. In practice, current single-photon detectors are far too slow to be used directly. Fortunately, there exists a two-photon interference effect [Hong et al., 1987], described in Chap. 5, which permits one to measure timing to within a fraction of the coherence times of the down-converted photons. Moreover, due to the energy-entangled nature of the photons, there arises a dispersion-cancellation effect, so that normal group-velocity broadening of the photon wavepackets does not degrade the resolution [Steinberg et al., 1992a]. Using this technique, we have demonstrated that single photons propagate through glass at the group velocity [Steinberg et al., 1992b], and have measured the tunneling time of single photons through a one-dimensional photonic bandgap structure[9] [Steinberg et al., 1993]. We found that on average, the tunneling photon arrives *sooner* than it would have if no tunnel-barrier were present, implying an effective tunnel velocity of nearly *twice* the speed of light. In this last experiment, time resolutions of less than 1

[8] In this sense, the output of our crystal is a limiting case of squeezed light.

[9] This is nothing more than very fancy jargon for a very common object--a dielectric mirror.

femtosecond were achieved.

As a final use for the photon twins, we cite the measurement of the absolute detection efficiency of single-photon detectors [Klyshko, 1980; Rarity et al., 1987; Kwiat et al., 1993a,b], which is the subject of Chap. 9. Because the photons are always produced in pairs, and have a strong angular correlation, we can effectively use one of the photons as a trigger for the detector of the conjugate photon. If we arrange our optics such that we are certain to catch (if not detect) this conjugate photon, then the efficiency of the detector is simply the ratio of the coincidence rate to the trigger detector singles rate.

Chapter 3: Single-Photon Berry's Phase

> Let us start with the extended right arm pointing straight in front of us, and the thumb pointing upwards. BACK STRAIGHT. Now move the arm upward toward the vertical, keeping the thumb parallel to itself at all times [i.e., in the plane of motion]. TUMMY IN. After the arm reaches the vertical position, move it downward along a perpendicular vertical arc, so that the arm extends to the right. HOLD IT--FEEL THE BURN. Complete the cycle by moving the arm in a horizontal arc, so that the arm is brought back to its starting position pointing forwards. AND BREATHE.
>
> Excerpted from the never-to-be-released Raymond Chiao Exercise Video, and also from [Chiao et al., 1990].

3.1 Introduction

Quantum interference possesses global geometrical features, which spring from the non-Euclidean properties of Hilbert space. One such feature is Berry's phase, which can be acquired by a quantum system whenever it evolves adiabatically and cyclically back to its initial state [Berry, 1984]. It has manifestations which range from low energy physics, e.g., in fiber optics [Chiao and Wu, 1986; Tomita and Chiao, 1986], to high energy physics, e.g., in chiral anomalies of gauge field theories [Niemi and Semenoff, 1985].

In its optical manifestations, prior to the present experiment [Kwiat and Chiao, 1991], Berry's phase had appeared only at the classical level, as a phase shift of a classical electromagnetic wave [Chiao and Wu, 1986; Tomita and Chiao, 1986; Bhandari and Samuel, 1988; Simon et al., 1988; Chyba et al., 1988; Jiao et al., 1989; Tompkin et al., 1990]. There has been a controversy as to whether one should view optical Berry's phases as originating at the quantum or the classical level. [Chiao and Tomita, 1987;

Haldane, 1987; Segert. 1987; Kugler and Shtrikman, 1988; Cai et al., 1989]. Here we point out that there can exist nonclassical manifestations of this phase when the state of the light is nonclassical. Using an entangled state of light, we have observed on the quantum level one form of Berry's phase, Pancharatnam's phase, which is generated after a cycle of polarization states [Pancharatnam, 1956]. We believe that this experiment will settle the controversy [Chiao et al., 1993]. In Sect. 3.II we briefly discuss Berry's phases in optics, including the particular phase examined in our experiment. The setup used is described in Sect. 3.III; our results and an inequality demonstrating their nonclassicality, in Sect. 3.IV. Conclusions are given in Sect. 3.V.

3.II. Berry's Phases in Optics

In its simplest form, Berry's phase is a *topological* phase acquired by a system after it is taken through some closed adiabatic cycle such that it is returned to its original state. Most of the physical variables characterizing the system obviously return to their original values; certain ones may *not* however. A good classical analog is a bicycle wheel, of which we paint one spoke red. We initially orient the wheel so that the painted spoke is down. If the wheel is rolled (without slipping) in a circle on the floor, then even though the wheel may end up at the same spot on the floor, in general the painted spoke will no longer point down. (It is easy to see that the necessary condition to return the spoke downward is that the circumference of the circle must be an integer multiple of the circumference of the wheel.) The term *anholonomy* is used to describe this phenomenon.

In studying the quantum adiabatic theorem, Berry [Berry, 1984] has shown that after a cycle C in the parameter space of the hamiltonian $H(\mathbf{R})$ (where $\mathbf{R}$ denotes some slowly varying parameters which return to their starting values), a wavefunction $|n, \mathbf{R}\rangle$ can acquire an extra phase $\gamma_n(C)$; n is the quantum number labelling the energy

eigenstate, i.e., $H(\mathbf{R})\,|\,n, \mathbf{R}\rangle = E_n(\mathbf{R})\,|\,n, \mathbf{R}\rangle$. This *topological* phase factor[1] exists in addition to the standard *dynamical* phase factor: $\exp(-i\int E_n dt/\hbar)$. Specifically, the Berry's phase is

$$\gamma_n(C) = \oint_C \mathbf{A}_{\text{eff}} \cdot d\mathbf{R} \ , \tag{3.1}$$

where $\mathbf{A}_{\text{eff}} = i\langle n;\mathbf{R}\,|\,\nabla_{\mathbf{R}}\,|\,n;\mathbf{R}\rangle$. Aharanov and Anandan have generalized Berry's result by removing the adiabatic restriction [Aharonov and Anandan, 1987]. The notion of *parameter* space is then replaced by *state* space, the projective space of rays in Hilbert space, and $\mathbf{R}$ is interpreted as coordinates in this state space. A cycle of changes in the system corresponds to a closed curve in its state space, whose geometry and topology determine Berry's phase [Chiao, 1990].

To date at least four distinct Berry's phases in optics have been discussed [Chiao et al., 1990]. The first is the "Chiao spin-redirection phase", which is acquired when the direction of light is caused to make a cycle C in $\mathbf{k}$-space [Chiao and Wu, 1986]. In particular, $\gamma(C) = -\sigma\Omega(C)$, where $\sigma = \pm 1$ is the helicity of the photon and $\Omega(C)$ is the solid angle subtended by C with respect to the center of the $\mathbf{k}$-sphere (we assume that only the *direction* of the light is changing, not the magnitude of its wave vector). The phase manifests itself as follows: if linearly polarized light (which we can always represent as a superposition of the two helicity states) is taken through a helical path (using a coiled fiber, for example), then the angle of linear polarization will in general

[1] The term "topological" is used to stress that the phase depends on the topology of the parameter space, and on the cycle of parameter changes. It is independent, for example, of the speed at which the changes are made. In our bicycle analogy, the final position of the painted spoke does not depend on the speed at which the wheel was rolled. In particular, we could stop the wheel halfway through and even roll it *backwards* before finishing the cycle.

rotate (because the two helicity components acquire opposite, but equal, phase shifts) [Tomita and Chiao, 1986]. One can understand this result very simply at a classical level as arising from *parallel transport* of the electric field vector of the light along the fiber [Ross, 1984; Haldane, 1986].

In the second type of Berry's phase in optics, the direction of the light is kept fixed, but the polarization is cycled (essentially the opposite of the first phase discussed). This phase was first investigated by Pancharatnam [Pancharatnam, 1956], and hence is sometimes known as "Pancharatnam's phase". The relevant state space is the Poincaré sphere, whose surface describes all possible pure polarization states of light. If the cycle C encloses a net area on the surface of the Poincaré sphere, then the light will acquire a geometrical phase given by $\gamma(C) = -\Omega(C)/2$, where again $\Omega(C)$ is the solid angle subtended by C with respect to the center of the sphere.[2] It is this Berry's phase that we will demonstrate below at the single-photon level.

Another type of optical Berry's phase, for which the state space has the topology of a hyperboloid of revolution, instead of a sphere, is the squeezed-states Berry's phase, discussed by Chiao and Jordan [Chiao and Jordan, 1988]. For example, if light is squeezed in one quadrature and subsequently in an orthogonal quadrature, and then compared with the same light squeezed in the opposite order (effectively completing a closed cycle), a Berry's phase will be present. Needless to say, due to severe experimental constraints, this phase has not yet been demonstrated experimentally.

Finally, Kitano and Yabuzaki have pointed out that if one cycles light through a sequence of *partial* linearly-polarized states, the light acquires a phase which is formally identical to the above squeezed-states Berry's phase [Kitano and Yabuzaki, 1989].

---

[2] Note that there is a factor of two difference associated with this phase compared to the spin-redirection phase. The underlying reason is that the polarization state space describes an effectively spin-1/2 system, whereas the momentum space reflects the bosonic nature of the photon.

Using tilted glass plates as partial linear polarizers, they have experimentally demonstrated this phase. It is an example of the extension of Berry's phase to statistical mechanics, since partially polarized light is a prototype of *mixed* states. States must then be described using the density matrix instead of the wavefunction.

The question arises as to whether these phases are classical or quantal in nature. In all of the cases mentioned above, we can view them either as phase shifts of a classical electromagnetic wave or as phase shifts of the wavefunction of the photon. A controversy has consequently developed surrounding this issue [Chiao and Wu, 1986; Tomita and Chiao, 1986; Haldane, 1987; Segert, 1987; Kugler and Shtrikman, 1988; Cai et al., 1989]. However, there do exist situations in which purely quantal (i.e., nonclassical) Berry's phases appear. In particular, when the light is in a *nonclassical* state that does not possess any correspondence principle limit in which it turns into a classical wave, then we may claim that any Berry's phase acquired is purely quantal. A Fock (or number) state is such a nonclassical state, in contrast to a coherent state. Our goal will be to demonstrate that a Berry's phase can be acquired by these manifestly *quantum* states.

3.III. Experimental Setup

Fig. 3.1 shows a schematic of the experiment. We used the correlated photons (at ~702.2 nm) produced via down-conversion in a 10-cm KDP crystal, excited by an ultraviolet (uv) argon ion laser operating at λ=351.1 nm, normally incident on the KDP input face. The idler photon (upper beam) was directed to the detector D1, which was a cooled RCA C31034A-02 photomultiplier (see Appendix E for specifications). The signal photon (lower beam) entered a Michelson interferometer, inside one arm of which were sequentially placed two zero-order quarter waveplates Q1 and Q2. The fast axis of the first waveplate Q1 was fixed at 45° to the horizontal, while the fast axis of the second

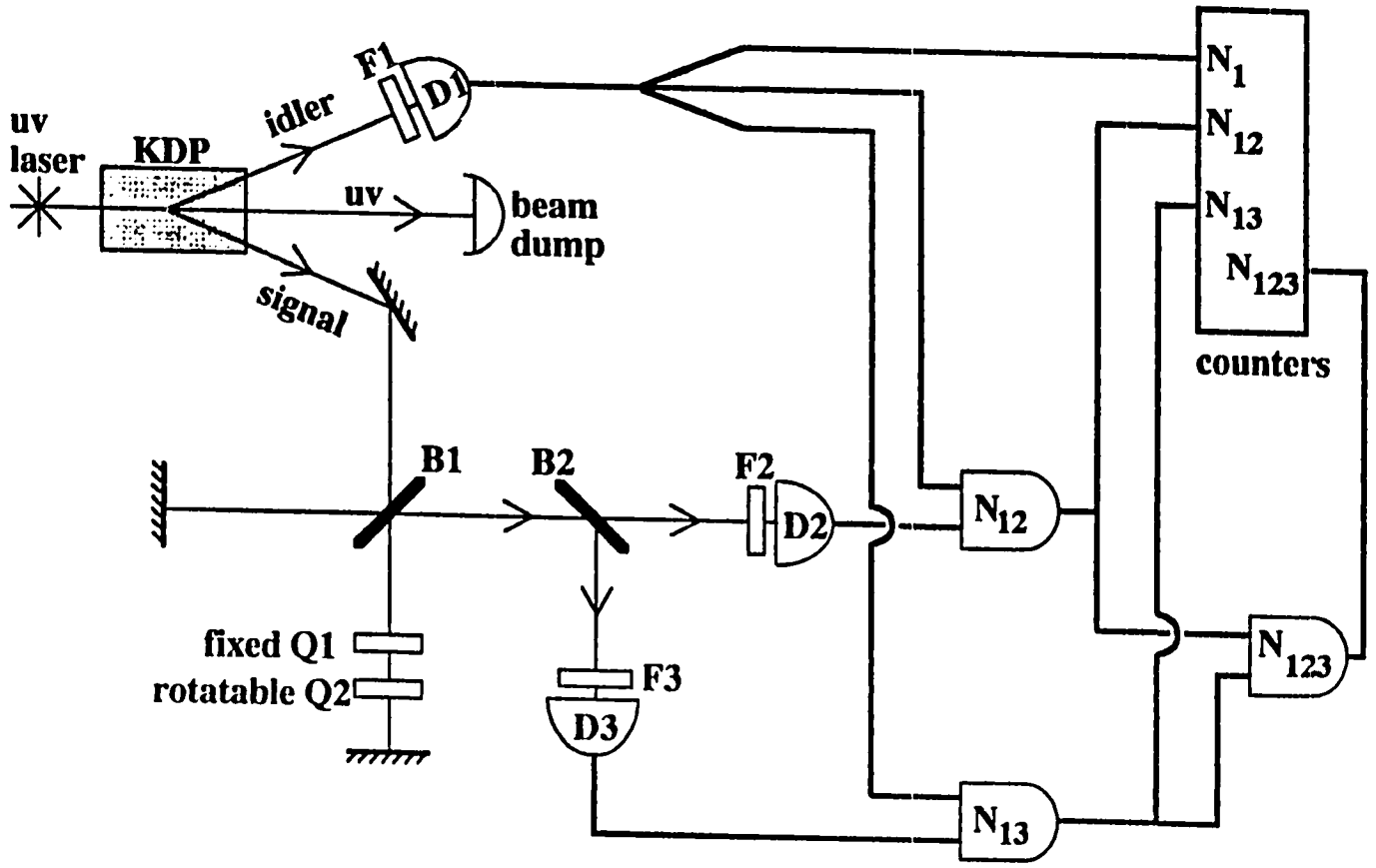


Figure. 3.1: Schematic of experiment to observe an optical Berry's phase on the quantum level. D1, D2, and D3 are photomultipliers, Q1 and Q2 quarter waveplates, and B1 and B2 beamsplitters. Logical "AND" symbols denote coincidence detectors.

waveplate Q2 was slowly rotated by a computer-controlled stepping motor. After leaving the Michelson interferometer the signal beam impinged on a second beam splitter B2. Detectors D2 and D3, essentially identical to D1, monitored the output ports of this beam splitter. (An interference filter located in front of each detector, indicated in Fig. 3.1, transmitted light centered at 702 nm. The bandwidths of these filters are an important parameter for the version of this experiment to be described in the next chapter. Filters F2 and F3 were identical and possessed a broad bandwidth of 10 nm FWHM [full-width at half-maximum]; filter F1 had either a 10-nm bandwidth or a 0.86-nm bandwidth.) Coincidences between D1 and D2 and between D1 and D3 were detected by feeding their outputs into constant fraction discriminators and coincidence detectors after appropriate delay lines. We used EG&G C102B coincidence detectors with coincidence window resolutions of 1.0 ns and 2.5 ns, respectively. Also, triple-coincidences between D1, D2 and D3 were detected by feeding the outputs of the two coincidence counters into a third coincidence detector (a Tektronix 11302 oscilloscope used in a counter mode). The various count rates were stored on a computer every second.

Our particular arrangement of quarter waveplates in the Michelson interferometer has been shown previously to generate Pancharatnam's phase at the *classical* level [Simon et al., 1988; Chyba et al., 1988]. Upon appropriate detection after the interferometer, we observed at the *quantum* level the interference fringes resulting from this phase. To calculate the phase, we use the generalized Poincaré sphere [Jiao et al., 1989; Tompkin et al., 1990] shown in Fig. 3.2, where polarization states are referred to space-fixed axes, and not to the direction of light propagation, as in the ordinary Poincaré sphere. (We do so in order to avoid extraneous discontinuities upon reflection from mirrors.) The first quarter waveplate Q1 converts horizontal linear polarization, represented by point A on the equator of the sphere, into circular polarization,

represented by B at the north pole. This transformation of polarizations is represented by a geodesic arc AB. Then Q2 converts the circular polarization back to linear polarization (C on the equator), but with an axis rotated from the horizontal by θ, the angle between the fast axes of Q2 and Q1, in real space. On the sphere, the azimuthal angle from A to C is 2θ. After reflection from the mirror, the linear polarization is unchanged with respect to space-fixed axes, and is again represented on the generalized Poincaré sphere by the same point C. After reentering Q2, this is converted to circular polarization represented by D (the south pole), completing geodesic arc BCD. Then

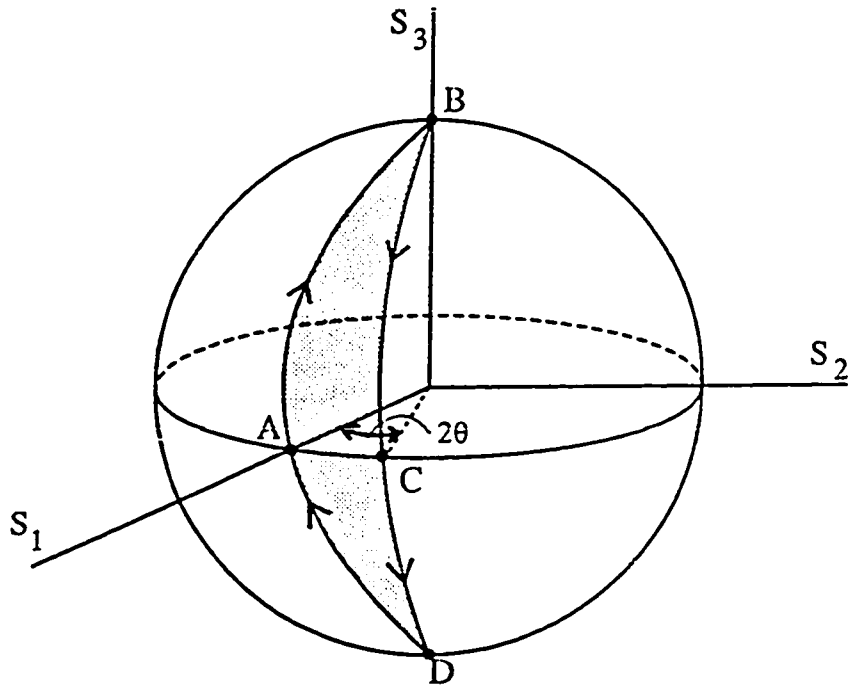


Figure. 3.2: Generalized Poincaré sphere, where S1, S2, and S3 are Stokes parameters. Circuit ABCDA represents a round trip through waveplates Q1 and Q2, where θ is the angle between their fast axes. Berry's phase, here Pancharatnam's phase, is 2θ for this circuit.

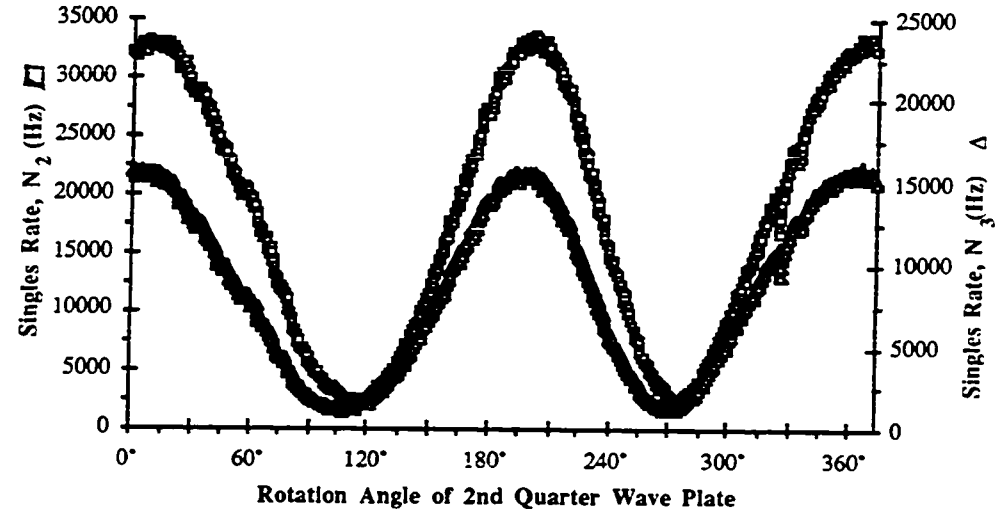
Q1 reconverts the circular polarization back to horizontal linear polarization, generating geodesic arc DA. Thus a cycle in polarization states is completed, represented by the circuit ABCDA. As stated earlier, Pancharatnam's phase is minus one-half the solid angle subtended by the circuit with respect to the center of the sphere [Simon et al., 1988; Chyba et al., 1988; Jiao et al., 1989; Tompkin et al., 1990]. For this circuit, the phase is equal to 2θ.

3.IV. Results

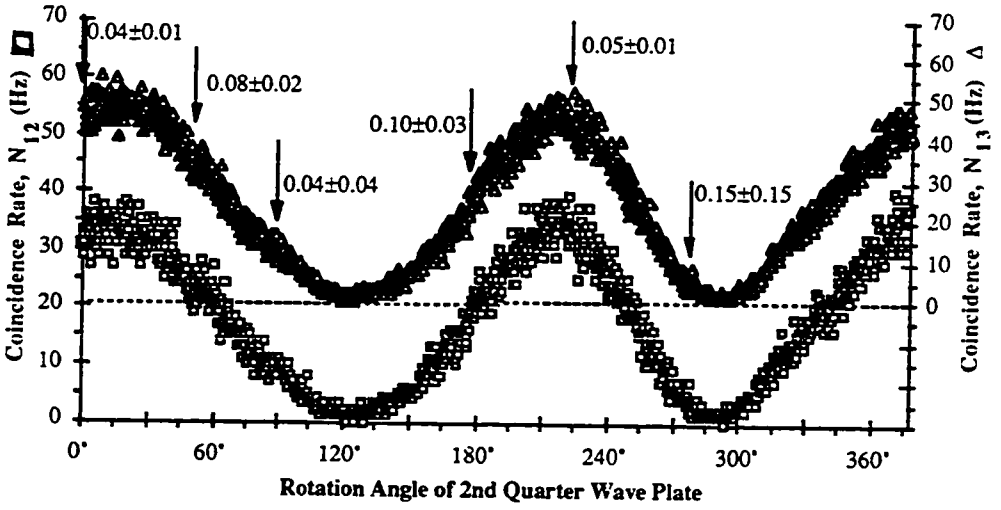
We took data both outside and inside the white-light fringe region where the usual interference in singles detection occurs. We report here only on data taken inside this region, where the optical path length difference was at a fixed value much less than the coherence length of the signal photons, determined by the filters F2 and F3. Hence, fringes are expected both in the singles rates at detectors D2 and D3 and in the coincidence rates.

In Fig. 3.3, we show data which confirm these predictions. Singles rates are plotted in Fig 3.3a, while coincidences are shown in Fig. 3.3b. The coincidences correspond to a 10-nm filter before detector D1. (The calculated coherence length of the *signal* photon wavepacket (50 μm) was much greater than the optical path length difference at which the Michelson was set (~0 μm). The slightly nonsinusoidal component in Fig. 3.3 can be explained by a slight wedge in Q2, in conjunction with the fact the signal beam was incident on Q2 off center.

The function of the second beam splitter B2 was to verify that the signal beam was composed of photons in an n=1 Fock state. In such a state, the photon, due to its indivisibility, will be either transmitted or reflected at the beam splitter, but not both. Thus coincidences between D2 and D3 should never occur, except for rare accidental occurrences of two pairs of conjugate photons within the coincidence window.



a.)



b.)

Figure 3.3 Interference due to a Berry's phase acquired in one arm of a Michelson interferometer, in the white-light fringe regime. a) Fringes in the signal singles rates. b) Fringes in the signal-idler coincidence rates, after accidentals have been subtracted; vertical arrows indicate where the anticorrelation parameter a was measured (see text). For classical fields, $a \geq 0.70 \pm 0.07$.

However, if the signal beam were a classical wave, then one would expect an equal division of the wave amplitude at the 50% beam splitter, and hence frequent occurrences of coincidences. An inequality, which was strongly violated in our experiment, places a lower bound on this coincidence rate for classical light (see below). This verifies the essentially n=1 Fock state nature of the light, and confirms an earlier result of Hong and Mandel [Hong and Mandel, 1986].

The vertical arrows in Fig. 3.3b indicate the points at which triple-coincidences were measured. Let us define the anticorrelation parameter [Grangier et al., 1986]

$$a \equiv N_{123}\, N_1 / N_{12}\, N_{13} \ , \tag{3.2}$$

where N_{123} is the rate of triple-coincidences between detectors D1, D2 and D3, N_{12} is the rate of double coincidences between D1 and D2, N_{13} is the rate of double coincidences between D1 and D3, and N_1 is the rate of singles detections by D1 alone. The inequality $a \geq 1$ has been shown to hold for any classical wave theory[3]. A rigorous proof is given in Appendix A. The equality a=1 holds for coherent states $|\alpha\rangle$, independent of their amplitude α. Since in our experiment the amplitude fluctuations in the double coincidence pulses led to a triple-coincidence detection efficiency ε less than unity, we should reduce the expected value of a accordingly[4]. The modified classical

[3] Similar higher-order criteria for nonclassical effects in photon statistics have been discussed by Lee [Lee, 1990].

[4] Specifically, what we actually measured was $N_{12\text{-}13}$, the number of coincident double-coincidences. We fed the output pulses of the two coincidence counters into a Tektronix 11302 oscilloscope, and if the sum of the two signals was greater than a preset threshold then a count was registered. Ideally, if the pulse heights were constant, then one could use a threshold just slightly greater than this constant, and be guaranteed of catching all triple-coincidences without ever falsely counting a double-coincidence. Because this was not the case, however, we were forced to use a higher threshold (roughly equal to 1.5 times the average pulse height from either counter, which was

inequality is $a \geq \varepsilon$. We calibrated our triple-coincidence counting system by replacing the two-photon light source with an attenuated light bulb, and measured $\varepsilon = 0.70 \pm 0.07$. During the data run of Fig. 3.3b, we measured values of a shown at the vertical arrows. The average value of a is 0.07 ± 0.08. This violates by seven standard deviations the predictions based on any classical wave theory[5], but is in complete agreement with the value $a_{QM} = 0.06$ predicted for our quantum-mechanical light, for which the triple-coincidences are due entirely to multiple simultaneous down-converted pairs[6], aside from background, which is negligible in this context. It is therefore incorrect to interpret these results in terms of a stochastic ensemble of classical waves, in a semiclassical theory of photoelectric detection [Clauser, 1974]. Classical waves with conjugate, but random, frequencies could conceivably yield the observed interference pattern, but they would also yield many more triple-coincidences than were observed.

---

greater than the maximum pulse height of either counter alone), to avoid falsely counting larger-than-average double-coincidence pulses.

[5] The uncertainties in a are rather large due to the fact that so few triple coincidences were observed. For example, in one run (corresponding to the third arrow from the left in Fig. 3.3b) only 1 triple-coincidence count was registered in 7,240 sec (18 ± 2 were expected for a classical [i.e., coherent state] source). A violation by about 9 standard deviations was observed in a slightly different version of this experiment, described in Chap. 4, where fringes were observed (in coincidence rates) outside the white light fringe region.

[6] To calculate a we calculate the expected rate of triple coincidences (given the measured singles rates and double-coincidence rates) using the formula: $N_{123} = w_{(12)3}\left[(N_{12} - A_{12})N_3 + (N_{13} - A_{13})N_2 w_{12} \wedge w_{13} + N_1 N_2 N_3 w_{12}\right]$, where w_{1j} is the gate window width for coincidence counts between detectors D1 and Dj, and $w_{(12)3}$ is the effective gate window for coincidence counts between detector D3 and the output of the coincidence counter measuring the rate N_{12}. This formula is valid so long as the actual detection efficiencies (here assumed to all be equal, for simplicity) are low, as they were in this experiment (otherwise the third term, which represents triple-coincidences arising from three separate down-converted pairs, needs to be reduced by the factor $(1 - \eta)\wedge 3$ [where η is the detector efficiency] to account for the fact that we need to *not detect* one member of each pair.)

3.V. Discussion and Conclusion

In fairness to the opposing side, we should present at least a glimpse of their reasoning [Haldane, 1987; Segert, 1987; Kugler and Shtrikman, 1988; Cai et al., 1989]. They would claim that although the states we used are undeniably quantum-mechanical (i.e., nonclassical), the calculation regarding the Berry's phase is done on the *fields* themselves, not the state. Although in quantum mechanics these fields are described by operators, the same Berry's phase would result in a classical description where they are simply c-numbers. Therefore, the phase is *classical*, even though the state is quantum-mechanical so that the number of triple-coincidences observed could not be explained by any classical-field approach.

Our answer to such arguments is two-fold. First, one must evaluate an experiment in terms of the *entire* system, not any of its parts. (We shall see that this philosophy is absolutely crucial if one is to understand another set of experiments we have performed, described in Chap. 5, demonstrating the phenomenon of quantum erasure.) It is misleading to say that the phase acquired in one part of an experiment is classical, even though the experiment as a whole does not possess any classical explanation. As a further example of this, we can consider the situation of a *two*-particle wavefunction, defined on the *configuration* space of the particles, which may be thought of as the relevant state space for the system. If this state space possesses a nontrivial topology (e.g., the state space will be a *torus* if the configuration space coordinates are restricted to two angular variables), and if these particles undergo a sequence of changes such that they return to their original state, then the system may acquire a Berry's phase. In particular, consider two photons prepared in an entangled state of energy, as discussed in the previous chapter. We will see in Chap. 8 that if each photon is directed to an unbalanced Mach-Zehnder interferometer, with a phase-shifter (possibly a Berry's phase-shifter, as was used in the present experiment) in each, then interference may be

discerned in the rate of coincidences between detectors looking at output ports of the two interferometers, even though no fringes can be seen in any of the singles rates. That is, the (Berry's) phase shift manifests itself in the *entire* two-particle wavefunction. These results are not only nonclassical, they are nonlocal as well.

Our second answer to doubters of quantum Berry's phases is that there *is* predicted to be one situation where the actual value of the phase depends on the state of light. In the squeezed-states Berry's phase, if one uses a Fock state $|n\rangle$, then the resulting Berry's phase acquired by the light is proportional to n + 1/2 [Chiao, 1990; Chiao and Jordan, 1988]. However, when a coherent state is taken through the same cycle, there is no such proportionality to the intensity of the light. Clearly, if this result is correct then one must indeed consider the quantum nature of the state when calculating a Berry's phase.

We have observed Berry's phase for photons in essentially n=1 Fock states, in a way which excludes with high probability any possible classical explanation. We conclude that this and other optical phases originate fundamentally at the quantum level, but under special circumstances, can survive the correspondence principle limit onto the classical level [Chiao and Tomita, 1987]. In other words, we feel that even ordinary two-slit interference is best viewed as a quantum phenomenon, which is still apparent at the classical level. Our main logic behind judging the quantum interpretation as "more fundamental" than the classical interpretation is that there exist situations, like that described above, where a quantum description yields the correct predictions, even though no classical description can. The reverse is never true.

Chapter 4: "Collapse of the Wavefunction"

When I hear of Schödinger's cat, I reach for my gun.
-Stephen W. Hawking

4.I. Introduction

There are a number of physicists who react nearly as violently as indicated above when mention is made of the "collapse" of the wavefunction. The two main problems are what causes the collapse (i.e., what really *is* a measurement, after all) and when does this occur. Herein we shall make no attempt to solve the long-standing controversy surrounding the collapse issue (see, for instance, [Wheeler and Zurek, 1983; Zurek, 1986; Peres, 1986]). Rather, we present experimental results which we feel are most readily understood using the collapse-of-the-wavefunction approach. For all its difficulties, this picture can be very useful in gaining predictive insight into otherwise subtle effects.

In this experiment we make use of the fact that the light from our down-conversion crystal is prepared in an entangled state consisting of a pair of photons whose energies (frequencies), although individually broad in spectrum, sum up to a sharp quantity because they were produced from a single (nearly monochromatic) pump photon whose energy E_p was sharp. According to the standard Copenhagen interpretation, the meaning of this entangled state is that when a measurement of the energy of one photon results in a sharp value E, there is a sudden collapse of the wavefunction such that instantly at a distance, the other photon, no matter how remote, also possesses a sharp value of energy E_p - E. We can observe this collapse phenomenon by sending one of the photons, say the signal photon, into an unbalanced Michelson interferometer,

operating outside the white-light regime, so that no interference is observable in the signal singles rate. Of course, if we were to put a narrow-band filter in front of the detector, then fringes would appear as long as the new coherence length were greater than the path-length difference in the interferometer. The remarkable fact is that we can instead put the narrow filter in front of the *idler* photon, which may be very distant; looking in coincidence will recover the fringes[1]. In Sect. 4.II we present both a simplied and a more rigorous analysis) of the experiment. Our setup and results are given in Sect. 4.III, with a discussion of other interpretations of the results and conclusions given in Sect. 4. IV.

4.II. Theoretical Analysis

The experiment described here is an extension of the previous one (presented in the previous chapter; see Fig. 3.1), in which the signal photons are directed into a Michelson interferometer, while the idlers are sent directly through a filter to detector D1. The basic difference is that here we operate *outside* the white-light fringe regime, so that no interference fringes are observed in any of the singles rates. First we present a simplified quantum analysis. Afterwards we will present a more comprehensive analysis based on Glauber's correlation functions. The energy-entangled state of the light after the Michelson interferometer is given by

$$|\psi\rangle_{out} = \int d\omega_s\, A(\omega_s)\, |\omega_s\rangle_s\, |\omega_p - \omega_s\rangle_i \frac{1}{\sqrt{2}}\left\{1 + e^{i\phi(\omega_s)}\right\}, \qquad (4.1)$$

where $\phi(\omega_s) = \Delta L\, \omega_s/c + \phi_{Berry}$ is the phase shift arising from the optical path-length difference ΔL of the Michelson for the signal photon with frequency ω_s, plus the

[1] Obviously, placing a filter in front of the remote idler detector cannot alter the signal *singles* rate, as this would permit sending superluminal messages.

Berry's phase contribution for this photon. $A(\omega_s)$ is the probability amplitude for generating a pair of signal and idler photons with frequencies ω_s and $\omega_p - \omega_s$, respectively.

The coincidence rate N_{12} between detectors D1 and D2 is proportional to the probability of finding at the same time t one idler photon at detector D1 placed at $\mathbf{r}_1$, and one signal photon at detector D2 placed at $\mathbf{r}_2$; similarly for N_{13}. When a narrow-band filter F1 centered at frequency Ω_i is placed in front of the idler detector D1, N_{12} and N_{13} become proportional to

$$|\psi'_{out}(\mathbf{r}_1, \mathbf{r}_2, t)|^2 = |\langle \mathbf{r}_1, \mathbf{r}_2, t | \, |\psi\rangle'_{out}|^2 \propto 1 + \cos\phi, \qquad (4.2)$$

where the prime denotes the output state after a von Neumann projection onto the eigenstate associated with the sharp frequency Ω_i upon measurement. Therefore, the phase ϕ is determined at the sharp frequency $\omega_p - \Omega_i$. In practice, the frequency width depends on the bandwidth of the filter F1 in front of D1[2], so that the visibility of the fringes seen in coincidences should depend on the width of this *remote* filter[3]. This fringe visibility can be high, provided that the optical path-length difference of the Michelson does not exceed the coherence length of the *collapsed* signal photon wavepacket, determined by F1. If a sufficiently broadband remote filter F1 is used instead, such that the optical path-length difference is much greater than the coherence length of the collapsed wavepacket, then the coincidence fringes should disappear[4].

---

[2] We assume here that the irises are sufficiently large that it is the filters which limit the bandwidth.

[3] One can also view the Michelson as a type of variable filter, with a comb-shaped spectral transmission function. This aspect is discussed in greater detail below and in Chap. 8.

[4] Note: For 1-mm irises about 1 m from the source, the maximum bandwidth of the down-converted photons is about 40 nm, due to the phase-matching constraints. Naturally, one cannot *increase* this bandwidth by using a broader bandwidth filter.

In other words, fringes will only be detected if the collapsed wavepacket overlaps with itself after reflection from the mirrors of the Michelson interferometer.

A more rigorous theoretical description of the experiment can be carried out within the Glauber correlation function formalism [Glauber, 1963]. The probability of joint detection of a signal-idler pair within the detector resolution window ΔT, after a total time $\mathcal{T}$, is then given by

$$P = \int_{-\mathcal{T}/2}^{\mathcal{T}/2} dt_s \int_{t_s - \frac{\Delta T}{2}}^{t_s + \frac{\Delta T}{2}} dt_i \, G^{(2)}(t_s, t_i; t_i, t_s) \, , \tag{4.3}$$

where the second-order correlation function $G^{(2)}$ is defined as in Eq. (2.4):

$$G^{(2)}(t_s, t_i; t_i, t_s) = \langle \psi \left| \, \hat{E}_s^{(-)}(t_s) \, \hat{E}_i^{(-)}(t_i) \, \hat{E}_i^{(+)}(t_i) \, \hat{E}_s^{(+)}(t_s) \, \right| \psi \rangle. \tag{4.4}$$

The negative- and positive-frequency parts of the electric field for the idler mode may simply be Fourier-expanded as in Eqs. 2.5c and 2.5d, assuming, as in Fig.3.1, that the idler photon is directed to detector D1. The effects of filter F1 are included in the factor η_1. Similarly, the signal mode field operators may be expanded, but these require an additional factor to account for the interferometer:

$$\hat{E}_s^{(+)}(t_s) = \int d\omega_s \, \eta_2^*(\omega_s) \, \hat{a}_s(\omega_s) \, e^{-i\omega_s t_s} \frac{1}{\sqrt{2}} \left\{ 1 - e^{i\omega_s \tau} \, e^{i\phi_B} \right\} \tag{4.5a}$$

$$\hat{E}_s^{(-)}(t_s) = \left(\hat{E}_s^{(+)}(t_s) \right)^\dagger, \tag{4.5b}$$

where $\tau = \Delta L/c$ is the optical delay time between the arms of the interferometer, and ϕ_B

is the geometrical/Berry's phase, whose value depends on the rotation angle of the second quarter waveplate (see Sect. 3.III).

In practice, in (4.3) the duration time $\mathcal{T}$ of any data point is essentially infinite (with respect to all relevant time-scales in the problem). In addition, for our experiment ΔT ($\approx$ 1 ns) was much greater than τ ($\approx$730 fs, for ΔL = 220 μm) and the reciprocal bandwidths ($1/\Delta\omega_i$ and $1/\Delta\omega_s$) of the filters F1 and F2. Hence, we are justified in setting the limits of integration in (4.3) to infinity. Substituting the energy-entangled state of the down-converted light [Eq. (2.3)], we have

$$P \approx \int_{-\infty}^{\infty} dt_s \int_{-\infty}^{\infty} dt_i \int d\omega \int d\omega' \, A(\omega)A^*(\omega') \int d\omega_s \int d\omega'_s \int d\omega_i \int d\omega'_i$$

$$\times \eta_1(\omega_i)\, \eta_1^*(\omega'_i)\, \eta_2(\omega_s)\, \eta_2^*(\omega'_s)\, e^{i(\omega_i-\omega'_i)t_i}\; e^{i(\omega_s-\omega'_s)t_s}$$

$$\times \frac{1}{2}\left(1 - e^{-i\omega_s\tau}e^{-i\phi_B}\right)\left(1 - e^{i\omega'_s\tau}e^{i\phi_B}\right)$$

$$\times {}_{s,i}\langle \omega',\ \omega_p-\omega' \,|\, \hat{a}_s^\dagger(\omega_s)\, \hat{a}_i^\dagger(\omega_i)\, \hat{a}_i(\omega'_i)\, \hat{a}_s(\omega'_s) \,|\, \omega,\ \omega_p-\omega\rangle_{s,i}\,, \quad (4.6)$$

where, for simplicity, we have assumed that the pump photon is monochromatic[5]. If we assume that the probability amplitude is essentially constant ($A(\omega) \approx A_0$) over the filter bandwidths $\Delta\omega_i$ and $\Delta\omega_s$, and that $\eta_2(\omega_s) \approx \eta_{20}$ over the bandwidth $\Delta\omega_s >> 1/\tau$ (i.e., a spectrally broad, flat bandpass filter F2 in front of detector D2), then (4.6) simplifies considerably:

---

[5] This approximation is very good in our system. The coherence length of our uv pump (not yet operating in single-line mode) was 3 cm, still much larger than the path-length difference (220 μm) in the interferometer.

$$P \approx |A_0|^2 |\eta_{20}|^2 \int d\omega_i |\eta_1(\omega_i)|^2 \{1 - \cos[(\omega_p - \omega_i)\tau + \phi_B]\}. \quad (4.7)$$

We now examine the behavior of this detection probability in two limiting cases of filter F1:

1. If $|\eta_1(\omega_i)|^2 \approx |\eta_{1d}|^2 \delta(\omega_i - \Omega_i)$ (i.e., a very narrow filter in front of detector D1), then

$$P \approx |A_0|^2 |\eta_{20}|^2 |\eta_{1d}|^2 \{1 - \cos[(\omega_p - \Omega_i)\tau + \phi_B]\}. \quad (4.8)$$

It is clear from (4.7) that in order to observe these fringes, it suffices to have $\Delta\omega_i << 1/\tau$. {If we allowed the pump to have a non-zero bandwidth in our calculation, we would have found that the coherence length of the pump also needs to be greater than the interferometer path-length difference to see interference. As discussed in Chap. 2, the frequency-correlation of the down-converted photons is only as sharp as the pump linewidth.} Note that since the filter F2 is relatively broadband (i.e. $\Delta\omega_s >> 1/\tau$), there are no fringes visible in the singles rate of detector D2, even though fringes in coincidence may be present.

2. If $|\eta_1(\omega_i)|^2 \approx |\eta_{1d}|^2 e^{-(\omega_i - \Omega_i)^2/\Delta\omega_i^2}$, where $\Delta\omega_i >> 1/\tau$ (we have previously stipulated the experimental condition $\Delta\omega_s >> 1/\tau$), then

$$P \approx |A_0|^2 |\eta_{20}|^2 |\eta_{1d}|^2, \quad (4.9)$$

a constant, with no fringes.

4.III Experiment and Results

The basic setup is the same as that used in the demonstration of Berry's phase at

the single photon level (presented in the previous chapter; see Fig. 3.1). The only difference is that here we operate *outside* the white-light fringe regime, so that no interference fringes are observed in any of the singles rates (see Fig. 4.1). In particular, the coherence length of the signal photons, determined by the 10-nm FWHM filters F2 and F3, is only 50 μm[6], whereas the optical path-length difference was fixed at 220 μm.

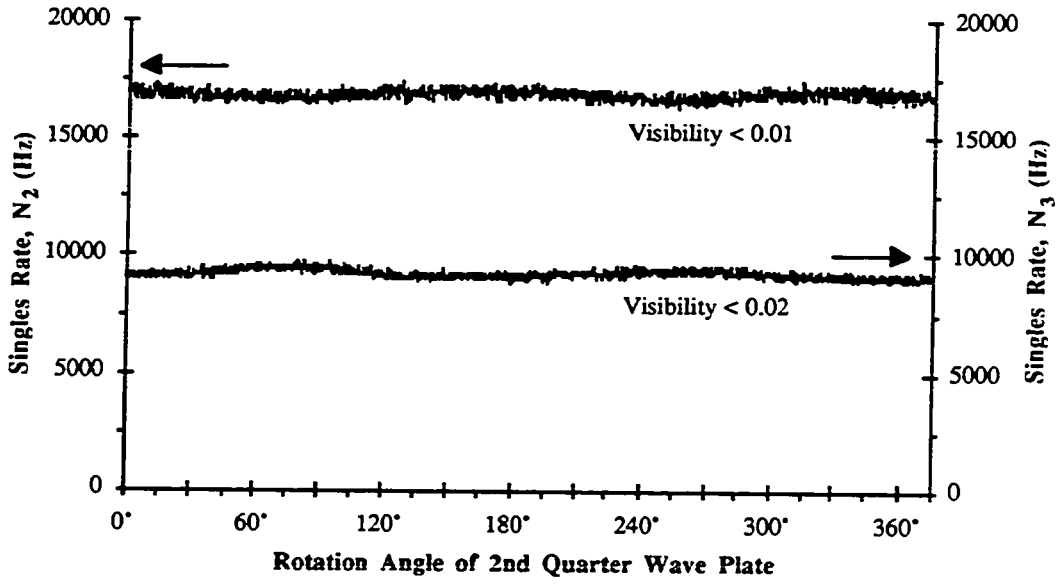


Figure 4.1 Far outside the white-light fringe region of the Michelson interferometer no fringes are observable in the singles rates. Variations in the rates are most likely due to slight beam deviations as the quarter waveplate Q2 is rotated, perhaps arising from a slight wedge in Q2, in conjuction with the fact the signal beam was incident on Q2 off center.

[6] The visibility of the signal single-photon interference close to the white light fringe region was measured to have a sinc-like variation with arm length difference (see Fig. 8.5a). We determined the coherence length from the position of the first null of this pattern.

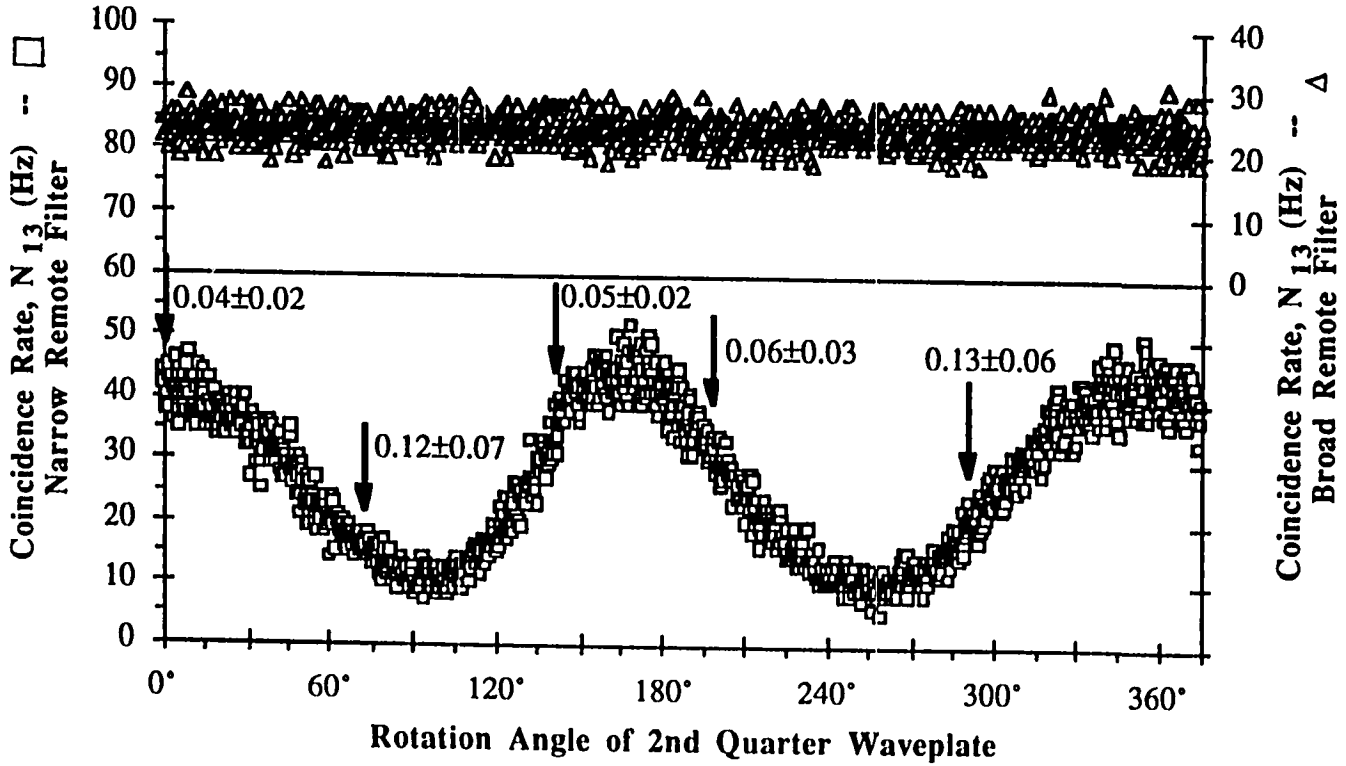


Figure. 4.2 Interference fringes outside the white light regime for an unbalanced Michelson with a slowly varying Berry's phase. Fringes are observed in coincidences between D3 and D1 with a narrow (0.86-nm FWHM) remote filter F1 (lower trace, squares); with a broad F1 (10-nm FWHM), these fringes disappear (upper trace, triangles). Accidental rates of ~1.2 s^{-1} and ~5.2 s^{-1} have been subtracted from the upper and lower traces, respectively. Vertical arrows indicate where the anti-correlation parameter a was measured (see Chap. 3).

In Fig. 4.2, we show data commensurate with the collapse picture. In the lower trace we display the coincidence count rate between detectors D1 and D3, as a function of the angle θ between the fast axes of waveplates Q1 and Q2, when the remote filter F1 was narrow (0.86-nm FWHM bandwidth). The calculated coherence length of the collapsed signal photon wavepacket (570 μm) was then greater than the Michelson optical path-length difference (220 μm), and the visibility of the coincidence fringes was thus quite high, viz., 60%±5%. For comparison, in the upper trace we display the coincidence count rate versus θ when a broad remote filter F1 (10-nm FWHM bandwidth) was substituted for the narrow one. The coherence length of the collapsed signal photon wavepacket was thus only 50 μm, and the coincidence fringes have indeed disappeared, as predicted.

As in the previous version of this experiment (in the white-light fringe regime), we have measured the anticorrelation parameter a [cf. Eq. (3.2)] at several data points (the values are indicated by the vertical arrows in Fig. 4.2). The average value of a we measured is only 0.08±0.04, which violates by nearly 9 standard deviations the predictions based on any classical wave theory ($a_{cl} \geq 0.7 \pm 0.07$), but is in perfect agreement with the quantum-mechanical prediction of 0.08. Classical waves with conjugate, but random, frequencies could conceivably yield the observed interference pattern (see below), but they would also yield many more triple coincidences than were observed.

4.IV. Interpretation and Conclusion

There are several ways to approach the results:

1) Just blindly calculate using the quantum mechanical rules, avoid thinking about the interpretation, and thus avoid any possible controversies. This is basically the precept of the "strict" Copenhagen interpretation (as distinguished by Stapp from the

"informal" interpretation [Stapp, 1986]), which states that the quantum formalism "merely offers rules of calculation for the deduction of expectations pertaining to phenomena obtained under well-defined experimental conditions" [Bohr, 1963] -- one is not supposed to ask "What actually happened?" We do not particularly care for this approach, even though it is espoused by a large number of physicists. In our view it has ever been the goal of the scientist to comprehend the *underlying nature* of the phenomena he observed. While there are certainly those who will claim that any questions regarding individual events are outside the scope of quantum theory, and that they are just not "fair questions", we find this attitude unhelpful. It is, in our opinion, natural and important to try to understand what is actually happening in a given event.

2) A second option is to interpret the filtering process as essentially selecting out photons with a particular color from a statistical ensemble describing the total bandwidth of the source. That is, one assumes that each photon emitted from the source has a well-defined, sharp frequency (although the frequencies of the photons in each down-converted pair are still tightly correlated), so that the coherence length of any given photon is very long. It is only in accepting photons of different colors, essentially summing their respective interference patterns, that the visibility is seen to decrease. In using a narrow filter in front of D1, and looking at the coincidence rates, we are post-selecting out only photons within a narrow range of colors. Fringes are then naturally observed in coincidences, without needing to resort to a nonlocal collapse.

This statistical ensemble approach is an example of a local hidden variable (LHV) model[7], about which we shall have much more to say in Chaps. 7 and 8. However, in light of the experimental violations of Bell's inequalities [Kwiat et al., 1993; Brendel et al., 1992], which refute all such LHV models[8], this approach is incorrect. Due to

[7] Note that if we had not measured a, then the results would be equally-well explicable by a simple correlated classical field model. Our low value of a excludes such a description.

the entangled nature of the state (2.3), energy (frequency) should not be viewed as a local, definite property carried by the photon *before it is actually measured.* We should point out, though, that the results of the present experiment alone are *not* sufficient to prove the non-"local-realism" implied by entangled states, i.e, one cannot violate a Bell's inequality, contrary to what some claim [Davis, 1989; Klyshko, 1989][9]. One must employ extra means to transform the state of light into something more akin to the singlet state of the Bohr version of the EPR paradox. This has been done in other experiments (see Chap. 8, and [Kwiat et al., 1993; Brendel et al., 1992]).

3) The third interpretation of our results invokes the "collapse of the wavefunction" -- the energy of the photons is not well-specified (i.e., to better than the broad bandwidth constraints of the phase-matching conditions) until a precise measurement is made on one of the photons of a pair. At that moment, the energy of the *conjugate* photon, no matter how remote, becomes equally well-defined: the state described by (2.3) collapses to a state with a much smaller spread in frequency[10]. The collapse is instantaneous (although our results in no way prove this) because the entangled state is nonlocal.

4) There are two main "alternatives" to a nonlocal collapse, although they are not widely accepted at present. One is the Everett-Wheeler many-worlds interpretation

---

[8] Several physically-reasonable auxilliary assumptions must also be accepted. These are discussed in Chap. 7.

[9] The simple reason is that one can explain all the results using the local hidden variable model described above. Underlying this is the fact that one can write a positive-definite Wigner function [Wigner, 1932] to describe the simultaneous time- and energy-correlations of the photons, and Bell has shown that under these conditions no *direct* violation of a Bell's inequality is possible [Bell, 1986]; i.e., the non-positive-definiteness of the Wigner distribution is a *necessary* condition for a locality paradox. Wódkiewicz has also discussed these issues [Wódkiewicz, 1988].

[10] Curiously, the effect of this collapse is actually to *increase* the coherence length of either of the photons, so that if one thinks in terms of wavepacket size, the term "collapse" may seem contradictory.

[Everett, 1957; DeWitt, 1973], according to which interactions that force the wavefunction to split into a multitude of components force the whole universe to split. A single specific measurement outcome is realized in each of the resulting "branches", without resorting to a non-unitary collapse. The other alternative is the Bohm-deBroglie pilot-wave picture [Bohm, 1952; Bohm and Hiley, 1984; Bohm et al., 1987; and deBroglie, 1960]. Here there is presumed to exist both a wavefunction and a particle, the former "guiding" the latter. This picture is a type of hidden variable theory (cf. Chap. 7), in which the specific, detailed initial conditions are the hidden variables. However, the model is explicitly *nonlocal*--changing a setting (e.g., polarizer, filter, etc.) in one location can instantly affect the wavefunction at distant locations. In this sense, a "collapse" of some sort still occurs. At present, all predictions of these alternate viewpoints are in agreement with those of the standard approach.

Throughout, we have spoken as though the measurement of the idler photon with filter F1 has "initiated" the collapse. This method led us in Sect. 4.II to quickly explain the results described above. However, we could have equally-well treated the interferometer as a sort of frequency filter (with a comb-like transmission function), which "measured" the frequency of the signal photons, collapsing the frequency spread of the remote idlers. Both approaches yield the same final result.

It is the tremendous strength of our quantum mechanical formalism that either approach gives the same result. For this reason, we think it unlikely that there will ever be a situation where the "collapse" of the wavefunction argument will be inconsistent with the formal mathematical calculation. Of course, as experimentalists we would be almost happy if this were not the case, for then we could perhaps distinguish between the at-present indistinguishable viewpoints, e.g., collapse, many-worlds, etc. In some ways, the situation at present is similar to that shortly after the introduction of the EPR "paradox" (discussed at length in Chap. 7): There were large philosophical differences

in the quantum-mechanical and the hidden-variable approach to explaining correlations of separated particles, but all real calculations seemed to yield identical predictions. This philosophical impasse extended until Bell [Bell, 1964] proved that there were situations in which the different viewpoints gave different experimental predictions (see Chap. 7), opening the door for *experimental refutation* of one of the theories[11]. A similar opportunity to actually rule out the "collapse" picture would be welcome.

In conclusion, using photons in essentially n=1 Fock states, we have observed nonclassical interference, the presence or absence of which depends on the value of a remote filter (but in any event one must correlate the results of the separated detectors to observe the interference, so it is not possible to send superluminal signals). These results can be explained in terms of the nonlocal collapse of the wavefunction.

[11] Unfortunately, as discussed in Chap. 7, a completely unambigous test has not yet been made, although in Chap. 10 we propose a setup for such a "loophole-free" test.

Chapter 5: Quantum Erasers -- Theory and Practice

But *why* must I treat the measuring device classically?
What will happen to me if I don't?
-Eugene Wigner

5.I. Introduction

The admonition (to treat measuring devices classically) that provoked the above response was perhaps intended to prevent the troublesome state of affairs in which a macroscopic apparatus (as well as the experimenters observing it, their friends, cats, etc.) is described by a quantum superposition. However, one must exercise great care in such matters, for it is sometimes *essential* that one treat a measuring apparatus quantum-mechanically, if correct predictions are to be made. The quantum eraser experiments described in this chaper (and the next) are just such situations.

Interference is arguably the most fundamental effect in quantum mechanics, the Young's two-slit experiment being the canonical manifestation of complementarity. As discussed by Bohr in his classic dialogue with Einstein, if one tries to determine which slit the particle traversed (e.g., by measuring the initial and final momentum of the recoiling slit mechanism), the uncertainty principle requires a sufficient uncertainty in the initial transverse position of the slits that the interference is lost. No *wave*-like behavior will be observed, although the "which-path" information is indicative of *particle*-like behavior. In a variant of this example, Feynman proposed to "watch" the passage of an electron through a particular slit by placing a light source immediately after the slits and scattering photons off the electron [Feynman et al., 1965]. Even if one does not observe the scattered light, the electron interference will be washed out (whenever the

light is scattered sufficiently to carry unambiguous information about which slit was traversed).

This loss of interference is commonly interpreted as arising from uncontrollable, irreversible interactions of the interfering system with a classical measuring apparatus (MA), which often takes the form of the environment [Peres, 1980]. The resulting measurement provokes an irreversible "reduction of the state vector," irrevocably introducing an uncertainty in the phase information carried by the particle, thereby eliminating the possibility of interference. In Feynman's example, scattering light off the electron changes its center-of-mass wave function in an uncontrollable manner, removing the phase coherence between the two paths. While it is true that many measurements are of this sort, there are situations where the measurement process need not be so uncontrollable. In these cases it is more helpful to view the loss of coherence as due to an entanglement of the system wave function with that of the MA[1], *which is itself a quantum system*; this is identical to the first step in von Neumann's measurement theory [von Neumann, 1983], but lacks the step in which the off-diagonal elements of the expanded density matrix are postulated to vanish. Through the entanglement of a quantum system to a MA, previously interfering paths can become *distinguishable* (assuming that the final MA states are orthogonal), such that no interference is observed (cf. the discussion of Feynman's rules in Chapter 1). This is true even though one may not actually make subsequent measurements on the MA to determine which path actually occurred, i.e., even if one does not look at the result of the MA. Whenever *welcher Weg* ("which way") information is available in principle about which possible path occurred, the paths are distinguishable, and no interference is

[1] Stern *et al.* have shown a general equivalence between this interpretation and one in which the environment "scrambles" the phase of the interfering system due to an uncertain interaction potential [Stern, 1990]. We find the second picture less helpful in the experiments under consideration, as the relation of this potential to the experimental system is often rather subtle. Thus we restrict ourselves here to the first interpretation.

possible. Interference may be regained, however, if one somehow manages to "erase" the distinguishing information, by correlating the results of measurements on the interfering particle with the results of particular measurements on the MA. This is the central concept of the quantum eraser [Hillery and Scully, 1983; Scully et al., 1991].

Scully *et al.*[Scully et al., 1978] discussed a simple experiment to see this effect, in which an atom is sent through a Stern-Gerlach interferometer[2]. Upon measurement of the atom's passage through one arm of the interferometer, the interference is made to vanish. This is true even if the measuring apparatus does not change the spin state of the atom, or affect the center-of-mass part of its wave function. Unfortunately, detailed calculations of the proposed experiment made clear that it would not be feasible in practice, due to the experimental difficulty of controlling the fields to the degree necessary to observe interference, even in the absence of a *welcher Weg* detector [Englert et al., 1988; Schwinger et al., 1988]. Another proposal using a two-slit type interference of neutrons, with micromaser cavities as *welcher Weg* detectors was also deemed very difficult [Scully and Walther, 1989]. To date, the most promising of the proposed experiments on particles involve the interference manifested in the quantum beat phenomenon [Scully and Walther, 1989; Zajonc, 1983]. However, in addition to also being rather difficult, though possibly feasible, these experiments suffer the conceptual disadvantage that there are not actually spatially distinct paths as in the double-slit versions. A somewhat different scheme with photons using optical

---

[2] The atom is initially polarized in the $+\hat{x}$ direction, using a Stern-Gerlach apparatus (SGA) along $+\hat{x}$ as a filter. An SGA measuring <S> along $+\hat{z}$ acts as the first beamsplitter. Two SGA's (along $-\hat{z}$) act as mirrors to deflect the spin-up atoms to the recombining beamsplitter, another SGA along $+\hat{z}$. Similarly, two other SGA's (along $-\hat{z}$) act as mirrors to deflect the spin-down atoms to the recombining beamsplitter. In principle, the SGA in the second half of the interferometer may be made to reverse the effects of the SGA in the first half; the atoms after the final SGA will then be polarized in the $+\hat{x}$ direction again. This may be checked by an analyzer (a second SGA along $+\hat{x}$) after the interferometer.

parametric amplifiers as welcher Weg detectors was also proposed [Raymer and Yang, 1991]. However, only partial erasure is possible, and even then a practical obstacle of lack of near-unit efficiency photon detectors must be overcome.

As described below, we have performed a comparatively simple experiment involving the interference of photons, which demonstrates the salient features of the quantum eraser phenomenon [Kwiat et al., 1991; Kwiat et al., 1992][3]. (In Chapter 6 several proposals for pedagogically-improved schemes are discussed.) The *welcher Weg* information is stored in the polarization states of the photons, which are made distinguishable by means of a half waveplate. The erasure is performed by means of polarizers placed before the detectors. In Sect. 5.II, we present a detailed discussion of the essential elements of a quantum eraser; our experimental setup is described in Sect. 5.III. The nonclassical interference effect we employ is reviewed in Sect. 5.IV, including the necessary quantum field theory formalism. The loss of this interference is investigated theoretically in Sect. 5.V, and experimental results are shown. A simple derivation of the quantum eraser effect is presented in Sect. 5.VI, as are experimental results. We show that not only is it possible to recover interference, but also to change the form of the interference pattern. A comparison of our experiment to the various proposals is made in Sect. 5.VII, along with a discussion of its relation to some Bell's inequalities experiments and other two-photon experiments. The main results are summarized in Sect. 5.VIII. Throughout we will try to explain the phenomena both at an intuitive level using Feynman's rules, and also at a more formal level, using the established quantum field-theoretic approach to photodetection and correlation.

---

[3] Two other down-conversion experiments have also been offered as demonstrations of quantum erasure [Zajonc et al., 1991]. However, for reasons discussed in Chap. 6. these are lacking some of the essential attributes.

5.II. Idealized Quantum Eraser

We first describe the relevant features that comprise an *ideal* quantum eraser. Such an experiment would begin with an interfering particle (or particles). Envision that the particle has two processes or paths, which we label "a" and "b," leading to the same outcome (such as striking a particular point on a screen, or exiting a particular port of an interferometer), with probability amplitudes ψ_a and ψ_b, and a variable phase ϕ between them. Thus the total amplitude for this particular outcome, in the absense of any *welcher Weg* detectors, could be written:

$$\psi_s = \psi_a + e^{i\phi}\,\psi_b \,. \tag{5.1}$$

For instance, ψ_S might represent the value of the wave function at a particular point on a screen after double-slits, or at one of the exit ports of a Mach-Zehnder interferometer. The squared modulus of ψ_S corresponds to the probability that this outcome would occur (i.e., that the particle would appear at the particular point on the screen, or would choose that interferometer exit port); interference arises from the cross-terms $e^{i\phi}\psi_a^* \psi_b + e^{-i\phi}\psi_a \psi_b^*$. We have implicitly assumed thus far that the "a" and "b" paths are not distinguishable (e.g., that we are using a very massive, rigid slit mechanism).

Inserting a MA into one or both paths causes the system wave function to become entangled with the wave function of the MA, enlarging the relevant Hilbert space for the problem. That is, ψ_S becomes $\psi_s\,|\,MA\rangle$, and we replace ψ_a by $\psi_a\,|\,A\rangle$ and ψ_b by $\psi_b\,|\,B\rangle$, where the states $|\,A\rangle$ and $|\,B\rangle$ span the Hilbert space of the MA, and are by definition orthogonal. The previous cross-terms therefore vanish, and interference is lost. Importantly, in principle the MA need not induce any disturbance in the center-of-mass wave function of the particle: loss of interference results from the distinguishability of the final states. The above description contains the essential elements of a

welcher Weg experiment.

In a quantum eraser we go one step further. By making a suitable measurement on the MA, and correlating the results to the detection of the original particle(s), one can revive the interference effect. Of course, if one post-selects only those events where either A or B was measured, then no interference fringes will be observed. But if we calculate the probability of obtaining the particular outcome of the initial system *and* finding the MA in the symmetric state $(|A\rangle+|B\rangle)/\sqrt{2}$, we find

$$\left|\left(\frac{\langle A|+\langle B|}{\sqrt{2}}\right)\psi_s|MA\rangle\right|^2 = \frac{1}{2}|\psi_a + e^{i\phi}\psi_b|^2$$

$$= \frac{1}{2}\left(|\psi_a|^2 + |\psi_b|^2 + e^{i\phi}\psi_a^*\psi_b + e^{-i\phi}\psi_a\psi_b^*\right) , \qquad (5.2)$$

a revival of the original fringes. For concreteness, we now assume that processes "a" and "b" are equally likely to lead to the outcome we are examining, and that all phase differences between ψ_a and ψ_b are included in ϕ. Then (5.2) is simply $(1 + \cos\phi)/2$. Moreover, depending on the precise measurement made on the MA, one can actually alter the form of the interference, yielding *anti*-fringes $(1 - \cos\phi)/2$ instead of the expected fringes. This is achieved by projecting onto the anti-symmetric state $(|A\rangle-|B\rangle)/\sqrt{2}$. (If we do not correlate to the MA results, then the fringes and antifringes will cancel.) More generally, projecting along $(|A\rangle+\exp(i\theta)|B\rangle)/\sqrt{2}$ yields $(1 + \cos(\phi-\theta))/2$. (As an aside, one may note from the previous expression that the system and MA play symmetric roles, i.e., one could equally well view the system as carrying "which-state" information about the measuring apparatus.)

5.III. Experimental Setup

A schematic of our apparatus is shown in Fig. 5.1. The correlated (horizontally-

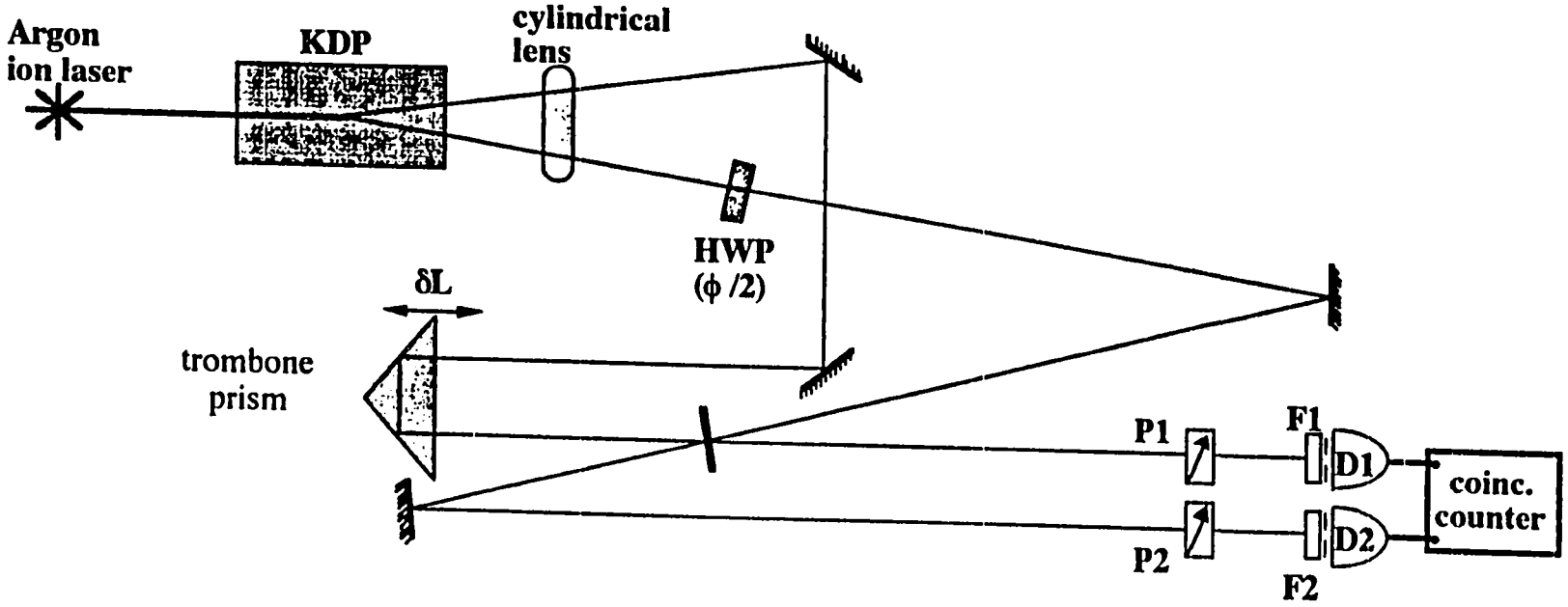


Figure 5.1. Schematic of experiment to observe Quantum Erasure. D1 and D2 are avalanche photodiodes, P1 and P2 are polarizers, F1 and F2 are bandpass filters, and HWP is a half waveplate whose optic axis is at an angle $\phi/2$ to the horizontal polarization of the down-converted beams.

polarized) photons from our KDP crystal were brought back together by means of mirrors, so that they impinged simultaneously on the surface of a beam splitter. (We shall henceforce describe this arrangement as a "Hong-Ou-Mandel (HOM) interferometer", after the inventors [Hong et al., 1987].) With irises (2-mm diameter) and filters (10-nm bandwidth, FWHM) at our detectors[4], we select out the nearly-degenerate signal and idler pairs at 702.2 nm. We measure singles and coincidence rates at the output ports (using a Stanford Research Systems SR400 Gated Photon Counter). As explained in the next section, if the beam splitter is placed such that the two photons reach it essentially simultaneously (i.e., within their coherence times), interference will result, in such a way that both photons always exit the same port of the beam splitter. Thus a null in the coincidence rate appears as the path length of one of the arms is slowly scanned, even though the singles rates remain unchanged. The width of the dip (≈40 μm FWHM) is determined by the filters in front of the detectors. In practice, it was preferable to vary the relative path length using an "optical trombone" in one arm of the interferometer, thus avoiding the problem of lateral walk-off associated with translating the recombining beam splitter directly. Translation of the prism was effected by a Burleigh Inchworm piezoelectric motion system; a Heidenhain optical encoder yielded a position resolution of 0.1 μm.

5.IV. Hong-Ou-Mandel Interference

A simplified schematic of the Hong-Ou-Mandel interferometer is shown in Fig. 5.2a. One can explain the coincidence null at zero path-length difference using the Feynman rules for calculating probabilities. The two indistinguishable processes (Fig. 5.2b) which lead to coincidence detection in the above setup are both photons being

[4] Each detector consists of an RCA C30902S avalanche photodiode (see Appendix E for specifications) cooled to -18°C, whose output is fed into an EG&G-Ortec #583 Constant Fraction Discriminator.

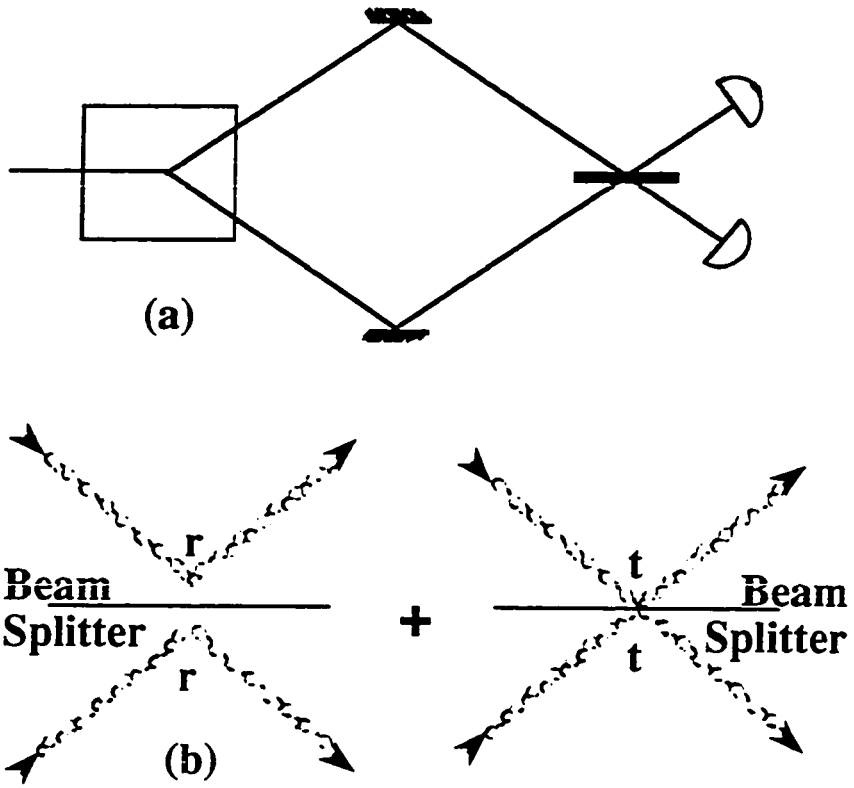


Figure 5.2. a) Simplified set-up for a Hong-Ou-Mandel interferometer.
b) Feynman paths for coincidence detection.

reflected at the beam splitter and both photons being transmitted. For simplicity, suppose we have a 50/50 beam splitter, and choose the amplitude transmission coefficient t to be real. The Feynman amplitudes are then $r \cdot r$ and $t \cdot t$, and the probability of a coincidence detection is

$$P_C = |r \cdot r + t \cdot t|^2 = \left|\frac{i}{\sqrt{2}} \cdot \frac{i}{\sqrt{2}} + \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}}\right|^2 = 0 \ , \qquad (5.3)$$

where the factors of i come from the phase shift upon reflection at a beam splitter. When the path-length difference is greater than the coherence length of the photons (i.e.

when the photon wave packets no longer overlap at the beam splitter), there is no such cancellation effect and coincidence events occur one half of the time, since each photon *individually* has a 50% chance of being reflected or transmitted.

More formally, we write the wave function after the beam splitter in terms of Fock states:

$$|\psi\rangle_{\Delta x=0} = \frac{1}{2}\left[\,|1_1\,1_2\rangle + i\,|2_1\,0_2\rangle + i\,|0_1\,2_2\rangle - |1_1\,1_2\rangle\right]$$

$$= \frac{i}{2}\left[\,|2_1\,0_2\rangle + |0_1\,2_2\rangle\right], \qquad (5.4)$$

where the subscripts denote the propagation modes to the two detectors, and the subscript of $|\psi\rangle$ indicates zero path-length difference. As discussed in Chap. 2, the conjugate photons actually have a relatively broadband frequency distribution, which is determined in practice by irises and filters in front of the detectors. However, since we operate near degeneracy, and since we are considering zero path-length difference, this generalization is an unnecessary complication for our purposes. As in Chap. 2, the coincidence counting rate is given by the second-order Glauber correlation function:

$$P_c = G^{(2)}(t_1,t_2;t_2,t_1) \propto \langle\psi|\,\hat{E}_1^{(-)}(t_1)\,\hat{E}_2^{(-)}(t_2)\,\hat{E}_2^{(+)}(t_2)\,\hat{E}_1^{(+)}(t_1)\,|\psi\rangle$$

$$\propto \langle\psi|\,\hat{a}_1^\dagger\hat{a}_2^\dagger\hat{a}_2\hat{a}_1\,|\psi\rangle\,, \qquad (5.5)$$

where we have used the Fourier representation of the field operators [cf. Eqs. (2.5a-d)], but have omitted the integrals over frequency for simplicity[5]. We have also neglected

[5] When the path-length difference is not zero, the field-operator integrals, along with those implicit in $|\psi\rangle$, determine the precise form of the interference dip as a function of path-length difference, when one integrates over the time resolution of the detectors. See also, Appendix B2, where the time dependence is explicitly taken into consideration, though for a somewhat different experimental arrangement.

polarization for the moment, which is justified because the photons are both horizontally-polarized, and our detectors are polarization independent. Clearly, the reduced coincidence operator $\hat{P}_{c,red.} \equiv \hat{a}_1^\dagger \hat{a}_2^\dagger \hat{a}_2 \hat{a}_1$ gives zero when it operates on $|\psi\rangle_{\Delta x=0}$, as given in Eq. (5.4).

When the path length difference is greater than the coherence length of the down-converted photons ($\Delta x \gg \tau_c$), the "transmission-transmission" and "reflection-reflection" coincidence possibilities are in principle distinguishable, so they do not interfere. In this limit, we find

$$P_c(\Delta x \gg \tau_c) \approx \frac{1}{4}\langle 1_1\, 1_2| \hat{P}_{c,red.} |\, 1_1\, 1_2\rangle + \frac{i^4}{4}\langle 1_1\, 1_2| \hat{P}_{c,red.} |\, 1_1\, 1_2\rangle = \frac{1}{2}. \qquad (5.6)$$

(The reduction by a factor of 2 reflects the fact that we are only considering *coincidence* counts, not cases where both photons go to the same detector.)

This demonstrates the coincidence dip at zero path-length difference to the beam splitter. Note that the singles rate at either detector, given by $P_s \approx G^{(1)}(t_j;t_j)$ $= \langle\psi| \hat{E}_j^{(-)}(t_j)\ \hat{E}_j^{(+)}(t_j)|\psi\rangle$, does not show this dependence on path length difference. It has been shown that as long as the visibility of the coincidence dip is greater than 50%, no semi-classical field theory can account for the observed interference [Ou, 1990; Franson, 1991].

5.V. Loss of Interference

In the spirit of the Feynman two-slit experiment, we ask if one can perform a "measurement" on the photons which will yield which-way information. Of course, we could place an APD or PMT directly in one of the input arms of the interferometer, but then lack of coincidence is a trivial consequence. We consider instead what happens when a half waveplate at an angle $\phi/2$ to the horizontal is inserted into one input arm of

the interferometer, as depicted in Fig. 5.1 (and we adjust the trombone to compensate for the optical path-length of the waveplate). The polarization of the photon in that arm is then rotated by ϕ, making the two Feynman paths partially distinguishable, thereby reducing the amount of interference. The degree to which the interference is lost depends on the angle ϕ, and is calculated below. In the extreme case ($\phi/2 = 45°$) the polarization states of the two different photons reaching the beam splitter are orthogonal. The two paths are now *completely* distinguishable and the amplitudes are squared before being summed. The result . . . no interference. These effects are shown in Fig. 5.3a.

In terms of our earlier formalism, we have entangled the number-state basis wave function with polarization information:

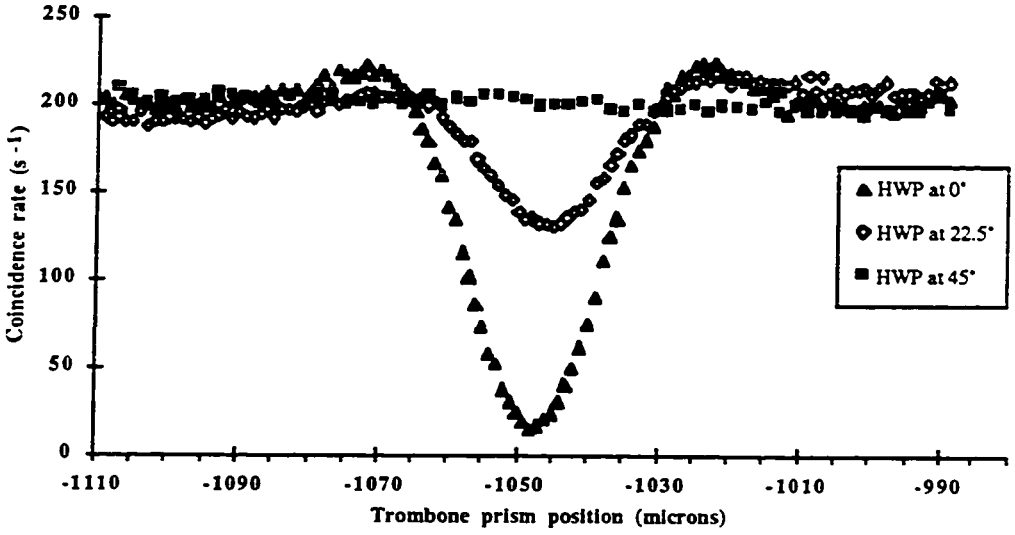
$$|\psi\rangle_{\Delta x=0} = \frac{1}{2}\left[\,|\,1_1\,1_2\rangle \otimes |\,H_1\,(H+\phi)_2\rangle - |\,1_1\,1_2\rangle \otimes |\,(H+\phi)_1\,H_2\rangle\right], \qquad (5.7a)$$

where the notation H_j indicates that the photon reaching detector *j* is horizontally polarized, and $(H+\phi)_j$ indicates that the photon is polarized at an angle ϕ to the horizontal. We have already omitted the kets in which both photons go to the same detector, since we are interested here in coincidence rates. We introduce the following simplified notation:

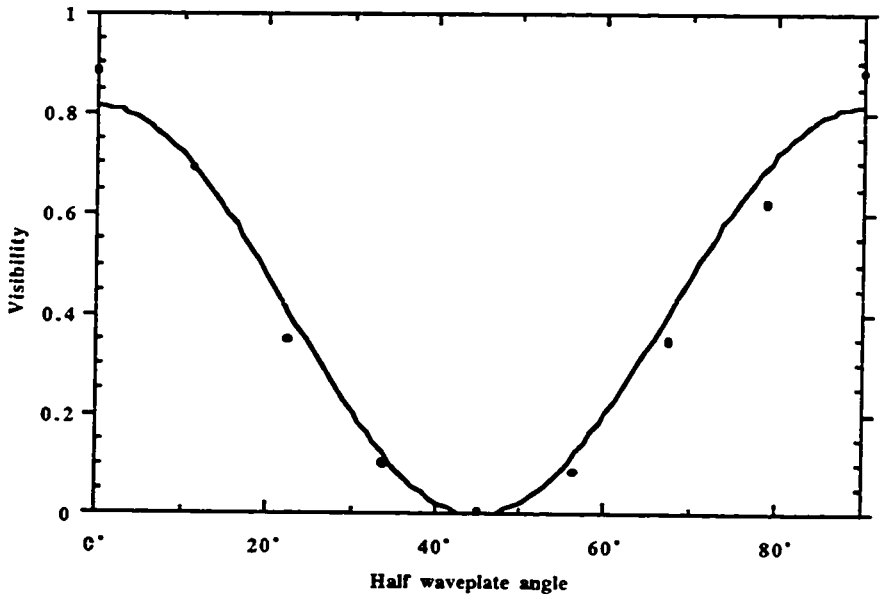
$$|\psi\rangle_{\Delta x=0} = \frac{1}{2}\left[\,|\,1_1^H\,1_2^{H+\phi}\rangle - |\,1_1^{H+\phi}\,1_2^H\rangle\right]. \qquad (5.7b)$$

To account for polarization, the field operators must now be generalized to the vector operators, $\hat{a}_j^\dagger = \hat{a}_{j,H}^\dagger\,\varepsilon_{j,H} + \hat{a}_{j,V}^\dagger\,\varepsilon_{j,V}$ and $\hat{a}_j = \hat{a}_{j,H}\,\varepsilon_{j,H} + \hat{a}_{j,V}\,\varepsilon_{j,V}$, where $\varepsilon_{j,H}$ and $\varepsilon_{j,V}$ are orthonormal polarization vectors (associated with detector index j=1,2) in the horizontal and vertical directions, respectively. The new reduced operator relevant for coincidence

Figure 5.3 (next page) a) Profile of interference dip in coincidence rate for three waveplate orientations. (Accidental coincidences have been subtracted, and rates far from dip have been normalized to the same value.) Note that the interference effect is seen to vanish when the waveplate is at 45°, i.e. when the input ports to the beam splitter are made distinguishable. b) Visibility as a function of half waveplate angle. The solid line is a fit to theory, with maximum visibility as the free parameter. The experimental points do not lie exactly on the same curve because slight fluctuations in alignment affect the visibility.


250
200
150
100
50
0
Coincidence rate (s -1)
-1110
-1090
-1070
-1050
-1030
-1010
-990
Trombone prism position (microns)
HWP at 0°
HWP at 22.5°
HWP at 45°

a.)


1
0.8
0.6
0.4
0.2
0
Visibility
0°
20°
40°
60°
80°
Half waveplate angle

b.)

Figure 5.3

counting is then

$$\hat{P}'_{c,red.} \equiv \sum_{\lambda_1,\lambda_2=H,V} a^{\dagger}_{1,\lambda_1} a^{\dagger}_{2,\lambda_2} a_{2,\lambda_2} a_{1,\lambda_1} = \sum_{\lambda_1,\lambda_2=H,V} a^{\dagger}_{1,\lambda_1} a_{1,\lambda_1}\, a^{\dagger}_{2,\lambda_2} a_{2,\lambda_2}$$

$$= \left(\hat{a}^{\dagger}_1 \cdot \hat{a}_1\right)\left(\hat{a}^{\dagger}_2 \cdot \hat{a}_2\right) = \left(\hat{a}^{\dagger}_{1,H}\,\hat{a}_{1,H} + \hat{a}^{\dagger}_{1,V}\,\hat{a}_{1,V}\right)\left(\hat{a}^{\dagger}_{2,H}\,\hat{a}_{2,H} + \hat{a}^{\dagger}_{2,V}\,\hat{a}_{2,V}\right) . \qquad (5.8)$$

Using the expansion $\left| 1_j^{H+\phi} \right\rangle = \left| 1_j^{H} \right\rangle \cos\phi + \left| 1_j^{V} \right\rangle \sin\phi$, we find

$$P_c(0) \propto \langle \psi | \hat{P}'_{c,red.} \left| \psi \right\rangle_{\Delta x=0} = \frac{1}{2} \sin^2 \phi \ . \qquad (5.9)$$

When the path-length difference is greater than the photon coherence length, the calculation of the coincidence rate proceeds as before:

$$P_c(\Delta x \gg \tau_c) \propto \frac{1}{4} \langle 1_1^H\, 1_2^{H+\phi} \left| \hat{P}'_{c,red.} \right| 1_1^H\, 1_2^{H+\phi} \rangle$$

$$+ \frac{i^4}{4} \langle 1_1^{H+\phi}\, 1_2^H \left| \hat{P}'_{c,red.} \right| 1_1^{H+\phi}\, 1_2^H \rangle = \frac{1}{2} . \qquad (5.10)$$

The visibility of the dip, defined as $V = \{P_c(\Delta x \gg \tau_c) - P_c(\Delta x = 0)\}/P_c(\Delta x \gg \tau_c)$, has the form $\cos^2 \phi$. The experimental demonstration of this relationship is shown in Fig. 5.3b. The lack of perfect visibility even at $\phi = 0$ results from imperfect alignment of the system, so that the signal and idler modes leaving the beam splitter are already somewhat distinguishable, regardless of their polarization.

Following Scully *et al.*[Scully et al., 1978], one can also approach the loss of interference in terms of the density matrix. When the photon-propagation states are entangled with the polarization states, the density matrix of the system is enlarged. It

still represents a pure state, however, with the quantum coherence of the entanglement manifested in the off-diagonal matrix elements. The "collapse" to a mixed state occurs when we trace over the polarization degrees of freedom, i.e., when we detect the final propagation direction of the photons *irrespective of their polarization*. This effectively removes the coherences between the formerly interfering states, although the full density matrix has undergone only unitary evolution and retains the original coherences. In this case the reduced density matrix has only diagonal elements, because the polarization states $|H_j\rangle$ and $|V_j\rangle$ (which are essentially the "environment" for our purposes) are orthogonal. This method of decoherence was recently discussed by Zurek, although he focussed on an environment which was either "uncontrollable" or possessed a large number of degrees of freedom [Zurek, 1991]. In either case, the process is effectively irreversible, which is certainly not the case in our experiment, as we shall see presently.

5.VI. Quantum Eraser

The essence of the quantum eraser can be understood relatively easily now in Feynman's language of distinguishability. As we have seen, with the half waveplate at 45° ($\phi = 90°$) the two paths leading to coincidence detection ("reflection-reflection" and "transmission-transmission") are distinguishable; they leave the light in each port in a different polarization state. For this reason, their probabilities are to be added incoherently, and there is no interference term. What if one could erase the information carried by the polarization, thus making the final states indistinguishable? This is precisely what happens when one places polarizers oriented at 45° to the horizontal in both output ports of the interferometer. (See Fig. 5.1.)

Both paths can lead to coincidence detection, and to the same final state. Therefore, their probability *amplitudes* are added, thus reviving the Hong-Ou-Mandel interference dip at equal path-length. Note that the insertion of a polarizer in only one of

the output ports is insufficient to erase the distinguishability of the final states, because the photon in the other port still possesses *welcher Weg* information. Hence, the only effect of a single polarizer is to reduce the relevant singles rate and the coincidence count rate by half.

The editing accomplished with two polarizers is not limited to erasure, as can be motivated by the following observation. Regardless of the rest of the system, the light in port 2 can always be broken up into its orthogonal polarization components. But we just saw that with both P1 and P2 at 45°, the interference dip reappeared (albeit attenuated by a factor of 4). Furthermore, we argued that the coincidence profile with polarizer P1 at 45° and P2 not in place was a flat line. It is clear then that if P1 is placed at 45° and P2 is placed at -45°, instead of a dip, a peak centered at zero path-length difference will now appear. These theoretical results are presented in Fig 5.4a and our data in Fig 5.4b. As shown below, this is merely a specific instance of a more general property of the two-photon state emitted by the interferometer. (It should be noted that the possibility of producing a *peak* at zero path-length difference greatly aids the alignment process for the Hong-Ou-Mandel interferometer.)

We now present a simplified analysis, limiting ourselves to the case $\phi = 90°$. A more complete calculation is presented in Appendix B1. The output of the interferometer is given by a special case of the entangled state in Eq. (5.7b):

$$|\psi(0)\rangle = \frac{1}{2}\left[\,|1_1^H\,1_2^V\rangle - |1_1^V\,1_2^H\rangle\right], \qquad (5.11)$$

where we have again dropped terms which could not lead to coincidence counts. Detection of a photon at one port with no polarizer collapses the remote photon into a *mixed state* with half polarized horizontally and half polarized vertically. However, if a linear polarizer is placed at an angle θ_1 to the horizontal in output port 1, a detection

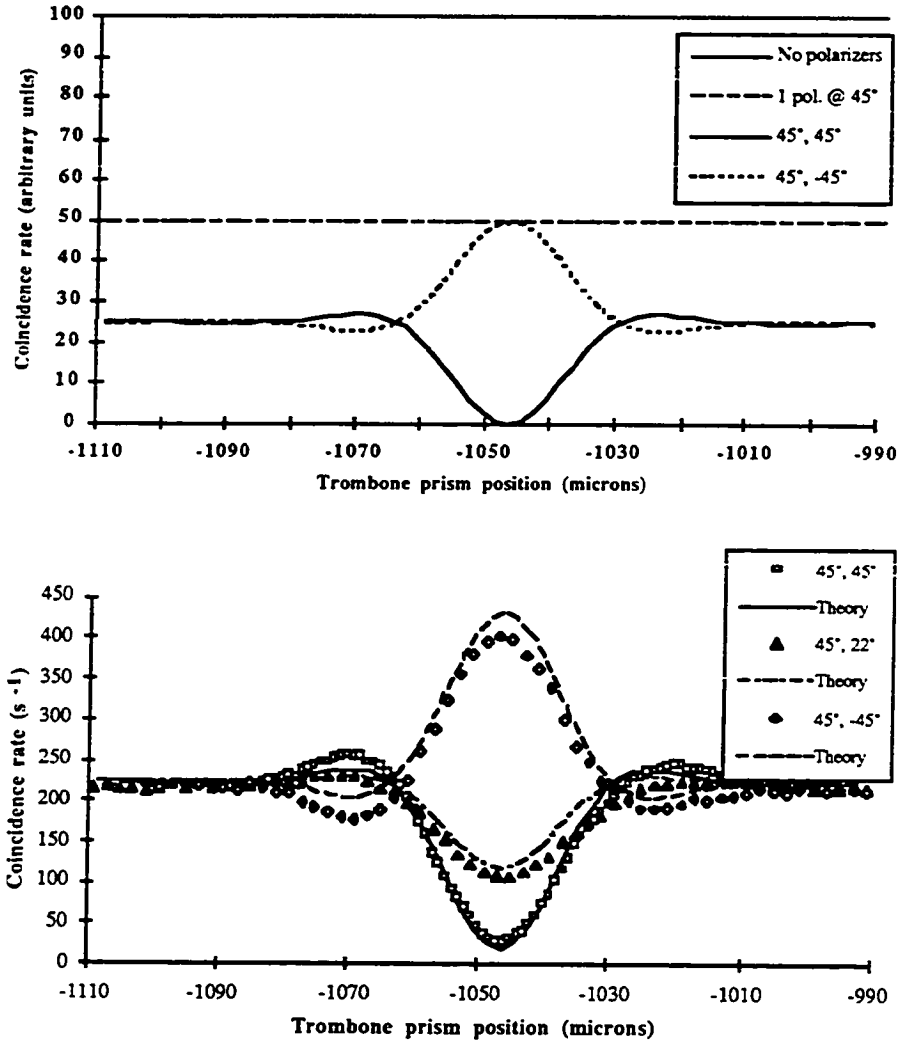


Figure 5.4. a) Theoretical curves showing how two polarizers at appropriately chosen angles can erase distinguishability, restoring an interference pattern. b) Experimental data and scaled theoretical curves (adjusted to fit observed visibility of 91%) with polarizer 1 at 45° and polarizer 2 at various angles. Far from the dip, there is no interference and the angle is irrelevant. At the dip, the nonlocal collapse of the polarization of photon 2 causes us to observe sinusoidal variation as predicted in Eq.(5.14). (Normalization is as in Fig. 5.3a.)

event at detector 1 corresponds to a von Neumann projection in the subspace corresponding to that port onto the state vector $|\theta_1\rangle = \left(|1_1^H\rangle\cos\theta_1 + |1_1^V\rangle\sin\theta_1\right)$. We are left with a pure state for the conjugate photon:

$$\langle\theta_1 \,|\, \psi\rangle_{\Delta x=0} = \frac{1}{2}\left(|1_2^V\rangle\cos\theta_1 + |1_2^H\rangle\sin\theta_1\right). \tag{5.12}$$

Examining output port 2 with another polarizer, we observe that the light in this mode is polarized orthogonal to θ_1; the probability amplitude is

$$\langle\theta_2\ \theta_1 \,|\, \psi\rangle_{\Delta x=0} = \frac{1}{2}\left(\langle 1_2^H \,|\cos\theta_2 + \langle 1_2^V \,|\sin\theta_2\right)\left(|1_2^V\rangle\cos\theta_1 + |1_2^H\rangle\sin\theta_1\right)$$

$$= \frac{1}{2}\left(\sin\theta_2\cos\theta_1 - \cos\theta_2\sin\theta_1\right) = \frac{1}{2}\sin\left(\theta_2 - \theta_1\right). \tag{5.13}$$

Thus,

$$P_c(0) \propto \left|\langle\theta_1\,\theta_2 \,|\, \psi\rangle_{\Delta x=0}\right|^2 = \frac{1}{4}\sin^2\left(\theta_2 - \theta_1\right). \tag{5.14}$$

Hence the interference dip can be revived, but depending on the *relative* angle of the polarizers (see Fig. 5.5), it may be phase-shifted and thus reappear as a peak, or any intermediate form. (Eq. (5.14) may be recognized as a typical prediction of quantum theory when applied to certain tests of Bell's inequalities (see Sect. 7.III, for instance). This is hardly a coincidence, for the same nonlocal effect is responsible for both phenomena. We will discuss the relationship between our quantum eraser experiment and similar Bell's inequalities experiments in the next section.)

On the other hand, when the path-length difference is great compared with the coherence length of the photons, the probabilities for the two different paths leading to coincident detection add incoherently regardless of any polarizers:

$$P_c(\Delta x \gg \tau_c) \propto \left|\frac{1}{2}\langle\theta_1\,\theta_2 \,|\, 1_1^H\, 1_2^V\rangle\right|^2 + \left|\frac{1}{2}\langle\theta_1\,\theta_2 \,|\, 1_1^V\, 1_2^H\rangle\right|^2$$

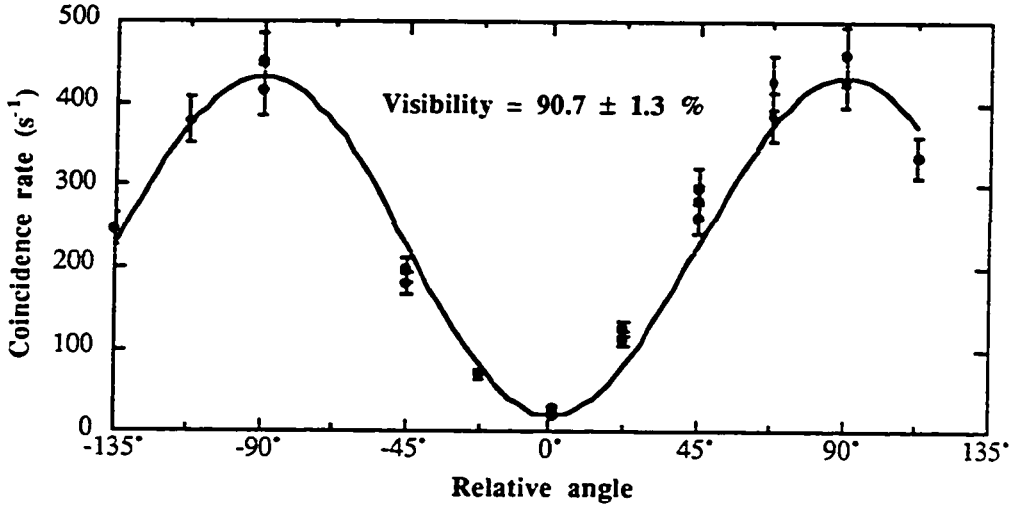


Figure 5.5. Plot of coincidence rate versus relative angle of polarizers 1 and 2, corrected for accidentals. The smooth curve is a fit to theory, with visibility as a free parameter.

$$= \frac{1}{4}\left\{\cos^2\theta_1 \ \sin^2\theta_2 \ + \sin^2\theta_1 \ \cos^2\theta_2\right\}$$

$$= \frac{1}{8}\left\{\sin^2(\theta_2 - \theta_1) \ + \sin^2(\theta_2 + \theta_1)\right\}. \qquad (5.15)$$

This varies with absolute angle, in contrast to Eq. (5.14), since horizontal and vertical components act independently (see Fig. 5.6).

5.VII. Discussion

The relationship between our experiment and other proposed quantum eraser schemes is rather subtle. For comparison, we will focus on a particular proposal by Scully *et al.*[Scully et al., 1991], in which excited atoms are made to interfere in a two-

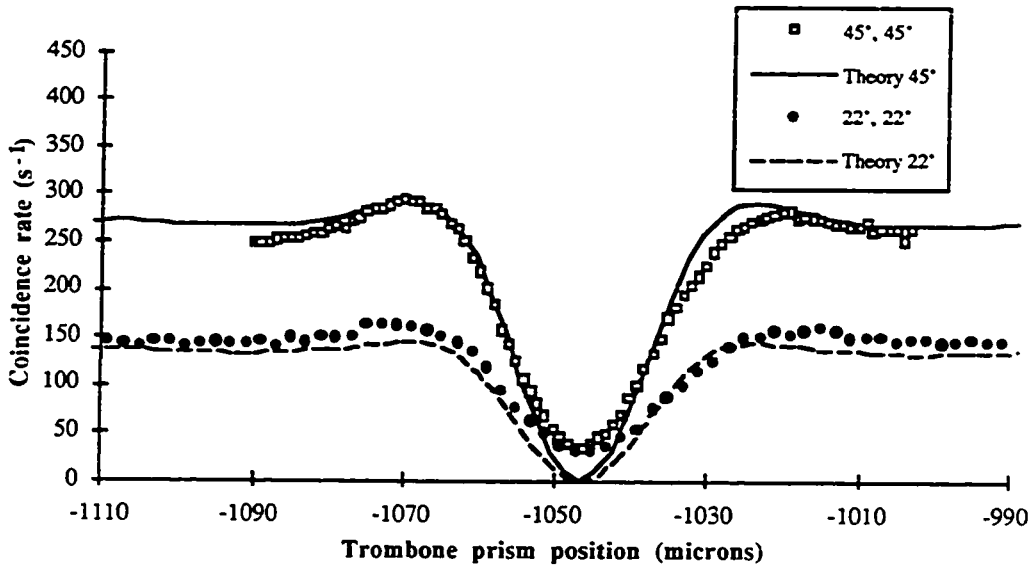


Figure 5.6. Erasure also occurs, but in a somewhat different fashion, if the two polarizers are kept at the same angle and scanned towards 45°. With perfect visibility, the absolute angle would affect only the count rates far from the dip (clearly, at 0° or 90° no coincidences can ever be observed), and a total null would be observed at the dip because the two photons in the "singlet" state of Eq.(5.11) are always orthogonal. (The data are normalized to singles and corrected for accidentals.)

slit type geometry. A micromaser cavity is placed in each of the interfering paths, and prepared so that an atom passing through will decay with near certainty, leaving a photon in the cavity. For certain initial states of the cavity fields (i.e., number states), the extra photon from the decay constitutes *welcher Weg* information, and the *first*-order interference effect (fringes visible in singles detection of the atoms) is washed out. By allowing the cavity fields to subsequently interfere at a suitably placed detector, quantum erasure may be accomplished. However, as the authors stressed, this erasure is fundamentally a second-order phenomenon, in that the fringes can only be seen by *correlating* the photon counts with data stored elsewhere. In principle the decision of

whether or not to erase could be postponed indefinitely, even beyond the time of detection of the atoms.

Our experiment differs somewhat from this proposal in that the basic Hong-Ou-Mandel interference effect is intrinsically a *second*-order, quantum mechanical effect. That is, fringes are never visible in singles detection, and coincidence fringe visibility above 50% defies semi-classical explanation. While this puts the pre- and post-eraser fringes on the same footing, it has a pedagogical disadvantage in that the distinguishing information we add and then erase is carried by the same photons which are to interfere. (Several proposals to avoid this deficiency are discussed in Chap. 6.) We maintain that this difference is not as fundamental as it may first appear. Although our photons themselves carry the information describing which trajectory they took, they do so only via their polarization vectors. We erase the information *after they have already left the interferometer*, and without affecting their center-of-mass wave function. In both Scully *et al.*'s proposal and the present experiment, the measurement of which-way information consists of coupling the interfering particle's spatial wave function to the disjoint Hilbert space describing the *welcher Weg* detection system (e.g., micromaser cavities or photon polarization space). While this does not affect the spatial wave function, it does enlarge the Hilbert space in which it resides. It is the enlargement of the Hilbert space through entanglement, and subsequent reduction, which is the central feature of the QE.

It is useful to consider a slight *gedanken* variant of our experiment, which is in principle identical to it. We employ polarizing beam splitters, rather than simple polarizers, so that both polarizations may be detected. A computer then stores in one file the times of photon detection events (regardless of polarization), and in another file the polarizations of the detected photons. (Note that by making a polarization-insensitive quantum nondemolition (QND) measurement *before* the polarizers, one could delay the

choice of polarizer orientations until after the coincident detection measurement.) Varying the orientation of the polarizing beam splitter then affects only the second file, and not the first; no interference is discernible in the first file until the data is correlated with that in the other file. As this may be performed long after the data is originally stored, we have a "delayed-choice" version of the quantum eraser. However, this is not really a *practical* proposal (hence the usage of the adjective "*gedanken*" in the first sentence), given the present status of QND results [Roch et al., 1992]. In the next chapter we propose three improved schemes, which have feasible delayed-choice versions.

Some of the results presented here have been observed previously by other researchers, in the context of non-local correlations and EPR experiments [Ou and Mandel, 1988; Shih and Alley, 1988]; see also Chap. 7. Our goal was to shift some of the focus to the phenomenon of quantum erasure, which is another striking manifestation of quantum entanglement. The central element in most tests of Bell's inequalities to date is the singlet state of the correlated photons [Shimony, 1990]. Although our photons are not produced in such a state in the down-conversion process, it effectively arises when their polarization states are entangled with their propagation modes (i.e. lower or upper arm)[6]. From this perspective, the quantum eraser and the Bell-type tests are just different approaches to investigating and understanding the character of the entangled states. One might then argue that some of the previous Bell-

[6] Strictly speaking this is not true, as is immediately apparent from the extra terms omitted from Eqs. (5.8), corresponding to both photons propagating to the same detector. Nevertheless, since we only detect coincidence counts, the *reduced* wave function is a singlet state (so long as the path-length difference is much less than the coherence times of the conjugate photons). In fact, other experimenters have used this exact arrangement in tests of Bell's inequalities [Ou and Mandel, 1988; Shih and Alley, 1988]. In Chap. 7 we will discuss in detail the issues involved with discarding counts, and the implications for a rigorous test of Bell's inequalities.

type tests *were* the first quantum eraser results. However, we believe that it is important to demonstrate the loss of interference before reviving it, an aspect that, to our knowledge, has not been covered in any Bell-type experiments. Furthermore, the general goals of the two viewpoints differ. While the Bell's inequalities experiments seek to disprove the reality of local hidden variable models, the quantum eraser stresses the loss of coherence through entanglement with the "environment", and the possibility of recovering that coherence in certain circumstances.

5.VIII. Conclusions

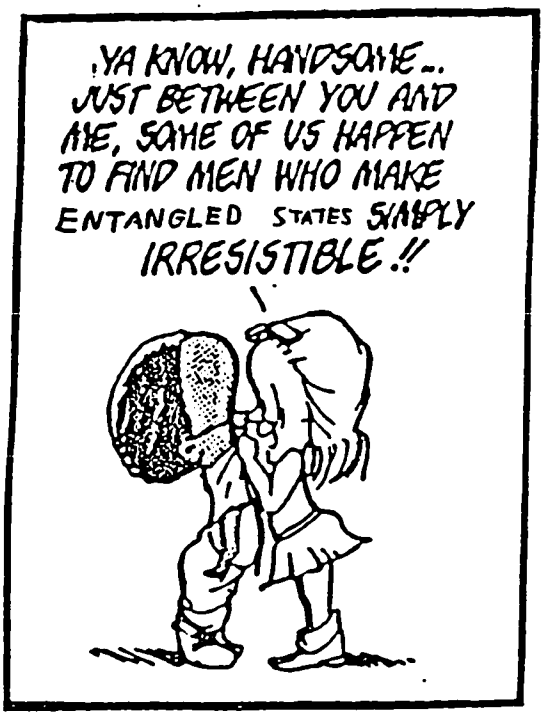
The quantum eraser offers a new perspective on interference and loss of quantum coherence in terms of (in)distinguishability of paths. Alternate paths are made distinguishable by correlating them to other parts of the "environment." Depending on how we reduce the resulting enlarged Hilbert space, we may opt to retain *welcher Weg* information and have no interference, or to reestablish indistinguishability and interference. We may make this choice long after the original interfering system has been detected, by correlating that data with the results of particular measurements on the "environment" with which the system was entangled. Of course, this demands that the coherence of the relevant environmental states be maintained.

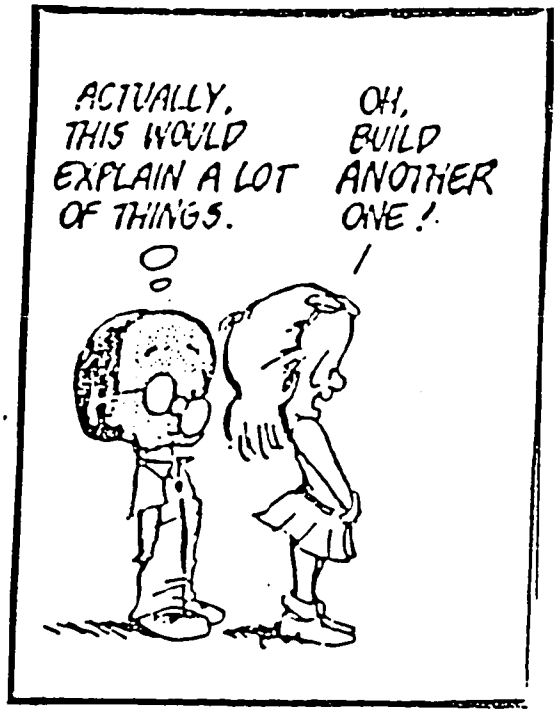
We have seen that it is possible to demonstrate the essential features of the quantum eraser using a comparatively simple arrangement involving the correlated photons produced in spontaneous parametric down-conversion. The interference normally present when the two photons are superposed at a beam splitter was made to vanish when the alternate processes leading to coincidence counts were made distinguishable. For this purpose a half waveplate in one arm of the interferometer served to entangle the photon spatial wave function with the polarization subspace. Using polarizers at the output, it was then possible to restore interference, and even to alter its form.

One of the things the quantum eraser teaches us is that the state involved in interference is the *total* physical state, which in addition to photon spatial wave functions may include photon polarization, or even distant atoms with which the photons have interacted. In all realizations of the quantum eraser, the "magic" comes about through entangling the interfering system with some other degrees of freedom. The eraser "meddles" with the interference only via this entanglement, regardless of whether the extra information is stored in states of remote atoms or in the polarization components of the photonic wave functions. The process allows the introduction of an arbitrary phase between different components of the entangled state; in this sense, the phenomenon is better described as quantum *editing*.


YOU! YOU'RE THE FELLOW WHO BUILT THIS WILD LITTLE QUANTUM ERASER, AREN'T YOU?!
5-23


OH, IT IS SO... SO RISKY! SO OUTRAGEOUS! SO... DELICIOUSLY DANGEROUS!


YA KNOW, HANDSOME... JUST BETWEEN YOU AND ME, SOME OF US HAPPEN TO FIND MEN WHO MAKE ENTANGLED STATES SIMPLY IRRESISTIBLE!!


ACTUALLY, THIS WOULD EXPLAIN A LOT OF THINGS.
OH, BUILD ANOTHER ONE!

Chapter 6: Three Proposed Quantum Erasers

6.I. Introduction

The fundamentals of quantum erasure were presented in the previous chapter, as well as the results of one experimental demonstration. There have been three such experiments discussed in connection with the quantum eraser [Ou et al., 1990a; Zou et al., 1991; Kwiat et al., 1992], all employing the correlated photon pairs produced in spontaneous parametric down-conversion[1]. For different reasons, however, none of these is an optimal demonstration of the phenomenon, each lacking one or more of its desirable attributes. In fact two are actually not quantum erasers at all in the strictest sense of the term, and the third, presented in the previous chapter, while incorporating the basic features of erasure, is pedagogically wanting. One may then ask whether it is possible to perform an experiment which is more complete. In this chapter we present three new experimental schemes for producing a quantum eraser, including delayed-choice mechanisms [Kwiat et al., 1993]. As these are all modifications of one of the earlier experiments [Zou et al., 1991], we believe them to be feasible. Our motivation for considering and performing such experiments (aside from that alluded to in the Introductory cartoon) is to emphasize as clearly as possible the non-separability inherent in quantum mechanics, and the role of the *total* physical state in determining coherence.

In Sect. 6.II we summarize those features of a quantum eraser that we feel need to be manifestly evident in an optimal demonstration. In Sect. 6.III we discuss the three previous experiments, concentrating on the shortcomings of each, also introducing

[1] As discussed in Sect. 5.I, several proposals have been made for demonstrating quantum erasers using atoms or neutrons [Scully, et al., 1978; Englert, et al., 1988; Schwinger, et al., 1988; Englert, et al., 1992], but all of these are at best very difficult in practice, and none have been carried out thus far.

techniques which will be relevant for the proposed experiments. These are presented in Sect. 6.IV, where we will show that each allows a truly nonlocal, delayed-choice aspect. Conclusions are in Sect. 6.V. A detailed calculation for one of the proposals is given in Appendix B2.

6.II. Experimental Requirements

Any quantum eraser experiment should contain certain basic elements (e.g., it should not be just a *welcher Weg* demonstration), as presented in Sect. 5.II; however, for an *optimal* demonstration (in a pedagogical sense), there are several aspects which should be emphasized. To underscore the non-separability (nonfactorizability, in the sense of Schrödinger, of two- (or more) particle wave functions) of quantum mechanics, one important feature of an ideal quantum eraser is that there be a possible element of "delayed choice": The measurement on the MA could be made *after* the interfering particle has been detected. In fact, our decision to use the MA as a *welcher Weg* detector or a quantum eraser, could also be made after detection of the interfering particle. It is only via the subsequent correlation of the results (of measurements on the original particle and the MA) that either interference or *welcher Weg* information may be recovered. In the original delayed choice discussion by Wheeler, he pointed out that the decision to display wave-like or particle-like aspects in a light beam may be delayed until after the beam has been split by the appropriate optics [Wheeler, 1979]. The situation with the quantum eraser is even more striking -- the decision to measure wave-like or particle-like behavior may be delayed until after the *detection* of the initial quantum (but of course *before* the result of the measurement on the measuring apparatus), an irreversible process[2].

[2] There is as yet no clear answer to the question of exactly *when* or even *how* the detection becomes an irreversible event. Presumably, once the detector has coupled to some large number of degrees of freedom, the process will be effectively irreversible.

A second desirable feature of a true quantum eraser is that it employ single particles (as opposed to coherent states, for example); the reason is that the entire discussion of "which way" information depends on the notion of the particle-like aspect of an indivisible quantum. Finally, although nowhere in Chap. 5 did we require that the distinguishing information be carried separately from the interfering particle, this is clearly preferable from a pedagogical point of view. As we stated in the previous chapter, it is the enlargement of the Hilbert space through entanglement, and subsequent reduction, which is the central point of the quantum eraser; however, the non-separability inherent in the process becomes more apparent when a system spatially distinct from the initial interfering system serves as the measuring apparatus.

6.III. Past Experiments

All three of the experiments previously discussed in the context of quantum erasure use down-converted photon pairs [Ou et al., 1990a; Zou et al., 1991; Kwiat et al., 1992; Zajonc et al., 1991]. The first experiment [Ou et al., 1990a] involves an interference effect which exists only in coincidence detection. Two down-conversion crystals are pumped by coherent cw pump beams (see Fig 6.1). The signal beams are mixed at a beam splitter, while the idler beams are mixed at a separate beam splitter, such that after the beam splitters there is no way of distinguishing from which crystal a given pair of photons originated. Fringes in coincidence are observed as any of the pre-beam splitter path lengths are varied, while no interference is seen in the singles rates. The "'delicate' change" which leads both to distinguishability and to erasure in that example is the removal and reinsertion of one of the beamsplitters[3] (e.g., if the idler beam

---

In practice, we can confidently claim that by the time the event has attained the level of a macroscopic pulse out of the detector, it is thermodynamically irreversible.

[3] The authors correctly point out that this beamsplitter seems to have no effect on the signal photons. Careful analysis, however, shows that the signal photons never exhibit interference on their own, but only when detected in coincidence with the idler photons.

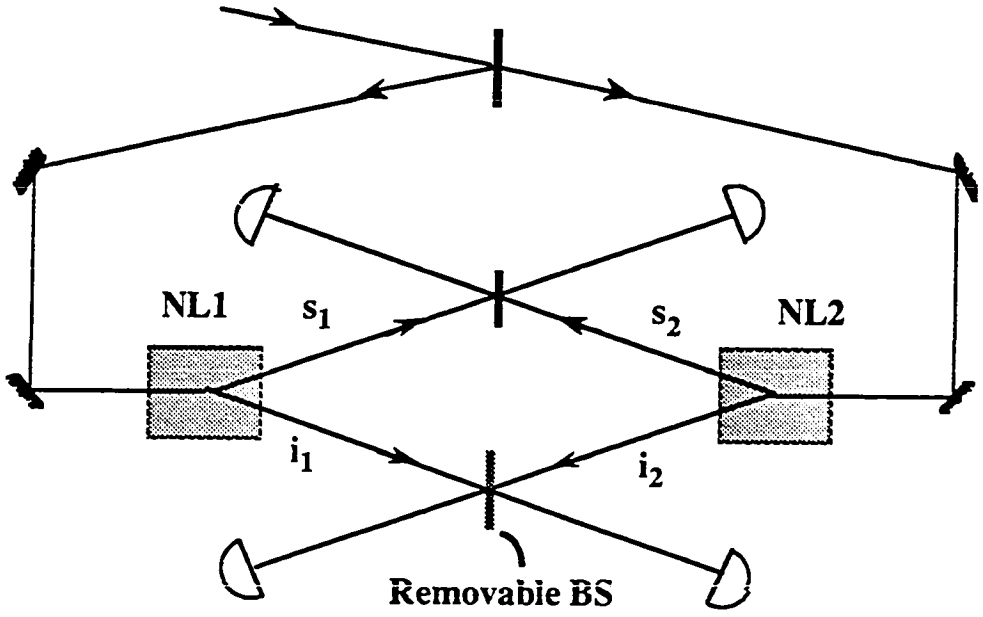


Figure 6.1 Schematic of setup used in [Ou et al., 1990a].

splitter is removed the idler beams then carry *which-crystal* information, and no interference occurs). In this sense, it is deficient as a quantum eraser since it is the structure of the interferometer itself, and not just the structure of the detection scheme, which determines once and for all the presence or absence of interference fringes. To put it differently, there is never interference unless a larger Hilbert space already including the idler photons is considered; removal of the idler beam splitter does not enlarge the Hilbert space.

The experiment of [Zou et al., 1991], while a remarkable demonstration of complementarity in its own right, differs fundamentally from the quantum eraser proposal, in that it is entirely a first-order (one-photon), not a second-order (two-photon), interference effect, and no delayed choice version would be possible. Again

As the beamsplitter does directly affect these idlers, it must be considered as part of the interferometer.

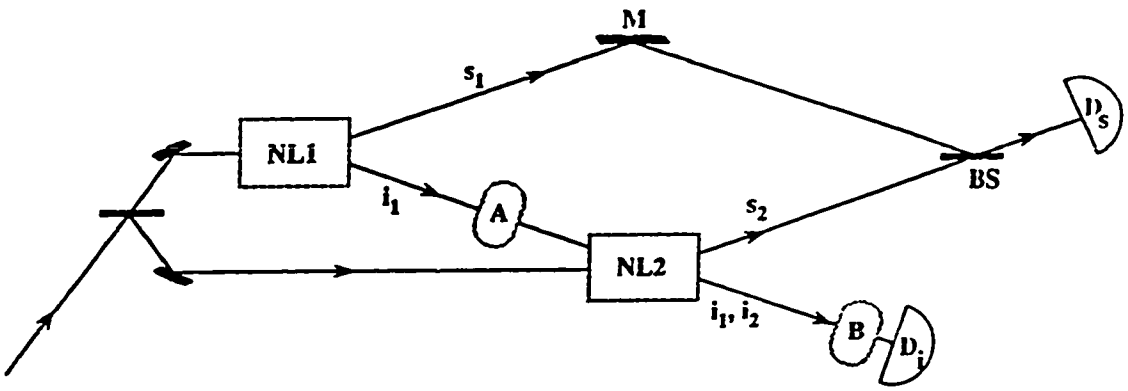


Figure 6.2 Schematic of setup used in [Zou et al., 1991], with the possible inclusion of additional elements to make it suitable for a quantum eraser. In the absence of any waveplates, the identically-polarized idler photons from either crystal are indistinguishable. Consequently, interference fringes may be observed in the signal singles rate (at detector D_S) if any of the path-lengths in the interfometer are varied. A half waveplate at A distinguishes the idler photons; the which-crystal information may be erased with a polarizer at B. Fringes may be recovered by correlating measurements at the two detectors, i.e., in the coincidence rate.

the trajectories of the idler photons from each crystal overlap. A beam splitter acts to superpose the trajectories of the signal photons. The basic interference effect arises between these signal photons, as the path length from either of the crystals to the beam splitter is varied slightly. If the path lengths are adjusted correctly, and the idler beams overlap precisely, there is no way to tell, even in principle, from which crystal a photon detected at D_S originated -- there results interference in the signal singles rate at D_S (and thus trivially in the coincidence rate between D_S and D_i). If the idler beam from crystal NL1 is prevented from entering crystal NL2 (or even if the two idler beams are only slightly misaligned), then the interference vanishes, because the presence or absence of an idler photon at D_i then "labels" the parent crystal. However, at this stage the *welcher*

Weg measurement is effectively irreversible. There is no way at all in practice to "erase" the distinguishability by any transformation on the idler state alone, and certainly no possibility of a delayed choice, even in principle [4]. Because detection events (of the signal photons) are never compared with measurements on the "measuring apparatus" (the idler photons, here), there is no way to recover fringes or anti-fringes.

In Chap. 5 we discussed the most recent experiment [Kwiat et al., 1992], which also employs a nonclassical second-order interference effect, so that coincident detection of the two photons is required, but uses only a single crystal. This experiment is pedagogically superior as a quantum eraser to the obvious "classical" analog[5] (the canonical two slit-experiment with a classical light field, a polarization-rotator in front of one slit, and a variable polarizer/analyzer in front of the detection screen), because its use of single-particle states permits the notion of *welcher Weg* information. However, it is not pedagogically optimal because the *welcher Weg* information is carried by the interfering particles themselves, not stored in some external measuring system. As a result, performing a delayed choice version of the experiment would be very difficult, requiring quantum nondemolition (QND) detection of the photons before a subsequent measurement on the polarization part of their wave function. It would be preferable if the entanglement were to an external system, not to an internal degree of freedom of the interfering particles.

---

[4] One might imagine recombining the two idler beams at a second beamsplitter after the second crystal. But this is then topologically equivalent to the first pseudo-eraser discussed above [Ou, et al., 1990a].

[5] We use the term "classical" to indicate that the effect has an explanation using a classical field model. In such a theory there is no meaning to the concept of "which-path" information, since a wave can follow two paths simultaneously; even the quantum mechanical description of such an experiment has many particles, some of which follow each route, and which are indistinguishable from one another.

6.IV. Proposed Quantum Erasers

The basic setup is the same as that of the second experiment discussed above [Zou et al., 1991]. The difference is the inclusion of the additional elements A and B (see Fig. 6.2), which we shall discuss presently. Note that for the signal beams to interfere, it is crucial that the idler photons be indistinguishable after crystal NL2, even though we need not detect them to observe the interference at the signal detector D_S. In particular, the idler photons from the two crystals must have the same polarization and color, and must arrive at the idler detector D_i at the same time (to within the two-photon correlation time). If this were not the case, then one could *in principle* determine which crystal emitted the detected *signal* photon by making a careful measurement at D_i of the polarization or energy or arrival time of the conjugate *idler* photon. We shall purposely violate these constraints to distinguish the idler photons from NL1 and NL2, removing the first-order interference at D_S. This scheme is superior to the previous experiment (of Chap. 5) because the *welcher Weg* information is not carried by the interfering particles, the signal photons. Correlating the counts at D_S with subsequent measurements on the idler photons, we can implement a quantum eraser.

There are several ways to proceed. The simplest is to insert a half waveplate between the two crystals (at A, in Fig. 6.2), rotating the polarization of the idler from NL1 by 90°. Since the idlers are now distinguishable, the interference of the possible signal paths disappears. Just as above [Kwiat et al., 1992], however, we can use a polarizer before D_i (at B, in Fig. 6.2) to erase the distinguishing information, and correlate the counts at D_S and D_i. If the polarizer is aligned along either of the idler-polarization directions, no interference will be seen in singles or coincidence. If aligned between the two (i.e., at ±45°), we will obtain either fringes or anti-fringes in coincidence (but not in singles). An experiment closely related to this is currently in progress in A. Zeilinger's lab at the University of Innsbruck. The delayed-choice

feature is that we could in principle decide on the polarizer setting, and hence whether we would see fringes, anti-fringes, or no fringes, *after* the signal photon was already detected. Experimentally, this could be accomplished by having an optical delay before the polarizer, and using a Pockels cell to effectively rotate the polarizer very quickly.

A second way of making the idler photons distinguishable is to increase the path length between the two crystals, by adding a delay line of length d_A (greater than the photons' coherence length) at position A (see Fig. 6.3). To subsequently make the idler

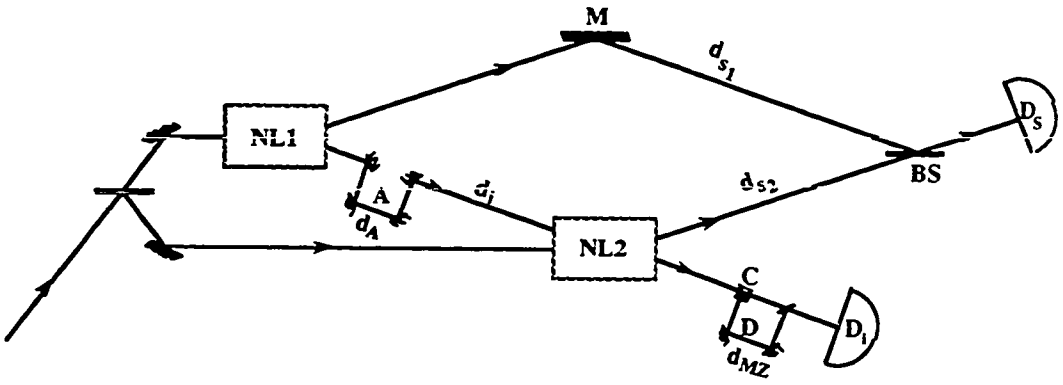


Figure 6.3 Two other quantum eraser versions rely on the same basic setup as in Fig. 6.2, but incorporate a delay line, of length d_A (much greater than the coherence length of the idler photons) between the two crystals. Erasure is performed after crystal NL2 using an unbalanced Mach-Zehnder, whose path-length difference d_{MZ} is essentially equal to d_A. Fringes in the correlated detection events between D_S and D_i will have 50% or 100% visibility, depending on the detector time resolution (see text). In the second version of this type, where the visibility is always 100%, a half waveplate at A rotates the NL1 idler polarization so that it is orthogonal to the NL2 idler polarization. In addition, the first beam splitter (at C) of the Mach-Zehnder is a polarizing beam splitter, oriented so that idler photons from NL1 (NL2) take the short (long) path. Finally, a second waveplate in one of the arms (e.g., at D) makes identical the idlers' polarizations.

photons indistinguishable again, we essentially need to add two delay lines after the second crystal, the difference in whose path lengths is equal to d_A. In practice, we can achieve this using an unbalanced Mach-Zehnder interferometer before detector D_i. The choice of whether to detect fringes or anti-fringes is set by the detailed phase difference in the Mach-Zehnder, which could in principle be chosen after the signal photons had been detected.

We now present a simplified calculation demonstrating the above claims. (A detailed formal calculation is given in Appendix B2.) For clarity we assume that the down-converted photons are degenerate at frequency ω_0. Let $\tau_s = \tau_{s_1} - \tau_{s_2}$ be the extra time required by a signal photon from crystal 1 (to reach the beam splitter before D_S) relative to a signal photon from crystal 2; similarly, let τ_i be the idler photon propagation time from crystal 1 to crystal 2, *when no delay line is present.* (These times are related to the distances labelled in the figure by $\tau = d/c$.) Further, assume that with no delay line, fringes are observable in the signal singles rate, i.e., the difference in these two times is much less than τ_c, the two-photon correlation time of the down-converted photons. Adding an extra delay τ_A ($>>\tau_c$) to τ_i then makes the two interfering processes distiguishable and singles fringes are not observed. Finally, let τ_{MZ} be the difference between the propagation times through the long and short paths in the Mach-Zehnder. There are four possible ways for the detectors D_S and D_i to register photons: 1) photons from NL1, idler takes short path in Mach-Zehnder (denoted $s_1\ i_{1,S}$); 2) photons from NL1, idler takes long path ($s_1\ i_{1,L}$); 3) photons from NL2, idler takes short path ($s_2\ i_{2,S}$); and 4) photons from NL2, idler takes long path ($s_2\ i_{2,L}$). For arbitrary values of τ_A and τ_{MZ} these four possibilities are in general distinguishable. However, if we consider $|\tau_A - \tau_{MZ}| << \tau_c$, then processes $s_1\ i_{1,S}$ and $s_2\ i_{2,L}$ are indistinguishable. To calculate the probability of coincidence, we sum the amplitudes of these indistinguishable processes, take the absolute square, and add the absolute square

of each of the distinguishable processes:

$$P_c \propto \left| e^{i\omega_o \tau_s} \frac{i}{\sqrt{2}} e^{i\omega_o(\tau_i+\tau_A)} \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} \frac{i}{\sqrt{2}} e^{i\omega_o \tau_{MZ}} \frac{i}{\sqrt{2}} \right|^2$$

$$+ \left| e^{i\omega_o \tau_s} \frac{i}{\sqrt{2}} e^{i\omega_o(\tau_i+\tau_A)} \frac{i}{\sqrt{2}} e^{i\omega_o \tau_{MZ}} \frac{i}{\sqrt{2}} \right|^2 + \left| \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}} \right|^2$$

$$= \frac{1}{2} + \frac{1}{4} \sin\left(\omega_o(\tau_s+\tau_i+\tau_A-\tau_{MZ})\right), \qquad (6.1)$$

where we have included coefficients of $1/\sqrt{2}$ ($i/\sqrt{2}$) for transmission (reflection) at each of the beamsplitters. (In the calculation in Appendix B2, we treat a much more general case, in which the extra path length between the two crystals comprises the long arm of an unbalanced Mach-Zehnder interferometer whose beam splitters may have arbitrary reflection and transmission amplitudes. There results a very complicated set of possibilities, with a rich set of results.) Implicit in (6.1) is the additional condition that our detection scheme could not exclude the contributions of non-interfering counts arising from the in-principle distinguishable $s_1\, i_{1,L}$ and $s_2\, i_{2,S}$ processes. This will be the case if we have slow detectors, and the visibility of the coincidence fringes will be limited to 50%. However, if the electronic resolution time ΔT of the coincidence equipment is less than τ_{MZ}, the visibility can be as high as 100% (see Appendix B2 for a calculation that incorporates such time-dependent effects). We recently observed this same sort of detector-dependent effect in connection with a Franson experiment to violate a Bell's inequality based on energy and time [Kwiat et al., 1993]; see also Chap. 8.

As an aside, yet another way to restore indistinguishability to the idler photons in this delay-line configuration is simply to use a narrow-band interference filter (instead of a Mach-Zehnder) before D_i. If the resulting coherence length is greater than d_A, then

interference will be restored in the coincidence rate. Correlating the signal photons with the idler photons transmitted through the interference filter will recover the fringes; correlating with the idlers *reflected* from the filter (assuming a non-absorbing filter) will yield anti-fringes. This same technique was used by us in a previous experiment, discussed in Chap. 4, to demonstrate the "collapse" of the wavefunction [Kwiat and Chiao, 1991].

The third quantum eraser method we propose is in some sense a hybrid of the other two. Once again, a half waveplate is inserted at position A between the two crystals, in addition to the optical delay line of length d_A (see Fig. 6.3). The polarization of the idler beam from NL1 is rotated so as to be orthogonal to the polarization of the idler beam from NL2. Next, a polarizing beam splitter is used in the unbalanced Mach-Zehnder interferometer. The polarizing beam splitter is oriented so that idler photons from NL1, with the extra optical delay d_A, take the short Mach-Zehnder path, while idler photons from NL2 take the long Mach-Zehnder path. Finally, a second half waveplate inserted at position D in the long path undoes the polarization rotation from the waveplate at A. Thus polarization no longer labels the parent crystal. As before, if the difference in the Mach-Zehnder path lengths is equal to d_A, then the parent crystal of a given idler photon is unknowable, and interference will be observed in coincidence. The advantage of this scheme over the previous one is that the visibility of the interference fringes can be shown to be 100%, *regardless of the speed of the idler detector*, because the use of polarization permits us to force the idler photons to take the appropriate paths in the Mach-Zehnder, eliminating the non-interfering processes.

6.V. Conclusion

The quantum eraser offers an important perspective on interference and loss of quantum coherence in terms of (in)distinguishability of paths. The loss of wave-like

behavior is not necessarily due to the uncertainty principle, but may be due solely to an entanglement of the interfering system with a measuring apparatus (or the entire "environment"). If the coherence of the measuring apparatus is maintained, then interference may be recovered by correlating results of measurements on the original system with results of particular measurements on the measuring apparatus. The state involved in interference is the *total* physical state, which in addition to photon spatial wave functions may include photon polarization, or even distant photons or atoms. Because we may employ spatially-separate entangled systems, a delayed-choice aspect is possible: The decision to measure which-path information or interference may be made after the initial particle is detected.

Proposed experiments using atoms or neutrons, while intellectually engaging in principle, are at best very difficult in practice. Fairly simple experiments are possible using correlated down-conversion photons, and several have been performed. As reported, however, none of these possesses all the attributes of a true quantum eraser. This deficiency is remedied by modification of one of the setups. The simplest extension uses polarization to provide a distinguishable label on the contributing paths; a suitable polarization measurement on the idler photons can then serve to yield *welcher Weg* information or interference fringes, *after* the interfering signal photons have been detected. Instead we may make use of the simultaneity of the down-converted twin photons by introducing a time delay into one of the idler arms. An unbalanced Mach-Zehnder can then be used to reduce the enlarged Hilbert space by projecting onto various linear combinations of the measuring apparatus states (here, the states of the idler photon). If the polarization and delay-line techniques are combined, one may achieve 100% visibility even with slow detectors. All of these schemes may incorporate a delayed-choice feature.

Chapter 7: The Einstein-Podolsky-Rosen "Paradox" and Bell's Inequalities

The Bell paradox and its solution.

The paradox. By firing tachyons you can commit a "perfect murder". Suppose that A purposes killing B, without risking prosecution. When he happens to see B together with a "witness" C, he aims his tachyon pistol at the head of B, until B and C (realizing the danger) start running away with speed, say, u. Then, A chooses to fire tachyonic projectiles T having a speed V such that $uV > c^2$. In the A rest frame, tachyons T reach B soon and are absorbed by B's head, making him die. Due to the fact that $uV > c^2$... however, the witness C--when questioned by the police--will have to declare that actually he only saw anti-tachyons $\overline{T}$ come out of B's head and be finally absorbed by A's pistol. The same would be confirmed by B himself, were he still able to give testimony.

-E. Recami, Riv. Nuovo Cim. **9**, 71 (1986)

A vast time bubble has been projected into the future to the precise moment of the end of the universe. This is, of course, impossible.

-D. Adams, *The Hitchhiker's Guide to the Galaxy*

7.I. Introduction

In this chapter, we discuss the famous *gedanken* experiment of Einstein, Podolsky, and Rosen, and the subsequent work of Bell and others. The chapter "intro-quotes" were chosen for the following reasons. First, the quote regarding the "Bell-paradox" (which, surprisingly, has very little at all to do with either the EPR-paradox or Bell's inequalities), hints at the general problems with nonlocality, and is fairly

indicative of the confusion surrounding this general topic. The second quote, from *The Hitchhiker's Guide to the Galaxy* by Douglas Adams, is very appropriate, for it points out the seeming impossibility of having nonlocal influences, which, if they could be used to signal superluminally, would then open up possibilities for time travel and other paradoxical phenomena. Nevertheless, we will see that nonlocality is in some sense a *cornerstone* of quantum mechanics, one which obeys a tenuous, yet peaceful coexistence with the tenets of special relativity.

After presenting two versions of EPR's argument and two derivations of Bell's inequalities, we will introduce the various "loopholes" that have plagued all experimental tests to date. Next, we will discuss briefly how one might interpret an experimental violation of a Bell's inequality. Finally, the basics of quantum cryptography are presented, as the only *practical application* to date of this entire topic of Bell's inequalities.

7.II. The EPR-"Paradox"

To begin, Einstein, Podolsky, and Rosen did *not* consider their *gedanken* experiment to be paradoxical. They viewed it as indicative of the inherent incompleteness of quantum mechanics. We will present two versions of the argument here: the first is the original of EPR, dealing with a position-momentum entangled state [Einstein et al., 1935]; the second is a somewhat simpler version by Bohm, using a spin-spin correlated state [Bohm, 1983]. Our motivation for including the first version, aside from historical completeness (certainly *not* an overwhelming consideration in this thesis), is that some of the most recent Bell's-inequality experiments have used an entanglement in energy and time, and are therefore quite similar to the originally envisioned scheme [Chap. 8; Brendel et al., 1992; Kwiat et al., 1993].

The three central elements that constitute either version of the EPR argument are 1) a belief in some of the quantum-mechanical predictions concerning two separated

particles, 2) a very reasonable definition of an "element of reality" [namely, that "if, without in any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity"], and 3) a belief that Nature is local (i.e., there are no spook-like actions at a distance, *spukhafte Fernwirkungen*). In the original EPR version [Einstein et al., 1935], the system under consideration is a pair of particles described by the wavefunction $\delta(x_1 - x_2 - a)$. This is an eigenstate of the operator $(x_1 - x_2)$ [with eigenvalue a], and of the operator $(p_1 + p_2)$ [with eigenvalue 0]. In other words, the *sum* of the particles' momenta is well-defined, as is the *difference* in their positions. Therefore, if we measure the momentum of one particle, we can predict with certainty the momentum of the other, even though it may be sufficiently far away that no signal could be transferred between them. Momentum must therefore be an element of reality. And if we measure the position of one, then we can predict with certainty the position of the (possibly distant) other particle. Hence, position is also an element of reality. But QM does not allow a precise specification of both the position and the momentum of a particle. This was not a paradox to EPR; rather, they concluded that QM must be incomplete.

In the modified argument by Bohm (and modified slightly further here), we consider two photons (travelling off back to back) which are described by the entangled polarization singlet-state:

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|H_1, V_2\rangle - |V_1, H_2\rangle\right) , \qquad (7.1)$$

where the letters denote horizontal (H) or vertical (V) polarization, and the subscripts denote which photon. (7.1) is symmetric, in that measurement of any polarization component for one of the particles will yield a count with 50% probability. Moreover,

according to QM, the polarization state of either particle alone is not specified. Nevertheless, if we measure the polarization component of particle 1 along some direction, we can predict with certainty the value of a measurement of the same polarization component for particle 2, seemingly without disturbing it (since it may be far away). Therefore, according to EPR's intuitive definition, we ascribe an element of reality to that component of polarization. The same logic applies for all polarization components, implying that the polarization is well-defined along all component directions simultaneously. Of course, a quantum mechanical state cannot specify that much information, and is consequently an incomplete description, according to the EPR argument.

To put the argument slightly differently, there are basically two ways to explain the strong correlations of two entangled particles. First, one can believe that QM is "the whole story" and that although the particles do not possess well-defined properties before measurement, upon measurement there is a nonlocal influence between the particles (the "collapse", discussed in Chap. 4). The second, and more intuitive explanation (which was implied by EPR, though not explicitly stated as such) is that the particles leave the source with definite, correlated properties; these properties would be determined by some local theory, more complete than QM, by incorporating variables that were "hidden".

At the level of *two* entangled particles, a local hidden variable (LHV) theory to explain correlations is trivial[1] -- an explicit example is given in Sect. 7.VI. Therefore, if one limits the discussion to situations of perfect correlations or anti-correlations (i.e.,

[1] Recent work has shown that it is not possible in some systems involving three or more entangled particles to explain even perfect correlations with a local description [Greenberger, et al., 1989; Mermin, 1990]. In particular, Greenberger, Horne, and Zeilinger (GHZ) have shown an experimental difference between QM and local realism *without* the need for inequalities (assuming perfect apparatus). It is my belief that if the GHZ argument had been shown to EPR, the latter would have been forced to recant.

using analyzers which measure only the same or orthogonal components), the predictions of an LHV theory and QM are identical, and the theories are indistinguishable. Choice of one over the other is a philosophical decision, not a *physical* one. This is how things remained from the introduction of the EPR argument until 1964. It was then that John Bell [Bell, 1964; Bell, 1987] brilliantly discovered that QM gives different *statistical* predictions than does any theory based on local realism, for situations of non-perfect correlations (i.e., analyzers at intermediate angles). In particular, the LHV theories predict that certain combinations of coincidence rates (and possibly singles rates) satisfy an inequality, which quantum mechanical predictions sometimes violate[2].

7.III. Bell's Inequalities, à la Shimony

The derivation of a Bell's inequality that follows is based on that given by Shimony [Shimony, 1990]. It is my personal "favorite" because it is particularly easy to understand the underlying assumptions, the mathematical proof of the inequality, and the role of detector efficiency. We begin with an ensemble of correlated pairs. (See Fig. 7.1.) The particles are described by the state λ, which belongs to a space Λ of complete states[3]; the distribution function for the states λ over Λ is given by $\rho(\lambda)$. We make no

---

[2] Hardy has also presented a clever gedanken experiment using electron-positron annihilation to achieve a contradiction with local realism without the need for inequalities [Hardy, 1992a], and has recently proposed an optical analog which may allow a feasible experimental implementation [Hardy, 1993]. Violations of Bell's inequalities with macroscopic (but non-classical) states of light have been discussed by Munro and Reid [Munro and Reid, 1993] and Franson [Franson, 1993].

[3] We make no restrictions on λ at this point; it may have discrete and/or continuous parts, and may even include the quantum mechanical description ψ. Recall that EPR claimed that QM was incomplete, because it could not specify the results of measurements on non-commuting observables. The state λ is presumed to be a more complete specification of the physical state, such that it allows one to define the probabilities of all possible measurements on the particles. If these probabilities only

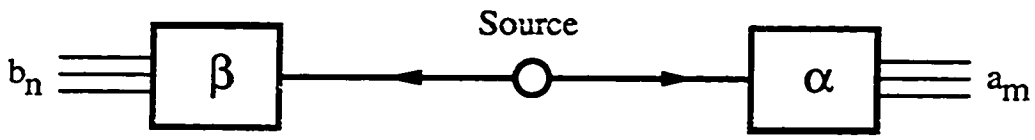


Figure 7.1 Simplified schematic of general setup used to derive Bell's inequalities.

restrictions on the form of $\rho(\lambda)$, other than that it be non-negative and normalized:

$$\int_\Lambda d\lambda\, \rho(\lambda) = 1. \tag{7.2}$$

One member of each pair is directed to analyzer 1 on the right, with a parameter setting α; the other member is directed to analyzer 2 on the left, with a parameter setting β. For example, the parameters may correspond to orientation angles of polarizers, or to phase settings in an interferometer. The possible outcomes of analyzers 1 and 2 are a_m (m = 1, 2, ...) and b_n (n = 1, 2, ...), respectively, where we assume these values to lie in the interval [-1, +1]. For instance, +1 might correspond to a detection in the ordinary-channel of a polarizing beam splitter, -1 to the extraordinary-channel, and 0 to nondetection of the particle. Alternately, ± 1 might correspond to the two output ports of an interferometer. We now assume that when λ and α and β are specified, the following probabilities are well-defined (and positive[4]):

---

assume the values 0 and 1, the LHV is "deterministic".

[4] It has been pointed out [M. O. Scully, private communication] that *negative* probabilities may well be a way "out" of the disturbing conclusions of a violation of Bell's inequalities, in that one need not abandon the intuitive notion of local realism, if one is willing to accept the notion of *negative probability*. While we do not dispute this claim, we find it unhelpful, since we have no understanding at all of what "negative probabilities" would mean.

$p(m \mid \lambda, \alpha, \beta)$	the probability that the outcome a_m is obtained, given the state λ, and the parameter settings α and β.
$p(n \mid \lambda, \alpha, \beta)$	the probability that the outcome b_n is obtained, given the state λ, and the parameter settings α and β.
$p(m \mid \lambda, \alpha, \beta, n)$	the probability that the outcome a_m is obtained (for analyzer 1), given the state λ, the parameter settings α and β, and the outcome b_n (for analyzer 2)
$p(n \mid \lambda, \alpha, \beta, m)$	the probability that the outcome b_n is obtained (for analyzer 2), given the state λ, the parameter settings α and β, and the outcome a_m (for analyzer 1)
$p(m, n \mid \lambda, \alpha, \beta)$	the joint probability of the outcomes a_m and b_n, given the state λ and the parameter settings α and β

From standard (Bayesian) probability theory, it follows that

$$p(m, n \mid \lambda, \alpha, \beta) = p(n \mid \lambda, \alpha, \beta, m)\, p(m \mid \lambda, \alpha, \beta)$$
$$= p(m \mid \lambda, \alpha, \beta, n)\, p(n \mid \lambda, \alpha, \beta)\,. \qquad (7.3)$$

Until now, the only physics that has gone into our description is the assumption of an underlying "realism", described by the complete state λ. We now incorporate the assumptions of locality. There are two separate independence conditions[5] which

[5] Two essentially identical conditions have been used by Jarrett [Jarrett, 1984].

constitute what is generally known as the "Bell locality condition":

Parameter Independence: $p(m \mid \lambda, \alpha, \beta) = p(m \mid \lambda, \alpha)$ and $p(n \mid \lambda, \alpha, \beta) = p(n \mid \lambda, \beta)$

Outcome Independence: $p(m \mid \lambda, \alpha, \beta, n) = p(m \mid \lambda, \alpha, \beta)$

and $p(n \mid \lambda, \alpha, \beta, m) = p(n \mid \lambda, \alpha, \beta)$

(We shall in Sect. 7.VI discuss the content of these conditions.) The conditions may be combined with the Bayesian relation (7.3) to yield the locality condition:

$$p(m, n \mid \lambda, \alpha, \beta) = p(m \mid \lambda, \alpha)\, p(n \mid \lambda, \beta)\,. \qquad (7.4a)$$

The salient feature of (7.4a) is that the joint probability for the m,n-outcome *factorizes*: we shall see below that this is not the case for QM.

The next step in the Bell's inequality derivation is to define expectation values in terms of the measurement outcomes and probabilities:

$$E(\lambda, \alpha) = \sum_m p(m \mid \lambda, \alpha)\, a_m$$ (the expectation value of the outcome of analyzer 1)

$$E(\lambda, \beta) = \sum_n p(n \mid \lambda, \beta)\, b_n$$ (the expectation value of the outcome of analyzer 2)

$$E(\lambda, \alpha, \beta) = \sum_{m,n} p(m,n \mid \lambda, \alpha, \beta)\, a_m b_n$$ (the expectation value of the product of the outcomes of analyzers 1 and 2)

Combining these and (7.4a), one finds

$$E(\lambda, \alpha, \beta) = E(\lambda, \alpha)\, E(\lambda, \beta). \qquad (7.4b)$$

Next we prove the following simple but very useful lemma:

If x_1, y_1, x_2, and y_2 all lie in the interval [-1, 1], then $S = x_1y_1 + x_1y_2 + x_2y_1 - x_2y_2$ lies in the interval [-2, 2].

Proof: Since S is linear in each of its arguments, it must take on its extreme values when the arguments themselves are extreme, i.e., when $(x_1, y_1, x_2, y_2) = (\pm 1, \pm 1, \pm 1, \pm 1)$. Clearly, under these conditions, S must be an integer between -4 and 4. But we can also rewrite S as $(x_1 + x_2)(y_1 + y_2) - 2\, x_2\, y_2$. Since $(x_1 + x_2)$ and $(y_1 + y_2)$ can only equal 0 or ± 2, and $2\, x_2\, y_2$ can only equal ± 2, the value of S can only be ± 2 or ± 6 at the extrema of the arguments. Therefore the extreme values of S are ± 2. Q. E. D.

To finish our task, we associate x_i with $E(\lambda, \alpha_i)$ and y_i with $E(\lambda, \beta_i)$, which obviously satisfy the condition $E(\lambda, \alpha_i), E(\lambda, \beta_i) \in [-1, 1]$. Then the lemma becomes:

$$-2 \le E(\lambda,\alpha_1,\beta_1) + E(\lambda,\alpha_1,\beta_2) + E(\lambda,\alpha_2,\beta_1) - E(\lambda,\alpha_2,\beta_2) \le 2. \qquad (7.5)$$

Finally, by integrating this result over the space Λ, weighting by $\rho(\lambda)$, and defining the ensemble expectation value as

$$E(\alpha,\beta) \equiv \int_\Lambda d\lambda\, \rho(\lambda)\, E(\lambda,\alpha,\beta)\,, \qquad (7.6)$$

we arrive at a Bell's inequality:

$$|S| \leq 2, \tag{7.7a}$$

where

$$S = E(\alpha_1,\beta_1) + E(\alpha_1,\beta_2) + E(\alpha_2,\beta_1) - E(\alpha_2,\beta_2) \ , \tag{7.7b}$$

and we have used the normalization condition (7.2) to write the right-hand side of (7.7a).

In arriving at the inequality, we have implicity included one extra assumption, that the parameter settings α and β do not affect the emission properties of the source. We assumed this when we used the same distribution $\rho(\lambda)$ to integrate over each of the terms of (7.5). In any true test of Bell's inequalities, one must experimentally ensure that the source and analyzers are sufficiently separated that this assumption of "ρ-independence" is valid (assuming no information transfer faster than c). In practice, if the analyzers are on opposite sides of the source [6], this separation will also guarantee that there is no causal link between the detectors.

The remarkable fact, of course, is that for certain states, QM predicts a value of S in violation of (7.7a). In particular, for the singlet-like polarization state (7.1), the probability of joint detections (assuming for the moment an ideal, lossless system) at the ordinary-channels of the two analyzers is (see Appendix D3):

$$p_{oo}(\alpha,\beta) = \frac{\sin^2(\beta-\alpha)}{2} \ . \tag{7.8a}$$

Similarly, the probability of joint detections at the extraordinary-channels is also

[6] It is sufficient that the source and analyzers lie on the points of an equilateral triangle.

$$p_{ee}(\alpha,\beta) = \frac{\sin^2(\beta-\alpha)}{2} \quad . \tag{7.8b}$$

Finally the probability of joint detections at different channels is

$$p_{oe}(\alpha,\beta) = p_{eo}(\alpha,\beta) = \frac{\cos^2(\beta-\alpha)}{2} \quad . \tag{7.8c}$$

Combining these various coincidence results, the correlation function $E(\alpha,\beta)$ has the value $-\cos 2(\beta-\alpha)$. A violation of (7.7a) may then be achieved by choosing the parameter settings as shown in Fig. 7.2[7] (e.g., $\alpha_1 = 0°$, $\alpha_2 = 45°$, $\beta_1 = 22.5°$, and $\beta_2 = -22.5°$), for which S has the value -2.83 $(= -2\sqrt{2})$. In fact, it has been shown that $\pm 2\sqrt{2}$ are the extreme values of S, according to QM [Su and Wódkiewicz, 1990]. Note that the above coincidence rates all assumed a fringe visibility of 100%. If, for experimental reasons, the visibility V is less than this, then the maximum value of S will

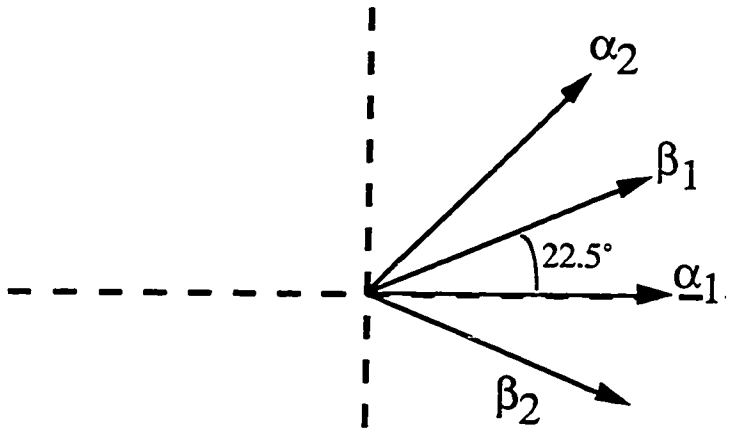


Figure 7.2 One set of polarizer angles which will yield an optimal violation of the Bell's inequality (7.7).

[7] From the symmetric nature of the singlet-state (7.2), any set of angles with the same *relative* spacing will also produce a maximum violation.

be accordingly reduced: $S_{max}(V) = V2\sqrt{2}$. Therefore, in order to violate the Bell's inequality (7.7a), one must have $V \geq 71\%$. This is a necessary, but not sufficient condition; it is also required that the fringes behave sinusoidally (in Sect. 7.VI we will examine a model which yields 100%-visibility *triangular* coincidence fringes.)

In the above coincidence probabilites (7.8), we omitted the effects of non-unity detection efficiency. Specifically, if each detector[8] has an efficiency of η, then the probabilities are reduced by η^2, as is $E(\alpha,\beta)$. Therefore, the maximum value of S is $\eta^2 2\sqrt{2}$, so to achieve a violation of $S \leq 0$, we need $\eta^2 > 1/\sqrt{2}$, or $\eta > 84\%$. With a different Bell's inequality, and a much more complicated argument, Mermin has shown that the requisite efficiency threshold is ~83% $[= 2(\sqrt{2} - 1)]$ [Mermin, 1986]. More recently, Eberhard has shown that the minimum efficiency may be as low as ~67%, in the limit of no background [Eberhard, 1993]. We will discuss the last two results below.

7.IV. Bell's Inequalities, à la Eberhard

We now present a different derivation of another Bell's inequality. This derivation is loosely based on an argument most recently used by Eberhard [Eberhard, 1993]; see also Sections 3.7 and 3.8 of [Clauser and Shimony, 1978]. In order to make the argument more clear, we shall initially make two simplifying assumptions: that the system under consideration is lossless, and that the underlying LHV theory is deterministic. (These assumptions can be relaxed after we understand the basic argumentation.) We have in mind again a system similar to that shown in Fig. 7.1; we will assume that there are only two possible measurement outcomes from an analyzer[9], corresponding to detection in the ordinary-channel (o) and detection in the

[8] For simplicity, we "lump" all losses of our correlated particles into the effective efficiency of the detectors. In experiments performed with gamma-ray photons (produced from positronium annihilation, for instance) the actual detector efficiencies are very high, but available polarizers are very poor [Clauser and Shimony, 1978].

extraordinary-channel (e) (the possibility of non-detection will be included later). In particular, we consider four different experimental setups, defined by two values of the parameter α and two values of the parameter β. By our notion of (deterministic) reality, every particle leaving the source must know what it will do at the analyzer for both choices of parameter setting (although it needn't know what the other particle of the pair will do). I find it helpful to imagine that each particle is given an "instruction list" at the source, detailing which channel of the analyzer it will take for one setting of the parameter, and which it will take for the other. Each *pair* of particles can then be characterized by a four-instruction list, which completely determines what each particle would do for the two settings of its parameter[10]. For example, the general list is

1. If $\alpha = \alpha_1$, then particle 1 take the $?_1$-channel;
2. If $\alpha = \alpha_2$, then particle 1 take the $?_2$-channel;
3. If $\beta = \beta_1$, then particle 2 take the $?_3$-channel;
4. If $\beta = \beta_2$, then particle 2 take the $?_4$-channel,

where $?_i$ (i = 1 - 4) is either the o- or the e-channel. We use the shorthand notation "$?_1?_2?_3?_4$" (e.g., "eeoe", "eooe", etc.) to represent a given list. There are a total of 16 ($= 2^4$) different lists. Locality is implicit in that the outcome of particle 1 depends only on the parameter α, and the outcome of particle 2 depends only on the parameter β.

Next we consider a large number of pairs, each with its own instruction list. We

---

[9] We choose the particular example of polarization correlations; identical arguments apply for an interferometric system, where the two outcomes correspond to the two output ports of an interferometer.

[10] Note: All reference to a deterministic model may be removed if one uses the terminology "what the particle *did*" instead of "what the particle *would do*". For simplicity, we will keep to the latter usage. In the former case, there are controversies involving assumptions of contrafactual definiteness (see the discussion in [Clauser and Shimony, 1978], for example); we will not discuss this controversy further here, except to say that there are other, non-controversial, ways to derive this Bell's inequality.

seek to prove the following relation:

$$n_{oo}(\alpha_2, \beta_2) + n_{oe}(\alpha_1, \beta_2) + n_{eo}(\alpha_2, \beta_1) - n_{oo}(\alpha_1, \beta_1) \geq 0 \,, \qquad (7.9a)$$

where $n_{oo}(\alpha_2, \beta_2)$ is the *number* of pairs with the instruction set "?o?o" (i.e., those pairs whose 2nd and 4th instructions are for "o-channel" outcomes; the 1st and 3rd instuctions can be anything); $n_{oe}(\alpha_1, \beta_2)$ is the number with the instruction set "o??e"; $n_{eo}(\alpha_2, \beta_1)$ is the number with the instruction set "?eo?"; and $n_{oo}(\alpha_1, \beta_1)$ is the number with the instruction set "o?o?". We can demonstrate (7.9a) explicitly by expanding each of the constituent terms as follows:

$$n_{oo}(\alpha_2, \beta_2) \equiv n_{?o?o} = n_{oooo} + n_{ooeo} + n_{eooo} + n_{eoeo} \qquad (7.10a)$$

$$n_{oe}(\alpha_1, \beta_2) \equiv n_{o??e} = n_{oooe} + n_{ooee} + n_{oeoe} + n_{oeee} \qquad (7.10b)$$

$$n_{eo}(\alpha_2, \beta_1) \equiv n_{?eo?} = n_{oeoo} + n_{oeoe} + n_{eeoo} + n_{eeoe} \qquad (7.10c)$$

$$n_{oo}(\alpha_1, \beta_1) \equiv n_{o?o?} = n_{oooo} + n_{oooe} + n_{oeoo} + n_{oeoe} \qquad (7.10d)$$

One can immediately check that all the terms of (7.10d) are contained within (7.10a-c), proving (7.9a). If one includes the third outcome possibility that a given particle may be undetected (u), then (7.9a) becomes

$$n_{oo}(\alpha_2, \beta_2) + n_{oe}(\alpha_1, \beta_2) + n_{ou}(\alpha_1, \beta_2)$$

$$+ n_{eo}(\alpha_2, \beta_1) + n_{uo}(\alpha_2, \beta_1) - n_{oo}(\alpha_1, \beta_1) \geq 0 \,, \qquad (7.9b)$$

as shown in [Eberhard, 1993]. This inequality can be connected to real counting rates if one makes the identifications

$$n_o^{\alpha}(\alpha_1) = n_o^{\alpha}(\alpha_1, \beta_2) \equiv n_{oo}(\alpha_1, \beta_2) + n_{oe}(\alpha_1, \beta_2) + n_{ou}(\alpha_1, \beta_2) \qquad (7.11a)$$

$$n_o^{\beta}(\beta_1) = n_o^{\beta}(\alpha_2, \beta_1) \equiv n_{oo}(\alpha_2, \beta_1) + n_{eo}(\alpha_2, \beta_1) + n_{uo}(\alpha_2, \beta_1) \,, \qquad (7.11b)$$

where the first equalities in (7.11a,b) are due to the locality condition. Making these substitutions in (7.9b), we finally have

$$n_{oo}(\alpha_1, \beta_1) + n_{oo}(\alpha_1, \beta_2) + n_{oo}(\alpha_2, \beta_1) - n_{oo}(\alpha_2, \beta_2) \leq n_o^{\alpha}(\alpha_1) + n_o^{\beta}(\beta_1) \,, \quad (7.12)$$

a Bell's inequality first derived by Clauser and Horne [Clauser and Horne, 1974]. One salient feature is that it contains actual *numbers* of counts (or equivalently, counting *rates*), instead of probabilities. Therefore, we do not need to concern ourselves with *pairs* we do not detect. Of course, inefficiencies which cause us to detect only one member of a pair will be detrimental -- if the detection efficiency is small, then the coincidence-terms on the left-hand side of (7.12) will always be much smaller than the singles-terms on the right-hand side, and no violation of the Bell's inequality will be possible.

From the previous section [cf. Eqs. (7.8)], the quantum-mechanical predictions for the coincidence counts, assuming particles correlated in a singlet-state (7.1), are given by

$$n_{oo}(\alpha, \beta) = \eta^2 N_o \frac{\sin^2(\beta - \alpha)}{2} \,, \qquad (7.13a)$$

(N_o is the total number of emitted pairs; it will cancel out) while the singles counts (assuming no background) are easily calculated to be

$$n_o^{\alpha}(\alpha) = n_o^{\beta}(\beta) = \eta \frac{N_o}{2} \quad . \tag{7.13c}$$

Inequality (7.12) then becomes

$$\eta^2 \left\{ \sin^2(\beta_1 - \alpha_1) + \sin^2(\beta_2 - \alpha_1) + \sin^2(\beta_1 - \alpha_2) - \sin^2(\beta_2 - \alpha_2) \right\} \leq 2\eta \quad . \tag{7.14}$$

The maximum value of the term in { }'s is $1 + \sqrt{2}$ (achieved, for example, using $\alpha_1 = 0°$, $\alpha_2 = 135°$, $\beta_1 = 67.5°$, and $\beta_2 = -67.5°$) so that an efficiency greater than $2/(1 + \sqrt{2}) \approx 83\%$ is needed to observe a direct violation of this Bell's inequality.

It has been shown that by using a non-maximally entangled state (i.e., one where the magnitudes of the probability amplitudes of the contributing terms are not equal), one may reduce the detector requirement to ~67% [Eberhard, 1993]. The basic idea is that by making one term of (7.1) have a greater amplitude than the other, one effectively *polarizes* the source. For example, if one uses the state

$$|\psi\rangle = \frac{1}{\sqrt{1 + f^2}} \left(|H_1, V_2\rangle - f\,|V_1, H_2\rangle \right), \tag{7.15}$$

where f is a real number ≤ 1, then photon #1, travelling to the α-analyzer, possesses a net horizontal polarization, while #2, travelling to the β-analyzer, appears somewhat vertically-polarized. By choosing α_1 nearly vertical, and β_1 nearly horizontal, one may reduce the contributions of the singles rates $n_o^{\alpha}(\alpha_1)$ and $n_o^{\beta}(\beta_1)$ to the right-hand side of (7.12). Nevertheless, by correctly choosing α_2 and β_2, one may still violate (7.12), even for η as low as 67%. We shall return to these issues in Sect. 10.IV.

7.V. Loopholes

While experimental tests of Bell's inequalities have been extended to new systems, some relying on energy-time or phase-momentum entanglement [Brendel et al., 1992; Kwiat et al., 1993c; Franson, 1989; Rarity and Tapster, 1990]. to date no true violation of Bell's inequalities has been observed. All experiments thus far have required supplementary assumptions (in addition to that of local realism), which although seemingly reasonable, severely reduce the true impact such an experiment might yield. These additional assumptions constitute several loopholes, which can be divided into three general categories: the angular correlation loophole, the detection loophole, and the space-like separation loophole. We propose in Chapter 10 a setup which should permit for the first time (simultaneous) closure of the first two of these loopholes; we will also discuss briefly how current technologies should allow an extension to close the third as well.

The angular correlation loophole was investigated by Clauser and Horne [Clauser and Horne, 1974], and in detail by Santos [Santos, 1991; Santos, 1992]. They showed that because of the cosine-squared angular correlation of the directions of photons emitted in an atomic cascade (of the type used in nearly all of the early Bell's-inequalities experiments [Clauser and Shimony, 1978; Aspect et al., 1982a,b]), there was an inherent polarization decorrelation, due to the transversality condition. More explicitly, since the photons do not necessarily fly off back-to-back in this three-body decay process, one must detect the photons emitted into a large solid angle in order to have a sufficiently high detection efficiency (see discussion below). But with this source the very polarization correlation which could result in a violation of Bell's inequalities is reduced for non-collinear photons, so that it is strictly impossible to disprove local realism by using a cascade source.

In order to essentially remove the angular-correlation problem, experimenters have switched from cascade sources to those using correlated photons produced in the process of spontaneous parametric down-conversion [Brendel et al., 1992; Kwiat et al., 1993c; Shih and Alley, 1988; Ou and Mandel, 1988; Rarity et al., 1987]. As discussed in Chapter 2, these photons can have an angular correlation of better than 1 mrad, although in general they need not be collinear [11]. The simplest of the down-conversion Bell's-inequality experiments [Shih and Alley, 1988; Ou and Mandel, 1988] used the same setup discussed in Chapter 5 as a quantum eraser (see Fig. 5.1) -- non-collinear correlated photons were directed through equal path lengths to opposite sides of a 50-50 beam splitter, aligned so that the transmitted mode of one photon coincided with the reflected mode of the conjugate photon, and vice versa. A half waveplate prior to the beam splitter was used to rotate the polarization of one of the photons (which were initially horizontally-polarized) by 90°. The output state of this "source" (including the down-conversion crystal, waveplate and beam splitter) was then

$$|\psi\rangle = \frac{1}{2}\left[|H\rangle_3 + i|H\rangle_4\right]\left[i|V\rangle_3 + |V\rangle_4\right]$$

$$= \frac{1}{2}\left[|H\rangle_3|V\rangle_4 - |V\rangle_3|H\rangle_4 + i|V\rangle_3|H\rangle_3 + i|V\rangle_4|H\rangle_4\right], \quad (7.16)$$

where the subscripts 3 and 4 denote the two output port modes of the beam splitter, and $|H\rangle_j$ ($|V\rangle_j$) denotes a single photon in mode j, horizontally (vertically) polarized. Coincidence rates between detectors looking at the two output ports were recorded, as a function of the orientation of polarizers at the detectors. In only measuring coincidence rates, the experimenters were able to effectively create a singlet-like state by discarding

[11] In contrast to the cascade situation, the polarizations remain well-defined even for non-collinear down-converted pairs. One consequence is that extra means must be used to prepare a polarization-entangled state. (Due to energy conservation, the photons are automatically produced in an energy-entangled state. See also Chapter 8.)

the last two terms of (7.16). However, it should be stressed that because of these discarded terms, the detection efficiency is inherently limited to 50%, and no indisputable test of Bell's inequalities is possible[12]. Similar problems arise in the Bell's-inequality experiments based on energy-time entanglement[13] (see Chapter 8) and phase-momentum entanglement [Rarity and Tapster, 1990a][14].

Low detection efficiencies have hindered experiments on Bell's inequalites from the outset. The detection loophole basically deals with the fact that with non-unity efficiency detectors (as before, we include all losses of our interfering particles in the effective efficiency of the detectors), only a fraction of the emitted correlated pairs is detected. If the efficiency is sufficiently low, then it is possible for the subensemble of detected pairs to give results in agreement with quantum mechanics, even though the *entire* ensemble satisfies Bell's inequalities.

Due to the non-existence of adequate detectors, experiments have so far employed an additional assumption, equivalent to the fair-sampling assumption that the fraction of detected pairs is representative of the entire ensemble, or to the no-enhancement assumption that insertion of a polarizer will not increase the probability of detection

---

[12] Hardy has pointed out that one could use the full state (7.16) if one's detectors could reliably distinguish between one- and two-photon states (e.g., if a two-photon detection resulted in a pulse twice as high as a single-photon detection), thereby allowing one to account for the cases in which both photons went the same way [private communication]. Although there is some indication that such detectors may be feasible/available [Petroff et al., 1987; Kwiat et al., 1993a,b], we feel that the non-collinearity of the beams will lead to reduced visibility, unless one uses small pre-detector irises. These, however, reduce the effective collection efficiency.

[13] In these tests as well, half of the counts are discarded (electronically) to eliminate a non-interfering background [Brendel, et al., 1992; Kwiat, et al., 1993c].

[14] Although in this scheme there is in principle no necessity of discarding half of the counts, in practice the small irises needed for high-visibility fringes severely reduce the effective detection efficiency.

[Clauser and Horne, 1974]. For example, to connect the probabilities [that go into the correlation functions $E(\theta_1,\theta_2)$, as in Sect. 7.II] to experimentally-observed rates in a scheme using two-channel analyzers (cf. the second Aspect experiment [Aspect et al., 1982a]), ideally one would define

$$p_{ij}(\theta_1,\theta_2) = \frac{R_{ij}(\theta_1,\theta_2)}{R_o}\,, \tag{7.17a}$$

where $p_{ij}(\theta_1,\theta_2)$ is the probability that the two particles in a given correlated pair will be detected at the "i-th" and "j-th" channels [extraordinary (e) and ordinary (o) channels] in each of the respective analyzers; $R_{ij}(\theta_1,\theta_2)$ is the experimentally-observed rate of coincidences; and R_o is the total rate of *emitted* pairs. But since the detection efficiencies are often small, the maximum value of $p_{ij}(\theta_1,\theta_2)$ as defined above would be close to zero, as would the value of S: No violation of $|S| \leq 2$ would be possible. To avoid this problem, an extra assumption is included so that one may relate the probability of coincidence counts to the number of pairs *detected* at any of the channels:

$$p_{ij}(\theta_1,\theta_2) = \frac{R_{ij}(\theta_1,\theta_2)}{R_{oo}(\theta_1,\theta_2) + R_{oe}(\theta_1,\theta_2) + R_{eo}(\theta_1,\theta_2) + R_{ee}(\theta_1,\theta_2)}\,, \tag{7.17b}$$

where the subscipts "o" and "e" denote the two channels of each analyzer. Essentially, one assumes that the pairs detected are representative (i.e., a "fair sample") of the entire ensemble. A more precise way to state the extra assumption mandated by (7.17b) is as follows:

> For every photon in the state λ, the sum of the detection probabilites in the "ordinary" and in the "extraordinary" beams emerging from a two-way polarizer does not depend on the polarizer's orientation. [Garuccio and Rapisarda, 1981]

However, it should be noted that physicists have invented local, realistic models which can explain all of the observed results--these models necessarily violate the fair-sampling or no-enhancement type assumptions [Clauser and Horne, 1974; Santos, 1992]. In order to experimentally close this loophole, one must have detectors with sufficiently high single-photon detection efficiencies. As discussed above, it was formerly believed that ~83% ($= 2\sqrt{2}-2$) was the lower efficiency limit, but Eberhard has reduced the detector requirement to ~67%, in the limit of no background [Eberhard, 1993][15]. Recently, we have measured the absolute single-photon detection efficiency of several detectors [Kwiat et al., 1993a; Kwiat et al., 1993b], and observed corrected efficiencies as high as 75%. (See a full description in Chapter 9.) However, there were additional losses (not corrected for in the above result) which should in principle be avoidable, leading us to believe that efficiencies in excess of 90% may be feasible; efforts toward this end are currently in progress. Combined with our proposal for a down-conversion source in which there is no need to reject half the counts (presented in Chapter 10), this may permit closure of the detection and the angular correlation loopholes simultaneously.

The final loophole concerns the space-like separation of the different parts of the experiment. Clearly, no claims about nonlocality can be made if the pre-detector analyzers are varied so slowly that a signal traveling at the speed of light could carry the analyzer-setting information back to the source or to the other analyzer before a pair was produced or detected[16]. To close this loophole, the analyzers' settings should be

[15] Taking a completely different approach, Braunstein and Mann [Braunstein and Mann, 1993] have also reduced the required efficiency by considering states of more than two correlated particles, a generalization of the three-particle example of Greenberger, Horne, and Zeilinger [Greenberger et al., 1989]. As the number of particles in the entangled state becomes large, the efficiency requirement can be reduced to 71%, also in the absence of noise.

changed *after* the correlated pair has left the source. (We assume here that the separation of the detectors is as least as great as the source-detector separation.)[17] Ideally, one would desire that the backward light cones of the detection apparatuses and the particle source not overlap, lest it be argued that the apparatus settings share some common cause. Practically, "this is, of course, impossible", and the best one could achieve would be to base the analyzer settings on random signals from astronomical objects outside each other's light cones (e.g., quasars on opposite sides of the universe)[18]. Only one Bell-type experiment, that of Aspect *et al.* [Aspect et al., 1982b], has made any attempt at all to address this locality condition, but even in that experiment the loophole remains. Although the experiment used rapidly-varying analyzers, the variation was not random, and it has been argued that the time of the polarization switching was not sufficient to disprove a causal connection between the analyzer and the source [Franson, 1985; Zeilinger, 1986]. Moreover, the inequality used by Aspect included coincidence rates when the polarizers were removed. This removal was certainly not performed with any alacrity.

---

[16] Nevertheless, researchers continue to make claims about nonlocality, even though no experiment has satisfied the requirement. The basic assumption is that no information other than that carried by the entangled particles is being shared between detectors (i.e., there is no hidden mechanism to produce correlations in the outputs of the detectors). Although correct according to quantum mechanics, the assumption is almost *circular*: In order to "prove" that Nature is nonlocal--that there are influences at a distance--one assumes that there is *no* influence between the detectors!

[17] Actually, this leaves a factor of two safety margin; the real constraint is that the information of analyzer setting should not reach the source before the particles are emitted.

[18] This compromise would rule out all local theories short of "superdeterminism," which amounts to the claim that there is no nonlocality since all events are predetermined, but they are predetermined to precisely mimic a nonlocal theory. Such a metaphysical viewpoint is clearly outside the scope of science

7.VI. Interpretation

The philosophical interpretation of experiments violating Bell's inequalities is a topic about which (far too) many papers have been written[19] (see, for instance, almost any issue of *Foundations of Physics*, or *Physics Essays*). However, we would be remiss if we completely omitted all discussion of the *implications* of such a violation, particularly because physical intuition about nonlocally-entangled states is elusive. There are two basic approaches to understanding Bell's inequality-violating results: First, one may consider what sorts of LHV models are disproved by them[20]; second, one may examine the common assumptions that comprise *any* LHV theory. We address each of these general approaches.

As mentioned earlier, one can easily concoct a local hidden variable (LHV) model to explain correlated behavior of separated particles (assuming they were not always separated, of course). For example, the simplest such source would produce pairs of particles with definite, orthogonal linear polarizations. To simulate an unpolarized appearance (i.e., the singles rate at any detector is independent of the angle of the polarization analyzer), we allow different pairs in the ensemble to have random (but *definite*) axes of polarization. We can describe such a source quantum-mechanically by using a mixture of product states, rather than a pure superposition state as in (7.2):

$$\rho = \int_0^{2\pi} \frac{d\lambda}{2\pi} \, |\lambda_1, (\lambda + \pi/2)_2\rangle\langle\lambda_1, (\lambda + \pi/2)_2| \quad . \tag{7.18}$$

[19] The opinions expressed herein are solely my own.

[20] It has been shown that one can derive a Bell's inequality without reference to hidden variables [Stapp, 1971; Stapp, 1991] at all. In some sense, the second derivation above (Sect. 7.IV) is of this type, since nowhere did we need to assume any sort of underlying model. All we assumed was that the measurements would have had a particular outcome for two parameter settings (even though only one can occur in reality); this is essentially an assumption of "counterfactual definiteness". See, for instance, Sect. 3.8 of [Clauser and Shimony, 1978].

This describes an ensemble of particle pairs; the members of each pair are orthogonally linearly-polarized; the polarization angle λ (with respect to horizontal polarization) is uniformly distributed from 0 to 2π over different members of the ensemble. One can easily calculate the result expected from an analyzer (at an angle θ with respect to the horizontal) looking at photon 1. We describe the polarizer by a projection operator $A_1(\theta) = |\theta_1\rangle\langle\theta_1|$. Expanding $|\lambda_1\rangle$ and $|\theta_1\rangle$ into the horizontal-vertical basis, the probability is then given by

$$P_1(\theta) = \mathrm{Tr}\,\rho A_1(\theta) = \int_0^{2\pi} \frac{d\lambda}{2\pi} \left(\cos\lambda \langle H_1| + \sin\lambda \langle V_1|\right) \langle(\lambda + \pi/2)_2|$$

$$\times \left(\cos\theta\, |H_1\rangle + \sin\theta\, |V_1\rangle\right)\left(\cos\theta \langle H_1| + \sin\theta \langle V_1|\right)$$

$$\times \left(\cos\lambda\, |H_1\rangle + \sin\lambda\, |V_1\rangle\right) |(\lambda + \pi/2)_2\rangle$$

$$= \int_0^{2\pi} \frac{d\lambda}{2\pi} \left(\cos\lambda\cos\theta + \sin\lambda\sin\theta\right)^2$$

$$= \int_0^{2\pi} \frac{d\lambda}{2\pi} \cos^2(\lambda - \theta) = \frac{1}{2}\ . \qquad (7.19)$$

There is no angular dependence to the singles rate. Similarly, one can look at the probability $P_{1,2}(\theta_1, \theta_2)$ of coincidences between two analyzers, one with a polarizer at θ_1, the other with a polarizer at θ_2. The calculation proceeds as above, with the result

$$P_{1,2}(\theta_1, \theta_2) = \int_0^{2\pi} \frac{d\lambda}{2\pi} \cos^2(\lambda - \theta_1) \cos^2\left(\lambda + \frac{\pi}{2} - \theta_2\right) \qquad (7.20a)$$

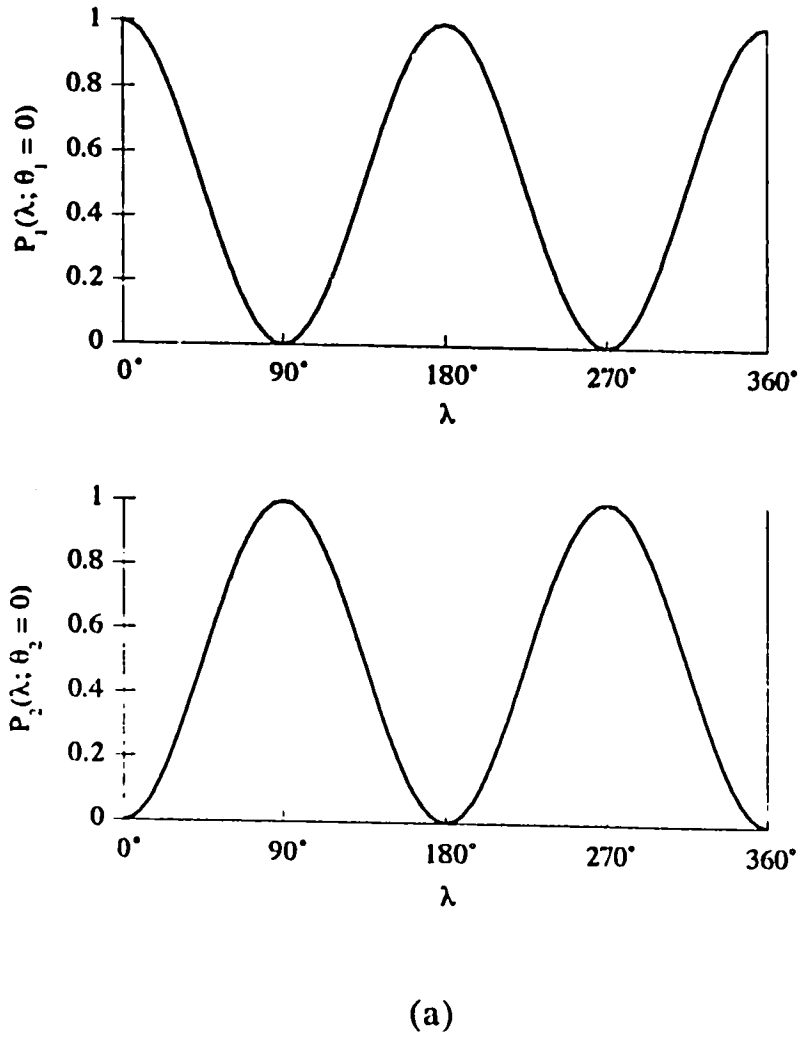
$$= \frac{1}{4}\left(1 - \frac{\cos 2(\theta_1 - \theta_2)}{2}\right) . \qquad (7.20b)$$

This displays 50%-visibility sinusoidal fringes (recall that the visibility must be less than 71% in any LHV model). From (7.20a), we see that the coincidence probability is just the *product* of the individual probabilities (as given by Malus' law) to pass either polarizer, averaged over λ [cf. Eq. (7.4a)]. In Fig. 7.3a, we indicate for particular choices of θ_1 and θ_2 the singles probabilities (as predicted by standard QM) as functions of λ. When these curves are integrated over λ, a constant value of 1/2 results [cf. Eq. (7.19)]. To get the coincidence probability as a function of θ_1 and θ_2, we essentially convolve the two singles probabilities; the resulting fringes [Eq. (7.20b)] have 50%-visibility (see Fig. 7.3b).

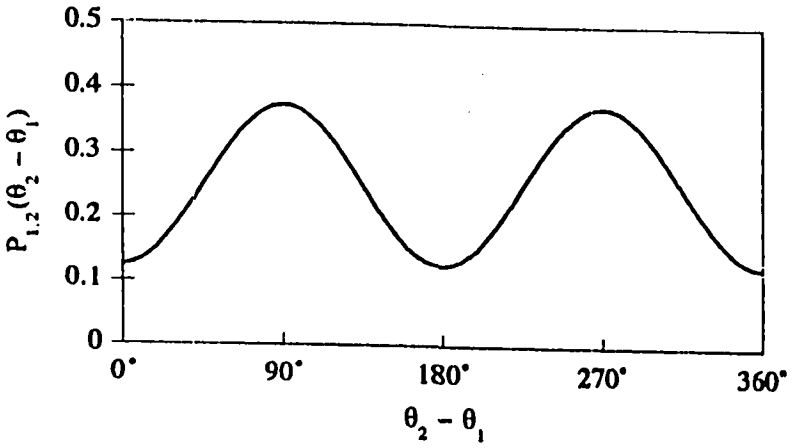
In a modification to this LHV model, we could instead postulate that the probability for a λ-polarized photon to pass a polarizer (at angle θ) is as follows:

$$\begin{aligned} P(\theta \mid \lambda) &= 1 \text{ if } \cos^2(\theta - \lambda) > 1/2 \\ &= 0 \text{ if } \cos^2(\theta - \lambda) \le 1/2 \ . \end{aligned} \qquad (7.21)$$

Plots of (7.21) for particular choices of θ_1 and θ_2 are shown in Fig. 7.4a. As before, integrating over λ yields a constant; the source appears unpolarized. Convoluting the two plots of Fig. 7.4a gives the coincidence probability as a function of θ_1 and θ_2 (see Fig. 7.4b). We see that it displays 100%-visibility *triangular* fringes. On the same graph we have included the result for a singlet-state source. The two plots agree at their extrema [i.e., for the cases of perfect correlation ($\theta_1 - \theta_2 = 0, \pi$) and anti-correlation ($\theta_1 - \theta_2 = \pi/2, 3\pi/2$)], and when $\theta_1 - \theta_2 = \pi/4, 3\pi/4$, etc. The largest discrepancy occurs when $\theta_1 - \theta_2 = \pi/8, 3\pi/8$, etc. It is not then surprising that these are the relative



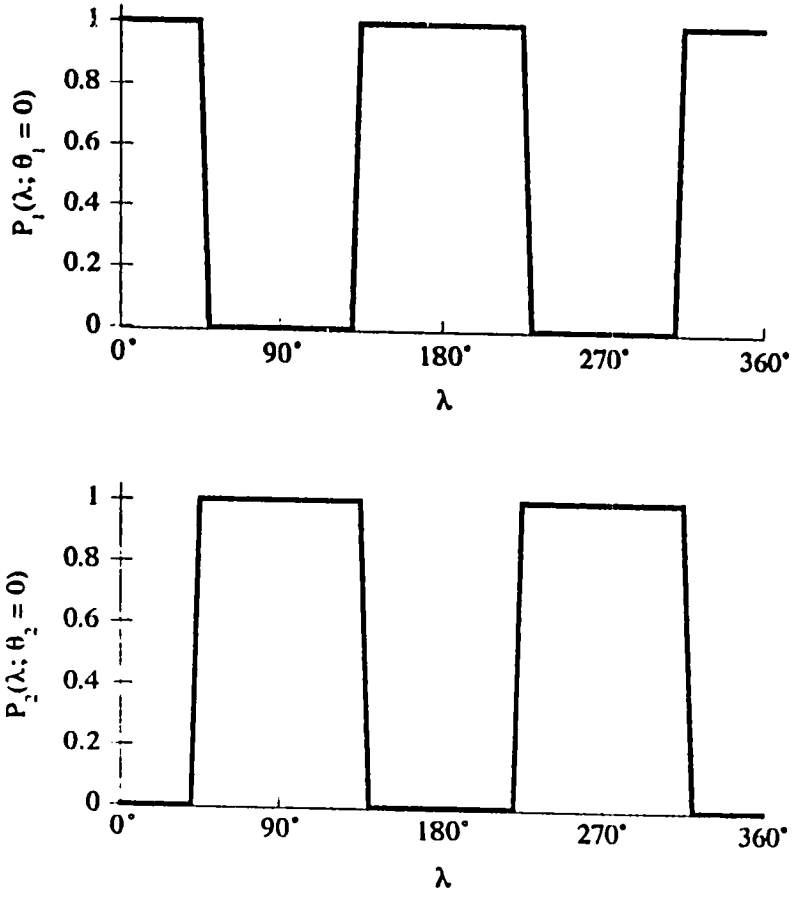
(a)



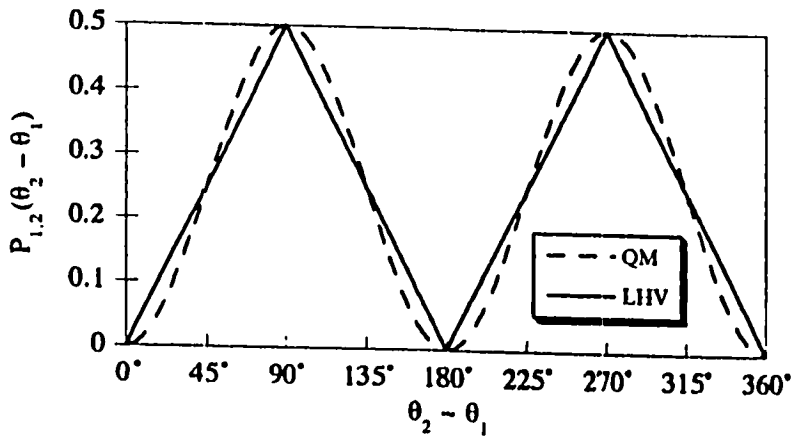
(b)

Fig. 7.3a Non-deterministic LHV predictions.
(a) The probability of a singles count vs. λ (hidden polarization variable) for specific value of parameter settings.

(b) The coincidence probability, averaged over λ, as function of $\theta_1 - \theta_2$; the sinusoidal fringes have only 50%-visibility for this LHV model.



(a)



(b)

Fig. 7.4a Deterministic LHV predictions.
(a) The probability of a singles count vs. λ (hidden polarization variable) for specific value of parameter settings. [See Eq. (7.21).]

(b) The coincidence probability, averaged over λ, as function of $\theta_2 - \theta_1$. The solid curve is the prediction of a deterministic LHV; the dashed curve the prediction for a QM singlet-state [Eq. (7.8a)].

angles used to demonstrate violations of Bell's inequalities (cf. Fig. 7.2).[21] The interpretation of such a violation would be that our simple LHV theory is untenable: The photons do *not* leave the source with definite polarizations. We will see in Chapter 8 that similar statements can be made for Bell's inequality experiments based on other sorts of entanglement, e.g., energy and time entanglement. Of course, the strength of Bell's inequalities is that they apply to a whole class of theories. We have offered just two simple examples, to help develop an intuition about the interpretation.

The second way of understanding the implications of a violation of Bell's inequalities is to examine the general assumptions that went into the derivation. Again, much has been written about this (see, for instance [Selleri, 1988]), so we will only highlight the argument. As discussed in Sect. 7.III, there are only a few assumptions necessary to derive a Bell's inequality. These are the Bell-locality condition, which consists of Parameter Independence (measurement results do not depend on the parameter setting of the distant analyzer) and Outcome Independence (measurement results for one particle do not depend on the results for the distant particle); the ρ-independence assumption (the source output does not depend on the parameter settings); and a notion that the results of measurements are governed by the complete state λ (see also footnote 3 in this Chapter, discussing the assumption of non-negative probabilities). We discuss the last of these first.

In the earliest derivations of Bell's inequalities, the underlying LHV theories were taken to be deterministic. This corresponds to probabilites $p(m \mid \lambda, \alpha, \beta)$ (and similar ones from Sect. 7.III) possessing only the values 0 or 1. Our second explicit LHV model, defined by (7.21), is an example of a deterministic theory. From a violation of these restricted Bell's inequalities, one might claim that the incorrect supposition in the

[21] In fact, calculation of either of the Bell's inequalities discussed earlier [i.e., (7.7a) and (7.12)], using the predictions of the deterministic LHV model at these angles will yield the limiting case, i.e., the equality-limits of (7.7a) and (7.12).

derivation was that of determinism; i.e., the photons described by a singlet-state do not "know" their polarization. Indeed, in the case of *deterministic* hidden variable theories, a "tangible" action-at-a-distance seems implied by a violation of Bell's inequalities: the possessed value of some attribute belonging to one member of a pair of spatially separated systems is instantaneously affected by varying the physical arrangement of an apparatus interacting with the other member. Here, "a direct conflict with special relativity seems difficult to avoid" [Redhead, 1986]. The Bohm-deBroglie guiding-wave model, discussed in Sect. 4.IV, is a hidden variable theory that *can* reproduce the quantum predictions because it is explicitly nonlocal.

However, the Bell's inequalities have been generalized to include non-deterministic sources, so that even if our LHV model is "relaxed" so that the photons are not assumed to possess definite polarizations, but are instead described by some sort of stochastic probability distribution, we still cannot explain the correlations implied by a violation. In other words, a violation implies some sort of nonlocality. It may be that the source somehow responds to the parameter settings (ρ-dependence). However, if the settings are changed randomly after the particles have left the source (i.e., if the space-like separation loophole is closed), this influence must be nonlocal[22]. The other possibility is that one (or both) of Parameter Independence and Outcome Independence is false.

Since they are at the heart of the entire argument, it behooves us to look more closely at the two independence conditions, and what QM has to say about them. The first, Parameter Independence (also known as "Jarrett-locality" [Jarrett, 1984]), states that the results on either side of the experiment do not depend on the parameter setting at the other side. Clearly, QM *satisfies* this condition; if it did not (and QM were the correct description of Nature) then it would be possible to send superluminal signals. For instance, the sender at analyzer 1 could send an instantaneous signal to the distance

[22] This sort of nonlocal influence is, of course, *not* predicted by QM.

receiver of particle 2 by appropriately adjusting the setting α, causing the distribution of the measurement outcomes b_n to change (e.g., causing all counts to appear at the ordinary-channel of 2's analyzer). The Outcome Independence condition (also known as "Jarrett-completeness" [Jarrett, 1984]) states that the results on one side do not depend on the *results* on the other side[23]. This condition is, in fact, violated by QM. Determination of the polarization/spin of one particle in an entangled state *nonlocally* determines the polarization/spin of the other particle. Nevertheless, there is no way to use the nonlocal correlation to send signals superluminally, for the results at each analyzer individually are completely random. It is only by bringing the data records together (a procedure which must be done at a speed less than or equal to c) that the correlations can be detected. (As we shall discuss next, some researchers have proposed to capitalize on the inherent randomness of the process in various quantum cryptographic schemes.)

In summary then, I would say that the interpretation of an experimental violation of a Bell's inequality is that Nature is nonlocal: The outcomes of measurements in spatially-separated regions of space depend on one another, and this dependence is not due to a "common cause". Moreover, if one wants outcomes in one location not to depend on apparatus settings in the other (which could imply a serious breach with the tenets of special relativity), he must relinquish an appealing, intuitive notion that an unmeasured observable nevertheless possesses an "element of reality".

7.VII. Quantum Cryptography

Although it is absolutely true that the EPR schemes cannot be used to send any signals superluminally, nevertheless, it has been proposed to utilize them for a different

[23] The condition of Jarrett-completeness identifies the hidden-variable description of the source as the *common cause* for the possessed values of correlated attributes such as spin- or polarization-components.

kind of communication, namely, cryptography. In the "one-time pad" scheme of classical cryptography [Brassard, 1988], it is supposed that two collaborators wish to share a secret "key", a random number (or string of binary digits) with which they may encode and decode a message. Such a key may provide an absolutely unbreakable code, provided that it is unknown to any potential eavesdropper. The problem arises in key distribution: any classical distribution scheme is subject to non-invasive eavesdropping, e.g., using a fiber-coupler to tap the line, without disturbing the transmitted classical signal. In the quantum cryptography proposals and demonstrations made to date, the security is guaranteed by using single-photon states [Ekert, 1991; Bennett et al., 1992; Ekert et al., 1992], some of the schemes employing particles prepared in an EPR-entangled state[24]. Both collaborators receive one member of each correlated pair, and measure the polarization (equivalently, some schemes use phase) in a random basis. After repeating the process many times, the two then discuss publicly which bases were used for each measurement, but not the actual measurement results. The cases where different bases were chosen are not used for conveying the key, and may be discarded (or perhaps examined to detect the presence of an eavesdropper [Barnett and Phoenix, 1993]), along with instances where one party detected no photon. In cases where the same bases were used, however, the participants will now have correlated information. From this, a random, shared key can be generated.

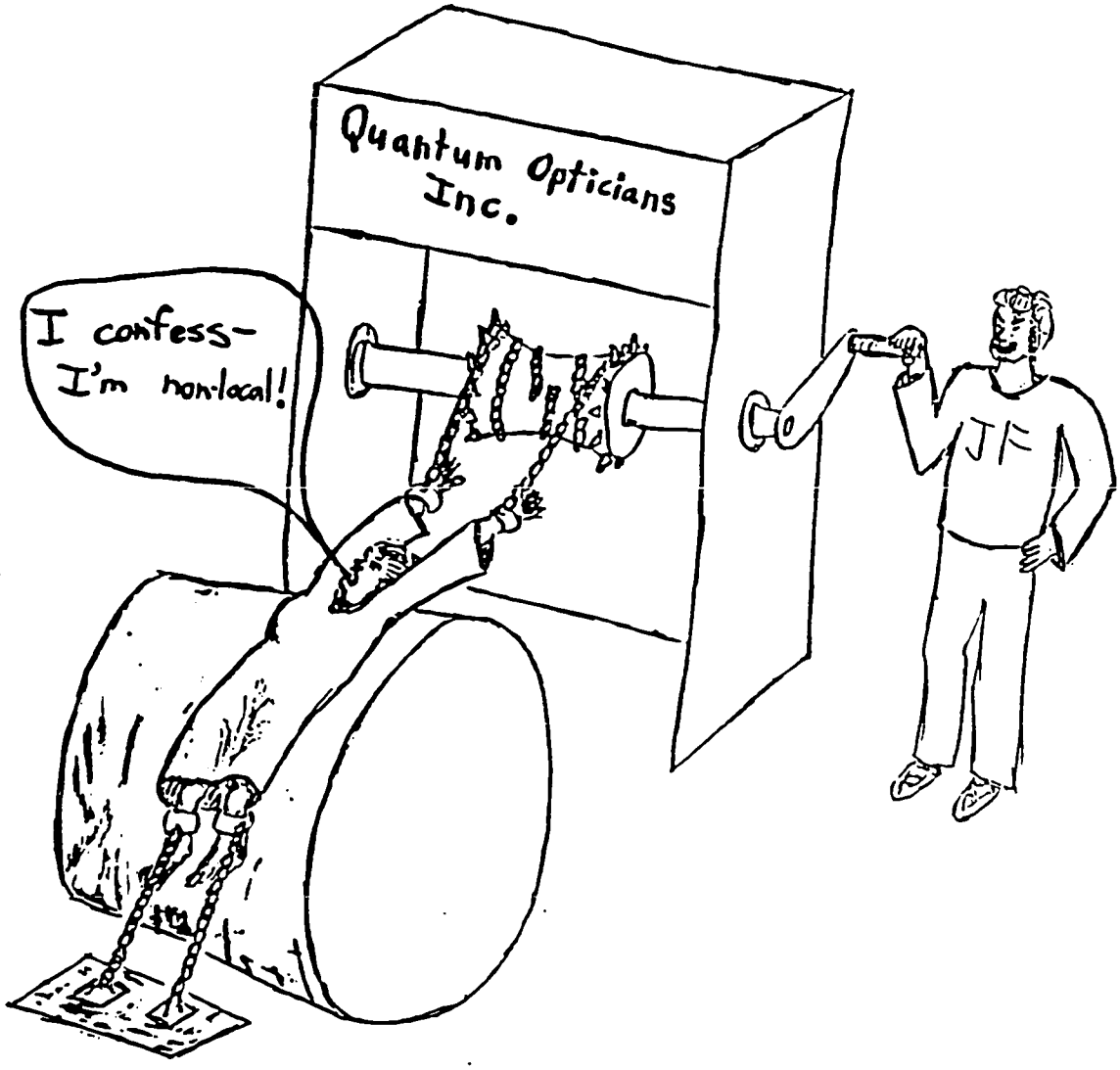
As long as single photons are used, any attempt at eavesdropping will necessarily introduce errors due to the uncertainty principle[25]. For if the eavesdropper uses the

---

[24] Alternately, the sender can send photons in a state of random, but *definite*, polarization (or phase). For this scheme, very weak coherent-state pulses, with an average photon number $<< 1$, can also be used [Bennett, 1992].

[25] Note that if classical pulses are used instead, as from an attenuated laser, they must be sufficiently weak that the average photon number per pulse is less than one, so that the probability of two photons in a pulse is much less than one; otherwise, an eavesdropper could make a measurement on one of the photons, sending on the other, undisturbed photon.

wrong basis to measure the polarization (phase) of a photon before sending it on to the real recipient, the very act of measuring will disturb the original state. Because of the "no-cloning" theorem [Wootters and Zurek, 1982], there is no way to copy the original quantum state and make measurements on the duplicate. By publically comparing of a subset of their correlated data (and subsequently discarding it), the collaborators may verify the security of their communication.



MOTHER NATURE
ON THE RACK

Chapter 8: The Franson Experiment

Cartoon, previous page: One mental image in performing these experiments is that as experimentalists in quantum optics, we are trying to "coax" Mother Nature into confessing Her nonlocality. The imagery is perhaps reversed, however, since it seems that most of the time it was the physicists who were under duress.

There is obviously no such limitation--I can measure the energy and look at my watch; then I know both energy and time!

-L. D. Landau, regarding the energy-time uncertainty principle

8.I. Introduction

In the previous chapter we introduced the EPR "paradox" and the Bell's inequalities which may be used to experimentally distinguish between theories based on the intuitive (but probably wrong) notion of local realism. Many experiments have been performed which, with certain reasonable auxiliary assumptions (discussed in Sect. 7.5), violated the Bell's inequalities derived for local hidden variable (LHV) theories, in support of quantum mechanics [Aspect et al., 1981; Aspect et al., 1982a,b; Clauser and Shimony, 1978; Selleri, 1988; Shih and Alley, 1988; Ou and Mandel, 1988]. The ramifications are unsettling; thus it is important to verify quantum mechanics in as many different realms as possible.

Nearly all experimental tests of the inequalities to date have involved the superposition of polarization (spin) states, along the lines of the Bohm version of the EPR paradox [Aspect et al., 1982a-c; Clauser and Shimony, 1978; Shih and Alley, 1988; Ou and Mandel, 1988]. A noteworthy exception is the recent experiment by Rarity and Tapster [Rarity and Tapster, 1990] based on "phase and momentum". The

experimental proposal by Franson [Franson, 1989] concerning a Bell inequality for non-polarization variables has received a fair amount of attention since its appearance several years ago [Kwiat et al., 1990; Ou et al., 1990b; Rarity et al., 1990; Ekert, 1991; Brendel et al., 1991; Rarity and Tapster, 1992; Rubin and Shih, 1992; Brendel et al., 1992; Shih et al., 1993; Kwiat et al., 1993a; Kwiat et al., 1993b]. The "Franson experiment" relies on the entanglement of a *continuous* variable, energy, and is thus closely related to the original EPR paradox. (We will contend that the present experiment is in some ways the "offspring" of both versions of the paradox.) Instead of polarizers or Stern-Gerlach analyzers, spatially separated Mach-Zehnder-like interferometers are used to investigate the nonlocal correlations.

In Sect. 8.II we present a simplified theory for the underlying 2-photon interference effect, as well as outline a more rigorous treatment. A first-version of the experiment, utilizing a dual-beam Michelson interferometer, is described in Sect. 8.III; the improved version with two Mach-Zehnder interferometers is given in Sect. 8.IV, along with a discussion of the effects of photon bandwidth and gate-window width. After showing in Sect. 8.V that we have violated a Bell's inequality based on energy and time entanglement, we will discuss the implications of this violation. Future directions of the Franson experiment are presented in Sect. 8.VI.

8.II. Theory

We have already seen that the down-converted photon pairs produced in parametric down-conversion possess strong frequency-correlations, due to energy conservation, $\omega_p = \omega_s + \omega_i$. The emitted photons may thus be described by an energy-entangled state [cf. Eq. (2.3)]. We select out pairs such that ω_s and ω_i are centered at $\omega_p/2$. Previous experiments [Hong et al., 1987; Steinberg et al., 1993b] have demonstrated that the photons also have very strong *temporal* correlations, so that if one photon is detected at time t, its conjugate will be detected within the two-photon correlation time τ_{tpc}, much

less than 1 ps. This is of the same order as the single-photon coherence time τ_c, usually determined in practice by filters before the detectors. Note that we now have a situation very similar to that originally proposed by EPR (see Sect. 7.II; [Einstein et al., 1935]). They described a system of two particles in a simultaneous eigenstate of the operators k_1+k_2 and x_1-x_2. If we make the transformations $k \rightarrow \omega/c$ and $x \rightarrow c\,t$, we essentially have this state. The signal and idler energies sum to a constant, and the difference in their times of emission is nearly zero. Bell has shown, however, that since there is a positive-definite Wigner function which can describe these properties, no *direct* violation of a Bell inequality is possible for these observables [Bell, 1986; see also Wódkiewicz, 1988]. Consequently, in order to investigate an inequality based on energy and time, it is necessary to produce a state more akin to the spin singlet state of the Bohm version of the EPR paradox. We shall see shortly how this arises in the present experiment.

We first give a simplified analysis of this experiment based on Feynman's notion of interference of indistinguishable processes. Let each conjugate photon enter a modified Mach-Zehnder interferometer (MZ); see Fig. 8.1. Each interferometer (j=1,2) has a short path of length S_j and a long path of length Lj. If the optical imbalance $\Delta L_j = Lj - Sj$ is less than the coherence length $c\tau_c$ of the incident photon, then fringes will be visible in singles rates as the long arms are moved slightly. We now restrict our

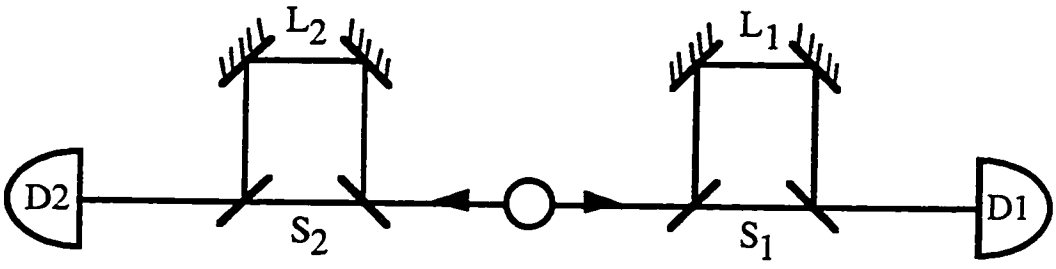


Figure 8.1 Simplified schematic of basic Franson experiment.

discussion to the case where $\Delta L_j >> c\tau_c$, in which there are no singles fringes. Fringes can nevertheless be observed in the rate of coincident detections between the two detectors. It becomes helpful to regard each MZ as consisting of two optical delay lines, "short" and "long", in parallel. Then for any incident photon pair, there are four processes (short-short, long-long, short-long, and long-short) leading to pair detection, where the first label refers to interferometer 1, and the second to interferometer 2. In the short-long and long-short processes, the two photons exit their respective MZ's having acquired a relative time-lag large with respect to τ_{tpc}. These processes, which account for one-half of the emitted photon pairs, are therefore *in-principle* distinguishable from each other as well as from the short-short's and long-long's. According to the Feynman rules for interference, they thus constitute a non-interfering background. We shall see that whether the electronic resolution ΔT is sufficient *in practice* to discriminate against the short-long and long-short counts will be crucial in determining the outcome.

If the difference *between* the path-length differences ($\Delta L \equiv \Delta L_1 - \Delta L_2$) is greater than $c\tau_{tpc}$, then the long-long's are also distinguishable from the short-short's, and no interference will result. We thus further restrict our discussion to the case $\Delta L << c\tau_{tpc}$. Then the long-long and short-short coincidence processes are indistinguishable from each other, because the absolute time of emission of the pair from the crystal is undetermined (for a cw pump)[1]. To calculate the probability of coincidences we need the probability amplitudes for the four coincidence processes. The phase accumulated in traveling a short path is[2] $\theta_j^S = \omega_j S_j/c$, while for a long path it is $\theta_j^L = \omega_j L_j/c$. The amplitudes for the various coincidences are then

---

[1] Note that we are making an assumption here. The fact that this assumption predicts interferences which are in fact observed implies that the assumption is correct. See Sect. 8.V.

[2] Although we are only including dynamical phases here, one could equally well include *geometrical* phases, i.e., Berry's phases (see Chap. 3).

$$A_{SS} = \left(\frac{1}{\sqrt{2}} e^{i\theta_1^S} \frac{1}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}} e^{i\theta_2^S} \frac{1}{\sqrt{2}}\right) = \frac{1}{4} e^{i(\theta_1^S + \theta_2^S)} , \tag{8.1a}$$

$$A_{LL} = \left(\frac{i}{\sqrt{2}} e^{i\theta_1^L} \frac{i}{\sqrt{2}}\right)\left(\frac{i}{\sqrt{2}} e^{i\theta_2^L} \frac{i}{\sqrt{2}}\right) = \frac{1}{4} e^{i(\theta_1^L + \theta_2^L)} , \tag{8.1b}$$

$$A_{SL} = \left(\frac{1}{\sqrt{2}} e^{i\theta_1^S} \frac{1}{\sqrt{2}}\right)\left(\frac{i}{\sqrt{2}} e^{i\theta_2^L} \frac{i}{\sqrt{2}}\right) = -\frac{1}{4} e^{i(\theta_1^S + \theta_2^L)} , \tag{8.1c}$$

and

$$A_{LS} = \left(\frac{i}{\sqrt{2}} e^{i\theta_1^L} \frac{i}{\sqrt{2}}\right)\left(\frac{1}{\sqrt{2}} e^{i\theta_2^S} \frac{1}{\sqrt{2}}\right) = -\frac{1}{4} e^{i(\theta_1^L + \theta_2^S)} \tag{8.1d}$$

where we have assumed a reflection (transmission) amplitude of $i/\sqrt{2}$ ($1/\sqrt{2}$) for the beamsplitters, and have assumed the particular output ports as shown in Fig. 8.1. The probabilities for short–long or long–short coincidences are just the absolute squares of the amplitudes of these in-principle distinguishable possibilities:

$$P_{SL} = |A_{SL}|^2 = \frac{1}{16} = |A_{LS}|^2 = P_{LS} . \tag{8.2a}$$

The long–long's and short–short's are indistinguishable, so their amplitudes must be summed first, then absolute squared:

$$P_{SS \text{ or } LL} = |A_{SS} + A_{LL}|^2$$

$$= \frac{1}{16}\left| e^{i(\theta_1^S + \theta_2^S)} + e^{i(\theta_1^L + \theta_2^L)} \right|^2$$

$$= \frac{1}{8}\left\{1 + \cos\left[(\theta_1^L - \theta_1^S) + (\theta_2^L - \theta_2^S)\right]\right\}$$

$$= \frac{1}{8}\left\{1 + \cos\left[\frac{\Delta L_1}{c}\left(\frac{\omega_p}{2} + \omega\right) + \frac{\Delta L_2}{c}\left(\frac{\omega_p}{2} - \omega\right)\right]\right\}$$

$$= \frac{1}{8}\left\{1 + \cos\left[\frac{\omega_p}{2c}(\Delta L_1 + \Delta L_2) + \Phi\right]\right\} , \qquad (8.2b)$$

where we have used energy conservation/entanglement to write $\omega_1 = \omega_p/2 + \omega$, $\omega_2 = \omega_p/2 - \omega$. Because we have assumed $\Delta L << c\,\tau_{tpc}$ (i.e., ΔL_1-ΔL_2 is arranged to be small relative to the inverse bandwidths of ω_1 and ω_2), the phase $\Phi \equiv \omega\,\Delta L/c$ is negligible. (As stated earlier, the fringes will wash out if this condition is not met.) The total probability for coincidence depends on the coincident gate time ΔT. For $\Delta T > \Delta L_j/c$, short–long's and long–short's contribute and we have

$$P_{tot,\,wide\ window} = \frac{1}{8}\left\{2 + \cos\left[\phi_1 + \phi_2\right]\right\} . \qquad (8.3)$$

These fringes have only 50% visibility, and have been observed experimentally (see Sect. 8.III) in slightly different, but equivalent geometries [Kwiat et al., 1990; Ou et al., 1990b; Brendel et al., 1991].

For $\Delta T < \Delta L_j/c$, the background of short–long and long–short coincidences is excluded and we are left with the rate of "true" coincidences:

$$P_{tot,\,narrow\ window} = \frac{1}{8}\left\{1 + \cos\left[\phi_1 + \phi_2\right]\right\} . \qquad (8.4)$$

This enhancement of the visibility by a factor of two has now been seen by several groups [Brendel et al., 1991; Shih et al., 1993; Kwiat et al., 1993c]. The 100%-visibility of (8.4) stems from the strong correlations between conjugate photons, and allows one to violate an appropriate Bell's inequality.

Equivalently, one could have calculated the coincidence probability by using the Schrödinger picture, and evolving the input state [Eq. (2.3)]. The output state after the interferometers is

$$|\psi\rangle_{out} \approx \frac{1}{4}\left[|\psi_S\rangle_1 |\psi_S\rangle_2 + |\psi_L\rangle_1 |\psi_L\rangle_2 - |\psi_S\rangle_1 |\psi_L\rangle_2 - |\psi_L\rangle_1 |\psi_S\rangle_2\right], \quad (8.5)$$

where we have omitted for simplicity the integral over frequency. For a wide coincidence gate window, all terms remain, and (8.5) can be rewritten

$$|\psi\rangle_{out} \approx \frac{1}{4}\left[|\psi_S\rangle_1 - |\psi_L\rangle_1\right]\left[|\psi_S\rangle_2 - |\psi_L\rangle_2\right]. \quad (8.6a)$$

The fact that the kets factorize leads to 50% visibility[3]. On the other hand, for a narrow coincidence window that last two terms of Eq. (8.5) will not contribute, and we have

$$|\psi\rangle_{out} \approx \frac{1}{4}\left[|\psi_S\rangle_1 |\psi_S\rangle_2 + |\psi_L\rangle_1 |\psi_L\rangle_2\right]. \quad (8.6b)$$

This state is similar to the familiar singlet state standardly used as a source for violations of Bell's inequalities.

As alluded to earlier, one cannot produce a violation of Bell's inequalities just using the entangled output of the down-conversion crystal and making measurements *directly* of the energy and time, for this state is describable with a non-negative Wigner distribution. How is it that we can *ever* use such a source to demonstrate a nonclassical effect, since a positive-definite Wigner function is equivalent to a classical probability distribution? The answer is that the Wigner distribution is not truly an LHV model, since it includes all the QM correlations. When measurements are made involving the coherent superposition of the field at different times, as in our Franson interferometers, it is possible to bring out the irreducible nonlocality. In this case the Wigner distribution

[3] Recall that the factorization condition is essentially the definition of the Bell-locality constraint, Eq. (7.4a).

acquires negative values.

The above simple calculations are sufficient if one only wishes to investigate specific limiting cases. However, to fully predict the functional form of the *transition* regions between the easy-to-understand limiting cases, a more complete theory is needed. We outline such a rigorous calculation here. The truly "interested" reader is directed to any of a number of lengthy papers for further details [Rarity and Tapster, 1992; Rubin and Shih, 1992; Shih et al., 1993; Campos et al., 1990]. Also, in Chap. 4 and Appendix B2 similar general calculations are given for slightly different down-conversion experiments. We start with the state out of the crystal [cf. Eq. (2.3)]

$$|\psi\rangle = \int d\omega_p \, A_p(\omega_p) \int d\omega_s \, A(\omega_s) \, |\omega_s\rangle_s \, |\omega_p - \omega_s\rangle_i \tag{8.7}$$

(where we have included an integral over the ω_p to allow for a non-monochromatic pump), and calculate the second-order Glauber correlation function relating the field amplitudes for the signal and idler modes after the interferometers at the times t_s and t_i, respectively:

$$G^{(2)}(t_s, t_i; t_i, t_s) = \langle \psi | \, \hat{E}_s^{(-)}(t_s) \, \hat{E}_i^{(-)}(t_i) \, \hat{E}_i^{(+)}(t_i) \, \hat{E}_s^{(+)}(t_s) \, | \psi \rangle \, , \tag{8.8}$$

where

$$\hat{E}_s^{(+)}(t_s) = \int d\omega_s \, \eta_s^*(\omega_s) \, \hat{a}_s(\omega_s) \, e^{-i\omega_s t_s} \frac{1}{\sqrt{2}} \left\{ 1 - e^{i\omega_s \Delta L_1} \right\} , \tag{8.9a}$$

$$\hat{E}_s^{(-)}(t_s) = \left(\hat{E}_s^{(+)}(t_s) \right)^\dagger , \tag{8.9b}$$

$$\hat{E}_i^{(+)}(t_i) = \int d\omega_i\, \eta_i^*(\omega_i)\, \hat{a}_i(\omega_i)\, e^{-i\omega_i t_i} \frac{1}{\sqrt{2}} \left\{1 - e^{i\omega_i \Delta L_2}\right\}, \tag{8.9c}$$

and

$$\hat{E}_i^{(-)}(t_i) = \left(\hat{E}_i^{(+)}(t_i)\right)^\dagger. \tag{8.9d}$$

We have sent the signal photon into interferometer 1, and the idler into interferometer 2. The probability of joint detection of a signal-idler pair within the detector resolution window ΔT, after a total time $\mathcal{T}$, is then given as in Eq. (4.7):

$$P = \int_{-\mathcal{T}/2}^{\mathcal{T}/2} dt_s \int_{t_s - \frac{\Delta T}{2}}^{t_s + \frac{\Delta T}{2}} dt_i\, G^{(2)}(t_s, t_i; t_i, t_s) \ . \tag{8.10}$$

In general, the resulting *10-integral* expression is intractable. However, if one assumes gaussian transmission spectra for the filters [represented by the $\eta(\omega)$'s], then it may be evaluated numerically. In Fig. 8.2 we present the results of such a numerical evaluation. The values of the various parameters were chosen to display the full character of the Franson experiment. We see that there are four main regions. For path-length differences (in time units) less than the coherence time of the down-converted photons, fringes will be observed in both singles rates, and hence trivially in the coincidence rate, which is essentially the product of the singles in this "white-light" regime. For path-length differences greater than the photons' coherence times, but less than the electronic resolution time, there will be no singles fringes, and coincidence fringes will be limited to 50%-visibility due to the presence of short-long and long-short coincidences. When the electronic resolution time becomes less than the path-length difference, coincidence visibility can reach 100% (assuming, as always, that the difference in the path length differences is less than the photons' coherence time).

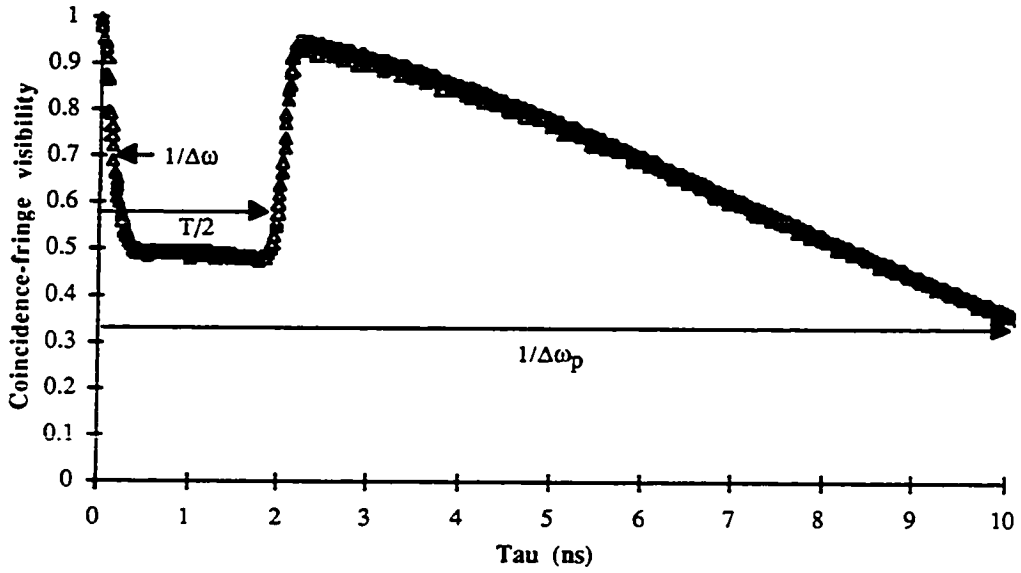


Figure 8.2 Numerical calculation of the coincidence fringe visibility in a Franson experiment, as a function of the path-length difference in either interferometer (the difference in path-length differences is set to be much less than the coherence length of the down-converted light). The various lengths and time-scales have been chosen to simultaneously display all three regimes of operation, and do not correspond to actual experimental parameters. In particular, for this calculation the pump frequency $\omega_p = 5 \times 10^6$/ns; the pump bandwidth (which determines the width of the entire profile) $\Delta\omega_p = 0.1$/ns; the bandwidth of the down-converted light (which determines the width of the initial 100%-visibility region, and the rise-time to the second 100%-visibility region) $\Delta\omega = 10$/ns; and the electronic gate window width (which determines the extent of the 50%-visibilility region) $\Delta T =$ 4 ns. Note that because the gate window is centered on the long-long and short-short coincidences, the transition from 50% to 100%-visibility occurs at 2 ns.

Finally, *no* fringes will be possible if the path lengths are much greater than the *pump* coherence time.

8.III. Dual-Beam Michelson Experiment

The original proposal by Franson involved an atomic cascade light source and two spatially separated Mach-Zehnder interferometers. Our experiments employed the parametric fluorescence light source; our first attempt used a single dual-beam Michelson interferometer, as shown in Fig. 8.3. The signal and idler beams were made parallel to each other by means of mirrors M1 and M2, and injected side-by-side into a single Michelson interferometer. They are thus both affected by the same path-length difference $\Delta L_1 = \Delta L_2 = 2\Delta L$[4]. Upon leaving the output port of the interferometer, photons in the two parallel beams passed through filters F1 and F2 and were detected by photomultipliers D1 and D2, while the travelling mirror of the Michelson was slowly scanned by a stepping motor. We calibrated the system by counting He-Ne laser fringes, and determined that one step of the stepping motor corresponded to an average motion of 6.101 ± 0.027 nm of the travelling mirror. Each detector consisted of an RCA C31034A-02 photomultiplier tube, which was cooled to approximately -30° C. The signals from the photomultipliers were amplified and directed into a Stanford Research Systems SR 400 Gated Photon Counter. The electronic delay between the signals was adjusted to maximize the coincidence count rate for a 5-ns-wide gate window.

When the Michelson imbalance was very small, fringes were observed in the singles rates, and trivially in the coincidence rate (which in this regime is essentially their product); see Fig. 8.4. Because the coincidence rate factorizes, violation of a Bell's

[4] We define ΔL as the actual arm-length difference in the Michelson; the optical path-length difference is *twice* this.

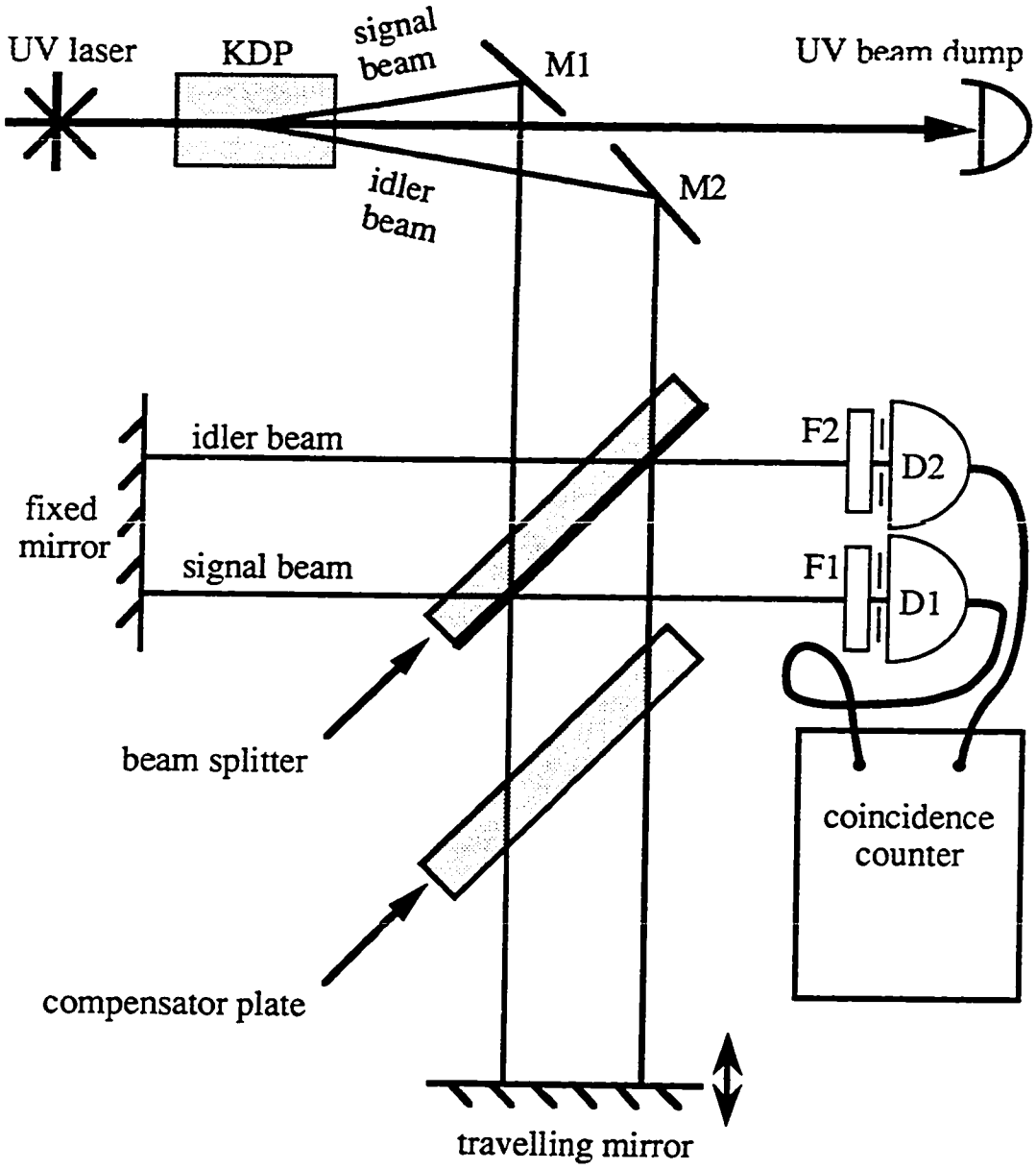


Figure 8.3 Experimental setup used in dual-beam Michelson version of the Franson experiment. See text for details.

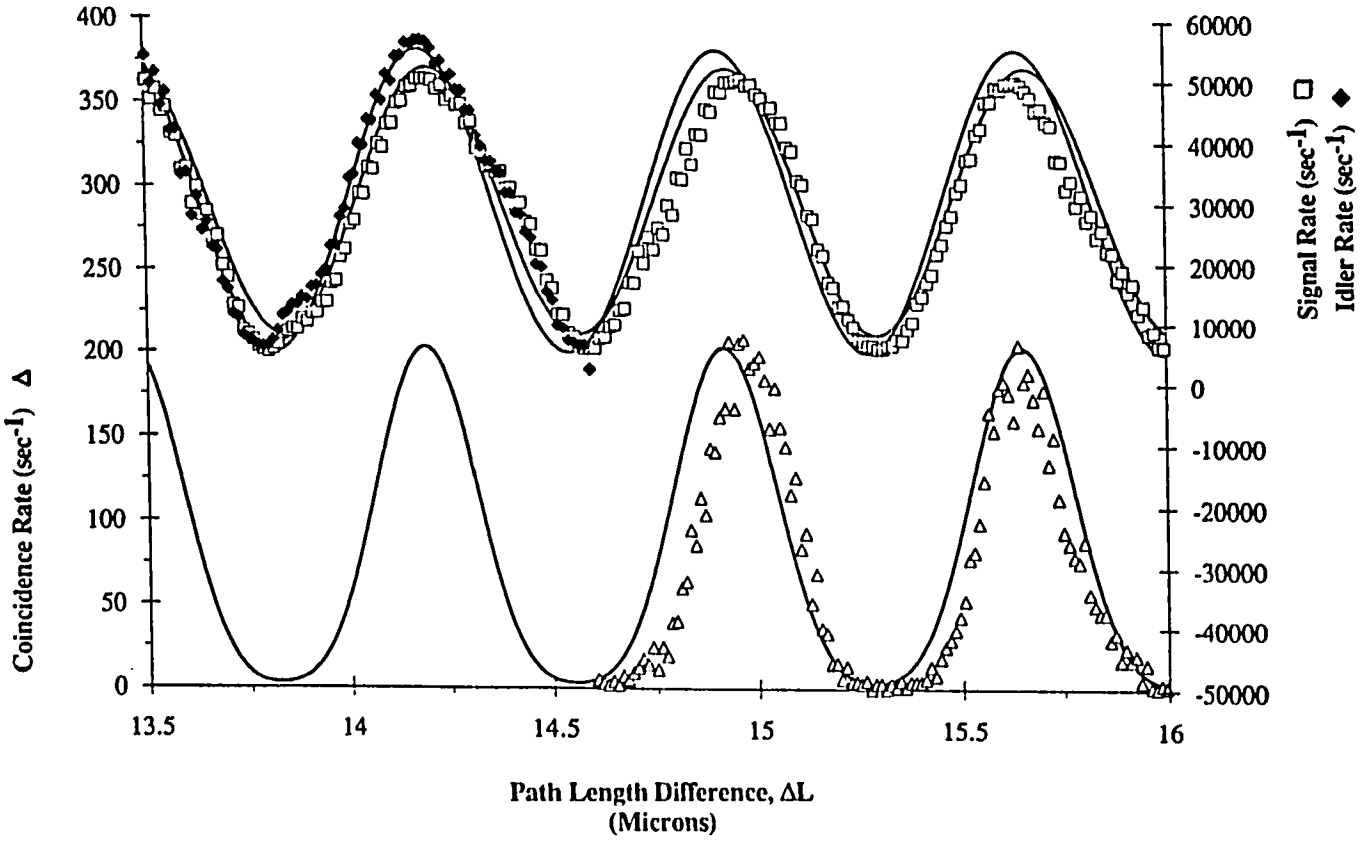


Figure 8.4 Singles (top traces) and coincidence fringes (bottom trace) in white-light regime of dual-beam Michelson experiment. The solid curves in the top trace are theoretical predictions, with visibility and overall phase shift adjusted to match data. Solid curve in bottom trace is the product of these curves.

inequality is clearly impossible. The visibility of the signal (or idler) single-photon interference close to the white-light fringe was measured to have a sinc-like variation with arm length difference (see Fig. 8.5a), with a maximum value of $93.0 \pm 1.0\%$. From the position of the first null of this pattern, we determined the coherence lengths of the down-converted light : $\Delta l_S = \Delta l_i = 50\ \mu m$, which are consistent with the 10-nm bandwidths of filters F1 and F2 centered at 702 nm (see Fig. 8.5b).

When the Michelson imbalance is greater than the coherence lengths (Δl_S, Δl_i), no singles-fringes are observed, although fringes may still be observed in coincidence. The results are shown in Fig. 8.6, where the singles count rate (upper trace) and the coincidence count rate R_C (lower trace) are plotted against the arm length difference ΔL. These data points were taken starting with $2\Delta L = 240\ \mu m$, as determined by counting the steps of the stepping motor starting from the position of the white light fringe. The visibility of the fringes in the singles count rate is very low ($<5 \times 10^{-3}$) in this region. However, the visibility of the fringes in the coincidence count rate is quite high: $46.0\% \pm 2.2\%$ (with 90% confidence). When we account for imperfect alignment of the Michelson arms (the 93%-visibility at zero path-imbalance, from Fig. 8.5a), the corrected coincidence visibility is $52.6\% \pm 3.0\%$. This agrees, within the experimental error, with the predicted value of 50%, which was used in the calculation of the solid sinusoidal curve shown in Fig. 8.6. The travelling mirror moved a distance of 176.1 ± 1.0 nm from one coincidence-rate maximum to the next, which also agrees, within the experimental error, with the predicted value, viz., $\lambda_p/2 = 175.6$ nm between adjacent interference maxima. Note that because both down-converted photons experience the same optical delay (and because their frequencies sum to the pump frequency), we observe fringes determined by the *pump* wavelength, even though the ultraviolet pump never enters the interferometer.

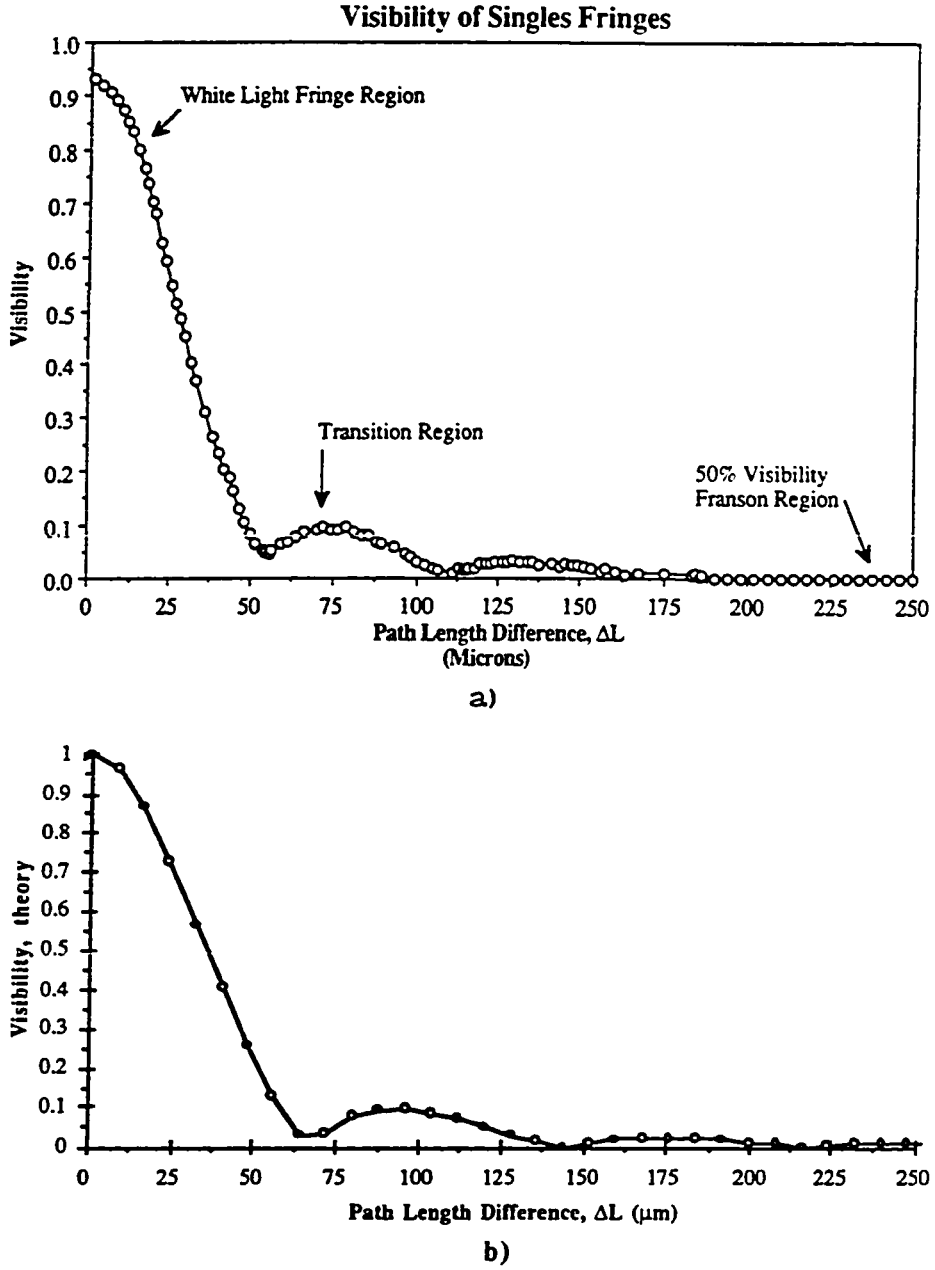


Figure 8.5 Sinc-like visibility of singles fringes, as a function of the optical path-length difference in the dual-beam Michelson. a) Experimental results with a 10-nm FHWM filter before the detector. b) Theoretical prediction based solely on the Fourier transform of the filter's spectral transmission data.

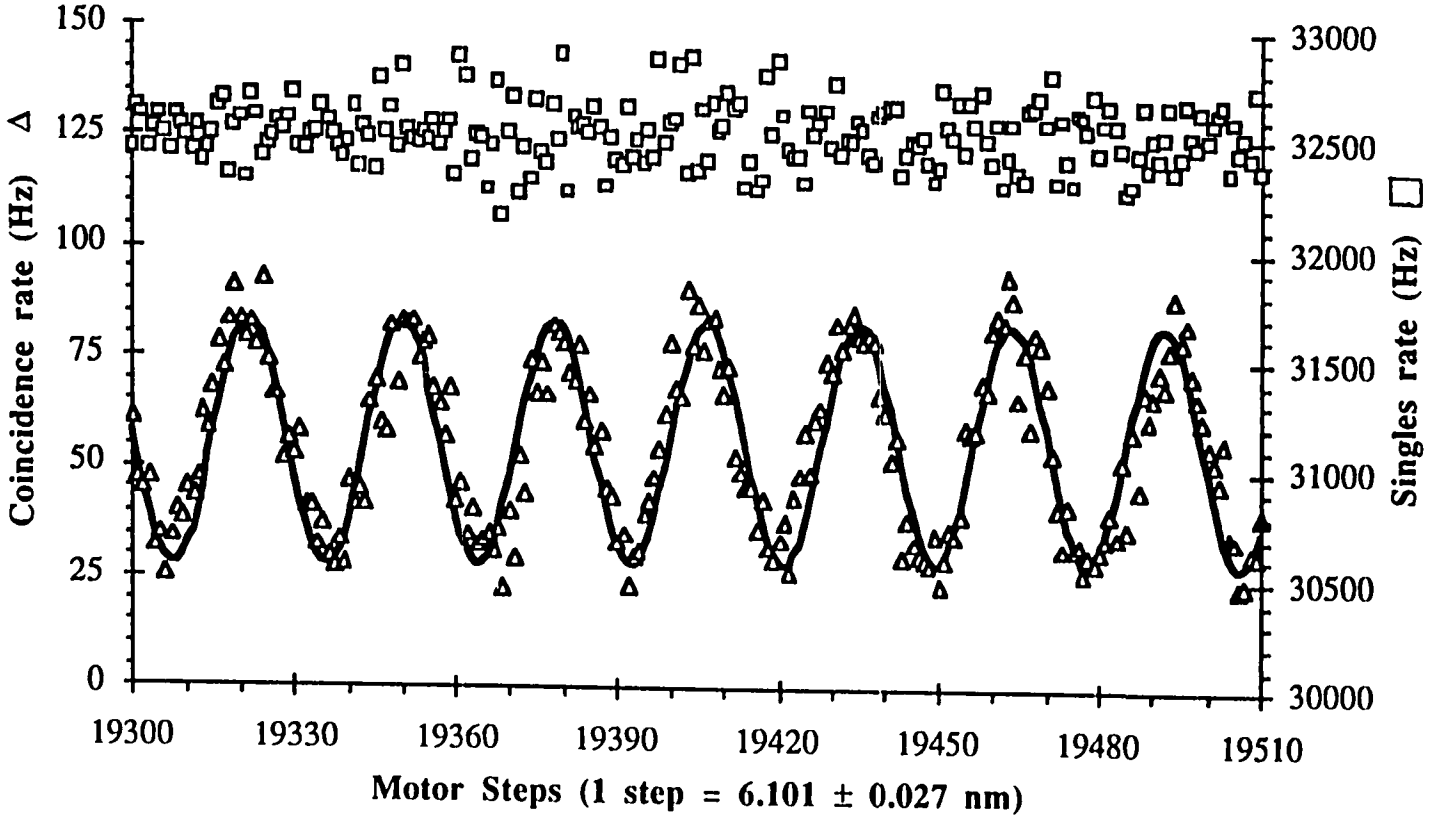


Figure 8.6 Two-photon interference in dual-beam Michelson interferometer: The coincidence count rate (left axis, lower trace) and the singles count rate (at detector D1) (right axis, upper trace) versus the arm length difference for the setup of Fig. 8.3. The solid line denotes the theoretical prediction (see text). These data points were taken in a region far away from the white light fringe (with $\Delta L = 240\ \mu m >> \Delta l_S = 50\ \mu m$). The integration time per step was 1 second. An accidental coincidence rate of $10.6\ s^{-1}$ has been subtracted from the coincidence count rate.

We therefore conclude that we have indeed observed a two-photon interference in the dual-beam Michelson interferometer [Kwiat et al., 1990] (Similar 50%-visibility results were reported by other groups as well [Ou et al., 1990b; Rarity and Tapster, 1992]). However, since the observed visibility is not significantly greater than 50%, we cannot claim that this is a nonclassical effect[1]. There exists a classical-field explanation in which the rates R_s, R_i of singles detection, and R_c of coincidence detection are ensemble averages in a stochastic classical field theory. In this theory, the wavenumbers k_s and k_i are classical random variables which are subjected, however, to the constraint that $k_s + k_i = k_p$, where k_p is a nonrandom variable[2]. Then

$$R_s \propto \langle 1 + \cos k_s \Delta L \rangle = 1 ,$$

$$R_i \propto \langle 1 + \cos k_i \Delta L \rangle = 1 ,$$

and

$$R_c \propto \langle (1 + \cos k_s \Delta L)(1 + \cos k_i \Delta L) \rangle$$

$$= \langle 1 + \cos k_s \Delta L + \cos k_i \Delta L + \frac{1}{2}\cos (k_s - k_i)\Delta L + \frac{1}{2}\cos (k_s + k_i)\Delta L \rangle$$

---

[1] There has been some confusion and controversy as to whether 50%-visibility is indeed a classical result. On the one hand, Mandel and Co. produced a mathematical proof (which few people understood) that *no* classical model could reproduce 50%-visibility Franson fringes. On the other hand, we suggested an explicit model, presented here, which *did* yield 50%. The resolution to these seemingly contradictory results came from Franson [Franson, 1991]; remarkably, *both* results are correct, when taken in the proper context. If one *only* considers the results of a Franson experiment, then a 50%-visibility model is possible. If, however, one incorporates the tight time-correlation results of *other* experiments, then no classical model can yield more than a few percent visibility.

[2] For example, one might envision a source of radio waves, with correlated side-bands.

$$= 1 + \frac{1}{2}\cos k_p \Delta L \,, \qquad (8.11)$$

where $\langle \cos k_s \Delta L \rangle = \langle \cos k_i \Delta L \rangle = \langle \cos (k_s - k_i)\Delta L \rangle = 0$, but $\langle \cos (k_s + k_i)\Delta L \rangle = \cos k_p \Delta L$. Thus this classical theory also predicts a 50% visibility. As in the Hanbury-Brown-Twiss effect, our interference effect in its present form possesses a classical explanation. However, unlike the Hanbury-Brown-Twiss case, it is possible to improve the experiment, so that if the visibility exceeds 50%, then a classical explanation is no longer possible. We have done this in the experiment discussed below.

8.IV. Mach-Zehnder Experiment

As seen in the previous section, initial attempts to perform the Franson experiment were limited by slow detectors and relatively small interferometers [Kwiat et al., 1990; Ou et al., 1990b; Rarity and Tapster, 1992]. The resulting extra background reduced the fringe visibility so that it was not possible in a single experiment to rule out classical field models, let alone *all* local realistic theories[3]. Brendel *et al.* were the first to succeed in removing the unwanted background [Brendel et al., 1991], reporting a visibility of 87%. However, their arrangement employed a *single-beam* Michelson interferometer, and no conclusions regarding locality could be made. Since this initial experiment, they have performed a similar two-interferometer experiment [Brendel et al., 1992].

A schematic of our experimental setup is shown in Fig. 8.7a. A 40-cm focal-length cylindrical lens was used to collimate the degenerate ($\lambda = 702$ nm) down-converted beams in the vertical direction. After the lens the conjugate beams traverse

[3] Two recent papers [Rubin and Shih, 1992; Shih, et al., 1993] even seem to imply that at large path-length differences, high visibility would still be unattainable, due to the finite size of the down-converting crystal. This is incorrect, however, as the present experiment explicitly demonstrates [Kwiat et al., 1993].

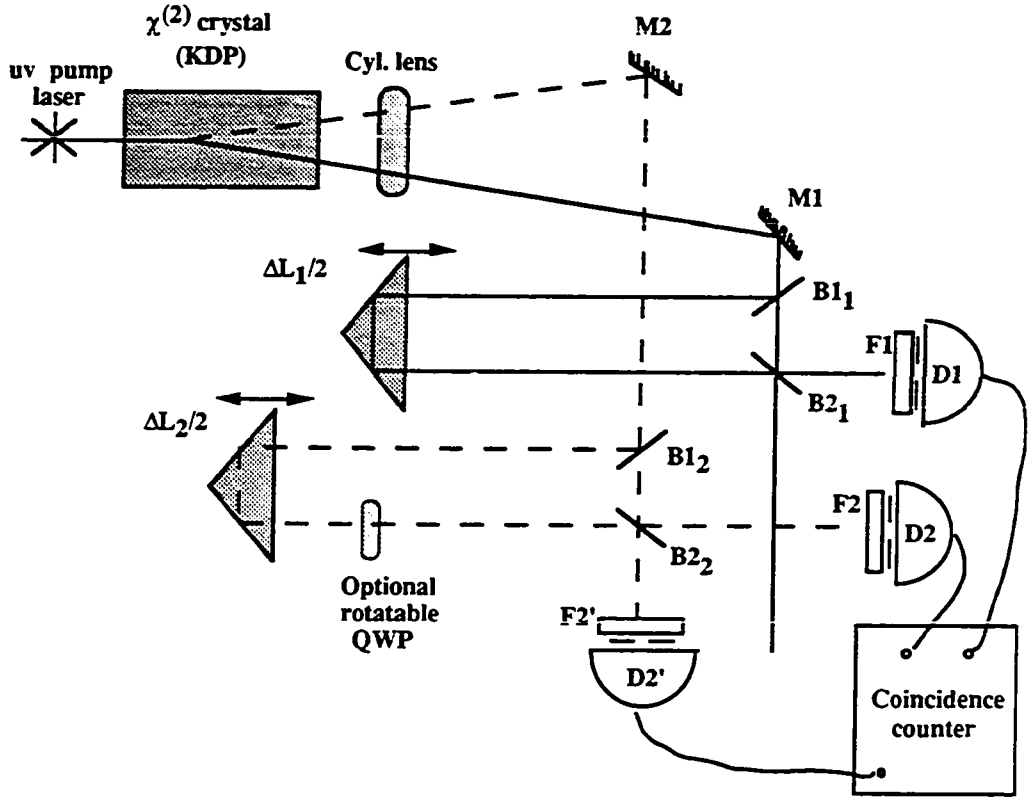


Figure 8.7a Schematic of experimental setup used in separated Mach-Zehnder version of the Franson experiment. See text for details. The axis of the quarter waveplate (QWP) could be reliably switched between horizontal and vertical.

similar optics, so we will describe only the path of beam 1. All optics had a nominal flatness of $\lambda/10$. Bending mirror M1 directs the beam into an unbalanced MZ, formed by two 50/50 beam splitters ($B1_1$ and $B2_1$) and a translatable right-angle prism. The beam splitters were specified to have R=45%/T=45% for p-polarization, and were wedged ($\Delta\theta = 0.5°$) to eliminate unwanted interfering paths. The prism was mounted on a dual-stage translation element, consisting of a standard optical translation stage for "coarse" adjustment, and a Burleigh Inchworm piezoelectric system, nominally capable of 4-nm single steps[4]. The position of the prism was monitored via a Heidenhain

[4] We actually observe an average step size of 6 nm, and individual steps which vary

optical encoder, with a 0.1-μm resolution. The optical path-length difference in each interferometer was approximately 63 cm, much smaller than the pump coherence length[5] ($c\tau_{pump} > 6$m), but much larger than the coherence length ($c\tau_c \approx 50\mu$m) of the down-converted light[6]. Following a filter and an adjustable iris, the beam was focussed onto detector D1, a customized silicon avalanche photodiode counting module (EG&G SPCM-200-PQ; see Appendix E for specifications). (We have measured the detection efficiency to be ~75% at λ=702 nm [see Chap. 9], but at present the filter and iris limit the effective efficiency to less than 1%.) We measured the time resolution (10%-90% region) of the devices to be 1.1ns. Detector D2 was another SPCM, while detector D2' (used in later versions of this experiment) was a standard passively-quenched silicon APD (EG&G C30902; see Appendix E for specifications). The outputs of the detectors were fed into a time-to-amplitude converter with 100-ps resolution. This was operated with a time window of 1.46 ns, thereby eliminating nearly all contribution from the long-short and short-long coincidences, which were displaced by ±2.1 ns relative to the short-short and long-long (due to the 63-cm path-length differences). The time profile and gate window are displayed in Fig. 8.8.

Typical results (see Fig. 8.9a) displayed sinusoidal coincidence fringes with a visibility V=80.4±0.6%, while no fringes were discernible in the singles rates. The less-than-unity visibility even with the short gate window is due to some combination of

---

over a 3 μm cycle from about 2 nm to 8 nm.

[5] For this experiment the argon-ion pump laser was operating in single-line mode.

[6] One of the primary difficulties of the arrangement in Fig. 8.7a is that one must ensure that the *difference* of the path-length differences is much less than ~50 μm, the photons' coherence length, as discussed in Sect. 8.II. In the previous experiment with a dual-beam Michelson, this condition was automatically satisfied. In general, however, it is not a trivial task to determine two lengths to within a fraction of a percent. The solution involved a clever trick by Aephraim Steinberg, in which one creates two white-light interferometers by temporarily adding extra optical elements. See the modified schematic of Fig. 8.7b.

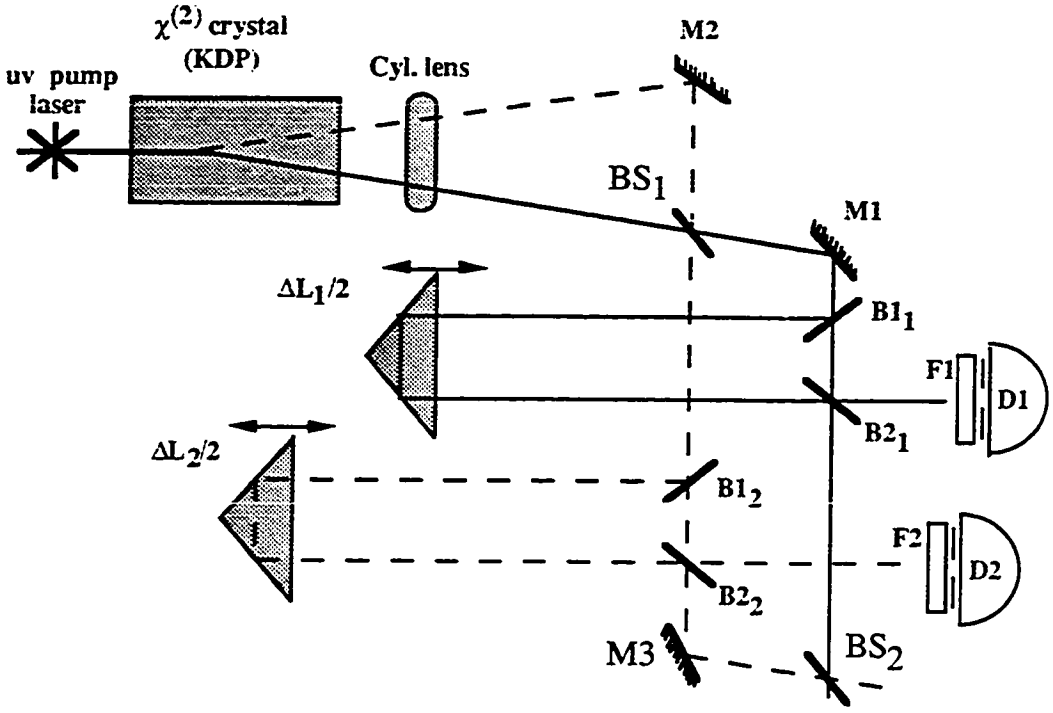


Figure 8.7b Temporary modification[7] of setup to achieve equal path length differences in two separated Mach-Zehnder interferometer (clever method due to A. M. Steinberg). An additional mirror M3 and extra beam splitters BS_1 and BS_2 create two additional interferometers. For example, with the trombone-prism paths blocked, the two paths BS_1-M1-$B1_1$-$B2_1$-BS_2 and BS_1-$B1_2$-$B2_2$-M3-BS_2 can be adjusted to make a white-light Mach-Zehnder interferometer. Similarly, with the short paths ($B1_1$-$B2_1$ and $B1_2$-$B2_2$) blocked instead, one can adjust one of the trombone arms to achieve white-light interference. In this way, the difference in the path-length differences is guaranteed to be less than a few microns.

[7] This modification was made before the detector D2' was installed.

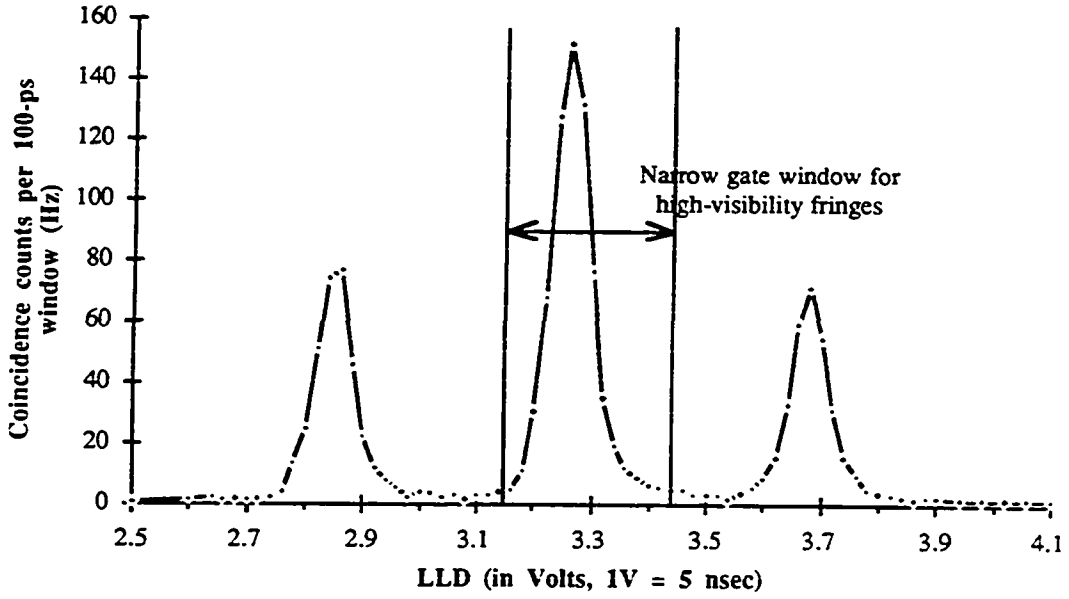


Figure 8.8 Coincidence timing profile, made with TAC and SCA operating in window mode (100-ps window), showing contributions from short-long (left peak), long-short (right peak) and long-long and short-short (central peak) coincidences. The vertical bars indicate the electronic gate window (1.46 ns) used to eliminate the non-interfering background of short-long and long-short coincidences in most of the experimental runs.

the following effects:

i.) Imperfect alignment of the bending mirrors, interferometers, and irises, such that short-short and long-long processes are not entirely indistinguishable.

ii.) Inevitable loss of visibility due to diverging input beams and large path-length imbalance. There were generally about twice as many short-short coincidences as long-long's (this ratio depended critically on alignment), which should reduce the maximum visibility to 94% [8]. (In a later version, described below, we compensated for this by

[8] The optimal visibility of interference between two processes if the probability of one process is reduced by a factor of α with respect to the other is $2\alpha^{1/2}/(1+\alpha)$; see Fig. 8.12)

using neutral density filters in the short paths.)

iii.) Finite size of irises, accepting light from various paths which acquire slightly different relative phases in the interferometers. As evidence for this effect, the approximate fringe visibility improved from 63% to 75% to 80% when the irises' sizes were reduced sequentially from ≈1.5mm to ≈0.8mm to ≈0.4mm in diameter.

iv.) Time averaging over slow drifts in laser frequency (the *instantaneous* linewidth of the pump should be negligible) and/or air temperature in the MZ's. From the fit in Fig. 8.9a one finds that the period of the fringes is ≈ 282 nm over the 1700-second run, differing from the expected value of 351 nm. This is consistent with the quarter-fringe drift over 10 minutes observed in a separate stability test. As a worst-case estimate, we treat the drift as a random walk, finding a phase diffusion coefficient of about $5°/s^{1/2}$, and a visibility reduction of about 1.3%. In runs where the counting times were shorter (~1s), the observed fringe spacing was 348±1nm, in much closer agreement with the expected value.

In Fig. 8.9b we show the simultaneous coincidence results using narrow (1.1ns) and broad (8ns) gating windows, corresponding to the exclusion and inclusion of long-short and short-long coincidences, respectively. In the former case the visibility of the best-fit curve is 76±1%, while in the latter it is 40±1%, in reasonable agreement with the ratio of two predicted by theory. These results demonstrate the cutoff effect which enables one to produce the Bohm singlet-like state of Eq. (8.6b). Comparing Eq.(8.3) with Eq.(8.4), we also expect the average value of these curves to differ by a factor of two. For our fringes the ratio was 2.3. This is partially explained by the fact that the 1.1-ns window actually cuts out about 10% of the long-long's and short-short's, in addition to 99% of the short-long's and long-short's (see Fig. 8.8). Interferometer misalignment also contributes to this discrepancy.

Figure 8.9 (next page) Interference fringes in a Mach-Zehnder Franson experiment. a) Constant singles rate (left axis) and coincidence fringes (right axis, with the accidental rate of approximately 10 (counts per 10 s) subtracted). For this run, the irises were at their minimum size (~0.4-mm diameter). The position data is interpolated from the Heidenhain encoder's 0.1μm-resolution output. The solid curve is a sinusoidal best fit, and has a visibility of $80.4 \pm 0.6\%$.
b) Gate window-size influence on coincidence-fringe visibility [same basic conditions as in a)]. With a narrow window (1.1ns), only the interfering short-short and long-long processes contribute (lower trace, triangles); for an ideal system, 100%-visibility is predicted. With a broad window (8.0ns), the background of non-interfering short-long and long-short coincidences increases the count rate (upper trace, diamonds) reducing the visibility by a factor of two.

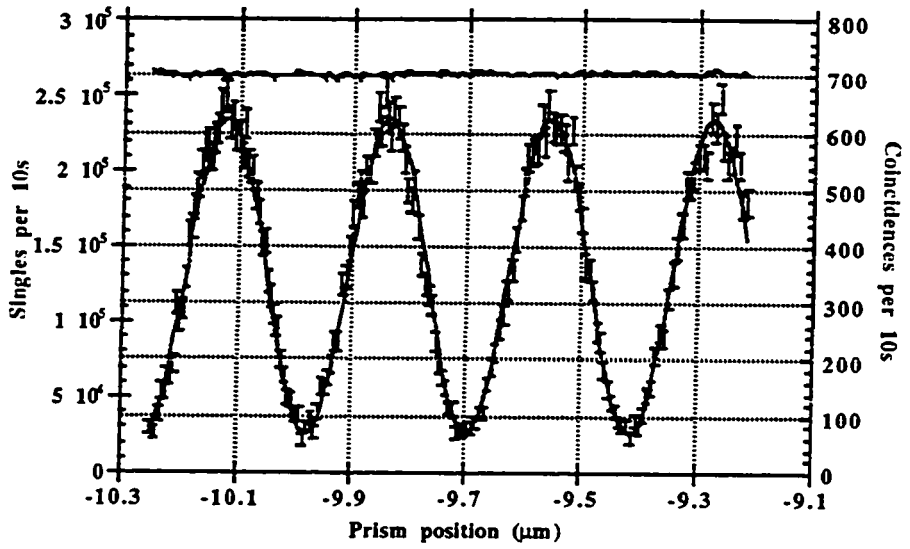


Figure 8.9a

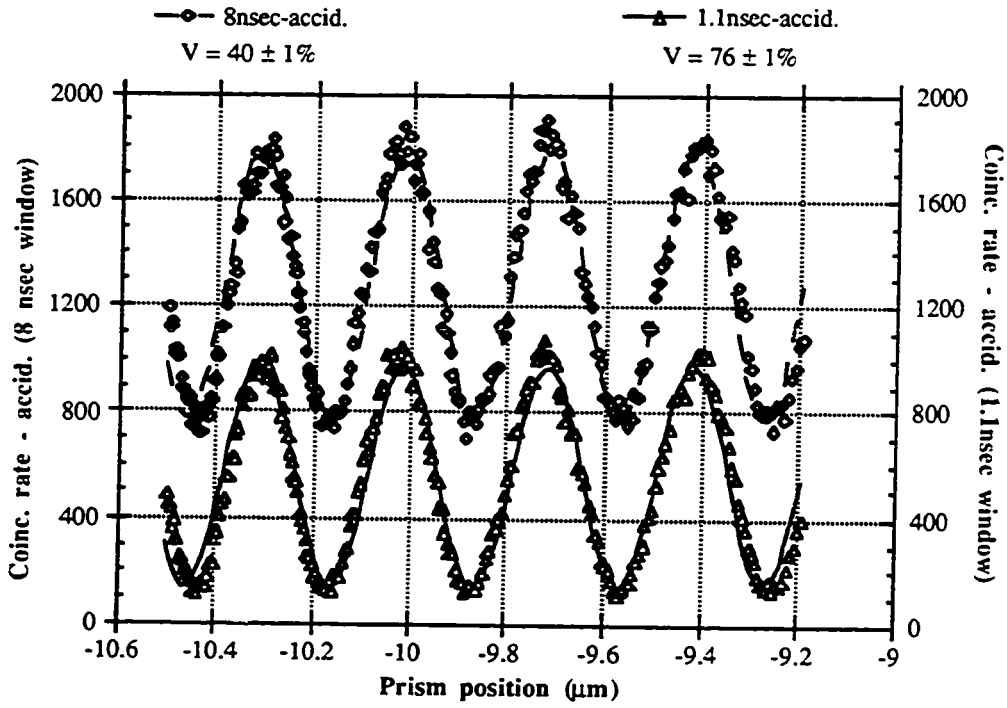


Figure 8.9b

As discussed earlier, coincidence fringes are only expected when $\Delta L < \Delta l_s, \Delta l_i$, the down-converted coherence lengths. By scanning the interferometer out further than about 20 μm, it is possible to observe this degradation (see Fig 8.10). The use of small irises will tend to narrow the effective bandwidth, thereby broadening the envelope slightly. When a 0.86-nm FWHM filter was substituted for one of the 10-nm filters, the envelope was found to broaden to ~350 μm. On the other hand, when both interference filters were replaced by colored-glass cutoff filters, the width did not change appreciably from the 10-nm case, indicating that the effective bandwidth imposed by the 0.4-mm irises is about 10 nm.

In advanced versions of the present experiment both output ports of interferometer 2 were used, to verify that the coincidence rates N_{12} and $N_{12'}$ displayed complementary fringes (and to improve the case for a Bell's-inequality violation, by removing the need for one of the supplementary assumptions; see discussion below). Typical results are displayed in Fig. 8.11.

By placing a variable attenuator in one arm of interferometer 1, we were able to investigate briefly the effect on fringe visibility. Our results, shown in Fig. 8.12, are in fair agreement with the simple theory discussed in Chap. 1, namely, that the optimal visibility of interference between two processes if the probability of one process is reduced by a factor of α relative to that of the other is $2\alpha^{1/2}/(1+\alpha)$.

Believing that the above results (e.g., Figs. 8.9a and 8.11) were limited significantly by an inequality in the number of short-short coincidences and the number of long-long coincidences (the former were typically as much as twice the latter, probably due to beam divergence over the large path imbalance), we installed a neutral-density filter (N.D. ~0.1) in the short arm of each interferometer, for a total short-short attenuation factor of ~0.6. The resulting contributions from the four coincidence processes were then found to be roughly equal. After realigning to compensate for the

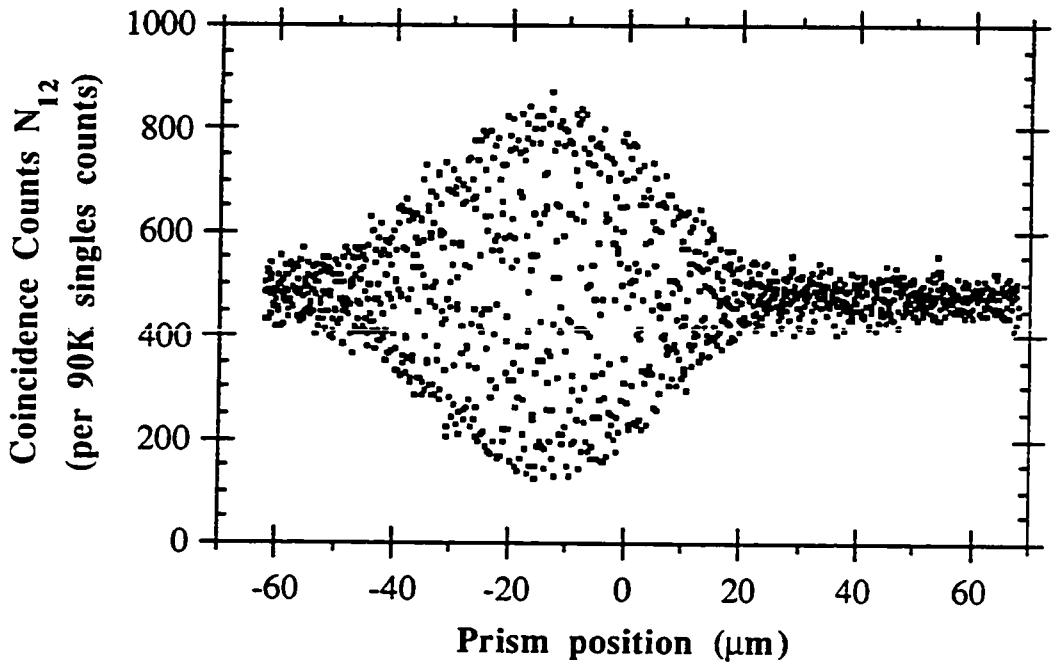


Figure 8.10 Envelope profile of Franson interference fringes, with 10-nm FWHM filters before detectors. For this data the Burleigh system was moved 0.1mm every count (1 count ≈ 90,000/29,100 s collection time + 0.2s dwell time). At this speed it is not possible to discern individual fringes, but the envelope is clearly displayed. The width of the profile (~ 50 μm) is governed by the inverse-bandwidth of the detected photons, and agrees with theoretical predictions based on the filter spectral transmission (cf. Fig. 8.3).

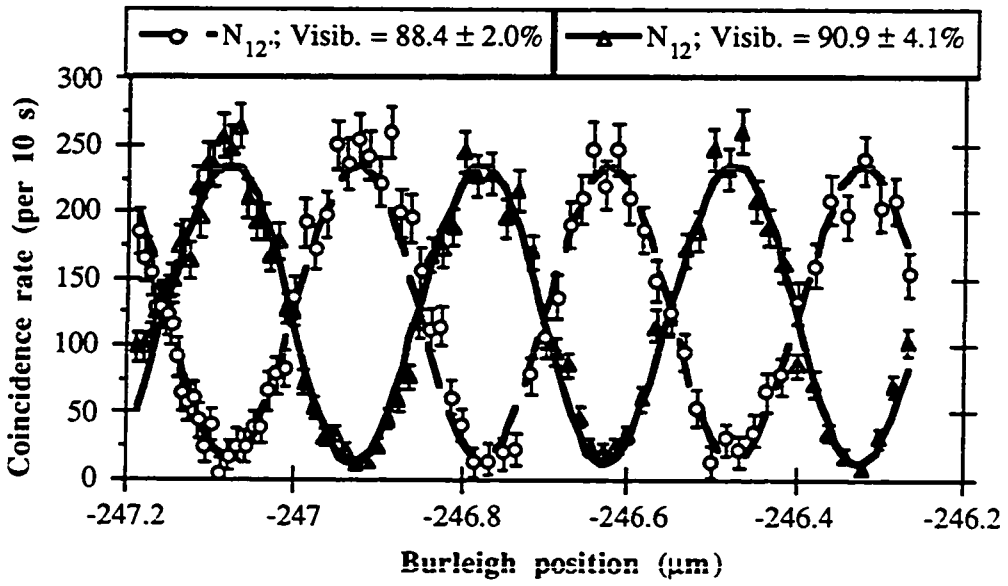


Figure 8.11 Coincidence rates N_{12} and $N_{12'}$ (i.e., using both output ports of interferometer 2) show complementary fringes. Singles rates display no fringes. The curves are sinusoidal fits, with visibilities indicated. Accidental rates of 0.53 s^{-1} (5.6 s^{-1}) have been subtracted from N_{12} ($N_{12'}$). (Attenuators were present in both short arms--see discussion in text.)

extra glass, and removing the cylindrical lens after the crystal (because it did not seem to be helping), we found significantly improved results.

In particular, we performed one final run, to allow a true test of Bell's inequalities. As will be discussed in the next section, as long as only one of the interferometers' phases is varied, then at best one can only infer a violation of a Bell's inequality. Recall from Sects. 7.III and 7.IV that two values of *each* parameter setting are needed. In practice this is not so easy to achieve due to the relatively large backlash present in any linear motion system[9]. The solution was to insert a zero-order quarter waveplate

[9] For example, experimental points in which the first *three* of the four required combinations of parameters (ϕ_1, ϕ_1', ϕ_2, ϕ_2') are determined pose no severe difficulty.

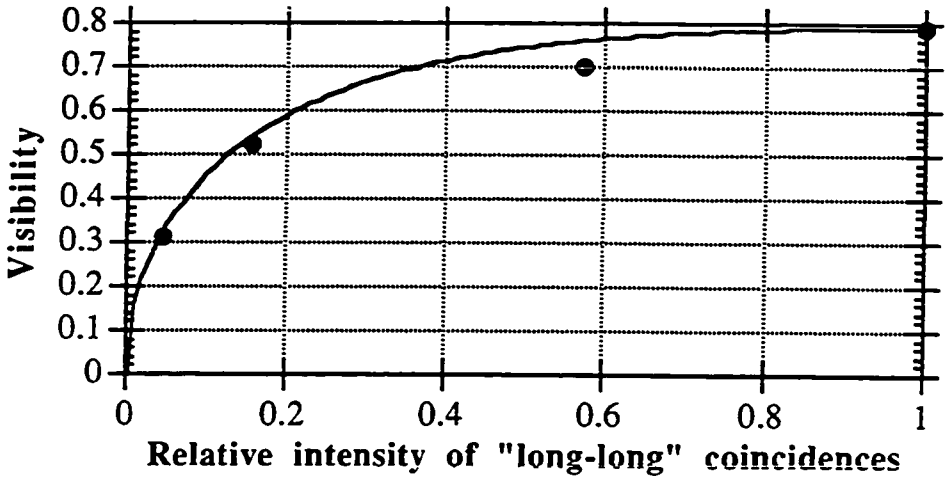


Figure 8.12 Plot of coincidence fringe visibility versus relative attenuation of the long-long processes. Solid curve is the theoretical prediction (cf. Fig. 1.1), normalized to agree at the maximum.

(QWP) into the long arm of interferometer 2 (where interferometer 1 was used for the actual prism-position scanning), as shown in Fig. 8.7a. The relative phase difference between the two possible orientations of the QWP fast axis (horizontal and vertical) is 90°, which is precisely what is needed for a Bell's inequality with our arrangement (see below). The procedure leading to the data displayed in Fig. 8.13 is straightforward--Interferometer 1 is (very) slowly scanned while the QWP is periodically switched from horizontal to vertical. From these results we can violate an energy-time Bell's inequality.

The problem is that to determine the final combination, one of the phase variables has to be reset to its initial value. This cannot be done reliably with typical motion systems (recall that the phase has heretofore been adjusted by piezoelectrically translating the trombone prism of interferometer 1).

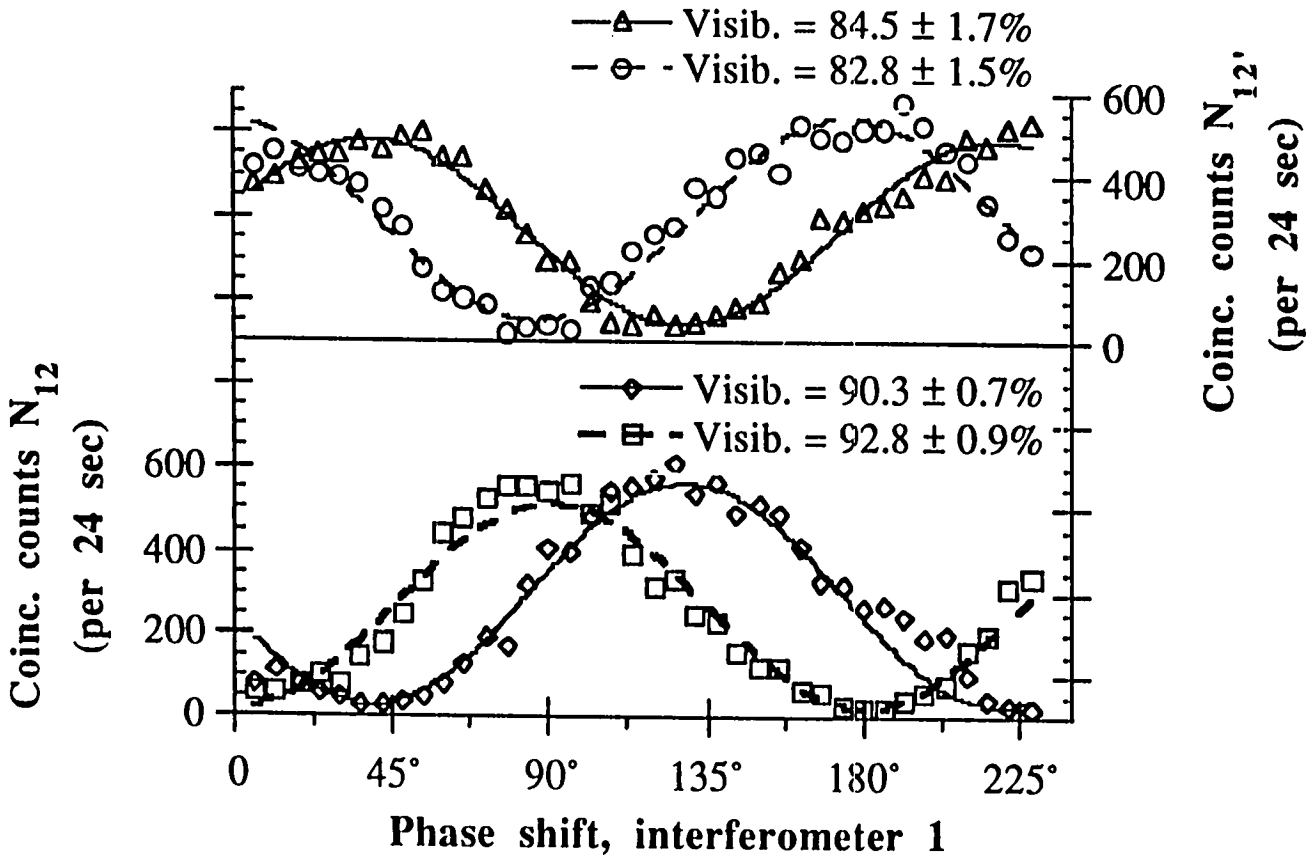


Figure 8.13 Plot of Franson fringes in the coincidence rates N_{12} (bottom curves, left axis) and $N_{12'}$ (top curves, right axis), versus phase shift from translating the trombone prism in interferometer 1 (see Fig. 8.7a). The data sets indicated by squares and circles correspond to one orientation of a quarter waveplate in interferometer 2; the data sets indicated by triangles and diamonds correspond to the orthogonal orientation of the same waveplate. Best-fit sinusoidal curves are shown, along with the visibilities. Accidental rates of 0.6 s^{-1} (5.6 s^{-1}) have been subtracted from N_{12} ($N_{12'}$).

8.V. Interpretations and Bell's Inequalities

As discussed in Sect. 8.III, 50% is the maximum visibility possible in a classical field approach to this sort of experiment. Our Mach-Zehnder results therefore preclude any classical description, without the need to include results from other experiments. Although we did not do this, it is clear that one could perform a different check of non-classicality by measuring the anti-correlation parameter a, as described in Chap. 3. Recall that a ($\Rightarrow \frac{N_1 N_{122'}}{N_{12} N_{12'}}$) must be greater than 1 for any classical source (see also Appendix A). In words, the number of triple coincidences for all three detectors times the singles rate (N_1) at the unpaired detector (D1) should be greater than the product of the double coincidences (N_{12} and $N_{12'}$) between this detector and the two paired detectors (D2 and D2'), for any classical source[10]. Instead, we have seen that the indivisibility of our single photons leads to an anti-correlation between the *individual counts* at the detectors D2 and D2', so that the number of triple coincidences is much less that than predicted by any classical field theory. This feature of non-classicality implied by a low value of a is very different than the feature implied by a violation of Bell's inequalities. The former results from the indivisibility of a single quantum, the latter from the nonlocal correlations implicit in an entangled state. We turn now to a

[10] The experimental arrangement is quite different here than in Chap. 3 for the single-photon Berry's phase experiment. In particular, the detectors D2 and D2' are in the two *different* output ports of interferometer 2, whereas in the earlier experiment the detectors both looked at the *same* output port of an interferometer (the fact that it was a Michelson instead of a Mach-Zehnder is not relevant here), after an extra beam splitter. Therefore, in the earlier experiment the two coincidence *rates* between these detectors and the unpaired detector behaved in a correlated fashion (i.e., the interference fringes were exactly in phase); in contrast, here the two coincidence *rates* N_{12} and $N_{12'}$ are anti-correlated (i.e., the interference fringes are exactly out of phase). Nevertheless, the value of a is still constrained to be at least 1 for any classical field theory. In fact, the present experimental setup is much more like that reported by Grangier, et al. in the article which introduced the anti-correlation parameter [Grangier et al., 1986].

discussion of Bell's inequalities for our experiment.

The results using the switchable quarter waveplate demonstrate that the coincidence rates depend only on the *sum* of the phase angles, as predicted by quantum mechanics. In general, however, the derivation of a *testable* Bell's inequality requires making additional reasonable assumptions, as discussed in the previous chapter. {In particular, for detection efficiencies below about 83%, the assumption must be made that the observed counts are "representative" of the entire ensemble, including those particles which escape detection. This "detection loophole" can never be evaded in the present version of the Franson experiment due to the necessity of discarding half the counts, i.e., the short-long's and long-short's (see, however, the proposal in the next section). The other principal assumptions are that the state of the particles emitted by the source is independent of the parameter settings at the analyzers, and that any correlations in the detector outputs arise solely from the entangled photons.} Once this is done, the inequality can be violated when the coincidence rate varies sinusoidally[11] as a function of the difference of the "parameter settings" at the two spatially separated analyzers, as in the QM prediction Eq. (8.4). (Typically, these settings are polarizer orientations, but in our experiment, one is $+\Delta L_1$, and the other is $-\Delta L_2$.) As discussed in Sect. 7.III, this violation requires the visibility to exceed $1/\sqrt{2}$, or about 70.7%. The fringes we have observed therefore imply a violation of a Bell's inequality by up to 16 standard deviations. When observations are only made at two or three of the four output ports, additional assumptions must also be made about the rates at the unused ports, based on the symmetry of the experiment. In particular, we must assume that the unobserved coincidence rate $N_{1'2'}$ (with a fourth detector looking at the remaining output port of

[11] Recall (Sect. 7.VI) the restriction that the rate vary *sinusoidally* with the sum of the two path-length differences is essential. It is fairly easy to concoct LHV models which predict 100%-visibility *triangular* fringes (Fig. 7.4), or fringes proportional to the *product* of separate sinusoidal functions of the two parameters (Fig. 7.3). These models do not violate the Bell inequality.

interferometer 1) displays complementary interference fringes to those displayed in $N_{12'}$, just as those of N_{12} and $N_{12'}$ are complementary (see Fig. 8.11).

Using the results of Fig. 8.13, we can calculate explicitly the value of S, as given in Eq. (7.7b):

$$S = E(\phi_1,\phi_2) + E(\phi_1,\phi'_2) + E(\phi'_1,\phi_2) - E(\phi'_1,\phi'_2), \qquad (8.12\text{-}7.7b)$$

where for any LHV theory, $|S_{LHV}| \leq 2$. Given that our net detection efficiencies are at best a few percent (due to small iris size), we must make a supplementary "fair sampling" assumption to calculate the correlation coefficients $E(\phi_1,\phi_2)$. We will use the following definition:

$$E(\phi_1, \phi_2) = \frac{N_{12}(\phi_1, \phi_2) - N_{12'}(\phi_1, \phi_2)}{N_{12}(\phi_1, \phi_2) + N_{12'}(\phi_1, \phi_2)} \qquad (8.13)$$

which is equivalent to that in Eq. (7.17b), if one assumes that the unobserved rates $N_{1'2'}$ and $N_{1'2}$ behave like the observed rates: $N_{1'2'}(\phi_1, \phi_2) = N_{12}(\phi_1, \phi_2)$ and $N_{1'2}(\phi_1, \phi_2) = N_{12'}(\phi_1, \phi_2)$, as predicted by quantum mechanics. The implicit assumption in using (8.12) is similar to that discussed in Sect. 7.V:

> For every photon in the state λ, the sum of the detection probabilities in the two output ports of each Mach-Zehnder does not depend on the phase shift in the interferometer.

The fact that we observed complementary fringes in N_{12} and $N_{12'}$ (Fig. 8.11) confirms this assumption for interferometer 2, *for the coincidence events we observe*[12]. We calculate (8.13) for the two points (the two waveplate orientations) at ~22.5° and for the two points at ~67.5°. Combining the results as in (8.12), we find $S_{exp} = 2.63 \pm 0.08$, which clearly violates the Bell's inequality. For comparison, 100%-visibility sinusoidal fringes would yield $S_{max} = 2.83$.

[12] The whole essence of a fair sampling assumption is that one assumes that the counts one doesn't observe behave like the counts one does.

There are several possible interpretations of the Bell's inequality for this experiment. The initial proposal named it a test of a "position and time" inequality. The related non-polarization-based experiment by Rarity and Tapster [Rarity and Tapster, 1990] describes a "momentum-phase" inequality. In a private communication, Caves and Braunstein have maintained that the variables in question are simply *which port* of the final beam splitter each photon exits. We describe the present experiment in terms of a Bell's inequality concerning energy and time. As discussed earlier, the underlying mechanism for the observed interference is the strong correlations of both of these variables in the conjugate photons. An example of an LHV model inconsistent with our results is one which ascribes to each photon some definite but unknown energy, but allows this energy to differ among the members of the ensemble, thus washing out any fringes in singles detection. This is in fact a general version of the classical model given in Sect. 8.III, for which the visibility was limited to 50%. If, instead of the standard sinusoidal transmission-probability, we postulate that the probability P for a photon (of well-defined, sharp frequency) to leave a particular interferometer port is given by the standard result, *rounded* to the nearest integer (so that P equals only zero or one; cf. Eq. 7.21), then we saw in Sect. 7.VI that *triangular* coincidence fringes result. Although they may display 100%-visibility, Bell's inequality is satisfied, and no nonlocality is implied.

In summary, if one accepts the reasonable supplementary assumptions (fair sampling, no "influence" of interferometer phase settings on the source, and symmetry considerations to account for the unobserved port), then the conclusion of the experiment is that Nature is nonlocal. Local hidden variable models in which the time of emission or energy of the down-converted photons are well-defined before detection cannot reproduce the observed correlations. In colloquial terms, we can say that the photons know neither their birthdays nor their color.

8.VI. Future Possibilities

Regarding future directions of the present experiment, we discuss only briefly two possible extensions of the current work. Other researchers have also studied the possibility of employing the nonlocal correlations in various communications and cryptography schemes [Ekert, 1991; Rarity and Tapster, 1992; Brassard, 1988; Bennett, 1992; Ekert et al., 1992]. In terms of a loophole-free Bell's inequality experiment (see Sect. 7.V), the present embodiment of the Franson experiment is insufficient, because in order to achieve the greater than 71% visibility needed for a violation, one must perforce discard the short-long and long-short coincidence events. This automatically reduces the maximum detection efficiency to 50%, which is insufficient to close the detection-loophole. In Chap. 10, we propose a means of generating a true singlet-like polarization-entangled state, *without the need of discarding half of the counts*. In Fig. 8.14, we show how such a source could be used in conjunction with a modified Franson scheme to perform an energy-time Bell's inequality experiment in which there are no short-long's or long-short's to discard. Basically, one relies on the polarization correlations to force the photons to always take the same-length paths in their respective interferometers.

As a second extension of the present work, consider the scheme depicted in Fig. 8.15, which would permit one to produce a pair of photons which were *doubly-entangled*, that is, simultaneously entangled in polarization and in energy. More precisely, the proposed setup would allow one to measure both polarization correlations (using polarizers before the detectors) and energy correlations (using the Franson interferometers). The polarization entanglement is established using the same technique as discussed in Chap. 5--one of two horizontally-polarized photons incident on opposite sides of a beam splitter is made vertically polarized. Neglecting cases in which both photons leave the same port of the beam splitter, an effective singlet-state (cf. Eq. 7.16)

is produced. It is not yet clear what, if any, advantages are gained with such a system, and more study is certainly needed. However, it is our hope that such a double-entanglement may lead to even greater violations of a Bell-type inequality.

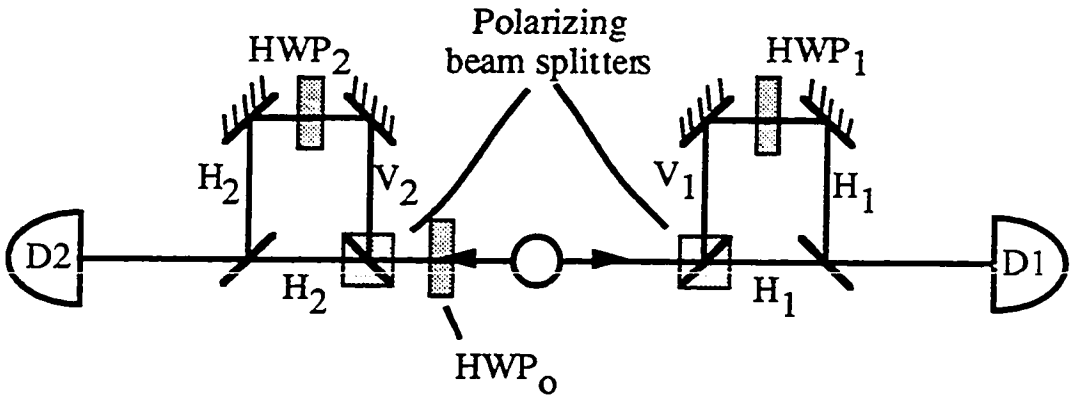


Figure 8.14 Use of a polarization-singlet state to perform a Franson experiment in which there is no background of short-long or long-short coincidences. The state after HWP_o is $|\psi\rangle \approx (|V\rangle_1 |V\rangle_2 - |H\rangle_1 |H\rangle_2)/\sqrt{2}$. Polarizing beam splitters force photons to take the same paths; i.e., either both take the long paths or both take the short paths. Half waveplates within the interferometers effectively remove polarization as a which-path label.

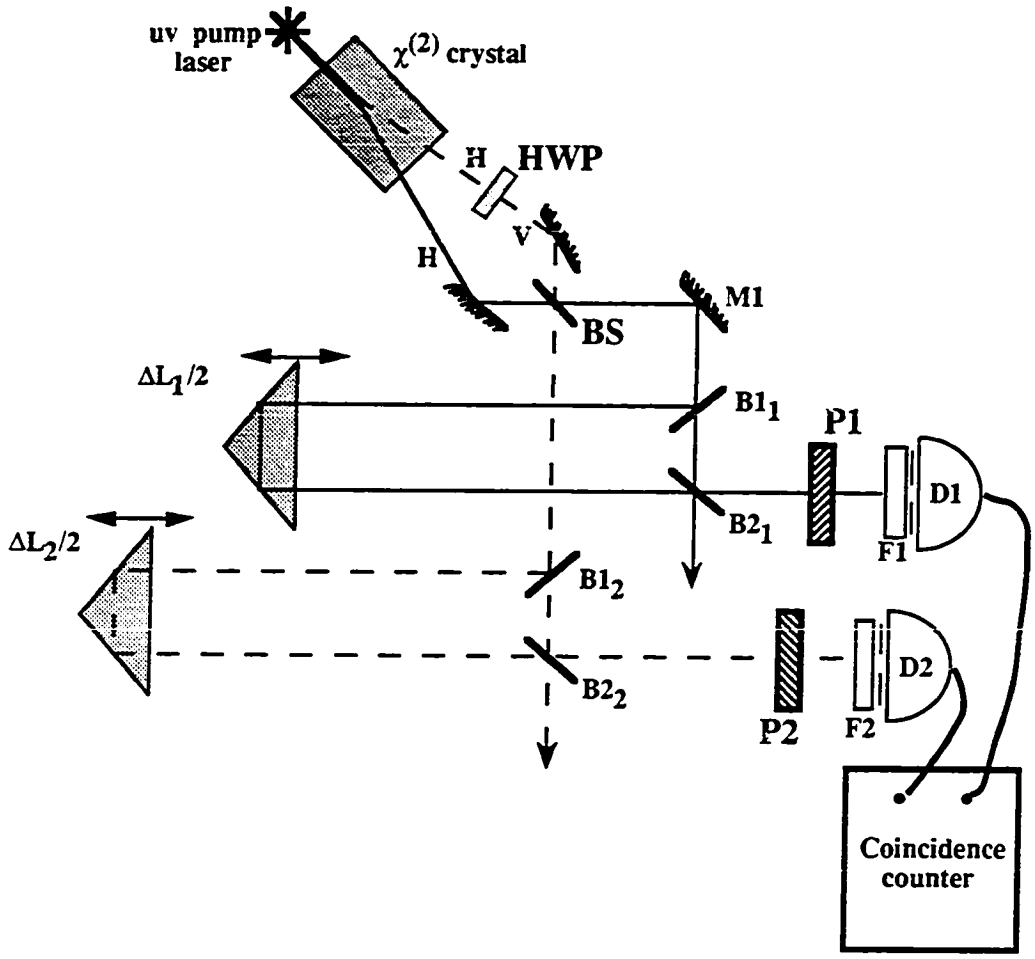


Figure 8.15 Possible setup to investigate combined polarization/energy-time entanglements. The source emits correlated pairs of horizontally- (H) polarized photons. The half waveplate HWP [which rotates the polarization to vertical (V)] and beam splitter BS produce the polarization-entanglement[13], which we examine using polarizers (or alternately, polarizing beam splitters) P1 and P2. The energy-time entanglement is again investigated with a Franson-type arrangement. Ideally, one would have detectors in all four output ports (detectors not shown are indicated by arrows).

[13] As in the other Bell's inequality experiments that use this scheme [Ou, 1988; Shih, 1988], processes where both photons leave the same port of the beam splitter must be discarded to produce an effective singlet state (cf. the discussion in Sect. 7.V).

"Waiter, there's a PROOF in my pudding!"
"But Sir, you haven't even TASTED it yet"

Chapter 9: High-Efficiency Single-Photon Detectors

9.I. Introduction

Recently there has been a growing interest in fast, high-efficiency single-photon detectors. High efficiencies are of course desirable in all applications relying on photon counting, such as photon correlation spectroscopy and velocimetry [Cummins and Pike, 1974; Cummins and Pike, 1977], optical time-domain reflectometry [Ripamonti et al., 1990], and laser ranging, as well as in investigations of novel quantum-mechanical interference phenomena at the single photon level. However, they are absolutely necessary for practical implementation of various quantum cryptographic schemes, as well as in proposed loophole-free experiments to demonstrate violations of Bell's inequalities, which until recently required detection efficiencies greater than 83% (by considering the use of modified quantum entangled states, this limit has now been reduced to 67% [Eberhard, 1993; see also Chaps. 7 and 10]. In a very real sense, the value of any loophole-free test (the "proof of the pudding") is determined by the detector efficiency (the "tasting").

Measuring the quantum efficiency of a detector at the single-photon level can be problematic. Typical tests involve a *relative* measurement of the efficiency, comparing to a "known" standard detector. In some schemes, the very weak photo-current out of a diode is measured, and arguments are made equating current to detection events. However, any such test is subject to uncertainties. Fortunately, as discussed in Chap. 2, the correlated photons from parametric down-conversion are ideal for making measurements of the *absolute* detection efficiency and temporal response of single-photon counting detectors, such as photomultipliers and avalanche photodiodes[1]. The

[1] It is important to distinguish between single-photon detectors and certain

basic idea is as follows: Because the photons are always produced in pairs, if one is detected at a "trigger" detector, then we can know with certainty that its conjugate is incident upon a second detector, whose efficiency we wish to measure. The highest single-photon detection efficiencies to date have been observed using avalanche photodiodes in the Geiger mode; until now these have been limited to about 40%. We have measured efficiencies as high as 76%, and there are indications that these may be improved to 80% or even 90%. In particular, we describe in this chapter a series of measurements [Kwiat et al., 1993a; Kwiat et al., 1993b] made on a pair of Solid State PhotoMultipliers (SSPM) from Rockwell [Petroff et al., 1987], and on a pair of Single Photon Counting Modules (SPCM-200-PQ) from EG&G [Lightstone et al., 1989; Lightstone and McIntyre, 1988; see also the specifications in Appendix E]. The highest adjusted efficiencies measured were 70.9±1.9% and 76.4±2.3%, with an SSPM and SPCM, respectively. However, subsequent tests on the SSPMs revealed possible damage in the input fibers, and there were substantial reflection losses within the SPCMs. At present it is difficult to exactly specify the appropriate correction factors. Nevertheless, the results are very encouraging, with efficiencies nearly twice those previously reported.

In Sect. 9.II we review briefly the need for high-efficiency single-photon detectors in quantum cryptography and loophole-free Bell's inequality experiments. Our down-conversion source has already been described in Chap. 2; Section 9.III discusses the use of such a source to measure absolute quantum efficiencies. Descriptions of the detectors, detailed experimental setup, and typical procedure are given in Sections 9.IV - 9.VI, respectively. In Section 9.VII, our efficiency results are presented; we also attempt to list the various sources of systematic error, corrections to apply to the data,

---

photodetectors (e.g. Si or InGaAs PIN photodiodes) which may be very fast and display efficiencies greater than 95% [Rarity et al., 1992], but which possess too much intrinsic noise to be useful at the single-photon level.

and the final corrected efficiencies. In Section 9.VIII the results of several related measurements are described, including the time-response characteristics of the detectors, the effects of saturation, and the presence of afterpulsing in the SPCMs. Our conclusions are presented in Section 9.IX. Two approaches to the statistical error analysis appropriate for our results are presented in Appendix C.

9.II. Motivation

As discussed in Chap. 7, quantum cryptographic schemes rely on using either single-photon states or very weak coherent-state pulses, with an average photon number $<< 1$. The discussion in Sect. 7.VII dealt only with perfect analyzers and detectors. Polarizers with cross-talk (or imperfect interferometers, for the phase example) and intrinsic noise within the detectors will introduce errors, and non-unity detection efficiency will increase the number of useless events where no photon was detected. Moreover, imperfect detectors will also require greater statistics to determine the presence of an eavesdropper, and place higher demands on any error-correcting codes. Clearly, high-efficiency, low-noise detectors are needed. Also, the data rate is obviously limited by the speed of the detectors.

While the cryptographic applications need high single-photon detection efficiencies to be practical, there is no intrinsic level below which the schemes categorically fail; in contrast, we have seen that certain experiments on quantum nonlocality possess a specific efficiency cutoff -- below this cutoff no completely rigorous tests may be made. In particular the lower limit on detector efficiencies necessary to close the detection loophole is 67% (assuming all other aspects of the experiment to be ideal). However, due to other limitations and non-idealities, efficiencies in excess of 90% are desired. Our present results demonstrate that such efficiencies are already available, and suggest that they may be made even higher.

9.III. The Method

Because photons produced in spontaneous down-conversion are *always* produced in *pairs*, with strong time correlation, they are ideal for absolute calibrations of photon detectors-- the detection of one of the photons guarantees with certainty the presence of a photon at the "conjugate" detector. The technique, first proposed by Klyshko [Klyshko, 1980], and used by Rarity *et al* [Rarity et al., 1987] to determine the efficiency of an RCA (now EG&G) silicon avalanche photodiode, is simple: Direct one photon of each pair to a "trigger" detector, and arrange the optics in order to catch all of the "conjugate" photons with the detector whose efficiency is to be determined (see Fig. 9.1). For some time interval T both the number of singles at the trigger detector and the number of coincidence counts between the detectors are measured. If the total number of emitted pairs is N, then the number of singles seen at the trigger detector is $N_t = \eta_t N$, while the number of coincidences is $N_{ct} = \eta_c \eta_t N$, where η_t (η_c) is the detection efficiency at the trigger (conjugate) detector[2]. Assuming no external losses on the path to the conjugate detector, one can determine the efficiency η_c by simply taking the ratio N_{ct}/N_t.

In practice the above simple formula needs to be modified to account for the presence of unwanted counts and the loss of desired counts. First, in addition to the desirable correlated photons, each detector will also possess a number BG of background counts, consisting of unwanted external light, dark counts within the detector, and possibly electronic noise. The true number of singles N_t is then the measured quantity S_t minus the background BG_t (determined in a separate measurement with the down-converted beams blocked for a duration T_{BG}), weighted by the ratio of the two measurement times, $r_{BG} \equiv T/T_{BG}$. Similarly, $N_c = S_c - BG_c r_{BG}$, where the subscripts refer to counts at the conjugate detector. Second, there will be spurious

[2] Due to saturation effects (arising from intrinsic deadtime in the devices), the efficiency actually depends on the incident light intensity. (See Section 9.VIII.) However, for simplicity we use η to denote the efficiency in the low-light limit.

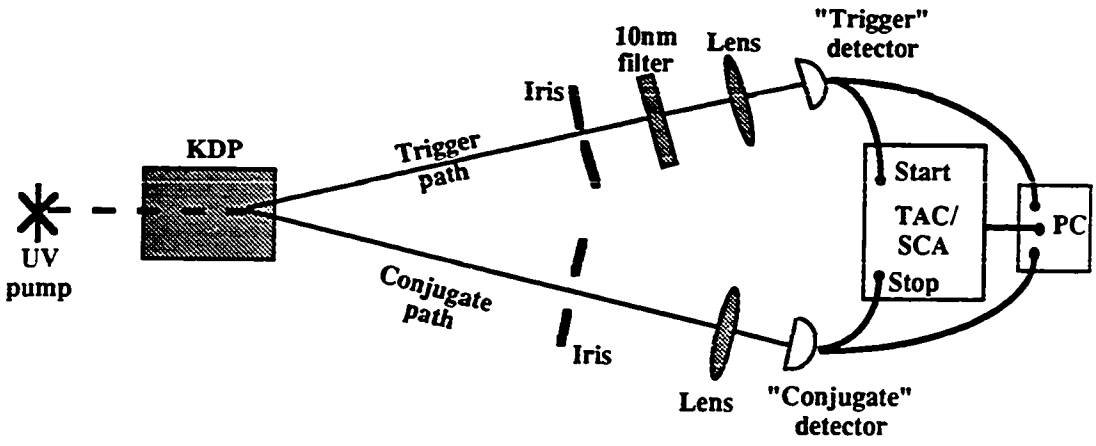


Fig. 9.1. Simplified schematic of setup to measure absolute efficiencies. One photon of a down-converted pair is directed toward the "trigger" detector; the other to the detector whose efficiency we wish to determine. If we are certain that all conjugate photons arrive at the detector (i.e., no intravening losses), then the efficiency may be obtained as a ratio of coincidence counts to trigger singles counts (after accounting for background and accidental coincidences).

coincidence counts due to the finite duration, w, of the coincidence window. The true number of coincidences, N_{ct}, is then the measured value C minus the "accidental" counts, $A \approx \left[S_c S_t - \eta_c S_c\left(S_t - BG_t r_{BG}\right)\right] w/T$ (see Appendix C). In practice, one uses as small a coincidence window as possible (on the order of tens of nanoseconds), while still catching all of the true coincidence counts. (Note: Some photodetectors are known to possess "trapping states", so that some fraction of the incident photons produce *extra* electronic signals at some undetermined time (which can be as large as seconds) after the typical time for such a signal [Cummins and Pike, 1974]. The presence of such afterpulses in the trigger rate reduces the measured conjugate efficiency, because they appear as "real" trigger photons *without* conjugate partners. For our trigger detector (an SPCM) less than 0.01% of the counts were echoed in this fashion, so the effect on the

efficiency measurements was negligible. See Section 9.VIII.)

The last major correction factor stems from loss of photons *before* they reach the detector, due to unwanted reflections, scattering, or absorption. (For example, any uncoated glass-air interface will yield about a 4% loss. These may be anti-reflection coated to reduce this loss significantly. The detector surfaces themselves, if uncoated, may cause sizable losses from Fresnel reflections.) Thus the net detection efficiency η_c for each photon may be written as the product of the detector's efficiency η^d_c and the path "transmissivity" η^p_c.

Combining all of these effects, one can now calculate the conjugate detector efficiency in terms of other, measurable, quantities:

$$\eta_c^d = \frac{1}{\eta_c^p} \frac{C - A}{S_t - r_{BG} BG_t} . \qquad (9.1)$$

Notice that the efficiency associated with the trigger photon does not appear in the above expression. Therefore, one strives to reduce BG_t as much as possible (using narrow-band interference filters and small irises before the trigger detector), even though this may reduce the path transmissivity to the trigger detector.

9.IV. The Detectors

We performed measurements [Kwiat et al., 1993a,b] on four single-photon detectors: two custom-modified Single Photon Counting Modules (SPCM-200-PQ), from EG&G Canada, Ltd., and two Solid State PhotoMultipliers (SSPM) from Rockwell International Corporation[3]. The SPCMs use thermoelectrically-cooled

[3] The SSPMs were developed at Rockwell Science Center, Anaheim; the specific devices used in the measurements reported here were prepared for UCLA as part of a DOE-sponsored program directed by M. Atac (of Fermilab and adjoint Professor of Physics, UCLA). We would like to thank Professor Atac for generously allowing the use of the SSPMs.

silicon avalanche photodiodes (APDs) in the Geiger mode, with passive quenching. These photodiodes have a small active area [~(100 mm)^2]; along with an improved gettering method in their production, this leads to a typical dark count rate of only 60 s^{-1}. Our model was customized for 30-V biasing over the breakdown voltage of 400 V, to provide faster time response and higher efficiency. A detection event gives rise to a 50-ns, 2-V pulse, followed by a dead time measured to be 0.5 μs. The SPCMs possess an effective "k" (the ratio of hole to electron ionization coefficients, averaged over the device structure) of only 0.002, to be contrasted with the k = 0.02 for typical photon-counting avalanche photodiodes (e.g., RCA 30902 silicon APD). It has been shown theoretically that the probability for a single photoelectron to cause an avalanche depends on the value of k, with lower k's leading to higher gains, other things (e.g., excess voltage, depletion layer width) being equal [Lightstone et al., 1989; McIntyre, 1973]. For our SPCMs, it is estimated that the ratio of δ (the field-dependent electron-ionization coefficient, integrated over the depletion width) at the operating voltage to δ at the breakdown voltage is greater than 1.1, so that detector efficiencies in excess of 80% are not unexpected [Robert McIntyre, EG&G Canada, private communication].

The SSPMs are silicon devices based on impurity-band-to-conduction-band impact-ionization avalanches. These avalanches contain about 5 x 10^4 electrons localized within areas several microns in size. Because impurity bands are involved, much lower field strengths are required for impact ionization than in standard avalanche photodiodes, which use valence-to-conduction-band impact ionization; the typical SSPM bias is only 7 V. Also, because they do not operate above the avalanche breakdown voltage, these devices are capable of distinguishing between single-, double-, etc. photon detections (i.e., two simultaneously detected photons will produce an output signal approximately twice as large as a single photon). SSPMs have previously been demonstrated (using a relative efficiency measurement) to possess greater than 50%

detection efficiency in the visible-light region, greater than 30% at 20 μm, and are predicted to have possible efficiencies greater than 90% [Petroff et al., 1987]. The dark counts were typically 7000 s^{-1} for these detectors, which have an active area ~$(1\ mm)^2$. (Note that the dark count rate per unit area is comparable to that of the SPCM.) For optimum performance the SSPMs require cooling to 6 K.

9.V. Experimental Setup

A schematic of one of the experimental setups used is shown in Fig. 9.2, with each component numbered; below we discuss each in turn, including Manufacturers and Part No's. in some cases. Losses associated with these components will be discussed in Section 9.VII. A slightly simpler setup, with no mirrors, (i.e., similar to the schematic in Fig. 9.1) was used for some of the measurements with the SPCMs.

1. An argon-ion laser [Coherent Innova 200] produced the vertically-polarized (polarization purity of 100:1) 351.1-nm pump beam, with a full-angle beam divergence of 0.3 mrad, and a beam diameter at the crystal of 2 mm. A "black glass" filter and an iris (2.2 mm) were used to removed unwanted laser fluorescence. Fixed and variable neutral density filters were used to control the input power. The net attenuation was typically 4 OD, so that the input to the crystal was typically 10 mW.

2. After passing through the crystal, the remaining, unscattered uv beam was directed by a small mirror to a beam dump.

3. A 10-cm long KDP crystal [Cleveland Crystal, custom order], cut for type I phase-matching, was oriented with its optic axis at 50.7° with respect to the pump beam. We measured the efficiencies using two sets of conjugate photons: 702 nm-702 nm and 633 nm-788 nm (the second wavelength is the color directed to the trigger detector).

The crystal itself was housed in a sealed aluminum container, with index-matching

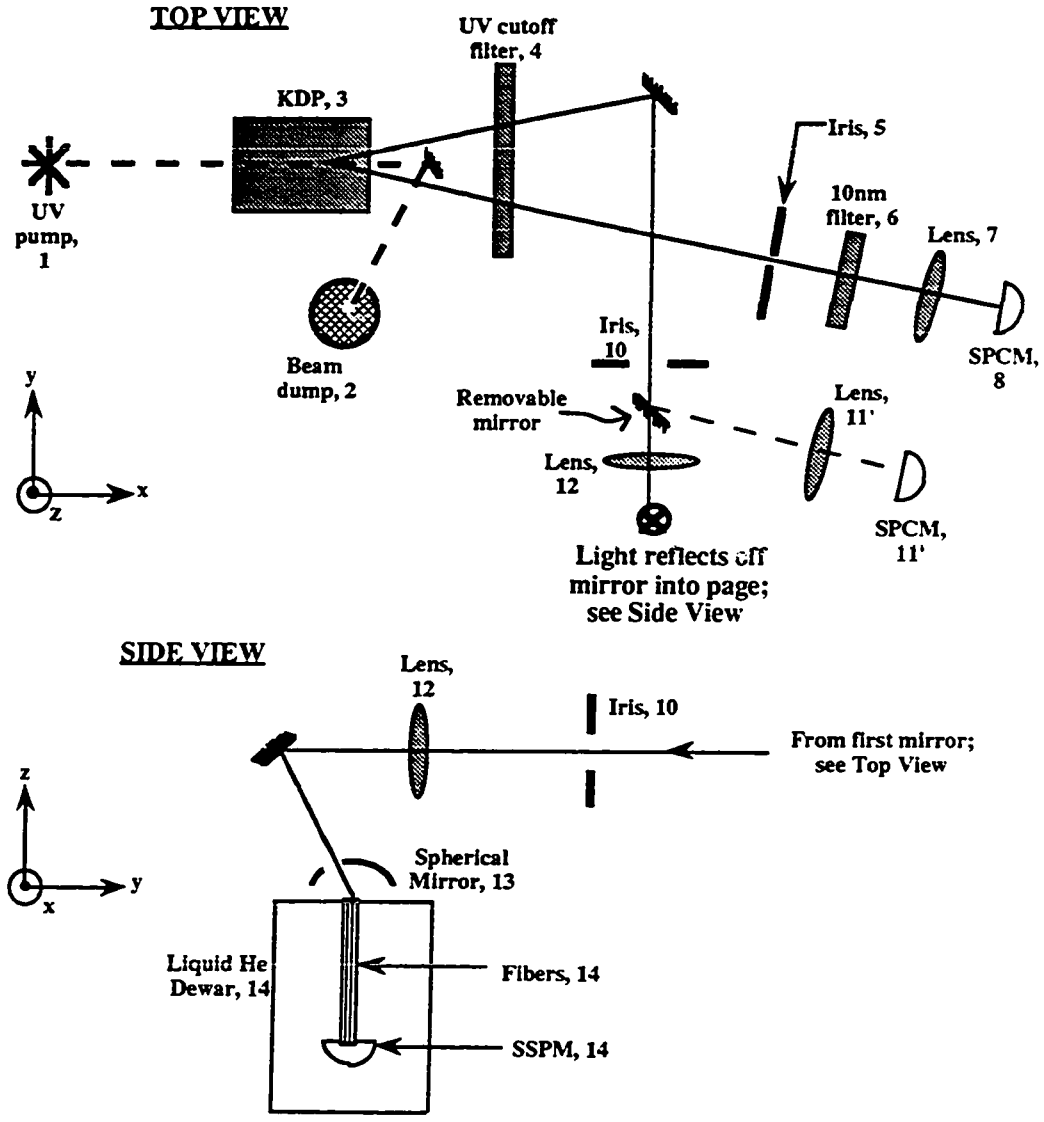


Figure 9.2 Detailed schematic of one of the setups used to measure absolute efficiencies. Specifics of each numbered component are given in the text.

fluid (n = 1.29) and fused silica end windows. The crystal input face was anti-reflection (AR) coated for the uv (single-layer of MgF_2, 1/4 wave at 351 nm); the output face was broadband anti-reflection (BBAR) coated to reduce output losses over the range 400-750 nm. Note: In some of the runs these directions were accidentally reversed, leading to an extra correction factor. See Section 9.VII.

4. A uv cutoff filter [Schott GG-475] was used to remove any scattered uv and discharge tube fluorescence from both down-converted beamlines. This filter was BBAR coated to have high external transmittance (>99%) in the range 600-800 nm.

We now trace the components on the path to the trigger detector, an SPCM.

5. A small iris (1.4 mm) 1.5 m from the crystal was used to select out the trigger photons. Due to phase-matching constraints in the down-conversion process, this iris determines the *direction* of the conjugate photons. That is, it selects out a small angular portion of the down-conversion cone.

6. A 10-nm FWHM interference filter [Melles Griot #03FIV024 for 702 nm; Ealing #35-4373 for 790 nm (≈788 nm)] was used to further select the trigger photons, and to remove most of the remaining background light. Due to energy conservation at the KDP source, this filter effectively determines the color (and hence, opening angle) of the conjugate photons.

7. A BBAR-coated 35-mm-focal length achromatic lens [Edmund Scientific #32319] was used to focus the light onto the small active area [~(100 mm)^2] of the trigger detector. The distance from the KDP crystal to the lens was ~152 cm, while the distance from the lens to the detector was ~3.5 cm. Therefore, in the geometrical optics limit, one expects the spot size at the detector to be ~3.5/152 x (2 mm) ≈ 50 μm, where we have used 2 mm for the size of the trigger beam at the crystal. We have verified this focusing technique by making transverse profiles with the SPCM. See Fig. 9.3. The

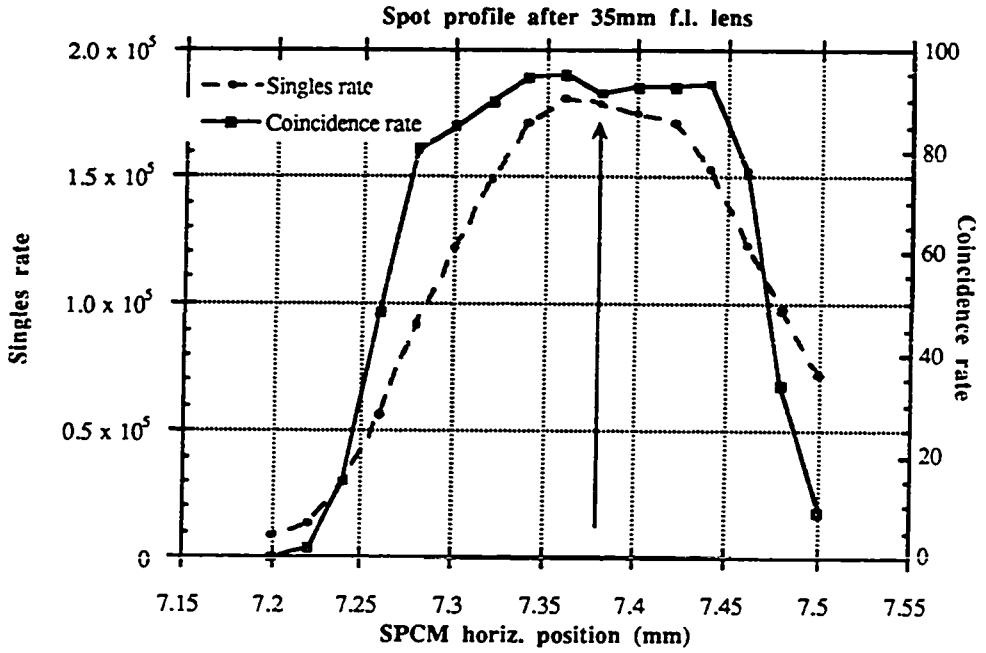


Figure 9.3 Transverse profile of singles and coincidence counts in the focal plane of a 35-mm lens. The 150-μm plateau (in singles) is from the detector size. (In coincidence two detector sizes are convolved.) A spot size of ~70 μm is implied.

profiles possessed a fairly flat plateau of ~150 μm (the size of the detector), with sharp, rapidly falling edges, implying a spot size of ~70 μm.

8. The trigger detector, a Single Photon Counting Module [EG&G #SPCM-200-PQ], was aligned transversely and longitudinally (via micrometers) to maximize the singles count rate. The silicon diode surface was AR coated with a quarter-wave layer (at 630 nm) of SiO; the glass window of the housing was uncoated.

The output of the SPCM was inverted [EG&G #IT100], split [Mini-Circuits #15542,ZSC-2-4 3dB splitter], and directed to a counter [Stanford Research Systems #SR400] (yielding the singles rate at the trigger detector), and to the START input of a Time to Amplitude Converter (TAC) [Tennelec #TC862].

9. To minimize the background a black cardboard box (not shown in figure) was placed over items 6-8, with a small opening to allow the trigger photons to enter. It was then found that the background rate was essentially indistinguishable from the intrinsic dark count rate.

We now trace the components on the path to the conjugate detector either an SSPM (unprimed numbers) or a second SPCM (11'), depending on the presence of a removable dielectric mirror [Melles Griot #02MFG001/001; with MAXBRIte high-reflectance coating]. The location of the SSPM was chosen in part for convenience in optical alignment, and in part to keep the electronics associated with the SSPM far from other sources of electrical noise, as the SSPM electronics seemed very sensitive to external noise pickup.

10. A large iris (9.0-mm diameter) was used to accept all of the photons conjugate to those passed by the 1.4-mm iris before the trigger detector. (Previous coincidence profile measurements had demonstrated that with this size iris, less than 1% of the conjugate photons were lost.)

11.' The focussing lens, the second SPCM, and electronics in the conjugate path were identical to those in the trigger path (#7,8), except that the signal was directed into the STOP of the TAC.

12. A 20-cm focal length plano-convex lens (2" diameter, BBAR for low reflection in the range 600-1000 nm) [Newport #KPX199-AR.16] was used to focus the light into an SSPM input fiber (#14 below). (The distance from the KDP crystal to the lens was ~152 cm, while the distance from the lens to the detector was ~20 cm. Therefore, in the geometrical optics limit, one expects the spot size at the detector to be ~20/152 x (2 mm) ≈ 260 μm, where we have again used 2mm for the size of the trigger beam at the crystal.) Fine adjustments in the final position of the focussed spot were

made using transverse and longitudinal micrometers on the lens. Another dielectric mirror [Melles Griot #02MFG001/001; with MAXBRIte high-reflectance coating] was used to direct the conjugate photons downward to the SSPM fiber. The mirror was aligned so that the beam impinged on the fiber end at an angle of 15° from the normal, to facilitate the use of a spherical retro-reflection mirror (#13).

13. For some of the measurements, a spherical dielectric mirror (35-mm radius of curvature; 2" diameter) [Melles Griot #01MCG013/001, with MAXBRIte high-reflectance coating] was used to recapture photons which were reflected from either end of the fiber or from the surface of the SSPM itself. A small hole (1/4") was made 15° off the symmetry axis to allow the beam to pass through. In practice the mirror was positioned with its center of curvature on the top of the fiber, so that any light reflected off the fiber end was reflected back onto the fiber. Similarly, some of the light reflected off the detector itself travelled back up the fiber, filling the numerical aperture, and was refocussed by the spherical mirror onto the fiber end. This led to substantial improvements in the efficiency. See Section 9.VII.

14. We examined briefly four Solid State PhotoMultipliers (SSPM), with greater detail on two of these. The efficiencies were similar for the four. Coupling light to the detectors was accomplished using 19-cm-long acrylic fibers (750-μm core) [Edmund Scientific #A2532 -> Mitsubishi ESKA SH4001], which also served as cold filters to remove the excess of infra-red background light. They have a rated attenuation of 0.43 dB/m at 633 nm, and 0.62 dB/m at 702 nm.

For the purpose of cooling the SSPMs to 6 K, a helium dewar was suspended from an X-Y translation stage, permitting coarse adjustments in the position of the relevant SSPM fiber, and easy switching from one fiber to another (and hence from one detector to another).

The output of each detector was amplified first by a fast internal charge-sensitive

pre-amplifier to 0.4 mV per photon, then twice more to make the signal appropriate for the following shaping circuitry [x 30 from in-house amplifier (1 to 1000MHz); x 5 from Stanford Research Systems SR445 (DC to 300MHz)]. The typical pulse corresponding to a single-photon detection had an amplitude of 60 mV, a rise-time of 10 ns, and a FWHM of 30 ns. The noise level was 35 mV.

After passing through an inverting transformer [EG&G #IT100], the pulse was directed into a constant-fraction discriminator (CFD) [EG&G #584], with a 254-cm external shaping-delay cable. The discriminator threshold was set at -40 mV for most of the efficiency tests, as this was found to be sufficient for catching all the coincidence counts. A 50-ns blocking window was used to eliminate ringing. One of the CFD outputs was directed to a counter, yielding the singles rate at the SSPM. The other timing output was directed into the STOP input of the TAC.

The TAC output (for both types of conjugate detectors) was then analyzed with a Single Channel Analyzer (SCA) [Tennelec #TC450], affording better than 100 ps time resolution. The SCA output was also directed to the counter, and data saved to a PC.

Note that by simply switching the size of the irises (#5 <-> #10) and moving the 10-nm filter (#6) from one SPCM path to the other, we were able to exchange which SPCM acted as a trigger, and which efficiency we were measuring. In this way the SPCM efficiency at 788 nm was also obtained (using a 10-nm FWHM filter at 633 nm in front of #11').

9.VI. General Procedure

The exact alignment procedure varied somewhat depending on a given run, but in general, the trigger detector, trigger iris, conjugate detector, conjugate iris and conjugate

lens were adjusted to maximize the coincidence rate, and the ratio of this rate to the trigger-singles rate. This was done cyclically until no further improvements were obtained.

The input pump intensity was reduced until saturation effects were no longer a concern. In one test we confirmed that beamline background (i.e. background counts from photons travelling along the main beamline to the trigger detector) was not an important concern by placing a second 10-nm FWHM filter at 702 nm before the trigger detector. (This was during the 702 nm-efficiency test, so the detector already had one such filter before it.) The efficiency was unaltered, to within statistical uncertainty.

The counts S_C, S_t, and C were collected (in one second "bins") for a time interval T, sufficient to allow adequate counting statistics; the background counts were collected for a similar interval. The accidental counts were calculated as stated in Section 9.III (we confirmed this method in separate measurements of the coincidence rate, setting the electronic delay between the conjugate detector and the trigger detector to 100 ns). The collected data were analyzed for the mean and standard deviation. Using Eq. (9.1) we then calculated the raw efficiency. The standard deviation was determined using the following formula, which is derived in Appendix C:

$$\frac{\Delta\eta_c}{\eta_c} = \sqrt{\frac{(1-\eta_c)C}{(C-A)^2} + \frac{2BG_t}{(S_t - BG_t)^2}} \quad . \tag{9.2}$$

9.VII. Efficiency Results

For the SSPM tests, typical counting rates were $S_C \equiv S_{SSPM} = 128{,}900\ s^{-1}$, $S_t = 363 \pm 2\ s^{-1}$, $BG_t = 73.6 \pm 1.4\ s^{-1}$, $C = 155 \pm 1\ s^{-1}$, and $A = 0.6\ s^{-1}$ (with a 20-ns gate window). Typical rates in the SPCM tests were roughly half of this, because of lower saturation rates, and the fact that the SPCM had a much lower dark count rate than the SSPM ($65\ s^{-1}$ versus $7000\ s^{-1}$). Note: The singles rates on the

trigger and conjugate detectors are very different. This is due to the small iris and 10-nm filter in front of the former, and the large iris and no filter (other than the uv-cutoff) before the conjugate detector.

We shall now attempt to list and account for the various systematic errors and their correction factors, which mainly stemmed from photon losses before the *conjugate* detectors. (Recall that losses of *trigger* photons do *not* influence the measured efficiency.) Using a photodiode and a Helium Neon laser (operating at 633 nm) aligned to coincide with the conjugate beam of photons, we measured the transmissivity η_c^P of the path leading to the conjugate detector. The light intensity just before the SSPM fiber (SPCM) was 91.3±1.0% (93.9±1.4%) of that just after the crystal, due to losses in the cutoff filter, mirrors, and the focussing lens. One would expect about 1% loss at each interface, so the measured value is not unreasonable. A smaller correction factor was obtained with the simpler setup used in several of the measurements with the SPCMs.

Two sources of loss before the detector remain: Reflection losses at the crystal, and scattering or absorptive losses within the crystal. We measured these, again using a HeNe laser and a photodiode. The reflection intensity off the output face of the crystal container was 1.02±0.12% [4]. The internal transmission factor (*i.e.*, subtracting out the reflection losses) for the *entire length* of the crystal was 0.954±0.020. If one assumes that the down-converted photons are produced roughly at the center of the crystal, one finds an attenuation (internal absorption and/or scattering loss) of 2.3±1.0%. Combining this with the reflective losses, the total crystal correction factor is then 0.967±0.010.

[4] A secondary measurement revealed that the reflection intensity off the crystal input face (coated for low loss at 351 nm) was 3.37±0.38%. In a few of the runs, the crystal was accidentally inserted backward; in these cases, the reflection losses are those from the 351-nm coating. This correction is included in the figures in Table 9.1, where appropriate.

At 702 nm and 788 nm there was no convenient way to measure directly the pre-detector losses. For both cases, the actual paths of the light were very similar to that of the 633-nm light (because the opening-angles of the down-converted light cone are similar). Therefore the only difference is the wavelength-dependence of reflection, absorption, and scattering losses. For simplicity, we assume the same path transmissivity $\eta^P(702\text{ nm}) = \eta^P(788\text{ nm}) = \eta^P(633\text{ nm})$. Note that this is probably a slight *under*estimate of the losses, as we expect the reflection losses to be greater for the longer wavelengths.

The effect of using a spherical mirror (#13, Fig. 9.2) with the SSPM was tested. The mirror was inserted and aligned to give the highest efficiency; then it was removed and the singles and coincidence rates remeasured. The mirror was found to increase the efficiency from 47.4±0.6% to 55.8±0.7% at 702 nm, and from 49.5±0.5% to 53.5±0.4% at 633 nm. Direct comparison of these results is difficult since varying the wavelength corresponded to entirely realigning the system. Clearly, however, use of the mirror gave a significant improvement.

In a few of the tests on the SSPM, the timing window accepted only the main coincidence peak, described below, thereby excluding 11.9±0.5% of the coincidences. Also, in examining the effects of SSPM bias (see Section 9.VIII), it was found that the maximum efficiency was obtained at a slightly higher bias voltage than was typically used. When these various systematic errors are accounted for, the corrected efficiencies are as listed in Table 9.1, where we have included the results for two SSPMs and two SPCMs.

It is important to note that associated with each of the detectors there are other sources of loss that may yet be reduced. Subsequent testing on the SSPM input fibers yielded a transmission of only 70±5% (including insertion losses). The expected loss due to normal transmission and Fresnel losses is only 10%, consistent with the

Table 9.1. Corrected single-photon absolute detection efficiencies of two SSPM's* and two SPCM's.

Wavelength (nm)	Corrected efficiency (%)			
	SSPM#1	SSPM#2	SPCM#1	SPCM#2
633	70.9±1.9	69.5±1.9	** 74.3±2.0	65.0±1.6
702	N.A.	66.3±1.4	** 76.4±2.3	** 75.4±1.5
788	N.A.	N.A.	53.7±1.4	54.4±1.0

* The results listed include the improvements of a spherical retro-reflection mirror on the SSPMs' efficiencies.

** These results were obtained with a slightly simpler setup than that in Fig. 9.2, with no mirrors.

improvements observed with the spherical mirror. If any of the excess measured transmission loss was present during the efficiency experiment, the actual SSPM detector efficiencies will be somewhat higher. Adjusting for *all* of the fiber losses would suggest device efficiencies as high as 93.7±7.2%. The problem of coupling light into the SSPMs is nontrivial. Because these device are sensitive in the infrared (out to ~20 μm), the count rate arising from black-body radition in the room is much larger than our typical signal. It is not sufficient to place a cutoff filter in front of the fiber, because even though this might not *transmit* infrared, it will thermally *radiate* it. The solution is to have some sort of filter which is cooled. The plastic fibers we used initially proved fairly adequate for attenuating the infrared, but not the signal (at 700 nm or 633 nm). However, we believe that the unexpected loss in these fibers was due to microscopic damage to the fibers during the cooling. (It is possible that new "dual-clad" plastic fiber may be more resistant to this damage.) Subsequent tests with quartz fibers revealed that these seemed to survive the cooling fairly well, but did not sufficiently attenuate the infrared. We are currently investigating the possibility of using a doped fiber. In

particular, silica fibers with a high OH content have a high absorption starting at around 1.3 μm. Unfortunately, there seems to be some sort of manufacturing limit to how much OH can be forced into a fiber. The highest number we have seen is ~1500 ppm (parts per million). A rough rule of thumb is that the infrared attenuation is about 40 dB/km/ppm, implying an attenuation of 60,000 dB/km for the high-OH fiber. At present the SSPMs employ only a short (~19 cm) length of fiber, implying a net transmission of ~7%, which is too high.

Another possible solution is to use the fact that bending loss in fibers can be wavelength-dependent, with longer wavelengths being more lossy. In particular, in a single-mode optical fiber, a very sharp cutoff can be induced via microbending which couples light out of the fundamental mode into higher-order, non-propagating modes. Of course, a large diameter fiber *can* support higher-order modes, so we would need to rely on the dependence of the indices of refraction (and consequently the critical angle for reflection) on the wavelength of light. Depending on the functional dependence, one may be able to substantially "leak out" the infrared while not significantly degrading the signal. Further investigation is needed. For example, a coiled fiber might not survive well as a cooled fiber. A final possibility being investigated is the wavelength-dependence of the evanescent wave in the cladding. By etching the cladding to a thickness of only several microns, and coating it with an absorptive material, strong filtering may be obtainable.

The SPCM detector is housed in a can with a glass window, which was not anti-reflection coated at all, implying 8% losses. The detector surface itself was BBAR coated, so that losses of 1% may be expected here. If these interfaces all had multi-layer coatings to essentially eliminate losses, then an SPCM detection efficiency of >82% should be achievable. It is also likely that further overbiasing would increase the efficiency, though other parameters such as dark counts will also be affected.

9.VIII. Timing Resolution and Related Results

In the course of the absolute efficiency measurements, a number of other effects were examined, including the effect of varying bias levels (for the SSPM), saturation (both), temporal response (both), and afterpulsing (SPCM). We now discuss these.

The SSPM is designed to operate with a voltage bias of about 7 V. A brief investigation of the effect of different bias voltages was made at 702 nm and 633 nm; the results are plotted in Fig. 9.4. One should not attach too much significance to these numbers as the amplifier bias and discriminator threshold were kept constant as the device bias was varied. It is likely that optimizing one or both of these at each device-bias level might have improved the effective efficiency. Nevertheless, for the settings used throughout most of our measurements, the device bias of 7.38 or 7.39 V seemed close to optimal, though the SSPM efficiencies in Table 9.1 have been adjusted upward slightly (by a 1.03±0.01 correction factor) using the above results.

By measuring the efficiency at several different pumping rates (controlled by changing the variable ND before the crystal), we evaluated the saturation characteristics of the detectors (see Fig. 9.5). The SSPM fared much better than the SPCM, which evinced a 3% efficiency reduction due to saturation effects for a singles rate of 100,000 s^{-1}. In a simplified model, the rate-dependent efficiency is given by $\eta(S_c;\delta\tau) = \eta_o(1 - S_c\,\delta\tau)$, where η_o is the no-saturation limit of the measured efficiency, S is the singles rate, and dt is related to the detector deadtime (including the subsequent electronics). Using the values from a linear fit to the saturation data, one finds δt_{SPCM} = 0.5 μs. (A direct measurement, described below, yielded a deadtime of 1.5 ms.) A similar calculation for the SSPM yields δt_{SSPM} = 53 ns, which is the same as the blocking window set with the CFD.

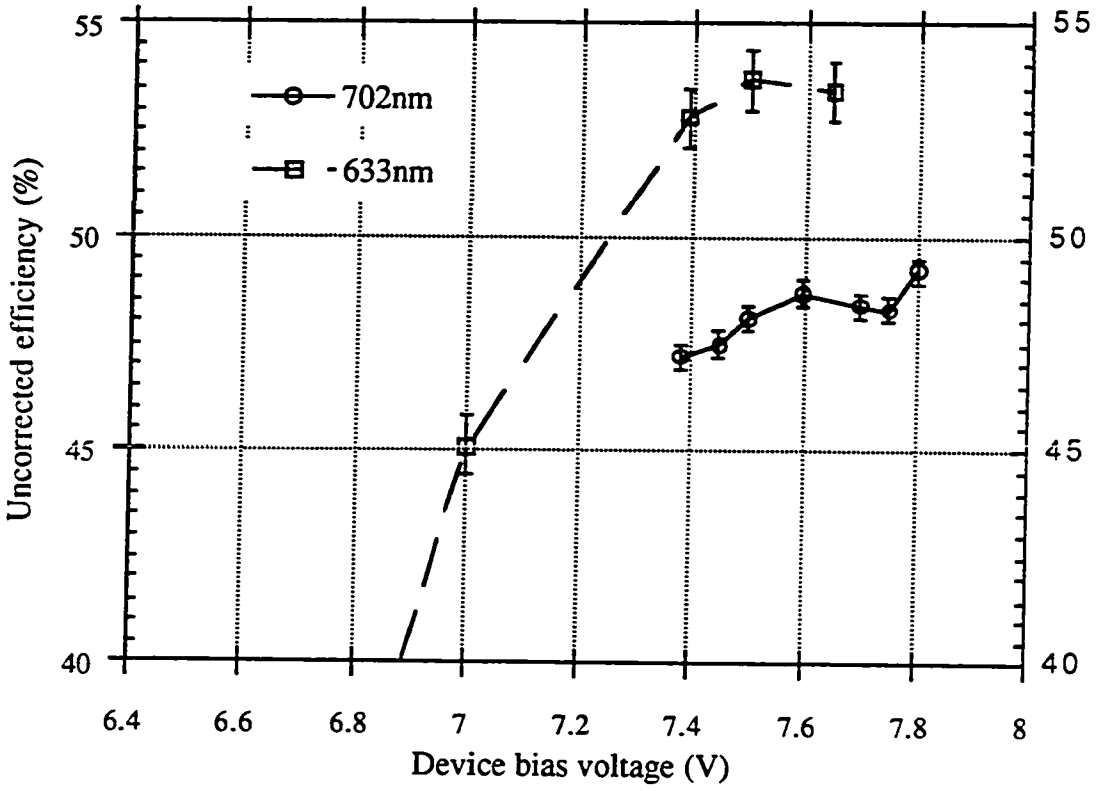


Figure 9.4 Uncorrected SSPM efficiency, as a function of device-bias voltage, for fixed amplifier bias and discriminator threshold. It was not possible to check higher bias voltages, as these caused the electronics to suffer frequent latch-up.

As discussed in Chap. 2, the correlated photons from our source are emitted with a time correlation of better than 40 fs [Steinberg et al., 1993; Steinberg et al., 1992]. Thus, by looking at the coincidence time-profiles in our experiment, we were able to accurately test the temporal response of the detectors at the single-photon level. Using

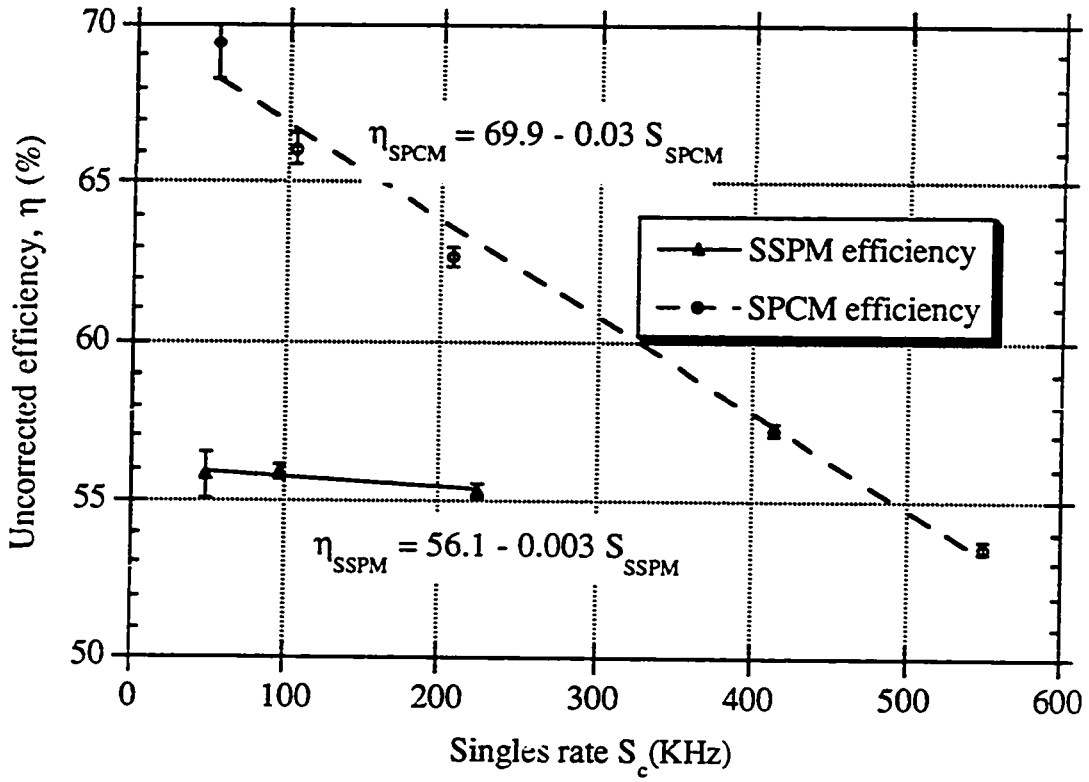


Figure 9.5 Plot of uncorrected detector efficiency (at 702 nm) versus singles count rate. SSPM had spherical mirror in place. Curves are linear fits. True efficiencies are obtained by correcting for crystal losses (due to reflection, scattering, and absorption), losses at intervening optics, and for the SSPM, non-optimal biasing.

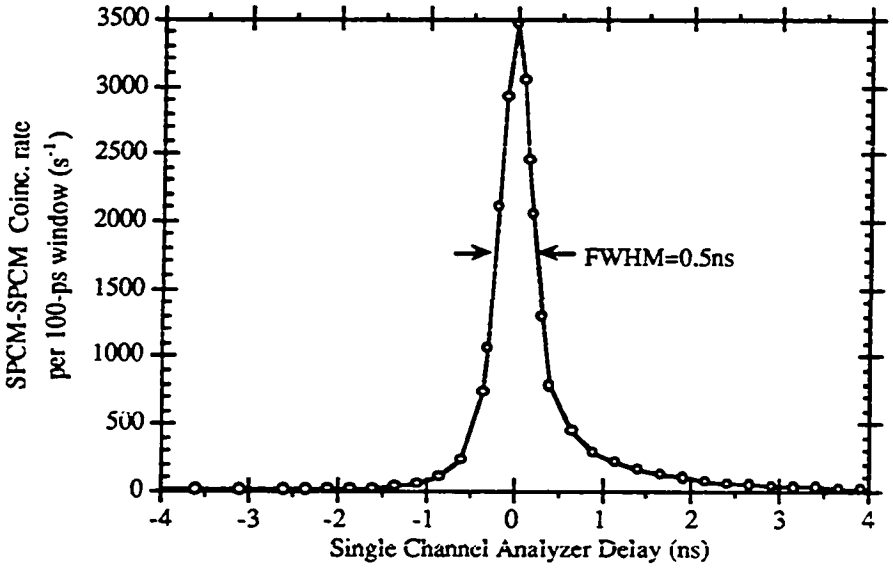
the TAC and SCA, several such profiles were taken at 702 nm with the SPCMs at various singles count rates. In particular, with the SCA in window mode, we were able to map out the temporal profile with a 100-ps window (a typical example is shown in

Figs. 9.6a). Our narrowest peak was 300 ps FWHM. Since this is essentially the convolution of two SPCMs, we conclude that the time response of each is 200 ps. (Other researchers have reported even shorter time peaks with this sort of device, using custom electronics specifically designed to give a faster time response [Li et al., 1992].) Increasing the counting rates led to a broader profile, mainly affecting the tails.

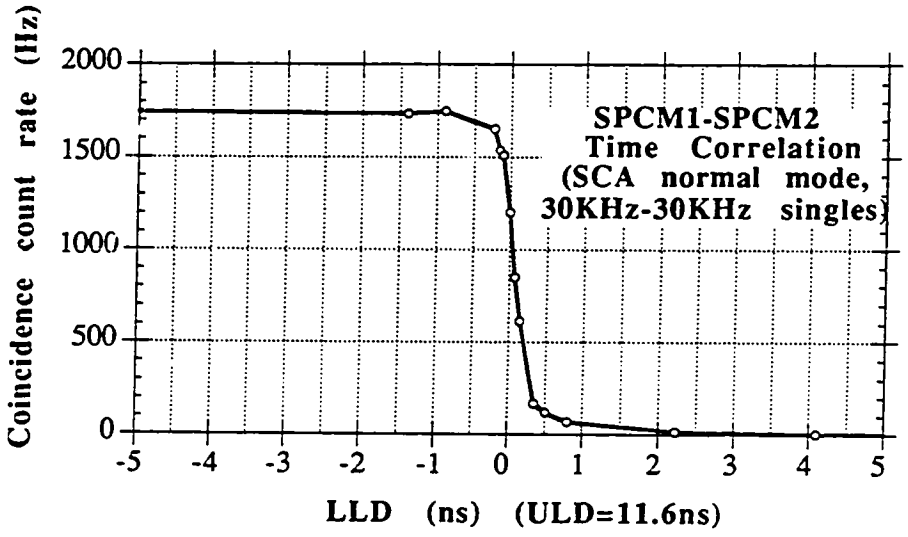
By using the SCA in normal mode, keeping the upper level discriminator (ULD) fixed, and varying the lower level discriminator (LLD), we obtained an *integrated* time profile (see Fig 9.6b). From this one can directly determine the time interval necessary to catch a certain percentage range of the counts. For example, the counts in the range 35-65% fall within a 150-ps window; those in the 10-90% range fall within a 500-ps window; and those in the 5-95% range fall within a 1000-ps window. We also observed that the time profile of the SPCM's broadened as the count rates were increased, so that the 10-90% range changed from 0.5 ns to 1.5 ns when the detected photon rates were increased from 30,000 s^{-1} to 140,000 s^{-1}. While it is possible that some of this saturation-like effect is due to the timing electronics (e.g. the TAC), one expects more time jitter as the count rate is increased, simply due to the SPCM deadtime (see below). More counts fall within the time interval during which the overbias voltage has not fully recharged; the output pulses in this case have a smaller amplitude, and thus cross the preset threshold (of the leading-edge discriminator within the SPCM internal electronics) at a slightly different time.

Similar coincidence time-profiles were taken (also at 702 nm) using the SPCM trigger as the START to the TAC, and an SSPM as the STOP, with a 1.0-ns SCA window. (See Fig. 9.7a.) At two different pump intensity levels (SSPM=50,000 s^{-1} and 165,000 s^{-1}, with corresponding trigger levels of 250 s^{-1} and 1,000 s^{-1}), the profiles were essentially identical. There were two peaks, a main peak (FWHM = 3.3 ns) and a smaller peak (FWHM = 4.5 ns) centered 11 ns earlier. The area under the

Figure 9.6 (next page) Timing results for SPCMs. a). Typical time-correlation profile with the SCA in window mode, displaying the coincidence rate between two SPCMs, with singles rates of 70 KHz and 250 KHz, as a function of gate-window (electronic) timing delay. Single channel analyzer window corresponded to 100 ps. Widths as low as 300 ps were seen at lower count rates. b.) By using the SCA in normal mode, keeping the upper level discriminator (ULD) fixed, and varying the lower level discriminator (LLD), we obtained an integrated time profile.


SPCM-SPCM Coinc. rate
per 100-ps window (s^{-1})
3500
3000
2500
2000
1500
1000
500
0
FWHM=0.5ns
-4
-3
-2
-1
0
1
2
3
4
Single Channel Analyzer Delay (ns)

a.)


Coincidence count rate (Hz)
2000
1500
1000
500
0
SPCM1-SPCM2
Time Correlation
(SCA normal mode,
30KHz-30KHz singles)
-5
-4
-3
-2
-1
0
1
2
3
4
5
LLD (ns) (ULD=11.6ns)

b.)

Figure 9.6

main peak was 89.6±0.5% of the total. A second time-window measurement at 633 nm yielded consistent results for the areas under the main and precursor peaks. [At one point a fast oscilloscope (triggered on the trigger-detector output) was used to observe the coincidence pulses. No precursors were observed, but two out of the ten coincident pulses were delayed by ~20 ns relative to the other eight (see Fig. 9.7b). The probability of both of these being accidental coincidences is very small. However, these "post-cursors" were not seen on a second scope trace with 14 coincidence events displayed.] The two peaks separated by 11 ns were not seen with two SPCMs in coincidence, and are not expected from the model for SSPM carrier generation and transport. The initial photo-carriers should appear within less than 1 ps, and their transport (ending in an avalanche) should end in less than 5 ns. The two-peak phenomenon may be due to the associated electronic circuitry used with the SSPM.

Any single-photon detection system will have an intrinsic deadtime, which will eventually limit the usable count rates, due to saturation. The SPCMs actually have two deadtimes associated with them. The first is a hard cutoff of 50 ns, due to the internal shaping circuity in the modules. The second deadtime involves the time necessary to regain the high overbias voltage once an avalanche breakdown has been quenched. This is achieved in the SPCMs by charging through a ballast resistor (passive quenching). In the Geiger mode, if the bias voltage is not beyond the breakdown voltage, then no avalanche can occur, and the net detector efficiency will be zero (assuming one is looking for avalanche output pulses). If the bias voltage is greater than the breakdown voltage, but not yet up to the standard operating point (30 V beyond avalanche breakdown, in our custom-modified detectors), then avalanche breakdown pulses will occur, but the photon detection efficiency will be reduced below its maximum value. Moreover, the pulse amplitude will be less, causing increased time jitter in the leading

Figure 9.7 (next page) a) Timing results for an SSPM (with an SPCM as a "trigger" detector) a.) Time-correlation profile, displaying the coincidence rate between an SSPM at a singles rate of 50 KHz, and a trigger SPCM at 250 Hz, as a function of gate-window (electronic) timing delay. Single channel analyzer window corresponded to 1.0 ns. The spherical mirror was not in place for these measurements. b) Oscilloscope trace (triggered on the trigger-detector output) showing coincidence pulses. No precursors are present, but two out of the ten coincident pulses are delayed by ~20 ns relative to the other eight.

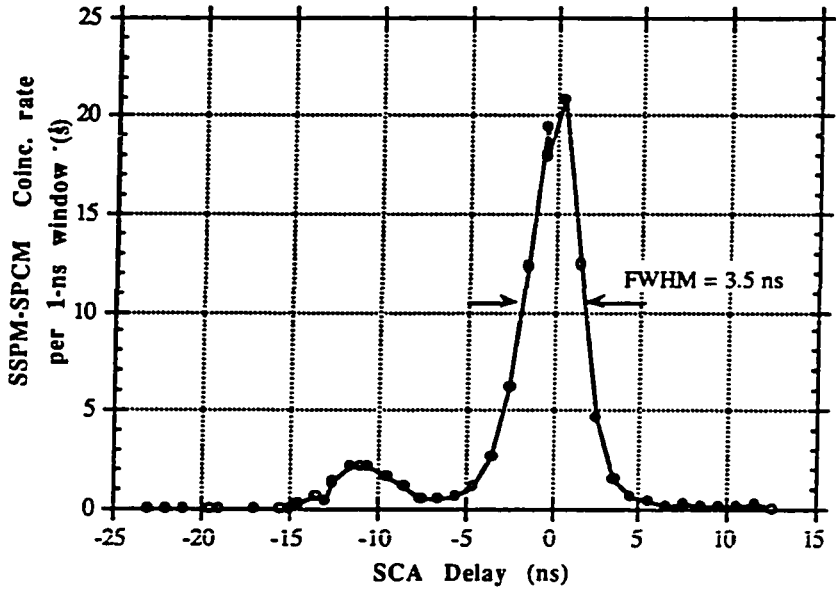


Figure 9.7a

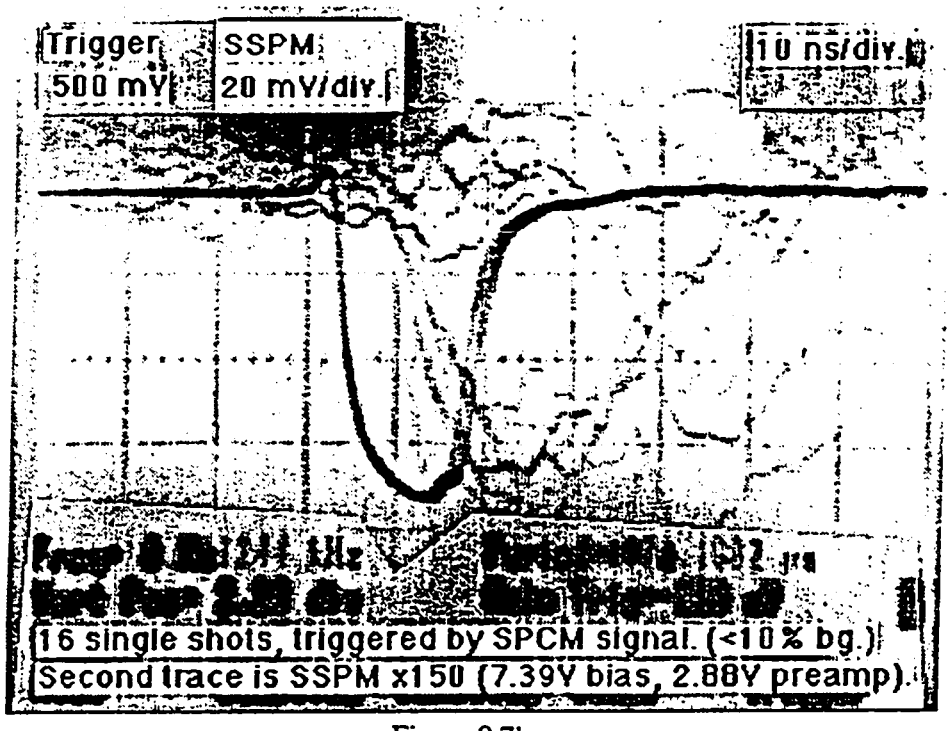


Figure 9.7b

edge discrimination. In the course of investigating the afterpulsing characteristics of these devices (see below), we have directly measured the time constant to be 1.5 μs, which we can compare to the 0.5-μs value inferred from the saturation data. The discrepancy is due to the fact that the efficiency during the deadtime is not constant, nor does it start at zero (after the 50-ns hard cutoff, during which no pulses are produced). For example, the efficiency near the beginning of the recharging deadtime may be two-thirds that of the maximum efficiency; then the "effective" deadtime, to which the saturation measurements are sensitive, will be reduced from the directly-observed deadtime by a factor of three.

The SSPM evinced no long-time recharging deadtime, since it is not operated above its avalanche breakdown voltage. We were unable to investigate intrinsic SSPM deadtimes less than the 50-ns CFD blocking window, which adequately accounts for the 53-ns deadtime inferred from the saturation measurements. The SSPMs are expected to be continuously operating detectors for counting rates less than about 3×10^7 s^{-1}, the saturation rate.

It is well known that avalanche photodiodes operating in the Geiger mode produce afterpulses. After an avalanche, an impurity in the silicon may act as a trap for one of the carriers; when the trap is emptied at some random later time, a new avalanche can occur, thus causing a second pulse to be detected. We measured the presence of afterpulses by performing an auto-correlation of an SPCM with itself. The output of the detector was fed into both the START and STOP inputs of the TAC. The START pulse was delayed so that it arrived just after the STOP pulse, to prevent the TAC from registering a "true" coincidence pulse and immediately resetting. A multi-channel analyzer (MCA) was used to record the profile over the ranges 2 μs, 20 μs, and 100 μs. Even if there were no afterpulses, one would still register counts on the MCA due to

"accidental" coincidences. The count rate from the SPCM was therefore adjusted (by varying the input light level) to keep the accidental rate small. If the singles count rate is S, and a count is registered at time t=0, then the chance of registering the next count within a small interval dt at time t is $P(t)\,dt = S\,e^{-St}\,dt$, assuming a Poisson process. For $tS << 1$, $P(t) \approx S(1 - St)$. For example, for the range 0-20 μs, we set $S=340\ s^{-1}$, and t was always less than 20 μs, so that $St < 7 \times 10^{-3}$, and we may even neglect the linear term. We counted for a total time of T = 95,029 s, and with a single-channel width of 20 μs/1024, we expect $S^2T\,dt$ to be 218 counts/channel. The presence of afterpulsing adds to this baseline a decaying structure, which is compatible with a falling exponential, with a time constant of t = 4.5±0.2 μs (see Fig. 9.8). If we integrate analytically the fitted exponential, we expect to find 650±30 counts. A numerical summation of the data yields 400 counts, which is less due to the initial deadtime. Using the larger of these results, we conservatively calculate that the fraction of afterpulses is $< 2 \times 10^{-5}$, and therefore the effect on η is negligible.

9.IX. Conclusion

Our results demonstrate that suitably modified Single Photon Counting Modules can possess very high photon detection efficiencies, and relatively fast time response. It seems likely that by improving the anti-reflection coatings on the detector surface and the glass window of the housing, efficiencies in excess of 80% should be realizable. By increasing the overbias voltage even further, still higher efficiencies may be possible. By switching to active quenching, it should also be possible to reduce the deadtime by about an order of magnitude. This may also help the time response at higher count rates. The primary drawback of the modules is the small active area [only ~$(100\ mm)^2$], which mandates careful focussing of the input light. Larger-area detectors are currently under development (as are detectors using active quenching) [Andrew MacGregor, EG&G

Canada, private communication], but it is unknown whether these will have as high efficiency.

The Rockwell Solid State PhotoMultipliers had a maximum corrected efficiency of 65.6±1.9% without the spherical mirror, and 70.9±1.9% with the mirror. After the measurements and dismantling of the apparatus, the input fibers were found to have deteriorated. If these subsequently measured losses were present during the experiment, and degraded the efficiency, one may conclude that the true device efficiency was as

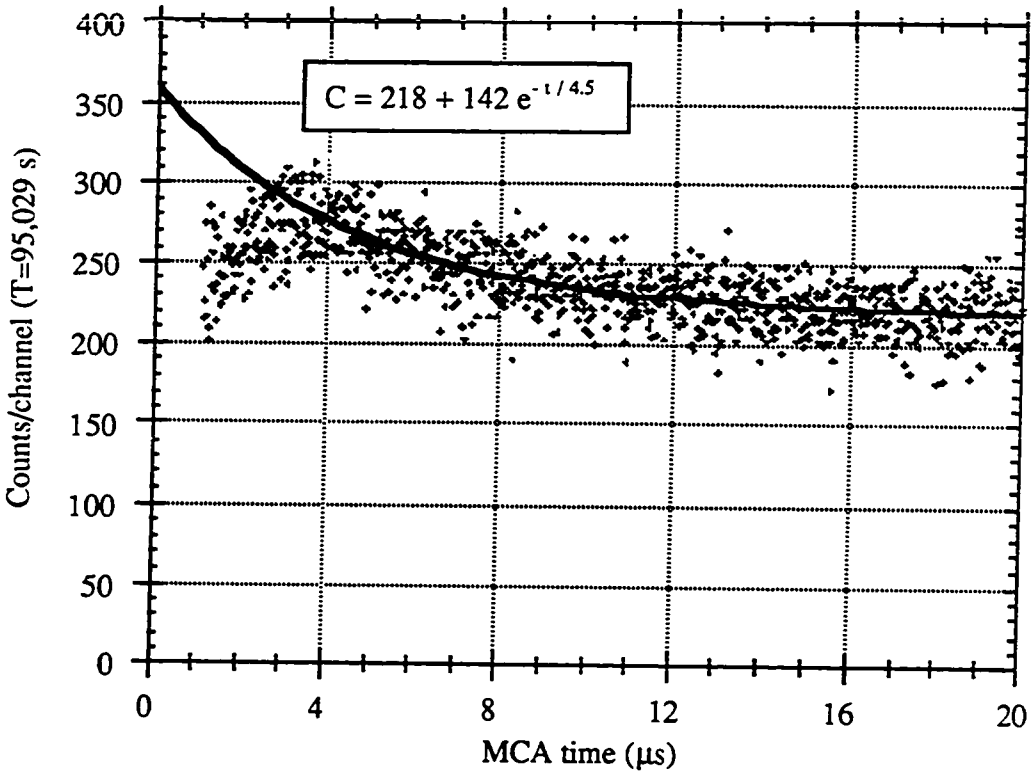
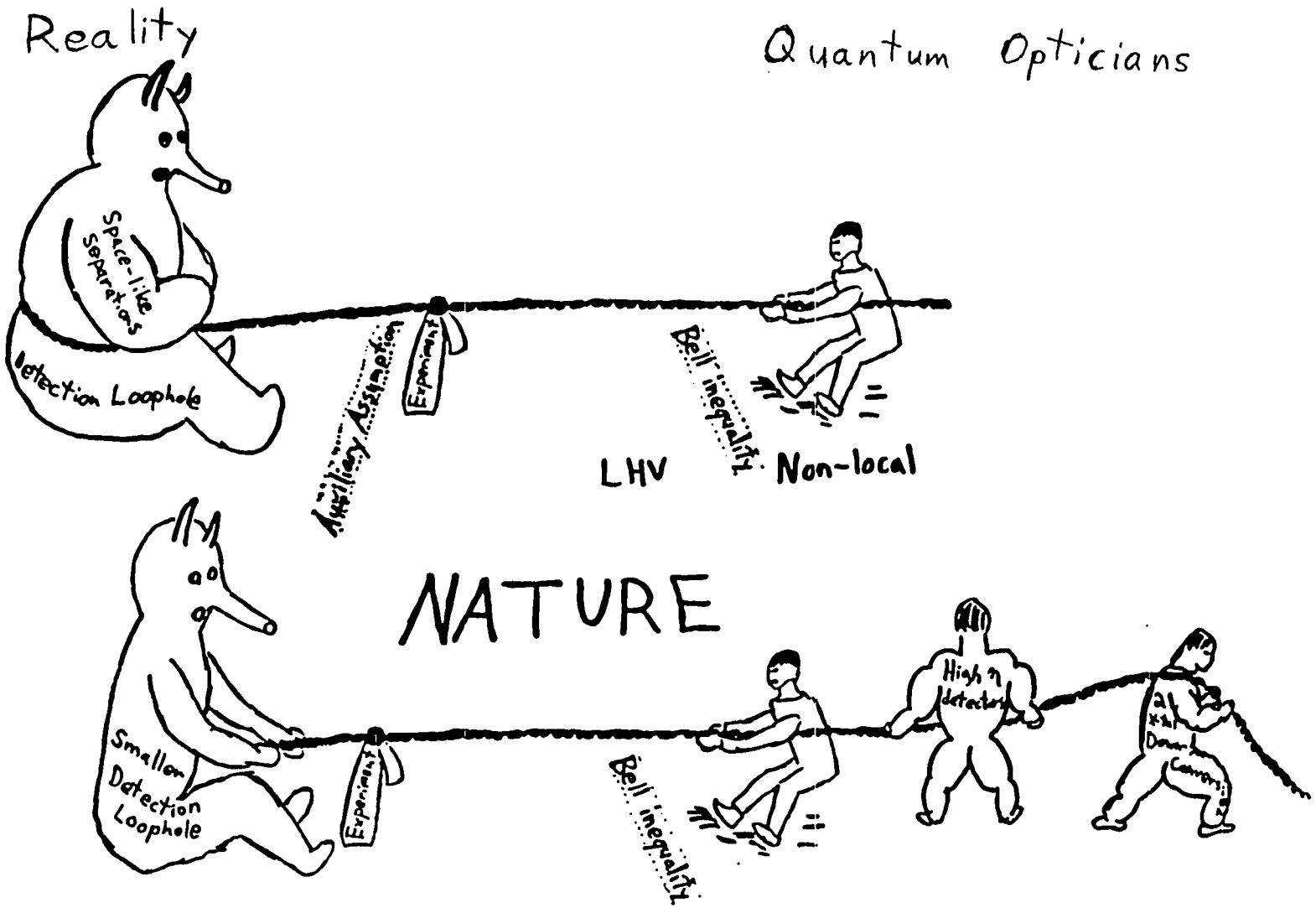


Figure 9.8 Temporal autocorrelation profile of an SPCM, demonstrating afterpulsing. Solid curve is an exponential fit to the data beyond the deadtime region (about 3 μs). Singles rate was 340 s^{-1}; total duration of measurement was 95,029 s.

high as 93.7±7.2%. Unfortunately, there is no way to ascertain how much of this loss should be corrected for.) The use of the spherical mirror definitely reduced reflection losses, increasing the efficiency by as much as 19% at 702 nm, but only 8% at 633 nm. The presence of a 10% "precursor peak" in coincidence timing was very surprising and is still not understood. The timing jitter we measured was probably limited by the amplifiers, so the SSPMs themselves may be somewhat faster. The main difficulty in using these detectors is the need to cool them to near liquid helium temperatures. Their high sensitivity in the infra-red necessitates the use of some sort of "cold-filter" (the fibers, in the system we used), which complicates input coupling. Before any further high-efficiency experiments can be performed, a better method of coupling light to the actual devices should be developed.


Reality
Quantum Opticians
Space-like Separations
Detection Loophole
Auxiliary Assumption
Experiment
Bell inequality
LHV
Non-local
NATURE
Smaller Detection Loophole
Experiment
Bell inequality
High η detectors

Chapter 10: Proposal for a Loophole-Free Bell's Inequality Experiment

A somewhat different view of the Bell's inequality affair is depicted in the previous cartoon, to be contrasted with the cartoon preceding Chap. 8. Here, Nature is not seen as an opponent, but rather as the ground on which the contest is waged. The top picture expresses the status of affairs in experiments to date. On one side is a lone Quantum Optician, on the other, horrible Reality, made almost unbeatable by the various loopholes yet to be closed in any experiment. Unable to actually pull the flag (the experiment) over the line (to violate a Bell's inequality, thereby proving that Nature is nonlocal), the clever experimentalist "redefines" winning (what inequality his results must violate) by introducing supplementary assumptions. The future state of affairs is represented in the bottom version. Reality is still formidable, but less so (due to reduced efficiency requirements), and the Optician is assisted by high-efficiency single-photon detectors (described last chapter) and a new 2-crystal down-conversion scheme (described below).

10.I. Introduction

As discussed in Chap. 7, quantum mechanics (QM) yields predictions which are inconsistent with the seemingly innocuous concepts of locality and reality. This was first shown by Bell in 1964 [Bell, 1964; Clauser and Shimony, 1978] for the case of two quantum-mechanically entangled particles, e.g., particles in a singlet state, which do not possess definite polarizations even though they are always orthogonally polarized. Since then, the theoretical results have been generalized to include entanglements of three or more particles [Greenberger et al., 1989; Mermin, 1990], macroscopic (but non-

classical) states of light [Munro and Reid, 1993; Franson, 1993], and even a clever gedanken experiment using electron-positron annihilation to achieve a contradiction with local realism without the need for inequalities [Hardy, 1992a], and a recently proposed optical analog which may allow a feasible experimental implementation [Hardy, 1993]. Unfortunately, none of these ingenious extensions and generalizations of the work of Bell reduces the experimental requirements for a completely unambiguous test. In fact, all of them seem to mandate even *stronger* constraints on any real experiment than the original two-particle inequalities[1]. One exception is the recent discovery, introduced in Chap. 7, that the detection efficiency requirement can be reduced by employing a state of two particles that are not maximally-entangled, i.e., with an unequal superposition of the two terms [Eberhard, 1993][2].

The situation for experimental tests has seen somewhat less growth: To date *no incontrovertible violation of Bell's inequalities has been observed.* As discussed in Sect. 7.V, all experiments thus far have required supplementary assumptions, severely reducing the true impact such an experiment might yield. We propose here a setup which should permit for the first time (simultaneous) closure of the angular-correlation loophole and detection loophole; we will also discuss briefly how current technologies should allow an extension to close the space-like separation loophole as well[3]. The

---

[1] For example, Braunstein and Mann have shown that the detection efficiency requirement can be reduced to 71%, but only in the limit of no background and a large number of entangled particles [Braunstein and Mann, 1993].

[2] Hardy also discusses non-maximally-entangled states, for use in a demonstration of nonlocality without inequalities [Hardy, 1993]. However, in the absence of supplementary assumptions, his scheme would require detector efficiencies greater than 98%.

[3] A very different proposal has been made by Edward Fry [Fry, 1991], using dissociated mercury dimers as the correlated particles. The advantage is that detection efficiencies of 95% are possible by photoionizing the atoms and detecting the photoelectrons. Although it may be that this scheme is a viable one, it has its own practical difficulties, which may be prohibitive; it should be noted that *none* of the Bell's inequalities experiments to date have relied on entangled atoms.

source may also find application in quantum cryptography (see Sect. 7.VII, and [Ekert, 1991; Ekert et al., 1992]), as it doubles the signal-to-noise ratio of most previous down-conversion EPR schemes.

In Section 10.II our proposed source is described, while several potential problems arising from various phase-matching considerations are discussed in Section 10.III, along with an experimental modification to mitigate these problems. The benefits of utilizing a non-maximal entanglement are presented in Section 10.IV. Section 10.V evaluates the detrimental effects of imperfect optical elements. A final analysis and conclusion are given in Section 10.VI.

In Appendix D1 we describe the relevant parameters and phase-matching calculation for what currently seems to be the best choice for a nonlinear crystal for this proposal, namely, BBO. The calculation to evaluate the effect of a non-plane-wave pump on coincidence detection efficiency, as a function of collection iris size, is presented in Appendix D2. Finally, a somewhat thorough calculation of the singles and coincidence probabilities relevant for a Bell's inequality, allowing for various sorts of entangled states, as well as for non-ideal polarizing beam splitters, is presented in Appendix D3.

10.II. Proposed Source

As discussed in Sect. 7.V, even with unity-efficiency detectors, the down-conversion schemes used until now are inadequate for a completely unambiguous test of Bell's inequalities, because they must perforce discard counts (although, as discussed in footnote 12 in that section, detectors capable of reliably distinguishing between one- and two-photon detections could in principle solve this problem). A schematic of our proposed source is shown in Fig. 10.1a.

Figure 10.1 (next page) Schematic of a novel arrangement in which a loophole-free test of Bell's inequalities is feasible. a) An ultraviolet pump photon may be *spontaneously down-converted* in either of two nonlinear crystals, producing a pair of orthogonally-polarized photons at half the frequency. One photon from each pair is directed to each output port of a polarizing beam splitter. When the outputs of both crystals are combined with an appropriate relative phase δ, a true singlet- or triplet-like state may be produced. By using a half waveplate to effectively exchange the polarizations of photons originating in crystal 2, one overcomes several problems arising from non-ideal phase-matching (see Sect. 10.III). An additional mirror is used to direct the photons oppositely to separated analyzers. b) A typical analyzer, including a half waveplate (HWP) to rotate by α the polarization component selected by the analyzing beam splitter, and precision spatial filters to select only conjugate pairs of photons. In an advanced version of the experiment, the HWP could be replaced by an ultrafast polarization rotator (such as a Pockels or Kerr cell) to close the space-like-separation loophole.

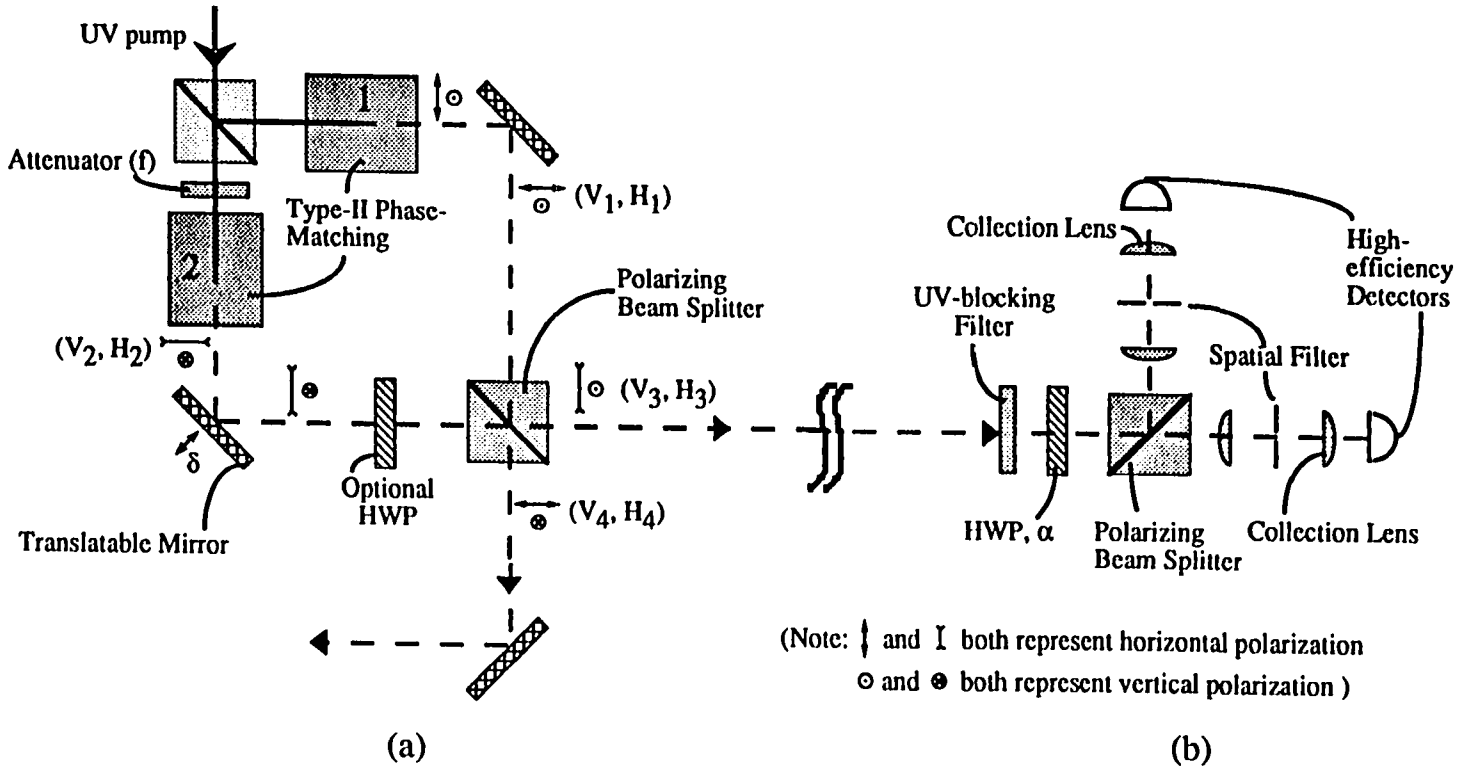

UV pump
1
2
Attenuator (f)
Type-II Phase-Matching
(V_1, H_1)
Polarizing Beam Splitter
(V_2, H_2)
(V_3, H_3)
δ
Optional HWP
(V_4, H_4)
Translatable Mirror
(a)
Collection Lens
High-efficiency Detectors
UV-blocking Filter
Spatial Filter
HWP, α
Polarizing Beam Splitter
Collection Lens
(b)
(Note: ↕ and I both represent horizontal polarization
⊙ and ⊛ both represent vertical polarization)

Figure 10.1

Two nonlinear crystals[4] are simultaneously pumped by a coherent pump beam to induce spontaneous parametric down-conversion; the pumping intensity can be independently varied at each crystal. The crystals are cut for type-II collinear degenerate phase matching (i.e., the down-converted photons are collinear and orthogonally polarized, with spectra at roughly twice the pump wavelength). For example, we envisage using 1-cm long crystals of beta-barium borate (BBO), pumped by the 325-nm line from a HeCd laser incident at 54° to the optic axis (see Appendix D1).

For clarity, we first assume a monochromatic pump beam (at frequency $2\omega_0$), and a single-mode treatment of the down-converted photons. The state after the crystals is

$$|\psi\rangle = \sqrt{1-|A|^2}\,|\text{vac}\rangle + \frac{A}{\sqrt{1+|f|^2}}\left(|H, V\rangle_{\text{crystal 1}} + f\,|H, V\rangle_{\text{crystal 2}}\right), \quad (10.1)$$

where we have omitted higher order terms (for the very unlikely case in which more than one pump photon down-converts; by reducing the pump intensity, the contribution of these terms can be made as small as desired). A includes the down-conversion efficiency into the modes we are considering, and also the pump field strength; f represents a possible attenuation of the pump beam incident on crystal 2. The state (10.1) describes a photon pair [one photon polarized horizontally (H), the other vertically (V)] originating with probability amplitude $A/\sqrt{1+|f|^2}$ in crystal 1 and with

[4] Hardy has also proposed a scheme which employs two nonlinear crystals, in a different geometry than that discussed here [Hardy, 1992b]. Specifically, he proposes directing the down-converted photons from one crystal through a second crystal, such that one cannot determine from which crystal a pair originated. However, as he points out, his scheme requires near-perfect alignment to function adequately (the best results reported so far with such a scheme displayed only 30% visibility [Zou et al., 1991]: to violate a Bell's inequality, one needs greater than 71% visibility, assuming 100% detection efficiency and no background.) The fact that we use photons which are (very nearly) collinear reduces our alignment difficulties considerably.

probability amplitude $Af/\sqrt{1+|f|^2}$ in crystal 2; it does not include that part of the pump beam which was not down-converted. [Physically, this implies that we have removed the unconverted pump photons, as with a cutoff filter before the detectors (see Fig. 10.1b).] We now combine the modes from the two crystals at a polarizing beam splitter. For an ideal polarizing beam splitter, incident p-polarized light (horizontal in Fig. 10.1) is completely transmitted, while incident s-polarized light [vertical (out of the plane of the paper) in Fig. 10.1] is completely reflected; therefore, one photon of each pair will be directed to output port 3, while the conjugate photon is directed to output port 4. (In Section 10.V the case of a non-ideal beam splitter is examined; a general calculation is given in Appendix D3.) Including a phase shift $\delta = 2\omega_0\Delta x/c$ (where Δx, the difference in path lengths, may be varied by moving one of the mirrors slightly) between the two non-vacuum terms of (10.1), we then have

$$|\psi\rangle \approx \left(|V\rangle_3|H\rangle_4 + fe^{i\delta}|H\rangle_3|V\rangle_4\right), \tag{10.2}$$

where we have omitted the (predominant, but uninteresting) vacuum term and the prefactor $A/\sqrt{1+|f|^2}$. For the balanced case ($f = 1$), and for $\delta = 180°$, (10.2) reduces to the familiar singlet-like state. (The case where $f \neq 1$ is discussed in Section 10.IV; Section 10.V examines the case $\delta \neq 180°$.) Note that this is different from Eq. (7.16), which additionally contains non-coincidence terms that must be intentionally discarded to prepare a singlet-like state. Consequently, our a scheme may find application in quantum cryptography (see Sect. 7.VII) as it has twice the signal to noise ratio of most other EPR-based schemes.

With the above source of correlated particles, one can now perform a *polarization* test of Bell's inequalities[5]. Polarization analysis is performed using an additional

[5] Using the 2-crystal source, one could in principle perform a Franson-type experiment, without the need to electronically discard counts that arise from the non-

polarizing beam splitter after each output port of the interferometer, and examining one or both channels of each analyzer with high efficiency detectors (see Fig. 10.1b). "Rotation" of these analyzers can be effectively accomplished by using a half waveplate before each one to rotate the polarization of the light.

The alignment of the entire setup is obviously somewhat critical. Fortunately, there are simple tests which can be made to verify the integrity of the system. For example, if a half waveplate (oriented with its axis at 22.5° to the horizontal) is used after crystal 2 to rotate the polarizations by 45°, then the new state of light from that crystal will be $|\psi\rangle_{\text{crystal 2}} \approx [(|H_2\rangle + |V_2\rangle)/\sqrt{2}][(|H_2\rangle - |V_2\rangle)/\sqrt{2}]$ $= (|H_2, H_2\rangle - |V_2, V_2\rangle)/2$. By blocking the light from the other crystal, one can easily verify this state at the analyzers -- either both photons go out on port 3 or both on port 4, and there should be no coincidence counts (assuming no background noise) between ports 3 and 4. More parameters can be checked with half waveplates after *both* crystals.

If the detectors are far separated from each other and from the source, and one uses some rapid, random means to rotate the light before the analyzers, (such as a Pockels or Kerr cell, whose voltage is controlled by a random signal), then one can close the space-like separation loophole. The signal could be derived, for instance, from the decay of a radioactive substance, or even from the arrival of starlight. Note that since the down-converted photons are emitted within tens of femtoseconds of one another (unlike the photons in an atomic cascade), the limiting time factors will be the detector resolution (expected to be less than 10 ns) and the switching time (which can also be on the order of nanoseconds)[6].

---

interfering processes where one photon travels the short path in its unbalanced Mach-Zehnder interferometer and the other photon travels the long path. Details are given in Sect. 8.VI.

[6] Franson has argued that the pump coherence time might also come into play [Franson, 1985]. However, this is also of the nanosecond timescale.

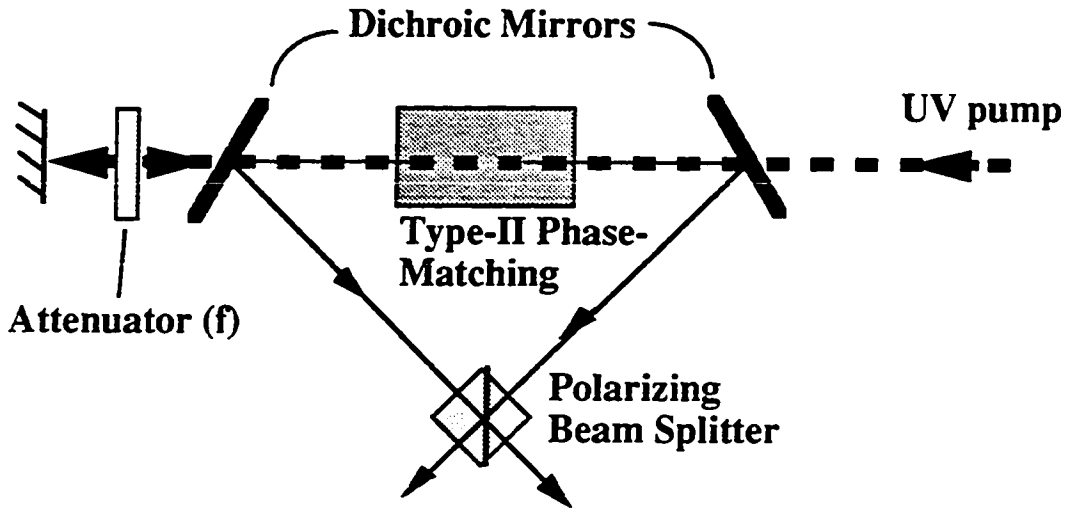


Figure 10.1c Proposed one-crystal scheme to produce EPR-state.

In Fig. 10.1c we show a possible means of producing the EPR-source using only *one* down-conversion crystal. The key is to reflect the pump beam back through the crystal. Although this scheme has the advantage that it reduces the number of optical elements (most importantly the number of crystals) it does have a few difficulties not present in the 2-crystal approach. First, one must use an isolator to prevent the pump from coupling back into the laser. Next, the return path must be kept short compared to the pump coherence length (typically 10 cm for a HeCd laser), to maintain coherence between the processes in which the down-converted pair originated from a left-going or a right-going pump photon. Finally, one must find very good dichroic mirrors to separate the pump beam from the down-converted photons. One solution is to use high-quality prisms (not shown in Fig. 10.1c).

10.III. Non-Ideal Down-Conversion Considerations

In practice, one must take into account several other features of the down-conversion process. First, the down-converted photons will have a finite bandwidth, so

that the members of each pair may have different frequencies (though energy conservation still constrains the *sum* of the frequencies to be equal to the pump frequency). We investigate the effect of this by letting the frequency of the horizontally-polarized photons be $\omega_0 + \omega$ and the frequency of the vertically-polarized photons be $\omega_0 - \omega$. We will always assume that the output spectra of the two crystals are identical, i.e., we do not need to further label our frequencies by the crystal number. Furthermore, because of the strong energy correlations, the phase $\delta = (\omega_0+\omega)\Delta x/c + (\omega_0-\omega)\Delta x/c = 2\omega_0\Delta x/c$ is independent of the spread in frequency. Then (10.2) becomes

$$|\psi\rangle \approx \int d\omega\, A(\omega) \left(|V_{\omega_0-\omega}\rangle_3 |H_{\omega_0+\omega}\rangle_4 + f e^{i\delta} |H_{\omega_0+\omega}\rangle_3 |V_{\omega_0-\omega}\rangle_4 \right). \quad (10.3a)$$

Note that the frequency of the light at each port is different in the two terms. Unless the bandwidth function $A(\omega)$ describing the amplitude for down-conversion production of the pair $|H_{\omega_0+\omega}\rangle |V_{\omega_0-\omega}\rangle$ is symmetric [i.e., $A(\omega) = A(-\omega)$], the photon color reaching a given detector can serve to *label* from which crystal a pair originated -- the first and second terms of (10.3a) arise from photons from crystals 1 and 2, respectively. To see this, consider the extreme case that $A(\omega) = \delta(\omega - \Omega)$; (10.3a) becomes

$$|\psi\rangle \approx |V_{\omega_0-\Omega}\rangle_3 |H_{\omega_0+\Omega}\rangle_4 + f e^{i\delta} |H_{\omega_0+\Omega}\rangle_3 |V_{\omega_0-\Omega}\rangle_4 \,. \quad (10.3b)$$

In principle, a precise frequency measurement at either detector could determine the *definite* polarizations of the photons. The distinguishability of the two terms of (10.3b) destroys their coherence even if no frequency measurement is actually made[7]; this

[7] The relationship between which-way information and interference was the subject of Chaps. 5 and 6; see also [Scully, et al., 1991; Kwiat, et al., 1991,19931; Tan and Walls, 1993].

weakens the correlations, making it impossible to violate a Bell's inequality.

This difficulty can be avoided, however, by inserting a half waveplate between crystal 2 and the polarizing beam splitter, oriented with its optic axis at 45° to the horizontal. This will effectively exchange the polarizations of the photons originating in that crystal ($|H_{\omega_o+\omega}, V_{\omega_o-\omega}\rangle_{\text{crystal 2}} \Rightarrow |V_{\omega_o+\omega}, H_{\omega_o-\omega}\rangle_{\text{crystal 2}}$), so that (10.2) instead becomes

$$|\psi\rangle = \int d\omega\, A(\omega)\left(|V_{\omega_o-\omega}\rangle_3 |H_{\omega_o+\omega}\rangle_4 + f e^{i\delta} |H_{\omega_o-\omega}\rangle_3 |V_{\omega_o+\omega}\rangle_4\right), \quad (10.4)$$

which shows that photon color no longer yields which-crystal information; interference persists, and violation of a Bell's inequality is possible.

For a plane-wave pump, the phase-matching constraints imply that with careful spatial filtering, one can in principle collect *only* conjugate pairs of photons from the crystals (i.e., essentially no unpaired photons, aside from stray light). Once we allow a more realistic, gaussian-mode pump, then this is no longer possible. For finite-sized collection irises, there will always exist situations where one photon is detected while the other is not (even aside from the problem of inefficient detectors). This effect is mitigated by collecting over a larger solid angle. (See Appendix D2.) Figure 10.2 shows a plot of inherent collection efficiency, versus the collection angle of the iris in units of the pump beam divergence angle[8]. We see that in order to keep the losses less than 2%, we must employ irises which accept light out to ~30 times the pump divergence angle. For example, if we employed a 325-nm pump with a beam waist radius ($1/e^2$) of 3.5 mm, then we would need to accept all half-angles up to 1 milliradian. (In practice, this could be accomplished by use of a precision spatial filter

[8] The theory underlying these results essentially involved (numerically) calculating integrals over truncated gaussians.

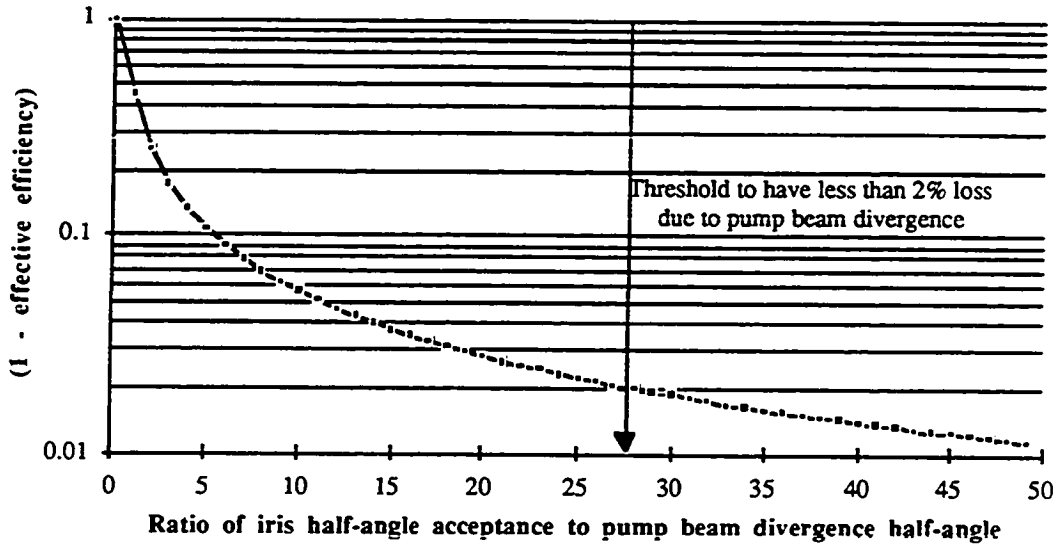


Figure 10.2 There is no way to guarantee that both photons will pass through irises of identical size (the optimal situation), due to the gaussian nature of the pump beam, which causes a slight spread in the angular correlation of the pair. The unavoidable losses may be reduced by using larger collection irises, however: When the acceptance angle of the irises is more than 30 times the divergence angle of the pump beam, losses from this effect are less than 2%. See Appendix D2.

system in each output port; see Fig. 10.1b.) If we assume that the length difference of the two paths is the minimum required to give a $\delta = 180^\circ$ phase-shift (i.e., $\Delta x = \lambda_{pump}/4$), then the correction factor to this phase shift for including a finite solid-angle of detection is at most 10^{-6}. This is completely negligible compared to the wavefront distortion from available optics, discussed in Sect. 10.V. Moreover, we will see there that one is much better off using $\delta \approx 0$ for other reasons anyway, in which case there is no correction from a finite detection solid-angle.

A third practical consideration is the effect of walk-off in the nonlinear crystals. While the birefringence of the nonlinear crystal is essential for achieving phase-

matching, it also results in a relative *displacement* of the two down-converted photons: they propagate in the same direction after exiting the crystal, but are separated by a distance $d = L \tan\rho$, where L is the propagation distance *inside* the crystal, and ρ is the intra-crystal angle between the ordinary and extraordinary beams [Dmitriev et al., 1991]. For a typical crystal length of 1 cm, and a typical ρ of 4° (appropriate for BBO, pumped at 54° to the optic axis; see Appendix D1), this separation is 0.7 mm, which can be a significant fraction of the beam width (cf. our previous example of a 3.5-mm beam radius). Consequently, after the polarizing beam splitter, the *position* of a detected photon partially labels its origin, degrading coherence. Consider, for example, horizontal polarization corresponding to the ordinary (undeviated) mode (i.e., in Fig. 10.1a, let the plane defined by the pump beam and the crystal optic axis be perpendicular to the plane of the paper) and vertical polarization corresponding to the extraordinary mode, *higher* by d. After the polarizing beam splitter, the two photon modes which travel to a given detector are therefore separated by the amount d. Just as in the situation of asymmetric spectra discussed above, this distinguishes the two otherwise coherent processes. Remarkably, insertion of an extra half waveplate after one of the crystals to rotate the polarization by 90° avoids this problem, in addition to solving the finite-bandwidth problem. Photons from either crystal exiting port 3 of the beam splitter would be initially extraordinary-polarized; photons exiting port 4 of the beam splitter would be initially ordinary-polarized. (Note that we could have arranged the optic axes of the crystals to lie in the plane of the figure. The waveplate will be an efficacious measure as long as the setup possesses mirror-symmetry about the plane perpendicular to the figure, as defined by the input and output beam splitters[9].)

Finally, even with a plane-wave pump, and a crystal cut for collinear phase-

[9] Actually, it is sufficient that the vectors $\mathbf{k}_1 \times \mathbf{n}_1$ and $\mathbf{k}_2 \times \mathbf{n}_2$ (where $\mathbf{k}_i$ and $\mathbf{n}_i$ are the pump beam and optic axis directions, respectively, associated with crystal i) be mirror-symmetric about this plane.

matching at degeneracy, there is no way to prevent *vector* phase-matching at small angles, for slightly different conjugate colors. In fact, a detailed calculation for BBO (see Appendix D1) has shown that even at degeneracy, there is a vector phase-matching solution, in addition to the collinear solution (see Figs. 10.3 and 10.4). However, because the angles of the conjugate ordinary and extraordinary modes (with respect to the pump beam) are nearly equal in magnitude (for small deviations from collinearity), the net effect on the phase-shift δ is completely negligible compared to the effect of imperfect optical elements, which we discuss later. Also, as with the walkoff problem, a half waveplate after crystal 2 can prevent the angle from distinguishing the interfering processes.

10.IV. Non-maximal Entanglement

We have until now ignored another parameter at our disposal, namely, the relative amplitude of the terms from the two crystals, governed in (10.2) by the factor f. As alluded to in Sect. 7.IV, one may reduce the required detector efficiency η from 83% to 67% by using a non-maximally entangled state [$|f| \neq 1$ in (10.1) - (10.4)], in the limit of no background[10] [Eberhard, 1993]. To gain some insight into this, we start with the standard Clauser-Horne Bell's inequality [Clauser and Horne, 1974], Eq. (7.12), rewritten here for convenience:

$$B \leq 0 \,, \tag{10.5a}$$

where

$$B \equiv n_{ss}(\alpha_1,\beta_1) + n_{ss}(\alpha_2,\beta_1) + n_{ss}(\alpha_1,\beta_2) - n_{ss}(\alpha_2,\beta_2)$$

$$- n_{3s}(\alpha_1) - n_{4s}(\beta_1) \,, \tag{10.5b}$$

[10] By "background", we mean all detector counts not arising from our correlated photons, i.e., stray light, detector dark counts, etc.

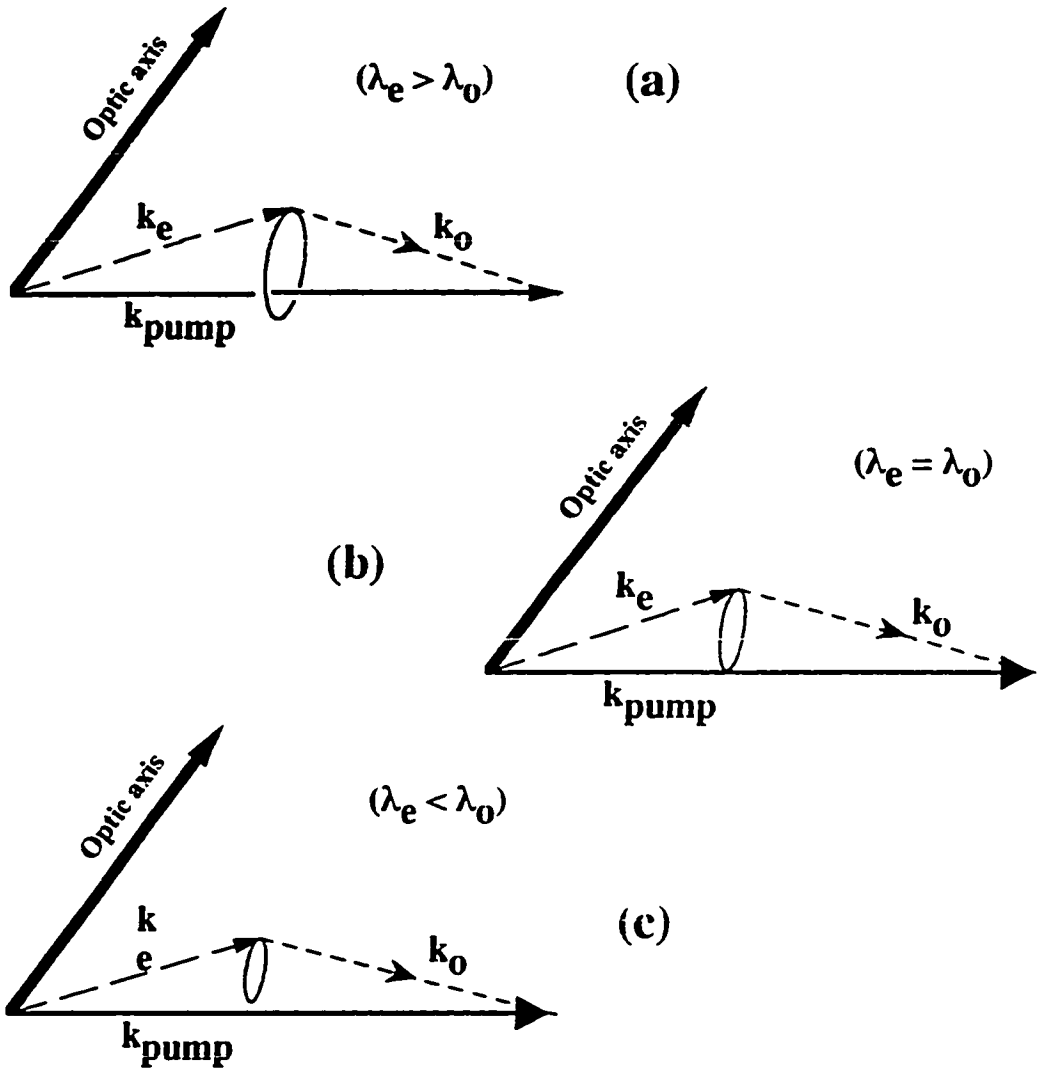
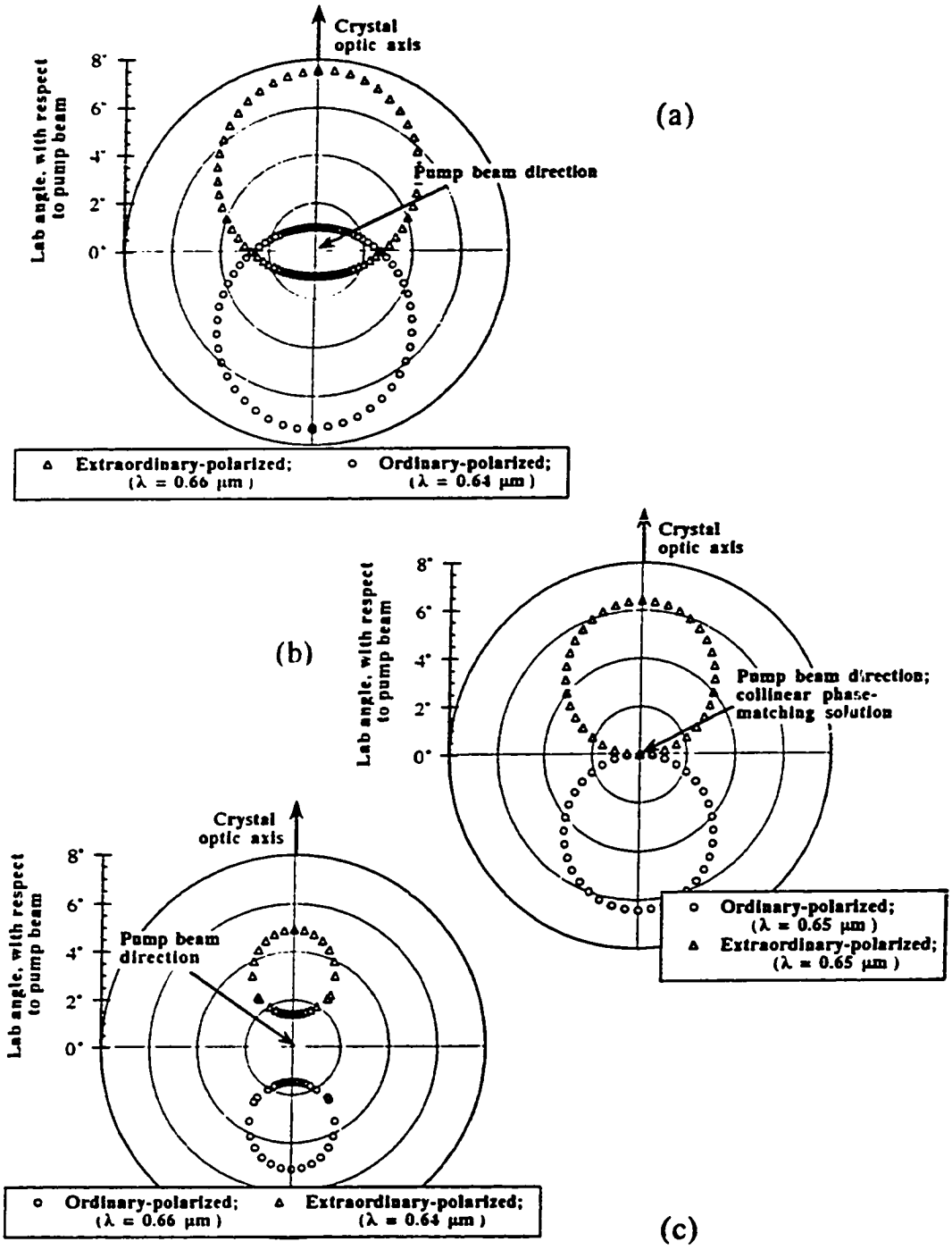


Figure 10.3. A schematic representation of spontaneous down-conversion into non-collinear modes (due to type-II vector phase-matching) for the case of a negative uniaxial crystal (such as BBO), in which the extraordinary index of refraction is less than the ordinary index (angles are exaggerated for clarity). The angle between the crystal optic axis and the pump beam is such as to allow collinear phase-matching for equal-frequency conjugate photons. Things to note: the angle between the extra-ordinary beam and the pump is slightly greater than the angle between the ordinary beam and the pump. These angles decrease as the wavelength of the ordinary (extraordinary) beam increases (decreases), as does the overall "phase-matching circle". a) The extraordinary beam tends to lie closer to the optic axis than does the ordinary beam, although both solutions are possible. b) Collinear and vector phase-matching are simultaneously possible. c) The extraordinary beam is always closer to the optic axis when $\lambda_e < \lambda_o$.

Figure 10.4 (next page) Plots showing Type-II vector phase-matching in BBO. The origin corresponds to the direction of the 325-nm pump beam (at 54° with respect to the BBO optic axis), and in (b) also to the collinear phase-matching solution. The left axis indicates the angular scale of the 4 concentric equal-angle contours. The points correspond to the transverse components (i.e., perpendicular to the pump vector) of the photons' momenta -- in other words, the plots are essentially what one would see looking into the pump beam, as shown in Fig. 10.3). The ordinary- and extraordinary-polarized photons of a given conjugate pair lie diametrically opposed, on either side of the origin.


Crystal optic axis
Lab angle, with respect to pump beam
8°
6°
4°
2°
0°
Pump beam direction
Extraordinary-polarized; (λ = 0.66 μm)
Ordinary-polarized; (λ = 0.64 μm)
(a)
Crystal optic axis
Lab angle, with respect to pump beam
Pump beam direction; collinear phase-matching solution
Ordinary-polarized; (λ = 0.65 μm)
Extraordinary-polarized; (λ = 0.65 μm)
(b)
Crystal optic axis
Lab angle, with respect to pump beam
Pump beam direction
Ordinary-polarized; (λ = 0.66 μm)
Extraordinary-polarized; (λ = 0.64 μm)
(c)

and $n_{3s}(\alpha_1)$ and $n_{4s}(\beta_1)$ are the singles count rate at the "s"-channels of the α–analyzer (in output port 3) and β–analyzer (in output port 4), respectively (we use here a notation appropriate for an ideal polarizing beam splitter analyzer which reflects all s-polarized light and transmits all p-polarized light). The n_{ss}'s are coincidence rates between the s-channels of the two analyzers. The parameters α_i and β_j are essentially the angles of the analyzers (though in practice, they might be the angles of half waveplates used to rotate the polarization of the incident light). Because the singles rates vary as η (e.g., $n_{3s} = \eta p_{3s}$, where p_{3s} is the singles rate for *unity*-efficiency), while the coincidence rates vary as η^2 [i.e., $n_{ss}(\alpha_i,\beta_j) = \eta^2 p_{ss}(\alpha_i,\beta_i)$, where p_{ss} is the coincidence rate for *unity*-efficiency], we must have

$$\eta \geq \frac{p_{3s}(\alpha_1) + p_{4s}(\beta_1)}{p_{ss}(\alpha_1,\beta_1) + p_{ss}(\alpha_2,\beta_1) + p_{ss}(\alpha_1,\beta_2) - p_{ss}(\alpha_2,\beta_2)} \tag{10.6}$$

in order to violate (10.5a). A straightforward calculation (see Appendix D3) shows that, in the case of perfect polarizing beam splitters, $p_{3s}(\alpha_1) = |A|^2\left[\cos^2\alpha_1 + |f|^2\sin^2\alpha_1\right]$. If $|f|$ is close to 1, then the singles rate is essentially a constant, independent of α_1; however, if we make f small (by attenuating the pump before crystal 2), then choosing α_1 close to 90° will substantially reduce the contribution of $p_{3s}(\alpha_1)$ to (10.6)[11]. A similar argument applies for $p_{4s}(\beta_1)$, which is reduced for β_1 close to 0°. Nevertheless, under these conditions, it is still possible to find values of α_2 and β_2 such that $B > 0$, even for η as low as 67%. The tradeoff is that the actual magnitude of B is reduced accordingly, however. For example, with $f = 0.311$, the maximum value of B is only

[11] Note that $p_{3p}(\alpha)$ varies as $\sin^2\alpha + |f|^2\cos^2\alpha$, so any attempt to simply average the signals from the p- and s-channels, for example by adding (10.6) to a similar inequality for the s-channel counts, as in the CHSH inequality [Clauser, et al., 1969], will remove the benefit of unbalancing the entanglement.

0.074, to be contrasted with the maximum value of 0.207 when f = 1. For this reason, background levels must be kept low for the method to be effective. We hope that background levels of 1% may be achievable.

A secondary effect of background counts is to place a lower limit on the number of "accidental" coincidences. Since these arise from *non*-correlated photons (or dark counts), they will reduce any violation of (10.5a). Accidental coincidence events result when only one member of a given conjugate pair is detected, simultaneously (within the coincidence resolution time) with another count (from a photon from a different pair or a stray-light photon or a dark count), or from simultaneous detection of two background counts. The former process is proportional to the rate of correlated-pair production, so in the absence of background, one can in principle make the accidental coincident rate as small as desired by reducing the pumping intensity to the two crystals (in practice, one eventually runs into another limitation: the system will thermally drift over long times). In the presence of a background rate B, however, the minimum accidental coincidence rate will be $B^2\Delta T$, where ΔT is the coincidence resolution time.

10.V. Imperfect Optical Elements

Thus far we have neglected the effects of imperfect analyzers, and an imperfect recombining beam splitter[12]. A general treatment is difficult (see Appendix D3); however, two special cases can be easily discussed. With no background, a maximally-

[12] It is reasonable to expect that we can obtain polarizing beam splitters with extinction ratios of 500:1; cf. Tom Bylack, at Prisms Unlimited [(714) (848-6007], who claims $|R_s|^2 = 0.995$, $|T_p|^2 = 0.98$-0.995 is a readily achievable specification. However, one must be certain to specify that (in the case of the recombining beam splitter) *both* input ports are to be used. Because of the gluing process, in general one input port will be better than the other--for one input port, the incident beam must propagate through the cement, bounce off the coated surface, and propagate through the cement again. Also, as below, wavefront distortion for the interferometer beam splitter must be minimal (i.e., better than $\lambda/20$).

entangled (f = 1) singlet-like ($\delta = 0°$ or 180°) state, and perfect beam splitters in the interferometer, the net detection efficiency must satisfy the following relation:

$$\eta\left(|R_p|^2 + |R_s|^2\right)\left[\sqrt{2}\left(\frac{|R_p|^2 - |R_s|^2}{|R_p|^2 + |R_s|^2}\right)^2 + 1\right] > 2 \ , \qquad (10.7)$$

where R_s is the reflection amplitude for s-polarized light (ideally, $|R_s| = 1$), and R_p is the reflection amplitude for p-polarized light (ideally, $|R_p| = 0$). (We have implicitly assumed a lossless beam splitter here.) This result is in agreement with that derived by Clauser and Horne [Clauser and Shimony, 1978; Clauser and Horne, 1974], and yields the well-known 83% "limit" for perfect analyzers.

Next we consider the case of perfect analyzers but an imperfect recombining polarizing beam splitter (in the interferometer); specifically, we consider a lossless splitter whose reflection and transmission amplitudes are related by $|r_s| = |t_p|$ and $|r_p| = |t_s|$ (i.e., the non-idealities for the two polarizations are equal). As before, we keep f = 1 and no background. We find that our results depend crucially on whether $\delta = 0°$ or 180°. In the former case, both the singles rates and the coincidence rates are independent of $|r_s| = |t_p|$; consequently, the value of B is also. In other words, even a non-polarizing 50-50 beam splitter could be used as the recombiner. The reason is that under these conditions there is quantum interference which prevents the two photons from exiting the same port of the beam splitter[13]; see Appendix D3. However, for the $\delta = 180°$ (triplet-like) case, the coincidence rates, and therefore the value of B, depend strongly on $|r_s| = |t_p|$. This increases the required detection efficiency to achieve a violation (B > 0):

---

[13] This is, in some sense, the complementary effect to the Hong-Ou-Mandel interference [Hong et al., 1987; Steinberg et al., 1993], in which the two-photons from a down-conversion crystal always take the same port of a 50-50 beam splitter (cf. Sect. 5.III.)

$$\eta > \frac{2}{(1+\sqrt{2})\left(|r_s|^2 - |r_p|^2\right)^2} \tag{10.8}$$

is required to close the detection loophole. A plot of the right-hand side of (10.8) is shown in Fig. 10.5, along with the minimum required efficiencies for $\delta = 0°$ and for f = 0.608. It is immediately apparent that one is much better off choosing $\delta = 0°$ (the

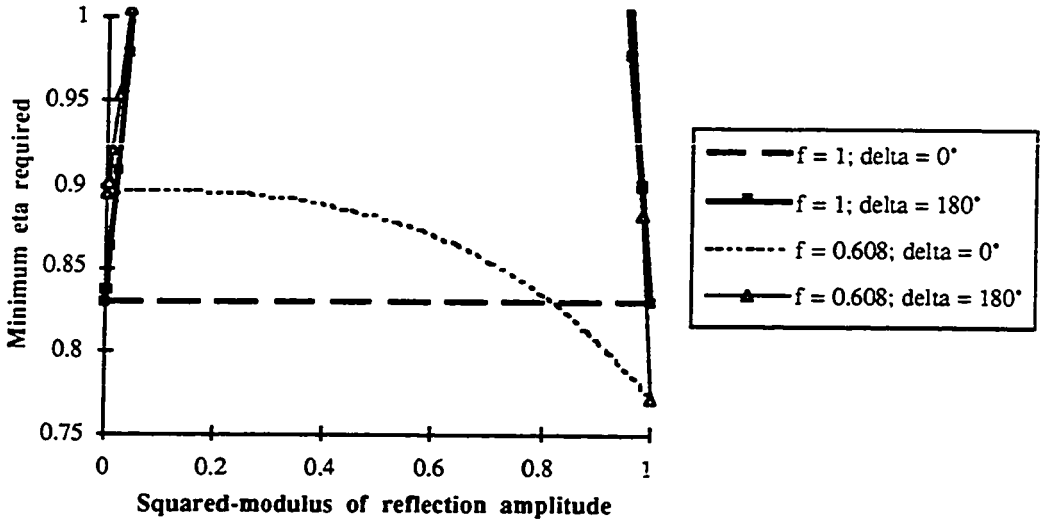


Figure 10.5 When $\delta = 180°$, the value of B [Eq. (10.5b)] is strongly dependent on the recombining polarizing beam splitter, placing strong constraints on the efficiency required to achieve a violation of $B \leq 0$. Unless the amount of "cross-talk" [i.e, the percentage of p-polarized light that exits the s-channel (equal to $|r_p|^2$), and the percentage of s-polarized light that exits the p-channel (equal to $|t_s|^2$)] is less than 4%, it is not possible to violate the Bell's inequality at all. (We also assume a lossless beam splitter, and that the cross-talk between polarization modes is symmetric, i.e., $|r_p|^2 = |t_s|^2$. A general calculation is given in Appendix D3.) However, if $\delta =$ 0°, the value of B is independent of $|r_p|^2 = |t_s|^2$ for f = 1, and not a strong function for $f \neq 1$. Hence, the requirements on detection efficiency are not nearly so severe.

efficiency's dependence on $|r_s| = |t_p|$ is less for f ~ 1, as is required at higher background levels [Eberhard, 1993]). Of course, any phase shift equal to an integer multiple of 360° is equivalent. However, to mitigate effects from unequal path lengths, one should strive to operate in a white-light configuration, with exactly equal optical path lengths in the two arms.

If instead of assuming that the cross-talk of the recombining beam splitter is equal for the two polarizations, we assume that the beam splitter is ideal for one polarization (e.g., $|t_p| = 1$), then the functional dependence of B or eta on $|r_s|$ does not depend on δ. See Fig. 10.6. The quantum interference effect that causes δ = 0° to be superior to δ = 180° depends on the beam splitter cross-talk being equal for both s- and p-polarization.

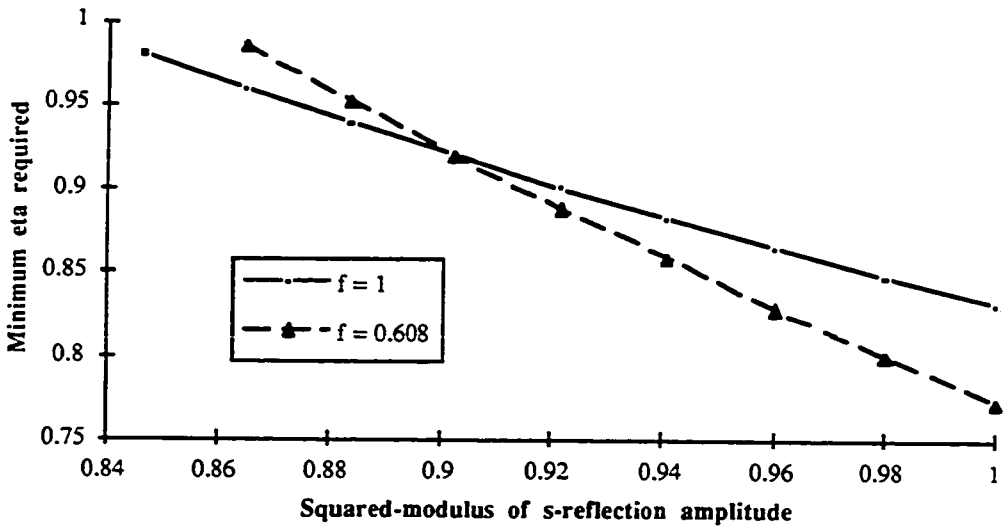


Figure 10.6 We assume for these plots a lossless beam splitter and that $|t_p| = 1$. The values of B and eta are independent of whether δ equals 0° or 180°, if only one channel of the recombining polarizing beam splitter is lossy (i.e, the beam splitter works perfectly for the other polarization). Apparently, the strong benefits of using a triplet-like state over a singlet-like state exist only if the the cross-talk between polarization modes is symmetric, i.e., $|r_p|^2 = |t_s|^2$ (cf. Fig. 10.5).

Ideally, one would like completely flat, homogeneous, optical elements to avoid wavefront distortion inside the interferometer. The effect of such distortion is to change the value of the relative phase shift δ for parts of the beam. The total detected rates will then be averaged over a range of δ-values (note that there will also be an averaging over δ if there is an uncorrected temporal *drift* of the optical path lengths). To estimate the effect of this, we have calculated the value of B, as function of δ. Typical results are shown in Fig. 10.7. We see that the value is not a strong function of the phase deviation from $\delta = 0°$. Therefore, it should suffice to specify a $\lambda/20$ flatness for the optics within the interferometer (a feasible requirement over a clear aperture of 1 cm), allowing a phase shift $\delta = 0 \pm 10°$.

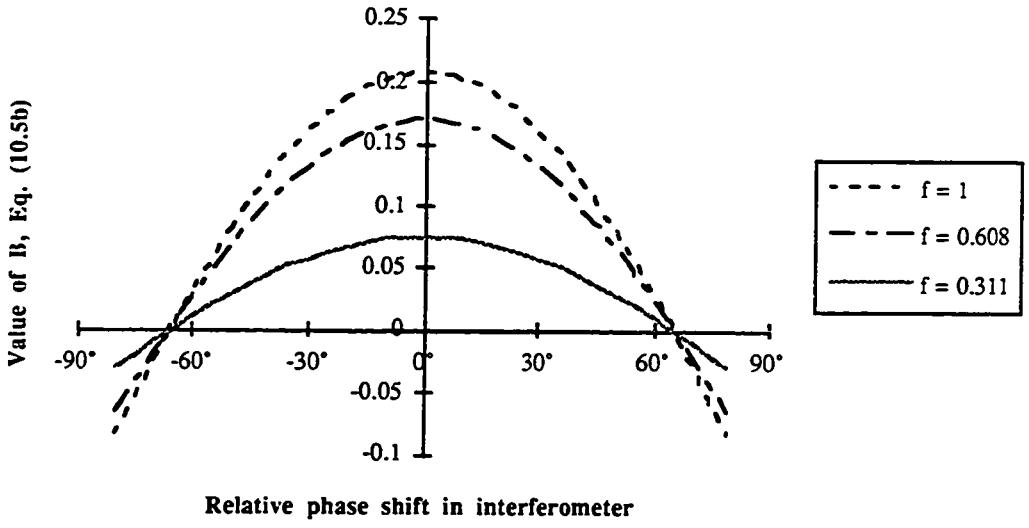


Figure 10.7 The dependence of B [Eq. (10.5b)] on δ, for various values of f (assuming $\eta = 1$, no background, and perfect polarizing beam splitters). Wavefront distortion arising from imperfect optics implies that the true value of B will be a sort of weighted average over a range of angles. Due to the weak dependence at small deviations from the ideal phase shift $\delta = 0$, this should not pose a serious problem if the deviations can be kept less than $\pm 10°$.

10.VI. Conclusion

For $\lambda/20$-flatness optics, a background level of 1%, and custom-selected polarizing beam splitters (with an extinction ratio of at least 500:1), numerical calculation predicts that a violation should be possible choosing $\delta \approx 0$, $\alpha_1 = 99°$, $\alpha_2 = 58°$, $\beta_1 = 9°$, $\beta_2 = -32°$, and $f = 0.55$, as long as the net detection efficiency is greater than 82.6%. Naturally, all optics would be anti-reflection coated to minimize reflection losses; including a 0.25% loss for each interface, and the 2% loss from the gaussian nature of the beam, this means that the bare detector efficiency needs to be at least 88.6%, which may be achievable in light of our recent measurements (Chap. 9, and [Kwiat et al., 1993a,b]). Of course, for a safety margin, one would like it to be even higher, if possible.

In conclusion, we have investigated a novel source of EPR-correlated particles, which makes use of a two-crystal interferometer. If one employs type-II phase-matching and a polarizing beam splitter to combine the outputs of the two crystals, a true singlet-like or triplet-like state may be produced without the need of discarding counts. Using a half waveplate in one of the interferometer arms to exchange the roles of the ordinary and extraordinary polarizations removes difficulties arising from finite bandwidth, walk-off, and vector phase-matching considerations. With the setup described herein it should be possible to produce an indisputable violation of a Bell's inequality. We have examined the effects of background, imperfect polarizing beam splitters, and phase-distorting optics. A loophole-free experiment still seems feasible, although detectors need to be improved somewhat.

Chapter 11: Conclusion

Nothing is forgotten; nothing is ever forgotten.

--Robin Hood, The Adventures of Robin Hood (1980's Television Series)

Our present journey through Quantum Wonderland is now essentially over. All that remains is to look back over the path we have travelled (cf. the parable that introduced Chap. 1), and see what we have learned. We shall first reminisce about some of the specific "places" visited, and then try to make some overall comments.

The phenomenon of Berry's phase has been demonstrated at a single-photon level, in a way which excludes with very high probability any possible classical explanation. We conclude that this phase originates fundamentally at the quantum level, as do all phases (e.g., the dynamical phase). Under special circumstances, these phases can survive the correspondence principle limit onto the classical level. In other cases, however, this is not the case. One such example is the Franson interference experiment (Chap. 8), in which the phase is "jointly shared" by two separated, but entangled, particles, i.e., the interference depends only on the sum of the phases accumulated by the constituent particles. There is no correspondence principle limit for this particular situation of a two-particle wavefunction defined on the configuration space of the particles. Based on these examples, classical interference is in some sense a "special case" of the more fundamental quantum-mechanical interference.

Related to the notion that we should treat our interfering systems quantum mechanically is the conclusion that in some experiments we must also treat our *measuring apparatus* quantum mechanically, as stressed by the quantum eraser

experiments. Only by doing so can we understand the revival of interference, which relies on the fact that the coherence of the initial particle is passed along to the measuring apparatus. If this coherence can be maintained, then one can at some later time (in principle long after the initial particle is detected) decide "posthumously" (i.e., after this detection, but before the measuring apparatus is measured) whether to gain which-path information or interference, by correlating the results of particular measurements on the measuring apparatus with the seemingly random detection record of the initial particle.

As we discussed in Chap. 5, the quantum eraser is just another means of gaining insight into the nature of entangled states. In the study of Bell's inequalities, entangled states are the means of gaining insight about Nature itself. In observing a violation of a Bell's inequality based on energy and time correlations (accepting for the moment the supplementary assumptions) we can conclude that Nature is nonlocal, and that the energy and time of emission of our down-converted photons are not "elements of reality". Of course, one need not abandon the objective reality of the energy and time if one is willing to accept that the outcomes in one location depend on the *parameter* settings in the other. Our personal preference is to avoid this parameter-dependence at the expense of well-defined properties.

The correlations observed in tests of Bell's inequalties are nonlocal--a measurement on one of the entangled particles causes the wavefunction of the other particle to "collapse" to a particular value. Using a system nearly identical to the Berry's phase setup we also saw a series of results also readily interpretable in terms of the collapse postulate. The status of the "collapse of the wavefunction" is somewhat similar to the status of local hidden variable theories between 1935 (EPR) and 1964 (Bell): There are no apparent situations where one can distinguish *experimentally* between this "intuitive" model and the other main contenders, the Bohm-deBroglie guiding-wave picture and Everett's many worlds picture. Although I think it unlikely that there will

ever be such a situation where these very different approaches, if applied correctly, will yield different predictions, such a circumstance would be welcome. Just as the work of Bell opened the door for experimental refutation of local hidden variable models (not yet accomplished unambiguously, although in Chap. 10 we proposed a novel scheme for such a "loophole-free" test), a similar opportunity to actually rule out the "collapse" picture or the "many-worlds" picture or the "pilot-wave" picture would be desirable.

It seems most appropriate to end with a general discussion of the nonclassicality in our results, given that that is (half of) the title of this work. Specifically, we have seen two very different *types* of non-classicality in the experiments; fortuitously, the photon pairs produced in spontaneous parametric down-conversion permit investigation of both of these nonclassical aspects. The first, demonstrated in the single-photon Berry's phase and Collapse experiments, arises from the *indivisibility* of a single quantum. This leads to an anti-correlation in the counts at detectors looking at the two output ports of a beam splitter, and consequently a near-zero value of the anti-correlation parameter a, which classically must be greater than or equal to 1 because any classical wave will split evenly at a 50-50 beam splitter.

This same notion of an indivisible quantum was also quite important in demonstrating the phenomenon of quantum erasure, where we wished to emphasize that even after an initial particle was detected, one could choose between observing interference or gaining which-path information by making a suitable measurement on a measuring apparatus with which it was entangled. The interpretation does not carry through if a classical wave is used: The notion of which-path information is meaningless since a classical wave can follow two paths simultaneously[1]. Of course,

[1] As we pointed out, it is also important, even with indivisible quanta, that we use *single*-particle states; otherwise one must contend with the situation in which some of the particles follow each route, and are indistinguishable from one another.

in many situations one hears similar statements. e.g., that a single photon actually travels through two slits at the same time. The key difference is that the quantum "particle" can exhibit both particle-like and wave-like behavior (though not both simultaneously, according to the principle of complementary), whereas a classical object can exhibit only one or the other.

The second non-classicality arises from the nonlocal nature of quantum-mechanical entangled states. This aspect was stressed in the Collapse of the Wavefunction experiment and the Bell's inequality experiment for energy and time. From violations of appropriate Bell's inequalities, we can conclude that no "local realistic" theory, of which classical-field models are a subset, can explain the correlations in Nature. Local hidden variable models based solely on standard classical theory are not even optimal, in the sense that other LHV variations could come closer to mimicking quantum mechanics. Specifically, we have seen that the modified version of a classical local hidden variable model, in which a particle exits a given port (of an interferometer or a polarizing beam splitter) if and only if the probability (calculated for a classical wave) is greater than 50%, yields 100%-visibility triangular coincidence fringes, and satisifes the Bell's *equality*. Curiously, the feature of this model which causes it to behave more quantum mechanically than does the classical wave model is the assumption of particle *indivisibility*, which is the physical interpretation of our modified probability at a beam splitter. Put differently, the maximum visibility in a Franson-type experiment is 50% for a classical wave. Incorporating the property of indivisibility produces the maximum correlations allowed in any theory constrained by local realism. Nevertheless, the correlations predicted by quantum mechanics are greater still. And Nature seems to be siding with quantum mechanics (the theory), even though it seems to have no great compassion for quantum mechanics (the people).

Appendix A: Triple Coincidence Inequality

The following formal calculation is presented courtesy of Dr. Ivan Deutsch. We feel it is important to include it here, though, since it is a main result used in one of our experiments (see Chap. 3), but is not published elsewhere, even in Dr. Deutsch's own dissertation [Deutsch, 1992].

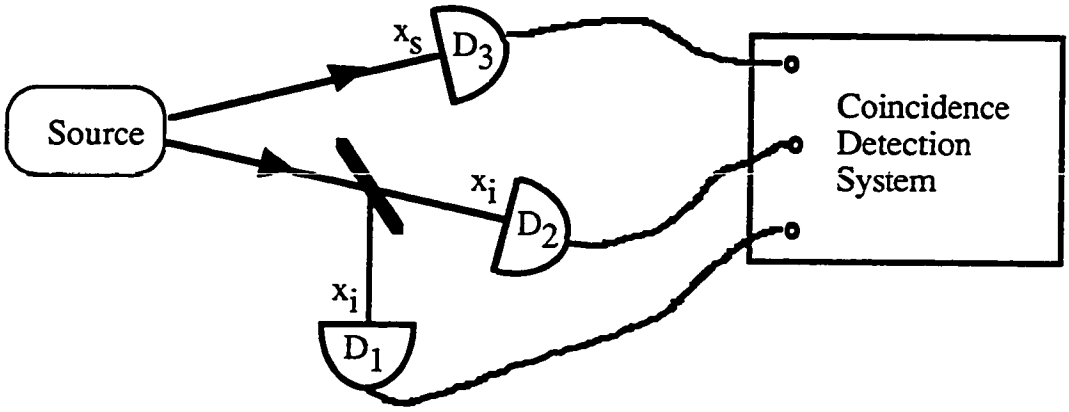


Figure A1

Herein we demonstrate that there is an inequality (involving the various count rates at three detectors) which is obeyed by classical fields (defined below), but not necessarily by manifestly quantum states of light. Our motivation is that we want to claim that particular results of experiments using the correlated photons from spontaneous parametric down-conversion do not possess any classical explanation. (See Chapter 3.) The basic system we are considering is shown in Fig. A1. Three detectors D1, D2, and D3 detect light from some source S. A beam splitter allows D1 and D2 to look at the same mode, while D3 looks at a second mode. A coincidence detection system permits measurement of the standard coincidence rates N_{13} (between

detectors D1 and D3) and N_{23} (between detectors D2 and D3), the rate of triple coincidences N_{123} (between all three detectors), as well as the singles rate N_3 at D3. We seek to prove that for classical fields

$$N_3 N_{123} \geq N_{13} N_{23} , \tag{A.1}$$

where, formally, the various N's are given in terms of correlation functions as

$$\begin{aligned} N_{j_1 j_2 \ldots j_n} &\equiv G^{(n)}(x_1, x_2, \ldots x_n; x_1, x_2, \ldots x_n) \\ &= \langle E^{(-)}(x_1) E^{(-)}(x_2) \cdot \cdot \cdot E^{(-)}(x_n) E^{(+)}(x_1) E^{(+)}(x_2) \cdot \cdot \cdot E^{(+)}(x_n) \rangle \\ &= \mathrm{Tr}\left(\rho_{\mathrm{field}} E^{(-)}(x_1) E^{(-)}(x_2) \cdot \cdot \cdot E^{(-)}(x_n) E^{(+)}(x_1) E^{(+)}(x_2) \cdot \cdot \cdot E^{(+)}(x_n)\right) . \end{aligned} \tag{A.2}$$

In words, the number of triple coincidences for all three detectors times the singles rate at the unpaired detector should be greater than the product of the double coincidences between this detector and the two paired detectors, for any classical source.[1] The inequality (A.1) was presented without proof by Grangier et al. [Grangier et al., 1986]; we now provide a formal proof.

We begin by assuming that D1 and D2 are placed equal distances from the beam splitter, so that they are affected by the field strength at the same position, here denoted x_i. Similarly, D3 is affected by the field strength at the position x_s. Now, if the field is classical in nature, the state is given by a density matrix that can be expressed in the "P-

[1]Of course, the inequality does not hold for the photon pairs produced in down-conversion, since a given idler photon (mode associated with x_i) can be detected at only one of D1 and D2; triple coincidences can only occur if more than one pair of photons is involved; by reducing the pumping intensity one can arbitrarily decrease the likelihood of this.

representation". For a single-mode field this is written

$$\rho = \int d^2\alpha\, P(\alpha)\, |\alpha\rangle\langle\alpha| \ , \tag{A.3}$$

where P is *positive definite*. Then any normally-ordered expectation value of $a^\dagger$ and a can be written

$$\langle (a^\dagger)^n a^m \rangle = \mathrm{Tr}\left(\rho\, (a^\dagger)^n a^m\right) = \int d^2\alpha\, P(\alpha)\, (\alpha^*)^n \alpha^m \equiv \overline{(\alpha^*)^n \alpha^m} \ , \tag{A.4}$$

where the "bar"-notation denotes integration weighted by the P-distribution.

For a multimode field we can write

$$\rho = \int d^2\xi\, P[\xi]\, |\{\xi\}\rangle\langle\{\xi\}| \ , \tag{A.5}$$

where $|\{\xi\}\rangle$ is a multimode coherent state ($E^{(+)}(x)\,|\{\xi\}\rangle = \xi(x)\,|\{\xi\}\rangle$) and the P-function is now a *functional* depending on the classical field ξ. Then

$$\begin{aligned} &\langle E^{(-)}(x_1)\cdots E^{(-)}(x_n)\, E^{(+)}(x_{n+1})\cdots E^{(+)}(x_m)\rangle \\ &= \int d^2\xi\, P[\xi]\, \xi^*(x_1)\cdots \xi^*(x_n)\, \xi(x_{n+1})\cdots \xi(x_m) \\ &= \overline{\xi^*(x_1)\cdots \xi^*(x_n)\, \xi(x_{n+1})\cdots \xi(x_m)} \ . \end{aligned} \tag{A.6}$$

In particular,

$$N_3 = G^{(1)}(x_s; x_s) = \langle E^{(-)}(x_s) E^{(+)}(x_s) \rangle = \overline{|\xi(x_s)|^2} , \tag{A.7a}$$

$$N_{13} = N_{23} = G^{(2)}(x_i, x_s; x_i, x_s)$$

$$= \langle E^{(-)}(x_i) E^{(-)}(x_s) E^{(+)}(x_s) E^{(+)}(x_i) \rangle = \overline{|\xi(x_i)|^2 |\xi(x_s)|^2} , \tag{A.7b}$$

and

$$N_{123} = G^{(3)}(x_i, x_i, x_s; x_i, x_i, x_s)$$

$$= \langle E^{(-)}(x_i) E^{(-)}(x_i) E^{(-)}(x_s) E^{(+)}(x_s) E^{(+)}(x_i) E^{(+)}(x_i) \rangle = \overline{|\xi(x_i)|^4 |\xi(x_s)|^2} . \tag{A.7c}$$

Thus, we seek to prove:

$$\left(\overline{|\xi(x_i)|^4 |\xi(x_s)|^2} \right) \left(\overline{|\xi(x_s)|^2} \right) \geq \left(\overline{|\xi(x_i)|^2 |\xi(x_s)|^2} \right)^2 . \tag{A.8}$$

To do accomplish this, we will use a version of the Schwartz-inequality, which we now present. Consider $\overline{fg} \equiv \int dr\, P(r)\, f(r)\, g(r)$, where P is some *positive-definite* function. Then

$$\overline{fg} = \int dr \left(\sqrt{P(r)}\, f(r) \right) \left(\sqrt{P(r)}\, g(r) \right) = \int dr\, h_1(r)\, h_2(r) , \tag{A.9a}$$

where $h_1(r) = \sqrt{P(r)}\, f(r)$ and $h_2(r) = \sqrt{P(r)}\, g(r)$. Since h_1 and h_2 are *real* functions,

$\overline{fg} = (h_1, h_2)$, the standard inner-product for functions. Similarly, we can calculate

$$\overline{f^2} = \int dr\, P(r)\, f^2(r) = \int dr \left(\sqrt{P(r)}\, f(r)\right)\left(\sqrt{P(r)}\, f(r)\right) = (h_1, h_1) \tag{A.9b}$$

$$\overline{g^2} = \int dr\, P(r)\, g^2(r) = \int dr \left(\sqrt{P(r)}\, g(r)\right)\left(\sqrt{P(r)}\, g(r)\right) = (h_2, h_2) \tag{A.9c}$$

We then use (A.9) and the standard Schwartz inequality, $(h_1, h_1)(h_2, h_2) \geq |(h_1, h_2)|^2$, to find

$$\overline{f^2}\,\overline{g^2} \geq \overline{fg}^2 , \tag{A.10}$$

which holds for averages of functions over a positive-definite probability function. The proof of (A.8), and thus of (A.1), is completed by using (A.10), and letting $f = |\xi(x_i)|^2 |\xi(x_s)|$ and $g = |\xi(x_s)|$.

Appendix B1: General Calculation for Polarization Quantum Eraser

We now present a fairly formal, general calculation of both the singles and coincidence rates expected from the setup shown in Fig. 5.1. As a preliminary to the general calculation of the quantum eraser effect, with arbitrary orientations of the half waveplate, polarizer P1, and polarizer P2, we write the effective projection operator for a polarizer at an angle θ to the horizontal, placed along the path corresponding to the propagation mode index j:

$$\hat{P}_{pol,j}(\theta) = \left| 1_j^{H+\theta} \right\rangle \left\langle 1_j^{H+\theta} \right|$$

$$= \left(\left| 1_j^H \right\rangle \cos\theta + \left| 1_j^V \right\rangle \sin\theta \right) \left(\left\langle 1_j^H \right| \cos\theta + \left\langle 1_j^V \right| \sin\theta \right) , \qquad \text{(B1.1)}$$

where we are considering only the effect on single-photon states. It will be useful to consider the effect of this operator on a state of arbitrary polarization:

$$\hat{P}_{pol,j}(\theta) \left| 1_j^{H+\phi} \right\rangle = \left| 1_j^{H+\theta} \right\rangle \left\langle 1_j^{H+\theta} \right| 1_j^{H+\phi} \rangle$$

$$= \left(\left| 1_j^H \right\rangle \cos\theta + \left| 1_j^V \right\rangle \sin\theta \right) \left(\left\langle 1_j^H \right| \cos\theta + \left\langle 1_j^V \right| \sin\theta \right) \left(\left| 1_j^H \right\rangle \cos\phi + \left| 1_j^V \right\rangle \sin\phi \right)$$

$$= \left| 1_j^H \right\rangle [\cos\theta(\cos\theta\cos\phi + \sin\theta\sin\phi)]$$

$$+ \left| 1_j^V \right\rangle [\sin\theta(\cos\theta\cos\phi + \sin\theta\sin\phi)] . \qquad \text{(B1.2)}$$

For example, we can then examine the rate of singles detection for a single-photon state,

horizontally polarized, passed through a polarizer:

$$P_s \propto G^{(1)}(t;t) = \langle \psi | \hat{E}^{(-)}(t)\ \hat{E}^{(+)}(t) | \psi \rangle$$

$$= \langle\, 1^H \,|\, \hat{P}_{pol,j}(\theta)\ \left(\hat{a}_H^\dagger \hat{a}_H + \hat{a}_V^\dagger \hat{a}_V\right) \hat{P}_{pol,j}(\theta) \,|\, 1^H \rangle$$

$$= \cos^4\theta + \cos^2\theta \sin^2\theta = \cos^2\theta \ , \qquad \text{(B1.3)}$$

which is the result expected from Malus's law.

We now show that using a single polarizer, before detector D1 for instance, is not enough to revive the interference dip. Using the reduced coincidence detection operator from Eq. (5.9), and the entangled state of the photons from (5.8b), we have

$$P_c(0) \propto \langle \psi \,|\, \hat{P}_{pol,1}(\theta_1)\, \hat{P}'_{c,red.} \hat{P}_{pol,1}(\theta_1) \,|\, \psi \rangle_{\Delta x=0}$$

$$= \frac{1}{2}\left[\langle 1_1^H\, 1_2^{H+\phi} \,|\ - \langle 1_1^{H+\phi}\, 1_2^H \,|\right] \ |\ 1_1^{H+\theta_1}\rangle \langle\, 1_1^{H+\theta_1} \,|\ \left(\hat{a}_{1,H}^\dagger \hat{a}_{1,H} + \hat{a}_{1,V}^\dagger \hat{a}_{1,V}\right)$$

$$\times \left(\hat{a}_{2,H}^\dagger \hat{a}_{2,H} + \hat{a}_{2,V}^\dagger \hat{a}_{2,V}\right) \,|\ 1_1^{H+\theta_1}\rangle \langle\, 1_1^{H+\theta_1} \,|\ \frac{1}{2}\left[\,|\, 1_1^H\, 1_2^{H+\phi}\rangle - |\, 1_1^{H+\phi}\, 1_2^H\rangle\right]$$

$$= \frac{1}{4}\sin^2\phi\left(\cos^4\theta + 2\cos^2\theta \sin^2\theta + \sin^4\theta\right) = \frac{1}{4}\sin^2\phi \ . \qquad \text{(B1.4)}$$

We see immediately that for $\phi = 90°$, when the two paths are maximally distinguishable there is no null in coincidence for any orientation of a single polarizer at the output.

We turn now to the general case with two polarizers set at arbitrary angles θ_1 and θ_2.

$$P_c(0) \approx \langle \psi \left| \hat{P}_{pol,1}(\theta_1)\, \hat{P}_{pol,2}(\theta_2) \hat{P}'_{c,red.} \hat{P}_{pol,2}(\theta_2) \hat{P}_{pol,1}(\theta_1) \right| \psi \rangle_{\Delta x=0}$$

$$= \frac{1}{2}\left[\langle 1_1^H\, 1_2^{H+\phi}\right| - \langle 1_1^{H+\phi}\, 1_2^H\left|\right]\left(\left|\, 1_1^{H+\theta_1}\right\rangle \left\langle 1_1^{H+\theta_1}\,\right|\right)\left(\left|\, 1_2^{H+\theta_2}\right\rangle \left\langle 1_2^{H+\theta_2}\,\right|\right)$$

$$\times \left(\hat{a}^\dagger_{1,H}\, \hat{a}_{1,H} + \hat{a}^\dagger_{1,V}\, \hat{a}_{1,V}\right)\left(\hat{a}^\dagger_{2,H}\, \hat{a}_{2,H} + \hat{a}^\dagger_{2,V}\, \hat{a}_{2,V}\right)$$

$$\times \left(\left|\, 1_2^{H+\theta_2}\right\rangle \left\langle 1_2^{H+\theta_2}\,\right|\right)\left(\left|\, 1_1^{H+\theta_1}\right\rangle \left\langle 1_1^{H+\theta_1}\,\right|\right)\frac{1}{2}\left[\left|\, 1_1^H 1_2^{H+\phi}\right\rangle - \left|\, 1_1^{H+\phi} 1_2^H\right\rangle\right]. \quad \text{(B1.5)}$$

Using Eq. (B1.2), one can expand out $\left|\, \tilde{\psi}\right\rangle_{\Delta x=0} = \hat{P}_{pol,2}(\theta_2)\hat{P}_{pol,1}(\theta_1) \left|\, \psi\right\rangle_{\Delta x=0}$.
After simplifying algebra one finds:

$$\left|\, \tilde{\psi}\right\rangle_{\Delta x=0} = \left|\, 1_1^H\, 1_2^H\right\rangle \cos\theta_1 \cos\theta_2 \sin\left(\theta_2 - \theta_1\right) \sin\phi$$

$$+ \left|\, 1_1^V\, 1_2^V\right\rangle \sin\theta_1 \sin\theta_2 \sin\left(\theta_2 - \theta_1\right) \sin\phi$$

$$+ \left|\, 1_1^H\, 1_2^V\right\rangle \cos\theta_1 \sin\theta_2 \sin\left(\theta_2 - \theta_1\right) \sin\phi$$

$$+ \left|\, 1_1^V\, 1_2^H\right\rangle \sin\theta_1 \cos\theta_2 \sin\left(\theta_2 - \theta_1\right) \sin\phi\ . \quad \text{(B1.6)}$$

It then follows that

$$P_c(0) \approx \langle \tilde{\psi} \left| \hat{P}'_{c,red.} \right| \tilde{\psi} \rangle_{\Delta x=0} = \sin^2\phi \sin^2\left(\theta_2 - \theta_1\right)\ , \quad \text{(B1.7)}$$

which is the more general case of Eq. (5.15).

Appendix B2: General Calculation for Quantum Eraser Proposal

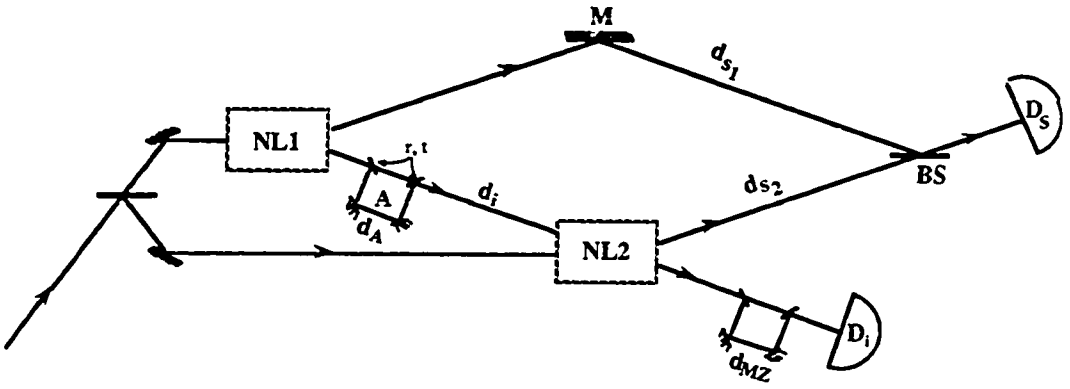


Figure B2.1 Schematic for general calculation given below. The two beam splitters of the inter-crystal Mach-Zehnder are assumed to have the same reflection coefficients and the same transmission coefficients. The beam splitters in the Mach-Zehnder before detector D_i are both assumed to be 50-50.

We now present a fairly formal, general calculation of both the singles and coincidence rates expected from the setup shown in Fig.B2.1. Both beam splitters of the inter-crystal Mach-Zehnder will be assumed to have the same reflection and transmission amplitudes (r and t, respectively); the beam splitters of the post-crystal Mach-Zehnder are assumed from the outset to be the 50-50 variety. Associated with the short idler path between the crystals (transmission at both inter-crystal beam splitters) is the delay τ_i, while the longer path (reflection at both beam splitters) has the delay $\tau_i + \tau_A$. In the calculations that follow, we assume a monochromatic pump beam at frequency $\omega_p = 2\omega_o$, and use a single spatial-mode treatment for each of the down-converted photons: the signal (idler) modes from crystal 1 and crystal 2 are labelled s_1 (i_1) and s_2 (i_2), respectively. Futhermore, we assume that the modes of the idler

photons from the two nonlinear crystals are spatially indistinguishable ($i_1 = i_2 = i$). Then we may write the wave function after the two crystals as

$$|\psi\rangle = \int_{-\infty}^{\infty} d\omega_1 \, A(\omega_1) \frac{1}{\sqrt{2}} \Big[|\omega_o+\omega_1, \omega_o-\omega_1\rangle_{s_1,i} \, e^{i(\omega_o-\omega_1)\tau_i} \left(r^2 \, e^{i(\omega_o-\omega_1)\tau_A} + t^2\right)$$

$$+ |\omega_o+\omega_1, \omega_o-\omega_1\rangle_{s_2,i} + |\omega_o+\omega_1, 0\rangle_{s_1,i} \, e^{i(\omega_o-\omega_1)\tau_i} \left(rt \, e^{i(\omega_o-\omega_1)\tau_A} + rt\right) \Big] , \quad \text{(B2.1)}$$

where we have omitted the vacuum term (for the predominant, but uninteresting, case in which neither crystal down-converts), and higher order terms (for the very unlikely case in which more than one pump photon down-converts). The state (B2.1) describes a signal-idler photon pair, satisfying energy conservation (enforced by the effectively infinite interaction time), and originating with equal probability in crystal 1 (first and third kets) or crystal 2 (second ket). (Note that we must include the process described by the third ket, in which the photons originate in crystal 1, but the idler photon is lost out the unused port of the inter-crystal interferometer. Although this term does not contribute to coincidences, it will be shown to affect the *signal singles* rate.) $A(\omega)$, the probability amplitude, which includes the down-conversion efficiency as well as the pump field strength, is determined ultimately by phase-matching constraints, but is limited in practice by filters and irises before the detectors.

The operator $E_S^{(+)}(t)$ for the positive-frequency part of electric field at the signal detector D_S may be expanded in terms of single-mode photon annihilation operators,

$$E_S^{(+)}(t) = \int_0^{\infty} d\omega_S \, e^{-i\omega_S t} \left(\frac{i}{\sqrt{2}} \, e^{i\omega_S \tau_s} \, \hat{a}_{s_1}(\omega_S) + \frac{1}{\sqrt{2}} \, \hat{a}_{s_2}(\omega_S) \right) , \quad \text{(B2.2)}$$

where we have omitted normalization constants, and assumed beam splitter (before D_S) transmission and reflection coefficients of $1/\sqrt{2}$ and $i/\sqrt{2}$, respectively.

$d_s = d_{s1} - d_{s2} = c\tau_s$ is the additional distance travelled by the signal field originating in crystal 1 relative to that originating in crystal 2. According to the standard Glauber theory for photodetection [Glauber, 1963], the probability per unit time of detecting a count with a unity-efficiency detector, D_S, during a duration T (assumed to be much longer than any other time scale in the problem, but short enough so we can neglect multiple pairs of photons) is proportional to

$$R_S = \frac{1}{T}\int_{-T/2}^{T/2} dt\, P_S(t)\ ,\ \ P_S(t) = \langle \psi \,|\, E_S^{(-)}(t)\, E_S^{(+)}(t) \,|\, \psi \rangle \ . \qquad \text{(B2.3)}$$

Using (B2.1) and (B2.2) in (B2.3) gives

$$R_S = \frac{1}{T}\int_{-T/2}^{T/2} dt \int d\omega'_1\, A^*(\omega'_1)\,\frac{1}{\sqrt{2}}$$

$$\times \Big[s_{1,i}\langle \omega_o+\omega'_1,\, \omega_o-\omega'_1 \,|\, e^{-i(\omega_o-\omega'_1)\tau_i}\left(r^{*2}\, e^{-i(\omega_o-\omega'_1)\tau_A} + t^{*2}\right)$$

$$+ s_{2,i}\langle \omega_o+\omega'_1,\, \omega_o-\omega'_1 \,| + s_{2,i}\langle \omega_o+\omega'_1,\, 0 \,|\, e^{-i(\omega_o-\omega'_1)\tau_i}\left(r^*t^*\, e^{-i(\omega_o-\omega'_1)\tau_A} + r^*t^*\right)\Big]$$

$$\times \int d\omega'_s \int d\omega_s\, e^{i\omega'_s t}\, e^{-i\omega'_s t}\,\frac{1}{2}\left(-i\, e^{-i\omega_s\tau_s}\,\hat{a}^\dagger_{s_1}(\omega'_s) + \hat{a}^\dagger_{s2}(\omega'_s)\right)\left(i\, e^{i\omega_s\tau_s}\,\hat{a}_{s_1}(\omega_s) + \hat{a}_{s2}(\omega_s)\right)$$

$$\times \int d\omega_1 A(\omega_1) \frac{1}{\sqrt{2}} \Big[|\omega_o+\omega_1, \omega_o-\omega_1\rangle_{s_1,i} e^{i(\omega_o-\omega_1)\tau_i} \left(r^2 e^{i(\omega_o-\omega_1)\tau_A} + t^2\right)$$

$$+ |\omega_o+\omega_1, \omega_o-\omega_1\rangle_{s_2,i} + |\omega_o+\omega_1, 0\rangle_{s_1,i} e^{i(\omega_o-\omega_1)\tau_i} \left(r e^{i(\omega_o-\omega_1)\tau_A} + r\right) \Big]. \quad \text{(B2.4)}$$

The canonical commutation relations yield the delta functions $\delta(\omega'_s - (\omega_o + \omega'_1))$ and $\delta(\omega_s - (\omega_o + \omega_1))$, while the integral over t enforces an effective delta function $\delta(\omega'_s - \omega_s)$. Combining these gives the additional constraint, $\delta(\omega'_1 - \omega_1)$, and (B2.4) reduces to a single integral:

$$R_S = K \int d\omega_1 A^*(\omega_1) A(\omega_1) \left\{ \left[e^{-i(\omega_o-\omega_i)\tau_i} \left(r^{*2} e^{-i(\omega_o-\omega_i)\tau_A} + t^{*2}\right)(-i) e^{-i(\omega_o+\omega_1)\tau_s} + 1 \right] \right.$$

$$\times \left[e^{i(\omega_o-\omega_1)\tau_i} \left(r^2 e^{i(\omega_o-\omega_1)\tau_A} + t^2\right) i e^{i(\omega_o+\omega_1)\tau_s} + 1 \right]$$

$$+ \left[e^{-i(\omega_o-\omega_1)\tau_i} r^* t^* \left(e^{-i(\omega_o-\omega_1)\tau_A} + 1 \right)(-i) e^{-i(\omega_o+\omega_1)\tau_s} \right.$$

$$\left. \left. \times e^{i(\omega_o-\omega_1)\tau_i} r \left(e^{i(\omega_o-\omega_1)\tau_A} + 1 \right) i e^{i(\omega_o+\omega_1)\tau_s} \right] \right\}$$

$$= K \int d\omega_1 |A(\omega_1)|^2 \Big[\left(r^{*2} e^{-i(\omega_o-\omega_1)\tau_A} + t^{*2}\right)\left(r^2 e^{i(\omega_o-\omega_1)\tau_A} + t^2\right) + 1$$

$$+ i r^2 e^{i\omega_o(\tau_A+\tau_i+\tau_s)} e^{-i\omega_1(\tau_A+\tau_i-\tau_s)} + i t^2 e^{i\omega_o(\tau_i+\tau_s)} e^{-i\omega_1(\tau_i-\tau_s)}$$

$$- i r^{*2} e^{-i\omega_o(\tau_A+\tau_i+\tau_s)} e^{i\omega_1(\tau_A+\tau_i-\tau_s)} - i t^{*2} e^{-i\omega_o(\tau_i+\tau_s)} e^{i\omega_1(\tau_i-\tau_s)}$$

$$+|r|^2|t|^2\left(1+1+e^{i(\omega_o-\omega_1)\tau_A}+e^{-i(\omega_o-\omega_1)\tau_A}\right)\Big] . \qquad \text{(B2.5)}$$

Here and below, K represents numerical constants, incidental to the calculation. As it stands, (B2.5) is completely general with regards to the various path lengths. We now choose settings which are of physical interest. In particular, we let τ_s and τ_i differ by no more than some incremental amount $\Delta\tau$: $\tau_i = \tau_s + \Delta\tau$. Furthermore, we assume $\tau_A >> 1/\sigma$, where σ is the characteristic bandwidth set by the filter function $A(\omega)$. The result is that terms of the form $e^{i\omega_1\tau_A}$ will vanish after integration over ω_1, and (B2.5) simplifies to

$$R_S = K\int d\omega_1 |A(\omega_1)|^2\left[|r|^4+|t|^4+1+\left(t^2ie^{i\omega_o(2\tau_s+\Delta\tau)}e^{-i\omega_1\Delta\tau}+\text{c.c.}\right)+2|r|^2|t|^2\right] . \text{(B2.6)}$$

If $\Delta\tau$ is much greater than the coherence time of the signal photons, no fringes will be seen after integration. Henceforth, we assume a balanced setup, so that $\tau_i \approx \tau_s$ (i.e., $\Delta\tau << 1/\sigma$), and (B2.6) becomes

$$R_S = \left[1+|r|^4+|t|^4+2|r|^2|t|^2-2|t|^2\sin\left[\omega_o(2\tau_s+\Delta\tau)+2\arg(t)\right]\right]$$

$$\times K\int d\omega_1 |A(\omega_1)|^2 . \qquad \text{(B2.7)}$$

This rate of singles detections at D_S displays fringes with a visibility given by

$$V = \frac{2|t|^2}{1+2|r\,t|^2+|r|^4+|t|^4} = \frac{2|t|^2}{1+\left(|r|^2+|t|^2\right)^2} = |t|^2 , \qquad \text{(B2.8)}$$

where we have used assumed lossless beam splitters, so that $|r|^2 + |t|^2 = 1$. For concreteness, we now assume a real transmission amplitude for the beam splitters in the inter-crystal Mach-Zehnder interferometer. There are several cases of interest:

1) $t = 1, r = 0$ $\qquad R_S \propto 2 - 2\sin\omega_0(2\tau_s + \Delta\tau)$ (B2.9a)

These are the 100%-visibility fringes from the original experimental setup of Mandel, which is expected since we essentially do not have an inter-crystal Mach-Zehnder if the beam splitters transmit everything.

2) $t = 0, r = i$ $\qquad R_S \propto 2.$ (B2.9b)

There are no fringes in this case, because all the photons from crystal 1 take the extra delay line (the long path through the Mach-Zehnder), and we have assumed $\tau_A >> 1/\sigma$. Physically, this implies that one could in principle determine the parent crystal of a given down-converted pair by making precise timing measurements. In other words, the inclusion of the extra delay τ_A serves to distinguish the previously interfering signal photons, yielding *welcher Weg* information by the timing of the idler photons; consequently, no interference is observed.

3) $t = 1/\sqrt{2}$, $r = i/\sqrt{2}$ $\qquad R_S = 2 - \sin\omega_0(2\tau_s + \Delta\tau)$ (B2.9c)

These fringes are limited to 50% visibility due to the background of non-interfering processes in which idler photons from crystal 1 take the long arm of the Mach-Zehnder.

We now examine the effect of adding an unbalanced Mach-Zehnder interferometer (with path-length imbalance $d_{MZ} = c\tau_{MZ}$) before the idler detector D_i. The transmission and reflection amplitudes of its beam splitters are assumed to have the typical values $1/\sqrt{2}$ and $i/\sqrt{2}$, respectively. The idler field at D_i may be expressed as follows:

$$E_i^{(+)}(t) = \int_0^\infty d\omega_i\, e^{-i\omega_i t}\,\frac{1}{\sqrt{2}}\left(1 - e^{i\omega_i\tau_{MZ}}\right)\hat{a}_i(\omega_i)\ . \qquad \text{(B2.10)}$$

Clearly, the presence of this post-crystal Mach-Zehnder will not alter the above calculation of signal singles, since it did not depend on $E_i^{(+)}(t)$. The probability of coincidences, i.e., correlations between detectors D_i and D_s, can show a revival of interference, however. The appropriate fourth-order correlation function is given by $P_C(t_1, t_2) = \langle \psi | E_s^{(-)}(t_1) E_i^{(-)}(t_2) E_i^{(+)}(t_2) E_s^{(+)}(t_1) | \psi \rangle$. In practice, one must include the coincidence time resolution of the detection system, set, for instance, by an electronic gate window of width ΔT. Also, due to the possibility of the extra idler delay τ_A between the crystals, we need to allow for an adjustable timing of the coincidence gate; this could be achieved by an electronic delay line of duration τ_e after detector D_i. The probability of a coincidence detection within a duration T is then

$$R_C = \frac{1}{T}\int_{-T/2}^{T/2} dt_1 \int_{t_1+\tau_e-\Delta T/2}^{t_1+\tau_e+\Delta T/2} dt_2\, P_C(t_1,t_2)$$

$$= \frac{1}{T}\int_{-T/2}^{T/2} dt_1 \int_{t_1+\tau_e-\Delta T/2}^{t_1+\tau_e+\Delta T/2} dt_2 \int d\omega'_1\, A^*(\omega'_1)\frac{1}{\sqrt{2}}$$

$$\times \Big[s_{1,i} \langle \omega_o+\omega'_1,\, \omega_o-\omega'_1 |\; e^{-i(\omega_o-\omega'_1)\tau_i}\left(r^{*2}\, e^{-i(\omega_o-\omega'_1)\tau_A} + t^{*2}\right)$$

$$+ s_{2,i} \langle \omega_o+\omega'_1,\, \omega_o-\omega'_1 | + s_{2,i} \langle \omega_o+\omega'_1,\, 0 |\; e^{-i(\omega_o-\omega'_1)\tau_i}\left(r^*t^*\, e^{-i(\omega_o-\omega'_1)\tau_A} + r^*t^*\right)\Big]$$

$$\times \int d\omega'_s \int d\omega_s\, e^{i\omega'_s t_1}\, e^{-i\omega_s t_1}\, \frac{1}{2}\left(-i\, e^{-i\omega'_s\tau_s}\, \hat{a}^\dagger_{s_1}(\omega'_s) + \hat{a}^\dagger_{s_2}(\omega'_s)\right)\left(i\, e^{i\omega_s\tau_s}\, \hat{a}_{s_1}(\omega_s) + \hat{a}_{s_2}(\omega_s)\right)$$

$$\times \int d\omega'_i \int d\omega_i\, e^{i\omega'_i t_2}\, e^{-i\omega_i t_2}\, \frac{1}{2}\left(1 - e^{-i\omega'_i\tau_{MZ}}\right)\hat{a}^\dagger_i(\omega'_i)\left(1 - e^{i\omega_i\tau_{MZ}}\right)\hat{a}_i(\omega_i)$$

$$\times \int d\omega_1 A(\omega_1) \frac{1}{\sqrt{2}} \Big[|\omega_o+\omega_1, \omega_o-\omega_1\rangle_{s_1,i} e^{i(\omega_o-\omega_1)\tau_i} \left(r^2 e^{i(\omega_o-\omega_1)\tau_A} + t^2\right)$$

$$+ |\omega_o+\omega_1, \omega_o-\omega_1\rangle_{s_2,i} + |\omega_o+\omega_1, 0\rangle_{s_1,i} e^{i(\omega_o-\omega_1)\tau_i} \left(rt\, e^{i(\omega_o-\omega_1)\tau_A} + rt\right) \Big] . \quad \text{(B2.11)}$$

Because $\hat{a}_i(\omega_i)\ |\omega_o+\omega_1, 0\rangle_{s,i} = 0$, these states do not contribute, unlike in the previous singles calculation, where they provided a background; this is physically intuitively obvious -- these terms do not contribute to coincidence counts. Furthermore, bra-ket orthogonality relations ($\omega_s \Rightarrow \omega_o+\omega_1$, $\omega'_i \Rightarrow \omega_o-\omega'_1$, $\omega_s \Rightarrow \omega_o+\omega_1$, and $\omega_i \Rightarrow \omega_o-\omega_1$) can be used to remove four of the integrals:

$$= \frac{K}{T} \int_{-T/2}^{T/2} dt_1 \int_{t_1+\tau_e-\Delta T/2}^{t_1+\tau_e+\Delta T/2} dt_2 \int d\omega'_1 \int d\omega_1$$

$$\times A^*(\omega'_1)\, A(\omega_1)\, e^{i(\omega_o+\omega'_1)t_1}\, e^{-i(\omega_o+\omega_1)t_1}\, e^{i(\omega_o-\omega'_1)t_2}\, e^{-i(\omega_o-\omega_1)t_2}$$

$$\times \left[-i\, e^{-i(\omega_o-\omega'_1)\tau_i} \left(r^{*2} e^{-i(\omega_o-\omega'_1)\tau_A} + t^{*2}\right) e^{-i(\omega_o+\omega'_1)\tau_s} + 1\right]$$

$$\times \left[i\, e^{i(\omega_o-\omega_1)\tau_i} \left(r2 e^{i(\omega_o-\omega_1)\tau_A} + t2\right) e^{i(\omega_o+\omega_1)\tau_s} + 1\right]$$

$$\times \left(1 - e^{-i(\omega_o - \omega'_1)\tau_{MZ}}\right) \left(1 - e^{i(\omega_o - \omega_1)\tau_{MZ}}\right) . \quad \text{(B2.12)}$$

Next we evaluate the time integrals, first the integral over t_2:

$$\int_{t_1+\tau_e-\Delta T/2}^{t_1+\tau_e+\Delta T/2} dt_2\ e^{i(\omega_o+\omega'_1)t_1}\, e^{-i(\omega_o+\omega_1)t_1}\, e^{i(\omega_o-\omega'_1)t_2}\, e^{-i(\omega_o-\omega_1)t_2}$$

$$= e^{i(\omega'_1-\omega_1)t_1} \int_{t_1+\tau_e-\Delta T/2}^{t_1+\tau_e+\Delta T/2} dt_2\ e^{-i(\omega'_1-\omega_1)t_2}$$

$$= \frac{\sin\left((\omega_1 - \omega'_1)\Delta T/2\right)}{(\omega_1 - \omega'_1)/2} e^{i(\omega_1-\omega'_1)\tau_e} . \qquad \text{(B2.13)}$$

Since the above expression no longer depends on t_1, integration over t_1 simply gives T. From (B2.13) we are led to make a change of variables to $\delta\omega = \omega'_i - \omega_i$ and $\Omega = \omega'_1 + \omega_1$. Also, we again make the substitution $\tau_i = \tau_s + \Delta\tau$ (i.e., a balanced initial setup). Finally, for concreteness we choose a definite form for our filter function:

$$A^*(\omega_1) = A(\omega_1) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\omega_1)^2}{2\sigma^2}} . \qquad \text{(B2.14)}$$

With the above variable substitution, this yields

$$A^*(\omega'_1)\, A(\omega_1) = \frac{1}{2\pi\sigma} e^{-\left(\frac{\Omega}{2\sigma}\right)^2} e^{-\left(\frac{\delta\omega}{2\sigma}\right)^2} . \qquad \text{(B2.15)}$$

Modulo overall constants, which we subsume into K, the coincidence probability is then (after considerable algebra)

$$R_C = K \int d(\delta\omega) \int d\Omega\, e^{-\frac{\Omega^2 + \delta\omega^2}{(2\sigma)^2}} \frac{\sin(\delta\omega\,\Delta T/2)}{\delta\omega} e^{-i\delta\omega\tau_e}$$

$$\times \left\{ e^{i\delta\omega\Delta\tau}\left[|r|^4 e^{i\delta\omega\tau_A} + |t|^4 + r^{*2}t^2 e^{-i\omega_o\tau_A} e^{i(\delta\omega+\Omega)\tau_A/2} + r^2 t^{*2} e^{i\omega_o\tau_A} e^{i(\delta\omega-\Omega)\tau_A/2}\right]\right.$$

$$+ 1 + ie^{i\omega_o(2\tau_s+\Delta\tau)}\, e^{i(\delta\omega-\Omega)\Delta\tau/2}\left(r^2 e^{i\omega_o\tau_A}\, e^{i(\delta\omega-\Omega)\tau_A/2} + t^2\right)$$

$$\left. - ie^{-i\omega_o(2\tau_s+\Delta\tau)}\, e^{-i(\delta\omega+\Omega)\Delta\tau/2}\left(r^{*2} e^{-i\omega_o\tau_A}\, e^{i(\delta\omega+\Omega)\tau_A/2} + t^{*2}\right)\right\}$$

$$\times \left[1 + e^{i\delta\omega\tau_{MZ}} - e^{-i\omega_o\tau_{MZ}} e^{i(\Omega+\delta\omega)\tau_{MZ}/2} - e^{i\omega_o\tau_{MZ}} e^{-i(\Omega-\delta\omega)\tau_{MZ}/2}\right]. \quad \text{(B2.16)}$$

After integrating over Ω using the general result

$$\int_{-\infty}^{\infty} d\Omega\, e^{-(\Omega/2\sigma)^2 + iX\Omega} = 2\sigma\sqrt{\pi}\, e^{-(\sigma X)^2}\ , \qquad \text{(B2.17)}$$

many of the terms (22 out of 36) are proportional to exp[-($\sigma\,\tau_A$)^2] or exp[-($\sigma\,\tau_{MZ}$)^2], and are thus negligible. (Again we assume here the case of interest that $\tau_A >> 1/\sigma$, so that there are no signal singles fringes, and also $\tau_{MZ} >> 1/\sigma$.) Consequently, (B2.16) reduces to

$$R_C = K\int d(\delta\omega)\ e^{-\frac{\delta\omega^2}{(2\sigma)^2}} \frac{\sin(\delta\omega\,\Delta T/2)}{\delta\omega} e^{-i\delta\omega\tau_e}$$

$$\times \left[e^{i\delta\omega\,(\Delta\tau+\tau_A)}|r|^4 + e^{i\delta\omega\Delta\tau}|t|^4 + 1 + e^{-(\sigma\Delta\tau/2)^2}\left(it^2 e^{i\omega_o(2\tau_s+\Delta\tau)}\, e^{i\delta\omega\Delta\tau/2} + \text{c.c.}\right)\right.$$

$$\left. + e^{i\delta\omega\,(\Delta\tau+\tau_A+\tau_{MZ})}|r|^4 + e^{i\delta\omega(\Delta\tau+\tau_{MZ})}|t|^4\right.$$

$$+ e^{i\delta\omega\tau_{MZ}} + e^{-(\sigma\Delta\tau/2)^2} e^{i\delta\omega\tau_{MZ}} \left(i t^2 e^{i\omega_o(2\tau_s+\Delta\tau)} e^{i\delta\omega\Delta\tau/2} + c.c. \right)$$

$$- e^{-(\sigma(\tau_A-\tau_{MZ})/2)^2} e^{i\delta\omega(\tau_A+\tau_{MZ})/2} \left(r2t^{*2} e^{i\omega_o(\tau_A-\tau_{MZ})} + c.c. \right)$$

$$- i r^2 e^{-(\sigma(\Delta\tau+\tau_A-\tau_{MZ})/2)^2} e^{i\omega_o(2\tau_s+\Delta\tau+\tau_A-\tau_{MZ})} e^{i\delta\omega(\Delta\tau+\tau_A+\tau_{MZ})/2}$$

$$+ i r^{*2} e^{-(\sigma(-\Delta\tau+\tau_A-\tau_{MZ})/2)^2} e^{-i\omega_o(2\tau_s+\Delta\tau+\tau_A-\tau_{MZ})} e^{-i\delta\omega(\Delta\tau-\tau_A-\tau_{MZ})/2} \Big] \quad \text{(B2.18)}$$

$$= K \int d(\delta\omega)\, e^{-\frac{\delta\omega^2}{(2\sigma)^2}} \frac{\sin(\delta\omega\,\Delta T/2)}{\delta\omega}$$

$$\times \Big\{ e^{i\delta\omega(\tau_A-\tau_e)} |r|^4 + e^{-i\delta\omega\tau_e} \left(|t|^4 + 1 - 2|t|^2 \sin\left[\omega_o(2\tau_s+\Delta\tau)+2\arg(t)\right] \right)$$

$$+ e^{i\delta\omega(\tau_A+\tau_{MZ}-\tau_e)} |r|^4 + e^{i\delta\omega(\tau_{MZ}-\tau_e)} \left(|t|^4 + 1 - 2|t|^2 \sin\left[\omega_o(2\tau_s+\Delta\tau)+2\arg(t)\right] \right)$$

$$- e^{-(\sigma(\tau_A-\tau_{MZ})/2)^2} e^{i\delta\omega(\tau_A+\tau_{MZ}-2\tau_e)/2} |r|^2 |t|^2 \, 2\cos\left[\omega_o(\tau_A-\tau_{MZ})+2\arg(r)-2\arg(t)\right]$$

$$+ e^{-(\sigma(\tau_A-\tau_{MZ})/2)^2} e^{i\delta\omega(\tau_A+\tau_{MZ}-2\tau_e)/2} |r|^2 \, 2\sin\left[\omega_o(2\tau_s+\Delta\tau+\tau_A-\tau_{MZ})+2\arg(r)\right] \Big\} , \quad \text{(B2.19)}$$

where in going from (B2.18) to (B2.19) we have once again assumed $\Delta\tau << 1/\sigma$, so that $e^{\pm i\delta\omega\Delta\tau} \approx 1$.

There are two regimes to consider, depending on the value of $|\tau_A - \tau_{MZ}|$; each has two main cases, depending on the time resolution ΔT. First we consider the situation in which the Mach-Zehnders' path imbalances are quite different; quantitatively, $|\tau_A - \tau_{MZ}| >> 1/\sigma$. Then the last two terms in (B2.19) are negligible. If the resolution

time is greater than any of the path lengths ($\Delta T >> \tau_A, \tau_{MZ}$), the sinc-function behaves like a delta function, and we are left with

$$R_C \propto 2 + 2|r|^4 + 2|t|^4 - 4|t|^2 \sin\left[\omega_0(2\tau_s + \Delta\tau + 2\arg(t)\right] , \qquad \text{(B2.20)}$$

with a visibility of $V = \dfrac{2|t|^2}{1+|r|^4+|t|^4}$. A plot of this visibility is shown in Fig. B2.2, along with a plot of the singles visibility (B2.8). As with the previous calculation of the singles rate at D_S, we can consider three special cases for beam splitters comprising the inter-crystal Mach-Zehnder:

1) $t = 1, r = 0$ $\qquad R_C \propto 4 - 4\sin\omega_0[2\tau_s + \Delta\tau]$ (B2.21a)

The 100%-visibility coincidence fringes are expected, since in this case the *singles* rate at D_S also displayed 100%-visibility fringes.

2) $t = 0, r = i$ $\qquad R_C \propto 4$ (B2.21b)

In this case, all of the idlers from crystal 1 have an extra delay, and no interference is observed. Again, this is expected since the timing of the idlers in principle now yields *welcher Weg* information (even though our coincidence time resolution cannot distinguish the contributing processes), and our second Mach-Zehnder (after crystal 2) is not adjusted to erase this information (recall that in this regime $|\tau_A - \tau_{MZ}| >> 1/\sigma$).

3) $t = 1/\sqrt{2}$, $r = i/\sqrt{2}$ $\qquad R_C \propto 3 - 2\sin\omega_0[2\tau_s + \Delta\tau]$ (B2.21c)

These 67%-visibility fringes are a slight improvement over the 50%-visibility fringes observed in the singles rate at D_S alone [cf. Eq. (B2.9c)]. The improvement arises because half of the idler photons which took the long path in the inter-crystal Mach-Zehnder (and hence have *in principle* distinguishable timing) exit the port of that interferometer which *doesn't* lead to crystal 2; hence they are not counted in a coincidence measurement between D_S and D_i.

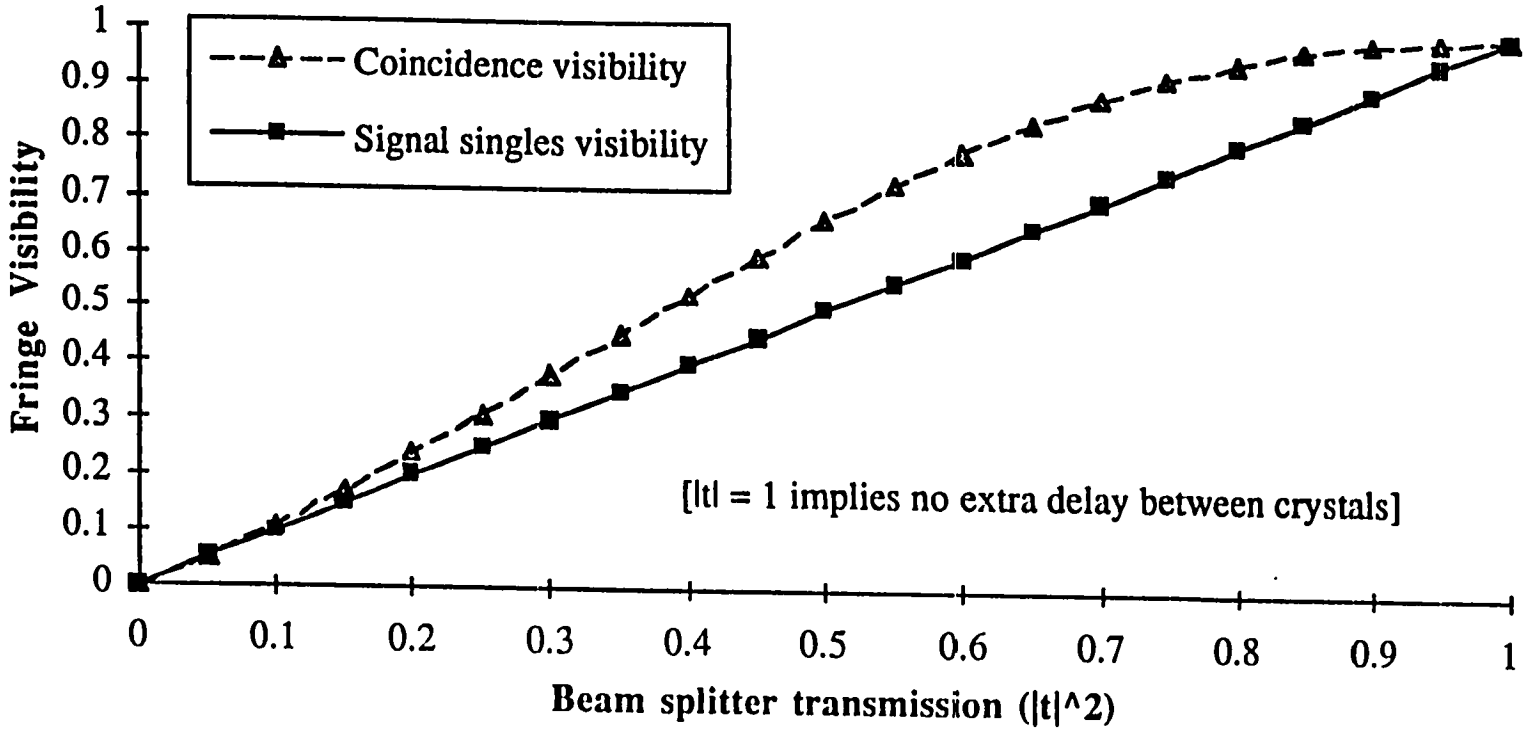


Figure B2.2 Plots of the fringe visibility for the signal singles rate and the coincidence rate, versus the transmissivity of the beam splitters in the inter-crystal Mach-Zehnder interferometer (see Fig. B2.1). The system path lengths are such as to allow maximum visibility when all the crystal 1 idler photons take the short path toward crystal 2 (i.e., when $|t|=1$).

Still in the regime $|\tau_A - \tau_{MZ}| >> 1/\sigma$, if the coincidence resolution time ΔT is less than τ_A and τ_{MZ} (but still satisfying $\Delta T >> 1/\sigma$), then integration over $\delta\omega$ will wash out terms with factors of the form $e^{i\delta\omega\tau}$, unless $\tau << 1/\sigma$ (i.e., unless the time is much less than the coherence time of the photons). We see from (B2.19) that the results will depend on the specific value τ_e of the electronic time delay, and on the values of the beam splitter reflection and transmission amplitudes:

1) $t = 1, r = 0$ $\qquad R_C \propto 2 - 2\sin\omega_0[2\tau_s + \Delta\tau]$ (B2.22a)

There are two timing arrangements which will yield the 100%-visibility coincidences indicated above. Specifically, we may choose $\tau_e = 0$ (coincidences between signal photons and idlers which travel the short arm of the post-crystal Mach-Zehnder) or $\tau_e = \tau_{MZ}$ (coincidences between signal photons and idlers which travel the long arm of the post-crystal Mach-Zehnder). [Note that we have indicated that the relative *value* of the rate is one half that given by (B2.21a); the longer coincidence window in that case included *both* of the timing arrangements which work here.] Other values of τ_e will yield no fringes.

2) $t = 0, r = i$ $\qquad R_C \propto 1$ (B2.22b)

In this case, for no value of τ_e is interference observed. The situation is essentially the same as that leading to (B2.21b), except that here our coincidence resolution is sufficient to determine *in practice* the parent crystal of the photons.

3) $t = 1/\sqrt{2}, r = i/\sqrt{2}$ $\qquad R_C \propto \frac{5}{4} - \sin\omega_0[2\tau_s + \Delta\tau]$ (B2.22c)

Depending on the choice of electronic delay, we can keep at most one of the first four terms in the curly bracket of (B2.19). A maximum visibility of 80% is obtained for $\tau_e = 0$ or $\tau_e = \tau_{MZ}$. For these cases, the use of the post-crystal Mach-Zehnder partially erases the distinguishability which limited the singles visibility to 50% (see (B2.9c)).

Henceforth, we restrict ourselves to the second regime: $\tau_A = \tau_{MZ} + \Delta\tau_{MZ}$, with

$\Delta\tau_{MZ} << 1/\sigma$. As before, a large value of ΔT (greater than τ_A) causes an effective delta-function, $\delta(\delta\omega)$. Under these conditions, (B2.19) simplifies to the following:

$$R_C \propto 2 + 2|r|^4 + 2|t|^4 - 4|t|^2 \sin[\omega_o(2\tau_s+\Delta\tau)+2\arg(t)]$$

$$- 2|rt|^2 \cos[\omega_o \Delta\tau_{MZ}+2\arg(r)-2\arg(t)]$$

$$+ 2|r|^2 \sin[\omega_o(2\tau_s+\Delta\tau+\Delta\tau_{MZ})+2\arg(r)] \quad . \qquad \text{(B2.23)}$$

Again we consider the three particular choices for r and t:

1) $t = 1, r = 0$ $\qquad R_C \propto 4 - 4\sin\omega_o[2\tau_s+\Delta\tau]$ (B2.24a)

2) $t = 0, r = i$ $\qquad R_C \propto 4 - 2\sin\omega_o[2\tau_s+\Delta\tau+\Delta\tau_{MZ}]$ (B2.24b)

The use of the post-crystal Mach-Zehnder has erased the distiguishability afforded by the extra delay experienced by the idler photons originating in crystal 1; however, due to insufficient time resolution to exclude the two non-interfering processes (in which an idler from crystal 1 takes the long path in the Mach-Zehnder, or an idler from crystal 2 takes the short path), the visibility is limited to 50%.

3) $t = 1/\sqrt{2}$, $r = i/\sqrt{2}$

$$R_C \propto 3 - 2\sin\omega_o[2\tau_s+\Delta\tau] + \frac{1}{2}\cos\omega_o[\Delta\tau_{MZ}] - \sin\omega_o[2\tau_s+\Delta\tau+\Delta\tau_{MZ}] \qquad \text{(B2.24c)}$$

This has a maximum visibility of 85.7% (6/7), achieved for $\Delta\tau_{MZ} = 0$ (i.e., $\tau_A = \tau_{MZ}$). For the case $\Delta\tau_{MZ}\omega_o = \pi/2$, the visibility is less: 74.4% ($= [1+2\sqrt{3}]/6$).

Finally, we consider the case where the post-crystal Mach-Zehnder is properly set to provide maximum erasure (i.e., $\tau_A = \tau_{MZ} + \Delta\tau_{MZ}$, with $\Delta\tau_{MZ} << 1/\sigma$), and the coincidence timing is sufficient to resolve the distances τ_A and τ_{MZ} (i.e. $\Delta T < .\tau_A$). As

before, the results will now depend on the specific value τ_e of the electronic time delay, and on the values of the beam splitter reflection and transmission amplitudes:

1) $t = 1, r = 0$ $\qquad R_C \propto 2 - 2\sin \omega_o[2\tau_s + \Delta\tau]$ (B2.25a)

Exactly as under the conditions leading to (B2.22a), there are two timing arrangements which will yield the 100%-visibility coincidences indicated above. Specifically, we may choose $\tau_e = 0$ (coincidences between signal photons and idlers which travel the short arm of the post-crystal Mach-Zehnder) or $\tau_e = \tau_{MZ}$ (coincidences between signal photons and idlers which travel the long arm of the post-crystal Mach-Zehnder). Other values of τ_e will yield no fringes.

2) $t = 0, r = i$

$\tau_e = 0, \tau_A + \tau_{MZ}$ $\qquad R_C \propto 1$ (B2.25b)

$\tau_e = \tau_{MZ}$ $\qquad R_C \propto 2 - 2\sin \omega_o[2\tau_s + \Delta\tau + \Delta\tau_{MZ}]$ (B2.25b')

Our system is now acting as a perfect quantum eraser (when $\tau_e = \tau_{MZ}$), as discussed earlier. Although no interference fringes are seen in singles under these conditions (see Eq. (B2.9b)), we have a complete revival by correlating with the idler detector. Moreover, by correctly choosing $\Delta\tau_{MZ}$, we can choose to observe fringes or anti-fringes. If this choice were made using a high-speed phase-shifter, then one would have a delayed choice version. We see explicitly that the high time resolution allows us to discard the background of non-interfering counts which would have otherwise limited the visibility to 50% (see Eq. (B2.24b). Physically, the background terms which appear in (B2.24b) (but not in (B2.25b)) are due to contributions from the following processes: 1) photons originate in crystal 1, but the idler takes the long path in the Mach-Zehnder; and 2) photons originate in crystal 2, but the idler takes the short path in the Mach-Zehnder. With a sufficiently short time resolution, these non-interfering contributions may be removed, yielding complete erasure.

3) $t = 1/\sqrt{2}$, $r = i/\sqrt{2}$

$\tau_e = 0, \tau_A+\tau_{MZ}$ $\quad R_C \propto \frac{1}{4}$ (B2.25c)

$\tau_e = 0$: $\quad R_C \propto \frac{5}{4} - \sin \omega_o[2\tau_s+\Delta\tau]$ (B2.25c')

$\tau_e = \tau_{MZ}$:

$$R_C \propto \frac{3}{2} - \sin \omega_o[2\tau_s+\Delta\tau]+ \frac{1}{2}\cos \omega_o[\Delta\tau_{MZ}] - \sin \omega_o[2\tau_s+\Delta\tau+\Delta\tau_{MZ}] \quad \text{(B2.25c'')}$$

One can immediately check that including all of these possible timing arrangements returns the value given by (B2.24c) for a poorer timing resolution system. For the case $\Delta\tau_{MZ}\omega_o = \pi/2$, (B2.24c'') reduces to $R_C \propto \frac{3}{2} - \sin \omega_o[2\tau_s+\Delta\tau] - \cos \omega_o[\tau_s+\Delta\tau]$, with a visibility of 94% ($= 2\sqrt{2}/3$). But for $\Delta\tau_{MZ} = 0$ (i.e., $\tau_A = \tau_{MZ}$), (B2.25c'') recovers 100%-visibility fringes. One physical interpretation is as follows: The inter-crystal Mach-Zehnder acts as a variable frequency filter. If the "analyzing" post-crystal Mach-Zehnder is set the same way (i.e., $\tau_A = \tau_{MZ}$), then it will pass all of the light which made it through the inter-crystal filter (i.e., half of the light from crystal 1), but only half of the light from crystal 2. In other words, after the second Mach-Zehnder, the amplitude of light from the two crystals is the same, and has the same spectrum. Therefore, the indistinguishability of the parent crystal is recovered, and coincidence fringe visibility is 100%.

Appendix C: Error-Analysis for Absolute Efficiency Measurements

C.I. Introduction

Calculation of the uncertainty in the efficiency is not entirely trivial. Below we present two approaches to the calculation. Both eventually yield the same result, but the underlying reasoning is very different. The first, fairly formal, calculation (Sect. C.II) relies on introducing the number of "anti-coincidences" (those cases where *only* the trigger detectors produces a count), and treating the number of coincidences and the number of anit-coincidences as *independent* random variables, governed by Poisson statistics. This only makes sense if one considers the total number of trigger-detector counts as a random variable.

In the second approach (Sect. C.III), the total number of trigger-detector counts is treated as a *nonrandom, but nonfixed* quantity, i.e., in any given run a definite number of trigger counts are seen. Although this number will vary from run to run, the number in any given run has *no* uncertainty. The probability of detecting the conjugate photon is then given by a *binomial* distribution.

C.II. Purely Poisson Approach

Each experimental run to measure the efficiency coefficient consisted of two measurement periods:

1. The "efficiency" period, of duration T, where the coincidences C and the singles counts S_c and S_t (equal to the number of coincidences C plus the number of "anti-coincidences" α) are counted.

2. The "background" period, of duration T_{BG}, where the source has been blocked and the trigger background counts BG_t are counted. For convenience we define

$r_{BG} \equiv T/T_{BG}$.

The accidental counts can be calculated from the measured quantities. There are four types of processes which can yield an accidental coincidence, depending on whether the trigger count was caused by a real photon (i.e., one with a conjugate partner that could have been detected) or background, and whether the conjugate detector count was caused by a real photon or background. In cases where a real photon is detected, its conjugate must *fail* to be detected (otherwise this would count as a true coincidence event), so that the number of accidental counts is given by

$$A = \left[N_t(1-\eta_c) + BG_t r_{BG}\right]\left[N_c(1-\eta_t) + BG_c r_{BG}\right]\frac{w}{T}$$

$$= \left[S_t - \eta_c N_t\right]\left[S_c - \eta_t N_c\right]\frac{w}{T}\,, \qquad \text{(C.1a)}$$

where w is the duration of the gate window, and we have used $S_t = N_t + BG_t r_{BG}$ and $S_c = N_c + BG_c r_{BG}$. To write (C.1a) in terms of experimentally-measured quantities, we use the fact that $S_c >> \eta_t N_c$ (the effective efficiency of the trigger detector is very small, due to the narrow-bandwidth filter and small iris before it), and again use $S_t = N_t + BG_t r_{BG}$:

$$A = \left[S_c S_t - \eta_c S_c\left(S_t - BG_t r_{BG}\right)\right]\frac{w}{T}\,. \qquad \text{(C.1b)}$$

Direct measurements of the accidental rate (using an extra 100-ns electronic delay to prevent any true coincidences) were in agreement with the values obtained from (C.1b).

The efficiency[1] is given by the formula

[1] As discussed in Sect. 9.VIII, the efficiency actually depends on the incident light intensity, due to saturation effects (arising from intrinsic deadtime in the devices). However, for simplicity we use η to denote the efficiency in the low-light limit.

$$\eta_c = \frac{C - A}{S_t - r_{BG}BG_t} \quad . \tag{C.2}$$

We consider the coincidences and anti-coincidences during the efficiency measurement, as well as the background events to be independent Poisson processes. In practice the uncertainty in the accidental counts was completely neglible compared with the other uncertainties. Thus, we derive the statistical error on

$$\eta_c = \frac{C - A}{C + \alpha - r_{BG}BG_t} \quad , \tag{C.3}$$

where $\alpha = S_t - C = ((1 - \eta_c)C - A)/\eta_c + r_{BG}BG_t$ results from an independent Poisson process:

$$(\Delta\eta_C)^2 = \left(\frac{\partial\eta_C}{\partial C}\right)^2(\Delta C)^2 + \left(\frac{\partial\eta_C}{\partial\alpha}\right)^2(\Delta\alpha)^2 + \left(\frac{\partial\eta_C}{\partial BG_t}\right)^2(\Delta BG_t)^2$$

$$= \left[\frac{C + \alpha - r_{BG}BG_t - C + A}{(C + \alpha - r_{BG}BG_t)^2}\right]^2 C + \left[\frac{C - A}{(C + \alpha - r_{BG}BG_t)^2}\right]^2 \alpha$$

$$+ \left[\frac{r_{BG}(C - A)}{(C + \alpha - r_{BG}BG_t)^2}\right]^2 BG_t \quad . \tag{C.4}$$

$$\left(\frac{\Delta\eta_c}{\eta_c}\right)^2 = \frac{1}{(C - A)^2}\left[(1 - \eta_c)^2 C + \eta_c^2\alpha + r_{BG}^2\eta_c^2 BG_t\right]$$

$$= \frac{1}{(C - A)^2}\left[(1 - \eta_c)C + r_{BG}(1 + r_{BG})\eta_c^2 BG_t\right] \quad . \tag{C.5}$$

$$\frac{\Delta\eta_c}{\eta_c} = \sqrt{\frac{1}{(C - A)^2}\left[(1 - \eta_c)C + r_{BG}(1 + r_{BG})\eta_c^2 BG_t\right]} \quad . \tag{C.6}$$

In practice the time intervals over which the trigger singles and background counts were collected were similar ($r_{BG} \approx 1$), so that Eq. (C.6) simplifies:

$$\frac{\Delta\eta_c}{\eta_c} = \sqrt{\frac{(1-\eta_c)C}{(C-A)^2} + \frac{2BG_t}{(S_t - BG_t)^2}} \quad . \tag{C.7}$$

C.III. Binomial Approach

The values recorded in a given experimental run were the means and uncertainties of the trigger singles rate (S_t, ΔS_t) and the coincidence rate (C, ΔC), measured simultaneously (for most runs) over some time interval T; the mean and uncertainty of the trigger detector background rate (BG_t, ΔBG_t), measured over a time interval T_{BG}; and the conjugate detector singles rate (for use in calculating the expected accidental rate). Using the fact that all coincidence counts are *predicated* on detecting a trigger count first, and that there is *no* uncertainty in the number of counts detected at the trigger detector during the interval T (another interval of equal duration might give a different result, but this does not matter), we conclude that ΔS never matters. However, the uncertainty in the coincidence rate needs to be modified, because the uncertainties in the trigger singles rate and the coincidence rate are not independent since they were collected simultaneously. This is obviously true if the uncorrected detection efficiency of the conjugate detector is 100%, for then any fluctuations in the trigger singles rate show up in the coincidence rate as well. As was shown by Rarity *et al.* [Rarity, 1987], in calculating the uncertainty in the efficiency, one should use the modified relative standard deviation for the coincidence rate (and hence the numerator of Eq. (C.2):

$$\sigma_{C-A} = \frac{\Delta C \sqrt{1-\eta}}{C-A} \quad . \tag{C.8}$$

We have not included the uncertainty in the accidental rate because $\frac{\Delta A}{C - A}$ is negligible compared to the other uncertainties. With the uncorrected detection efficiencies of ~50% seen in these measurements, this indicates a reduction factor of 0.7.

There is also an uncertainty in the background rate which must be subtracted from S, and thus an uncertainty in the denominator of Eq. (C.2):

$$\Delta(S_t - BG_t) = \sqrt{\sqrt{\frac{BG_t}{T}^2} + \sqrt{\frac{BG_t}{T_{BG}}^2}} \quad . \tag{C.9}$$

There are two terms because the background and trigger singles rate were not measured simultaneously (*i.e.*, there was an uncertainty in the mean value of the background that was measured, plus there was an uncertainty whether this number of background counts actually arose in the interval over which the trigger singles were measured). Since in practice the time intervals over which the trigger singles and background counts were collected were similar ($T_{BG} \approx T$),

$$\sigma_{(S_t - BG_t)} = \frac{\sqrt{2}\sqrt{\frac{BG_t}{T}}}{S_t - BG_t} = \sqrt{2}\frac{\Delta BG_t}{S_t - BG_t} \, . \tag{C.10}$$

The compounded uncertainties in efficiency can now be computed by adding the relative errors (C.8) and (C.10) in quadrature:

$$\frac{\Delta\eta_c}{\eta_c} = \sqrt{\left(\frac{\Delta C}{C - A}\right)^2 (1 - \eta_c) + 2\left(\frac{\Delta BG_t}{S_t - BG_t}\right)^2} \quad . \tag{C.11}$$

Using the results $\Delta C = \sqrt{C}$ and $\Delta BG_t = \sqrt{BG_t}$, this is identical to (C.7).

Appendix D1: BBO

We have investigated a rather large number of nonlinear crystals. Subject to the constraints that it must be possible to satisfy Type-II phase-matching at pump wavelengths of 325 nm or 316 nm, that the conversion efficiency must be appreciable (i.e., comparable to KDP), and that there must not be significant loss at the down-converted wavelengths (650 nm or 632 nm), beta-barium borate (BBO; β–BaB_2O_4) was found to be the best candidate. The crystals investigated were KDP, KD^*P, ADP, AD^*P, RDP, RD^*P, CDA, CD^*A, $LiIO_3$, $LiNbO_3$, $K_3Li_2Nb_5O_{15}$, LBO, BBO, and BANANA (barium sodium niobate). Herein we list the relevant parameters for BBO, which may be found in a suitable combination of [Dmitriev et al., 1991; Chen, 1989; Eimerl et al., 1987].

BBO is a negative uniaxial crystal. The indices of refraction are given by

$$n_o^2 = 2.7405 + \frac{0.0184}{\lambda^2 - 0.0179} - 0.0155\,\lambda^2 .$$

$$n_e^2 = 2.3730 + \frac{0.0128}{\lambda^2 - 0.0156} - 0.0044\,\lambda^2 .$$

The thermal coefficients over the range 0.4 - 1.0 μm are $\frac{\partial n_o}{\partial T} = -1.66 \times 10^{-5}\ (K)^{-1}$ and $\frac{\partial n_e}{\partial T} = -0.93 \times 10^{-5}\ (K)^{-1}$. For type I phase-matching, the effective nonlinearity is $d_{eff} \approx d_{31} \sin\theta - d_{11} \cos\theta \cos 3\phi$, while for type II phase-matching, it is $d_{eff} \approx d_{11} \cos^2\theta \sin 3\phi$. Here θ and ϕ are polar coordinates referring to z (= c) and x (=a), respectively. The actual values of the coefficients, both in units of the KDP-

nonlinearity and in "real" units, are as follows: $d_{11} = 4.1\ d_{36}(\text{KDP}) = 1.8 \times 10^{-12}$ m/V and $d_{31} < 0.3\ d_{36}(\text{KDP}) = 0.13 \times 10^{-12}$ m/V.

In the region of transparency, optical absorption is primarily due to trace impurities; therefore, the attenuation of a given sample may vary considerably depending on the crystal quality and crystal growth process. The transmission of a 4-mm sample (at an angle of 21° with respect to the optic axis, the type I phase-matching angle [for an unstated sum-frequency process]) was ~87.5% over the range 0.35 - 2.0 μm [Eimerl et al., 1987]. Assuming Fresnel losses of 5% at each interface, this implies an attenuation factor of less than 0.97/4mm -> 0.36dB/cm.

Calculation of the phase-matching angles proceeds similarly to the calculation outlined in Chap. 2 for KDP. There is a fairly nasty added complication, however. As we are interested in type-II phase-matching, (and because BBO is a negative uniaxial crystal), both the pump and one of the down-converted photons (henceforth assumed to be the idler) will be extraordinary-polarized. This has two related effects: First, the refractive index of the idler beam will depend on the phase-matching angle; second, it will depend in particular on both the angle with respect to the pump beam *and* on the azimuthal angle about the pump beam. An important consequence is that the distribution of down-converted photons of a particular color will not be symmetric about the pump beam, in contrast to the situation with type-I phase-matching, where a cone centered about the pump beam was emitted (see Chap. 2). This will become clear presently.

From the schematic in Fig. D1 we can immediately write down the longitudinal and transverse momentum conservation relations:

$$\kappa_p = \kappa_s \cos\theta_s + \kappa_i \cos\theta_i\ , \qquad \text{(D1.1a)}$$

and

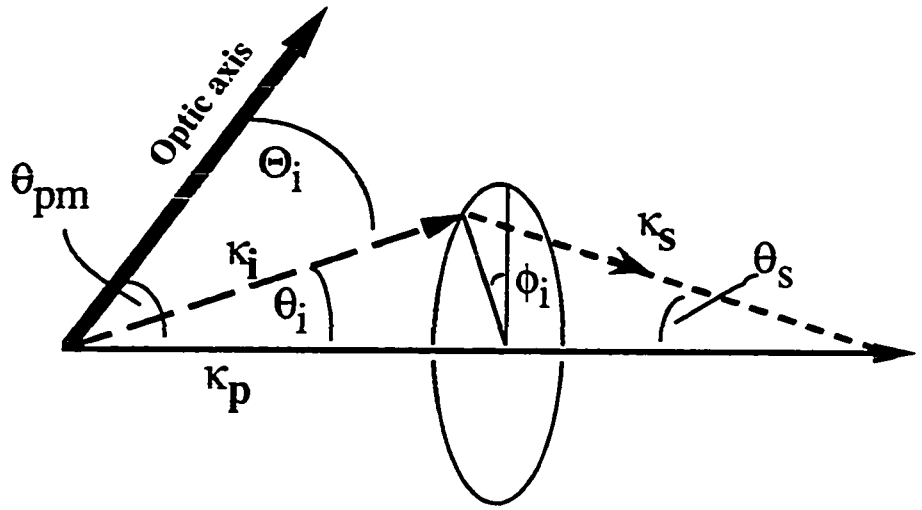


Figure D1. Schematic representation of type-II vector phase-matching (angles are exaggerated for clarity).

$$\kappa_s \sin\theta_s = -\kappa_i \sin\theta_i \ , \qquad \text{(D1.1b)}$$

where the κ's are the momenta *inside* the crystal, given by $\kappa_p = \omega_p\, n_e(\omega_p, \theta_{pm})$, $\kappa_s = \omega_s\, n_o(\omega_s)$, and $\kappa_i = \omega_i\, n_e(\omega_i, \Theta_i)$. Here θ_{pm} is the angle between the pump and the optic axis and Θ_i is the angle between the idler beam and the optic axis. If we define ϕ_i to be the azimuthal angle of κ_i, referenced to the optic axis (so that $\phi_i = 0$ means that κ_i lies between the optic axis and κ_p, and $\phi_i = \pi$ means that κ_p lies between the optic axis and κ_i), then we can express Θ_i in terms of θ_{pm}, θ_i, and ϕ_i:

$$\cos\Theta_i = \cos\theta_p \cos\theta_i + \sin\theta_p \sin\theta_i \cos\phi_i \ . \qquad \text{(D1.2)}$$

Squaring (D1.1b) and using the trigonometric identity, $\sin^2\theta = 1 - \cos^2\theta$, we have

$$\kappa_s^2 \cos^2\theta_s = \kappa_{i,\Theta_i}^2 \left(\cos^2\theta_i - 1\right) + \kappa_s^2 \,, \qquad \text{(D1.3)}$$

where the second subscript on κ_i stresses the angular dependence. Next we move $\kappa_i \cos\theta_i$ to the left-hand side of (D1.1a), and square this expression; substituting (D1.3) for the new right-hand side gives

$$\kappa_{p,\theta_{pm}}^2 + \kappa_{i,\Theta_i}^2 \cos^2\theta_i - 2\kappa_{p,\theta_{pm}} \kappa_{i,\Theta_i} \cos\theta_i = \kappa_{i,\Theta_i}^2 \cos^2\theta_i + \kappa_s^2 - \kappa_{i,\Theta_i}^2 \,, \text{(D1.4a)}$$

or

$$\kappa_{p,\theta_{pm}}^2 - \kappa_s^2 - 2\kappa_{p,\theta_{pm}} \kappa_{i,\Theta_i} \cos\theta_i + \kappa_{i,\Theta_i}^2 = 0 \,. \qquad \text{(D1.4b)}$$

This is a very transcendental equation, which we solve numerically (using an Excel spreadsheet). First we set θ_s, θ_i =0, and solve for the phase-matching angle θ_{pm} under the degeneracy condition, $\omega_s = \omega_i$. For BBO, we find that $\theta_{pm}(\lambda_p = 325\text{nm}) = 54.25°$ and $\theta_{pm}(\lambda_p = 316.5\text{nm}) = 56.39°$. [Note that this somewhat large angle implies a reduction of the effective nonlinearity by a factor of ~0.33.] Next we use our spreadsheet to find the value of θ_i [and thus θ_s, through (D1.3)], given a value of ϕ, over the idler wavelength range of interest. Lastly, we use Snell's law to obtain the laboratory angles θ'_s and θ'_i:

$$\sin\theta'_s = n_o(\omega_s) \sin\theta_s \,, \qquad \text{(D1.5a)}$$

and

$$\sin\theta'_i = n_e(\omega_i, \Theta_i) \sin\theta_i \,, \qquad \text{(D1.5b)}$$

where we have assumed that the pump is incident normal to the crystal face. It is these angles which are displayed in the plots of Fig. 10.4.

As alluded to in Sect. 10., the birefringence of the nonlinear crystal results in a relative *displacement* of the two down-converted photons: they propagate in the same direction after exiting the crystal, but are separated by a distance $d = L \tan\rho$, where L is the propagation distance *inside* the crystal, and ρ is the intra-crystal angle between the ordinary and extraordinary beams, given by [Dmitriev et al., 1991]

$$\rho(\Theta_i) = \arctan\left[\left(\frac{n_o(\omega_i)}{n_e(\omega_i)}\right)^2 \tan\Theta_i\right] - \Theta_i \quad . \tag{D1.6}$$

This has a value of 3.9° for our system.

Appendix D2: Iris Size Effect on Collection Efficiency

For a plane-wave pump, it is in principle possible to have 100%-collection efficiency[1], by which we mean that one could collect only down-converted photons whose conjugate partners were also collected. Once we allow a more realistic, gaussian-mode pump, then this is no longer possible. For identical, finite-sized collection irises, there will always exist situations where one photon is detected while the other is not (even aside from the problem of inefficient detectors). This effect is mitigated by collecting over a larger solid angle.

We approach the problem by assuming that our signal and idler irises collect light over the angular ranges $[-\Theta_s,\Theta_s]$ and $[-\Theta_i,\Theta_i]$, respectively[2]. For simplicity we will treat the problem first in only one transverse dimension. If the joint probability distribution for the signal and idler modes is $g(\theta_s, \theta_i)$, then we can write the coincidence probability C as

$$C = \int_{-\Theta_i}^{\Theta_i} d\theta_i \int_{-\Theta_s}^{\Theta_s} d\theta_s \; g(\theta_s,\theta_i) \; , \qquad \text{(D2.1)}$$

the signal singles probability as

---

[1] We are neglecting the (very) small spread in wave vectors due to the finite size of the crystal.

[2] Here we assume that the idler photons are sent to one detector, while the signal photons are sent to the other, as is possible in the modified version of the source with a half waveplate after one of the crystals. Also, we implicitly assume that we are collecting those down converted photons that are roughly collinear (and that these are essentially degenerate); otherwise, the angular ranges would possess an offset.

$$S_S = \int_{-\infty}^{\infty} d\theta_i \int_{-\Theta_s}^{\Theta_s} d\theta_S \; g(\theta_S, \theta_i) \tag{D2.2a}$$

and the idler singles probability as

$$S_i = \int_{-\Theta_i}^{\Theta_i} d\theta_i \int_{-\infty}^{\infty} d\theta_S \; g(\theta_S, \theta_i) \; . \tag{D2.2b}$$

As discussed in Chap. 9, we can define the efficiencies in terms of the coincidence and singles rates: $\eta_i = C/S_S$ and $\eta_S = C/S_i$ (the efficiency of each detector depends on the singles rate for the *other* detector). In practice, the angular ranges we will need to accept are relatively small, and for small angles the distribution $g(\theta_S, \theta_i)$ is symmetric in its arguments. Therefore, if the irises are chosen to be the same (i.e., $\Theta_S = \Theta_i$) then the singles probabilities S_S and S_i will be identical, and $\eta_S = \eta_i \equiv \eta$, where

$$\eta = \frac{\displaystyle\int_{-\Theta_i}^{\Theta_i} d\theta_i \int_{-\Theta_s}^{\Theta_s} d\theta_S \; g(\theta_S, \theta_i)}{\displaystyle\int_{-\infty}^{\infty} d\theta_i \int_{-\Theta_s}^{\Theta_s} d\theta_S \; g(\theta_S, \theta_i)} \; . \tag{D2.3}$$

We now turn our attention to the distribution function $g(\theta_S, \theta_i)$. This was calculated in the paraxial approximation explicitly by Ivan Deutsch (cf. p. 119 in [Deutsch, 1992]) for the case of degenerate collinear type-I phase matching. He found that if the pump beam consisted of a pulse with both a gaussian temporal as well as transverse profile, described by

$$\varepsilon_p(x_T,\tau) = \varepsilon_{p_o} \exp\left\{-\frac{x_T^2}{2w_o^2}\right\} \exp\left\{-\frac{\tau^2}{2\tau_o^2}\right\} , \qquad \text{(D2.4)}$$

(where w_o and $1/\tau_o$ are the pump beam waist and pump pulse bandwidth, respectively), then the joint distribution (in Fourier space) after propagating a distance z inside the nonlinear medium is

$$|\tilde{f}(z)|^2 = \left|\frac{V_o z}{2}\right|^2 \frac{w_o^4 \tau_o^2}{2\pi} \exp\{-|q_s + q_i|^2 w_o^2\} \exp\{-(\Omega_s + \Omega_i)^2 \tau_o^2\}$$

$$\times \operatorname{sinc}^2\left\{\frac{z}{2}\left|\frac{q_s^2 + q_i^2}{2k_o} - \frac{(\Omega_s^2 + \Omega_i^2)k''}{2}\right|\right\} , \qquad \text{(D2.5)}$$

where $V_o \equiv \frac{2\pi\omega_o}{n_o c}\chi^{(2)}\varepsilon_{p_o}$, $k_o \equiv \frac{\omega_o}{c} n(\omega_o)$, $\omega_o \equiv \omega_p/2$, and $k'' \equiv \left.\frac{\partial^2 k}{\partial\omega^2}\right|_{\omega_o}$. q_s and q_i are the transverse momentum components of the signal and idler photons, respectively, and Ω_s and Ω_i are the frequency shifts (away from degeneracy). Henceforth we assume a cw pump, so that $\tau_o \to \infty$ and we have $\Omega_s = -\Omega_i$. In the small angle approximation and near degeneracy, we can define $\theta_s \approx \frac{q_s}{k_p/2}$, $\theta_i \approx \frac{q_i}{k_p/2}$., so that $|\tilde{f}(z)|^2$ is essentially the function $g(\theta_s, \theta_i)$.

The sinc-term of (D2.5) basically governs the phase matching, and the angular spread of the down-converted photons *in singles*. In treating type-II phase-matching it is the argument of the sinc which would change somewhat (e.g., to include a k_e'' and k_o'', for the two polarizations). In any event, it is the gaussian that enforces the tight angular correlation of the signal and idler photons--for a plane-wave pump this correlation is exact, even though the sinc-function may ascribe a broad distribution to either of the

down-converted modes individually[3]. Consequently, we may write

$$g = A\, e^{-|q_s + q_i|^2 w_o^2} = A\, e^{-\left\{\left[(q_{s,x} + q_{i,x})^2 + (q_{s,y} + q_{i,y})^2\right] w_o^2\right\}} \qquad \text{(D2.6a)}$$

$$= A\, e^{-\left\{\left[(\theta_{s,x} + \theta_{i,x})^2 + (\theta_{s,y} + \theta_{i,y})^2\right] \frac{k_p^2 w_o^2}{4}\right\}} \qquad \text{(D2.6b)}$$

$$= A\, e^{-\left\{\left(\frac{\theta_{s,x} + \theta_{i,x}}{\theta_{diff,1/2}}\right)^2 + \left(\frac{\theta_{s,y} + \theta_{i,y}}{\theta_{diff,1/2}}\right)^2\right\}} , \qquad \text{(D2.6c)}$$

where in going from (D2.6a) to (D2.6b) we used the definitions of the angles, and in making the step to (D2.6c), we have used the following definition of the diffraction half-angle of the pump beam:

$$\theta_{diff,1/2} = \frac{\lambda_p}{\pi\, w_o} = \frac{2}{k_p w_o} . \qquad \text{(D2.7)}$$

From (D2.6c) it is clear that we can treat the two dimensions separately. Hence, we use

$$g(\theta_s, \theta_i) = A'\, e^{-\left(\frac{\theta_s + \theta_i}{\theta_{diff,1/2}}\right)^2} \qquad \text{(D2.6d)}$$

in (D2.3)[4]. A plot of the effective collection efficiency (in one transverse dimension)

[3] It is interesting to compare the relative widths of the gaussian and the sinc terms of (D2.5). The gaussian forces the angular correlation to be less than $k_p w_o$; the sinc (in the limit of $k'' = 0$) constrains the angles to be less than $\sqrt{k_p z}$. Assuming that the crystal length and the pump beam waist are of the same order, then the sinc-function is much broader than the gaussian, i.e., we can essentially treat it as a constant over the small irises needed according to the gaussian angular correlation.

[4] In neglecting to include the sinc-function, we have hidden any dependence which might make it desirable to use different *size* signal and idler irises, or non-circular

versus the iris sizes is shown in Fig. D2.1. To account for both dimensions, we simply square the result[5]. Figure 10.2 shows a plot of inherent collection efficiency (2-dimensional), versus the collection angle of the iris in units of the pump beam divergence angle. In order to keep the losses less than 2%, we must employ irises which accept light out to 30 times the pump divergence angle.

irises. However, our calculation should still be correct in terms of evaluating the effect of finite-iris sizes on effective collection efficiency.

[5] Note: One must also account for the fact that the above figuring implicitly assumed that the photons were *inside* the nonlinear medium. To convert the angles into the lab frame, we multiply by the index (because the angles are small).

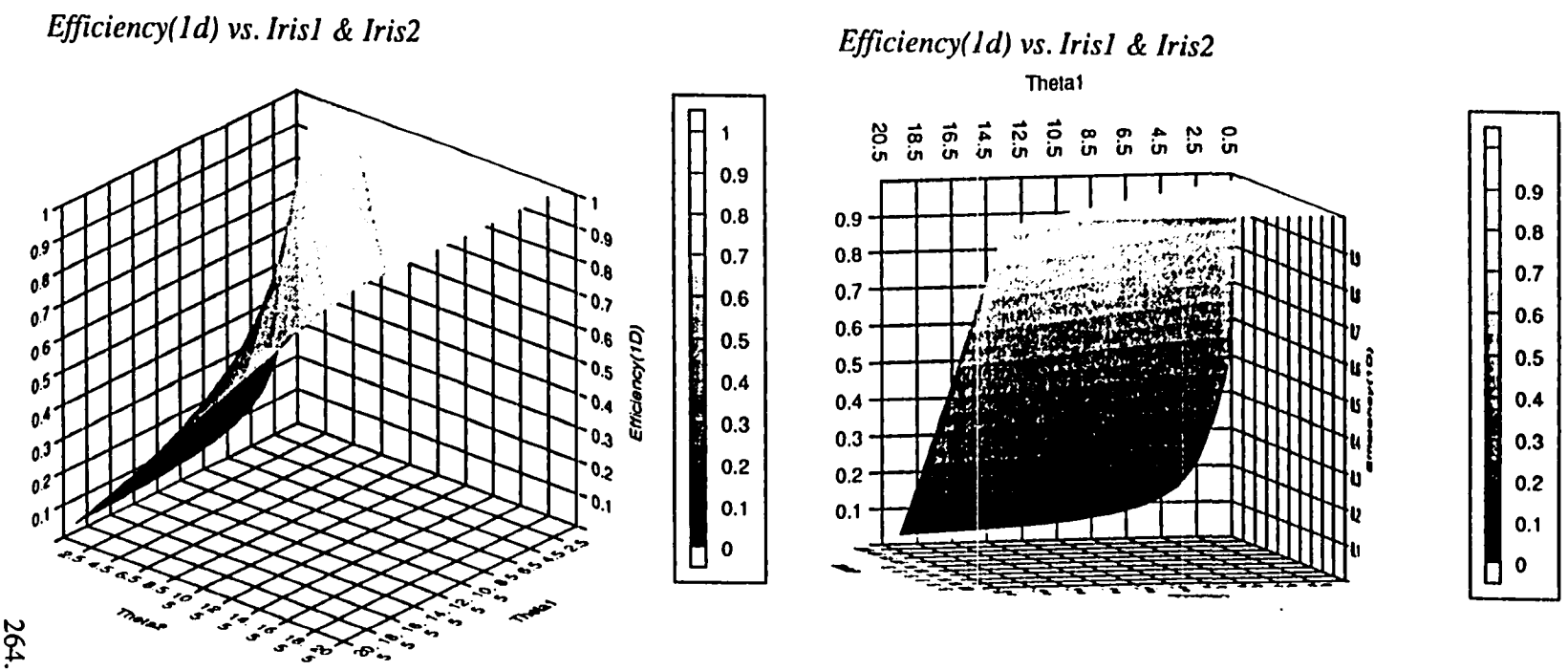


Figure D2.1 Two views of a plot of effective collection efficiency [(D2.3), in one transverse dimension], for a gaussian distribution function. The horizontal axes are the signal and idler iris collection half-angles, in units of the pump diffraction half-angle. Note that the results are not symmetric in θ_s (θ_1 in the plot) and θ_i (θ_2 in the plot), because (D2.3) writes the efficiency as C/S_s. Therefore, when $\theta_s >> \theta_i$ the efficiency will be small, while for $\theta_s << \theta_i$ the efficiency is high. This latter case is essentially the same arrangement used in the experiments to measure absolute detector quantum efficiency (see Chap. 9); the signal iris corresponds to the *trigger* detector, the idler iris to the detector whose efficiency we are measuring.

Appendix D3: General Calculation for Loophole-Free EPR Proposal

We present here a calculation of the various coincidence and singles rates expected from the setup shown in Fig. 10.1, when non-ideal polarizing beam splitters and non-unity efficiency detectors are allowed. We start with the state out of the two crystals, as given in Eq. (10.1):

$$|\psi\rangle = \sqrt{1-|A|^2}\,|\text{vac}\rangle + \frac{A}{\sqrt{1-|f|^2}}\left(|H, V\rangle_{\text{crystal 1}} + f\,|H, V\rangle_{\text{crystal 2}}\right). \qquad \text{(D3.1a)}$$

Henceforth, we will drop the vacuum term, as well as the prefactor A; physically, we must filter out the unconverted pump beam, which would otherwise give rise to an overwhelming background at our detectors. We write the state (D3.1a) in terms of photon creation operators:

$$|\psi\rangle \propto \frac{1}{\sqrt{1+|f|^2}}\left(\hat{a}^{\dagger}_{H,1}\hat{a}^{\dagger}_{V,1} + f\hat{a}^{\dagger}_{H,2}\hat{a}^{\dagger}_{V,2}\right)|0\rangle\ , \qquad \text{(D3.1b)}$$

where the subscript *letters* denote polarization, and the subscript *numbers* denote spatial mode.

Our strategy will be to transform the creation operators $\hat{a}^{\dagger}_{\lambda,1}$ and $\hat{a}^{\dagger}_{\lambda,2}$ into operators appropriate for the modes reaching the detectors. To this end, we state at the outset the transformation rules for a half waveplate and a general beam splitter. The transformation matrix $H_j(\theta)$ for a half waveplate (in spatial mode j) with its axis at an angle θ with respect to the horizontal is

$$H_j(\theta) = \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}, \tag{D3.2a}$$

so that

$$H_j(\theta)\hat{a}^\dagger_{V,j} = \hat{a}^\dagger_{V,j}\cos 2\theta + \hat{a}^\dagger_{H,j}\sin 2\theta,$$

$$H_j(\theta)\hat{a}^\dagger_{H,j} = \hat{a}^\dagger_{V,j}\sin 2\theta - \hat{a}^\dagger_{H,j}\cos 2\theta\ . \tag{D3.2b}$$

[We assume for simplicity that the wavelengths of the fields are equal to the design wavelength of the waveplate. This approximation is very good, given that the bandwidth of our photons (for the 0.05° half-angle iris acceptance calculated in Section 10.IV) is less than 1 nm, and zero-order waveplates have a broad performance window.] The mode transformation for a lossless symmetric beam splitter with input ports 1 and 2 and output ports 3 and 4 is

$$\begin{pmatrix} \hat{a}^\dagger_{\lambda,2} \\ \hat{a}^\dagger_{\lambda,1} \end{pmatrix} = \begin{pmatrix} t_\lambda & r_\lambda \\ r_\lambda & t_\lambda \end{pmatrix} \begin{pmatrix} \hat{a}^\dagger_{\lambda,3} \\ \hat{a}^\dagger_{\lambda,4} \end{pmatrix}, \tag{D3.3}$$

where t_λ and r_λ are the transmission and reflection amplitudes, respectively. We distinguish each polarization component λ to allow for the possibility of a polarizing beam splitter. In what follows, we use r_p to denote the reflection amplitude for p-polarized light (ideally equal to 1), and r_s to denote the reflection amplitude for s-polarized light (ideally equal to 0). For each polarization the amplitudes satisfy the standard relations, $|r_\lambda|^2 + |t_\lambda|^2 = 1$ and $\arg(r_\lambda) - \arg(t_\lambda) = \pm\pi/2$, derivable from unitarity. Without loss of generality, let $t_\lambda = |t_\lambda|$ and $r_\lambda = i|r_\lambda|$.

Using (D3.3), and including a relative phase shift δ between the two paths

(associated, say, with a variable path length in the interferometer arm with crystal 2), the state after the recombining beam splitter can be written

$$|\psi\rangle \approx \left(A_{33}\,\hat{a}^{\dagger}_{V,3}\hat{a}^{\dagger}_{H,3} + A_{44}\,\hat{a}^{\dagger}_{V,4}\hat{a}^{\dagger}_{H,4} + A_{34}\,\hat{a}^{\dagger}_{V,3}\hat{a}^{\dagger}_{H,4} + A_{43}\,\hat{a}^{\dagger}_{V,4}\hat{a}^{\dagger}_{H,3}\right)|0\rangle, \quad \text{(D3.4)}$$

where

$$A_{33} = \frac{1}{\sqrt{1+|f|^2}}\left(r_p r_s + f e^{i\delta} t_p t_s\right), \quad \text{(D3.5a)}$$

$$A_{44} = \frac{1}{\sqrt{1+|f|^2}}\left(t_p t_s + f e^{i\delta} r_p r_s\right), \quad \text{(D3.5b)}$$

$$A_{34} = \frac{1}{\sqrt{1+|f|^2}}\left(r_s t_p + f e^{i\delta} r_p t_s\right), \quad \text{(D3.5c)}$$

and

$$A_{43} = \frac{1}{\sqrt{1+|f|^2}}\left(r_p t_s + f e^{i\delta} r_s t_p\right). \quad \text{(D3.5d)}$$

Note that the terms in (D3.4) with coefficients A_{33} and A_{44} correspond to processes where both photons exit the same port of the interferometer. Only the A_{34} and A_{43} terms will contribute to coincidences.

Next we include the effect of a half waveplate in port 3, with its axis at angle $\theta_3 \equiv \alpha/2$ to the horizontal, and a half waveplate in port 4, with its axis at angle $\theta_4 \equiv \beta/2$:

$$|\psi'\rangle = H_3(\theta_3)H_4(\theta_4)\,|\psi\rangle$$

$$= \Big[\, A_{33}\left(\hat{a}^{\dagger}_{V,3}\cos\alpha + \hat{a}^{\dagger}_{H,3}\sin\alpha\right)\left(\hat{a}^{\dagger}_{V,3}\sin\alpha - \hat{a}^{\dagger}_{H,3}\cos\alpha\right)$$

$$+ A_{44}\left(\hat{a}^{\dagger}_{V,4}\cos\beta + \hat{a}^{\dagger}_{H,4}\sin\beta\right)\left(\hat{a}^{\dagger}_{V,4}\sin\beta - \hat{a}^{\dagger}_{H,4}\cos\beta\right)$$

$$+ A_{34}\left(\hat{a}^{\dagger}_{V,3}\cos\alpha + \hat{a}^{\dagger}_{H,3}\sin\alpha\right)\left(\hat{a}^{\dagger}_{V,4}\sin\beta - \hat{a}^{\dagger}_{H,4}\cos\beta\right)$$

$$+ A_{43}\left(\hat{a}^{\dagger}_{V,4}\cos\beta + \hat{a}^{\dagger}_{H,4}\sin\beta\right)\left(\hat{a}^{\dagger}_{V,3}\sin\alpha - \hat{a}^{\dagger}_{H,3}\cos\alpha\right)\Big]\,|0\rangle\ . \qquad \text{(D3.6)}$$

Lastly, we include the effects of the final analyzers, polarizing beam splitters with reflection and transmission amplitudes R_λ and T_λ, respectively. (Note: We use upper case for the analyzing beam splitters and lower case [cf. (D3.5)] for the interferometer recombining beam splitter.) The 16 terms of (D3.6) become a total of 64![1]

$$|\psi'\rangle = \Big[\ A_{33}\ \Big\{\ \cos\alpha\sin\alpha\left(R_s^2\hat{a}^{\dagger}_{V,3s}\hat{a}^{\dagger}_{V,3s} + T_s^2\hat{a}^{\dagger}_{V,3p}\hat{a}^{\dagger}_{V,3p} + 2R_sT_s\hat{a}^{\dagger}_{V,3s}\hat{a}^{\dagger}_{V,3p}\right.$$

$$\left. - R_p^2\hat{a}^{\dagger}_{H,3s}\hat{a}^{\dagger}_{H,3s} - T_p^2\hat{a}^{\dagger}_{H,3p}\hat{a}^{\dagger}_{H,3p} - 2R_pT_p\hat{a}^{\dagger}_{H,3s}\hat{a}^{\dagger}_{H,3p}\right)$$

$$+ \left(\sin^2\alpha - \cos^2\alpha\right)$$

$$\times\left(R_sR_p\hat{a}^{\dagger}_{V,3s}\hat{a}^{\dagger}_{H,3s} + T_sT_p\hat{a}^{\dagger}_{V,3p}\hat{a}^{\dagger}_{H,3p} + R_sT_p\hat{a}^{\dagger}_{V,3s}\hat{a}^{\dagger}_{H,3p} + T_sR_p\hat{a}^{\dagger}_{V,3p}\hat{a}^{\dagger}_{H,3s}\right)\Big\}$$

$$+ A_{44}\ \Big\{\ \cos\beta\sin\beta\left(R_s^2\hat{a}^{\dagger}_{V,4s}\hat{a}^{\dagger}_{V,4s} + T_s^2\hat{a}^{\dagger}_{V,4p}\hat{a}^{\dagger}_{V,4p} + 2R_sT_s\hat{a}^{\dagger}_{V,4s}\hat{a}^{\dagger}_{V,4p}\right.$$

$$\left. - R_p^2\hat{a}^{\dagger}_{H,4s}\hat{a}^{\dagger}_{H,4s} - T_p^2\hat{a}^{\dagger}_{H,4p}\hat{a}^{\dagger}_{H,4p} - 2R_pT_p\hat{a}^{\dagger}_{H,4s}\hat{a}^{\dagger}_{H,4p}\right)$$

$$+ \left(\sin^2\beta - \cos^2\beta\right)$$

---

[1] This is an exclamation, *not* a factorial. If it *were* a factorial, then we would have to write "64!!!"!

$$\times \left\{ R_s R_p \hat{a}^{\dagger}_{V,4s} \hat{a}^{\dagger}_{H,4s} + T_s T_p \hat{a}^{\dagger}_{V,4p} \hat{a}^{\dagger}_{H,4p} + R_s T_p \hat{a}^{\dagger}_{V,4s} \hat{a}^{\dagger}_{H,4p} + T_s R_p \hat{a}^{\dagger}_{V,4p} \hat{a}^{\dagger}_{H,4s} \right\} \right\}$$

$$+ \left\{ A_{34} \cos\alpha \sin\beta + A_{43} \sin\alpha \cos\beta \right\}$$

$$\times \left(R_s^2 \hat{a}^{\dagger}_{V,3s} \hat{a}^{\dagger}_{V,4s} + T_s^2 \hat{a}^{\dagger}_{V,3p} \hat{a}^{\dagger}_{V,4p} + R_s T_s \left[\hat{a}^{\dagger}_{V,3s} \hat{a}^{\dagger}_{V,4p} + \hat{a}^{\dagger}_{V,3p} \hat{a}^{\dagger}_{V,4s} \right] \right)$$

$$+ \left\{ - A_{34} \cos\alpha \cos\beta + A_{43} \sin\alpha \sin\beta \right\}$$

$$\times \left(R_s R_p \hat{a}^{\dagger}_{V,3s} \hat{a}^{\dagger}_{H,4s} + T_s T_p \hat{a}^{\dagger}_{V,3p} \hat{a}^{\dagger}_{H,4p} + R_s T_p \hat{a}^{\dagger}_{V,3s} \hat{a}^{\dagger}_{H,4p} + T_s R_p \hat{a}^{\dagger}_{V,3p} \hat{a}^{\dagger}_{H,4s} \right)$$

$$+ \left\{ A_{34} \sin\alpha \sin\beta - A_{43} \cos\alpha \cos\beta \right\}$$

$$\times \left(R_s R_p \hat{a}^{\dagger}_{H,3s} \hat{a}^{\dagger}_{V,4s} + T_s T_p \hat{a}^{\dagger}_{H,3p} \hat{a}^{\dagger}_{V,4p} + T_s R_p \hat{a}^{\dagger}_{H,3s} \hat{a}^{\dagger}_{V,4p} + R_s T_p \hat{a}^{\dagger}_{H,3p} \hat{a}^{\dagger}_{V,4s} \right)$$

$$+ \left\{ - A_{34} \sin\alpha \cos\beta - A_{43} \cos\alpha \sin\beta \right\}$$

$$\times \left(R_p^2 \hat{a}^{\dagger}_{H,3s} \hat{a}^{\dagger}_{H,4s} + T_p^2 \hat{a}^{\dagger}_{H,3p} \hat{a}^{\dagger}_{H,4p} + R_p T_p \left[\hat{a}^{\dagger}_{H,3s} \hat{a}^{\dagger}_{H,4p} + \hat{a}^{\dagger}_{H,3p} \hat{a}^{\dagger}_{H,4s} \right] \right) \Big] |0\rangle \quad \text{.(D3.7)}$$

As a check of this rather unwieldy expression, one can readily confirm that the normalization is as expected from (D3.4) (assuming lossless analyzing beam splitters):

$$\langle \psi' | \psi' \rangle = |A_{33}|^2 + |A_{44}|^2 + |A_{34}|^2 + |A_{43}|^2 \ . \qquad \text{(D3.8)}$$

From (D3.7) we can now calculate the various rates which constitute Bell's inequalities, using the standard Glauber theory of photodetection [Glauber, 1963]. For instance, the rate $n_{3s}(\alpha)$ of detection events at a polarization-*insensitive* detector (with efficiency η) at the s-channel of the analyzer in port 3 (described by the parameter α) is given by

$$n_{3s}(\alpha) \propto \eta \langle \psi' | \mathbf{E}_{3s}^{(+)} \cdot \mathbf{E}_{3s}^{(-)} | \psi' \rangle$$

$$\propto \eta \langle \psi' | \left(\hat{a}_{H,3s}^{\dagger} \varepsilon_{H,3s}^{*} + \hat{a}_{V,3s}^{\dagger} \varepsilon_{V,3s}^{*} \right) \cdot \left(\hat{a}_{H,3s} \varepsilon_{H,3s} + \hat{a}_{V,3s} \varepsilon_{V,3s} \right) | \psi' \rangle$$

$$= \eta \langle \psi' | \hat{a}_{H,3s}^{\dagger} \hat{a}_{H,3s} + \hat{a}_{V,3s}^{\dagger} \hat{a}_{V,3s} | \psi' \rangle$$

$$= \eta \left(\left| \hat{a}_{H,3s} | \psi' \rangle \right|^2 + \left| \hat{a}_{V,3s} | \psi' \rangle \right|^2 \right) \quad . \tag{D3.9}$$

Using (D3.7) in (D3.9), one finds, after many applications of the canonical commutation relations and much algebra,

$$n_{3s}(\alpha) = \eta \left\{ |A_{33}|^2 \left(|R_s|^2 + |R_p|^2 \right) + |A_{34}|^2 \left(\cos^2\alpha \, |R_s|^2 + \sin^2\alpha \, |R_p|^2 \right) \right.$$

$$\left. + |A_{43}|^2 \left(\sin^2\alpha \, |R_s|^2 + \cos^2\alpha \, |R_p|^2 \right) \right\} \quad . \tag{D3.10a}$$

Similarly, one can calculate the singles rate for the p-channel:

$$n_{3p}(\alpha) = \eta \left\{ |A_{33}|^2 \left(|T_s|^2 + |T_p|^2 \right) + |A_{34}|^2 \left(\cos^2\alpha \, |T_s|^2 + \sin^2\alpha \, |T_p|^2 \right) \right.$$

$$\left. + |A_{43}|^2 \left(\sin^2\alpha \, |T_s|^2 + \cos^2\alpha \, |T_p|^2 \right) \right\} \quad . \tag{D3.10b}$$

Identical expressions hold for $n_{4s}(\beta)$ and $n_{4p}(\beta)$, when the substitutions $3 \leftrightarrow 4$ and $\alpha \rightarrow \beta$ are made. (Note: In the numerical calculations which were performed, an additional variable background was added; this was *not* taken to rely on η, and so could include dark counts, as well as counts from stray light.) (D3.10) is very general, and therefore rather complicated. The results simplify greatly if one considers the special case of ideal polarizing beam splitters ($|r_s| = |R_s| = |t_p| = |T_p| = 1$, $|r_p| = |R_p| = |t_s| = |T_s| = 0$), for then A_{33} and A_{44} vanish. Moreover, $|A_{34}|^2 = 1/(1+|f|^2)$ and $|A_{43}|^2 = |f|^2/(1+|f|^2)$, so that we have

$$n_{3s}(\alpha) = \frac{\eta}{1+|f|^2}\left\{\cos^2\alpha + |f|^2 \sin^2\alpha\right\}, \qquad \text{(D3.11a)}$$

$$n_{3p}(\alpha) = \frac{\eta}{1+|f|^2}\left\{\sin^2\alpha + |f|^2 \cos^2\alpha\right\} . \qquad \text{(D3.11b)}$$

The various coincidence rates between detectors in positions i and j can be calculated in like fashion, using

$$n_{i,j}(\alpha,\beta) \propto \eta^2 \langle \psi' | \, \mathbf{E}_i^{(+)} \mathbf{E}_j^{(+)} \mathbf{E}_j^{(-)} \mathbf{E}_i^{(-)} \, | \psi' \rangle$$

$$\propto \eta^2 \langle \psi' | : \left(\hat{a}_{H,i}^\dagger \hat{a}_{H,i} + \hat{a}_{V,i}^\dagger \hat{a}_{V,i}\right)\left(\hat{a}_{H,j}^\dagger \hat{a}_{H,j} + \hat{a}_{V,j}^\dagger \hat{a}_{V,j}\right) : | \psi' \rangle \ , \qquad \text{(D3.12)}$$

where again we have assumed polarization-insensitive detectors with efficiency η. For example,

$$n_{3s,4s}(\alpha,\beta) \equiv n_{s,s}(\alpha,\beta) = \eta^2 \left\{ \left| A_{34} \cos\alpha \sin\beta + A_{43} \sin\alpha \cos\beta \right|^2 |R_s|^4 \right.$$

$$+\left(\left|A_{34}\cos\alpha\cos\beta - A_{43}\sin\alpha\sin\beta\right|^2 + \left|A_{34}\sin\alpha\sin\beta - A_{43}\cos\alpha\cos\beta\right|^2\right)\left|R_S R_p\right|^2$$

$$+\left|A_{34}\sin\alpha\cos\beta + A_{43}\cos\alpha\sin\beta\right|^2\left|R_p\right|^4\Big\} \ . \qquad \text{(D3.13)}$$

As with the single-event rates, this fairly complicated form simplifies greatly if, for example, the analyzers (described above by R_s and R_p) are ideal. Then only the first term remains:

$$n_{s,s}(\alpha,\beta) = \eta^2\left|A_{34}\cos\alpha\sin\beta + A_{43}\sin\alpha\cos\beta\right|^2 \ . \qquad \text{(D3.14)}$$

If we further specialize to the familiar case of an equal superposition of the contributions from the two crystals (i.e., f = 1), and consider an ideal recombining beam splitter, we are simply left with

$$n_{s,s}(\alpha,\beta) = \frac{\eta^2}{2}\left|\cos\alpha\sin\beta + e^{i\delta}\sin\alpha\cos\beta\right|^2 \ , \qquad \text{(D3.15)}$$

which yields the familiar result $\frac{\eta^2}{2}\sin^2(\alpha-\beta)$ when $\delta = 180°$.

However, as discussed in Sect. 10.VI, it is preferable to use $\delta = 0$ for an implementable, loophole-free test of Bell's inequalities, because the required detection efficiency is much less affected by non-idealities of the recombining polarizing beam splitter than when $\delta = 180°$. In fact, by examining the structure of A_{33} and A_{44} [Eq's (D3.5a) and (D3.5b), respectively] one can readily see the destructive interference that prevents both photons from exiting the same port of the recombining beam splitter, under the conditions f = 1, $|r_s| = |t_p|$ and $|r_p| = |t_s|$, and $\delta = 0$. (Remember that $\arg(r_\lambda) - \arg(t_\lambda) = \pm\pi/2$!)

Appendix E: Detector Trivia

In this Appendix we include copies of the various specifications sheets for the three types of detectors used in our tests, as well as the circuit layout for the (in-house built) temperature controller used with one of them. The reason for documenting this information is that it is not generally available except through the company which makes the products.

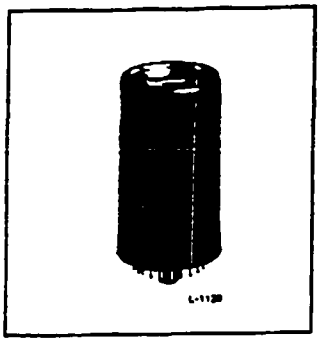
The contents of this appendix are as follows:

[1] The companies that supplied these detectors have gone through several changes of ownership.

[2] These devices are identical to the C30902S, but come with an integral thermoelectric cooler.

[3] Our devices were custom-modified (by EG&G) to provide higher than usual overbiasing. In particular, our overbias voltage was about 30 V, compared to the standard 7 V. This substantially increased the time resolution and the efficiency (cf. the specified values and those reported in Chap. 9).

C31034 Series Photomultipliers



51-mm (2-inch) Diameter 11-Stage, End Window Quantacon® PMTs

- **Typical Cathode Responsivity - Luminous -**
  C31034 : 440 uA/lm
  C31034A : 720 uA/lm
  Radiant -
  C31034: 52 mA/W at 860 nm
  C31034A: 81 mA/W at 860 nm
- **Broad Spectral Response Range 185 to 930 nm**
- **Small Photocathode Area (Projected) - 4 mm x 10 mm minimum**
- **Low Dark Noise at -30 °C - Maximum Dark Pulse Summation**
  C31034, C31034A : 50 cps
  C31034-06, C31034A-05: 25 cps
  C31034-02, C31034A-02: 12 cps

The BURLE C31034 Series are 51-mm (2-inch) diameter end-window, 11-stage Quantacon photomultipliers. They have GaAs:Cs-O photocathodes, ultraviolet-transmitting glass windows, and in-line copper beryllium dynode structures. The C31034A is a variant of the C31034 Series which exhibits higher photocathode responsivity.

The C31034 Series is designed specifically for use at reduced temperatures, e.g., -30° C. When cooled to such temperatures, these tubes are highly useful in photon counting applications such as Raman and fluorescence spectroscopy and astronomy. Cooling reduces dark noise to a minimal value and allows taking full advantage of the performance capabilities of the tube.

Variants with lower dark pulse summations are shown under Performance Data.

Absolute-Maximum Ratings[1]

	Limiting Values	
Average Anode Current:		
Averaged over any 30 second interval	100[2]	nA
DC Supply Voltage:		
Between anode and cathode	2200	V
Between anode and dynode No.11	250	V
Between adjacent dynodes	250	V
Between dynode No.1 and cathode	200	V
Temperature Range[3]:		
Storage	-80 to +50	°C
Operating	-80 to +30	°C

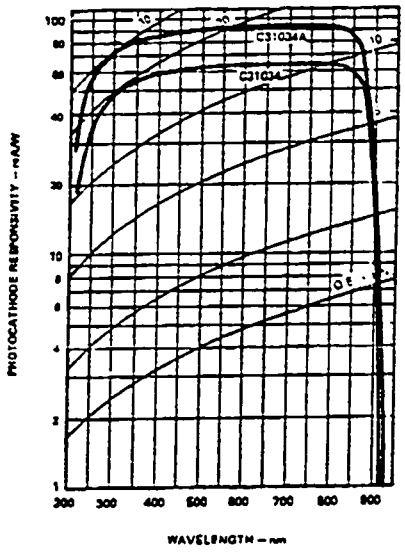


Figure 1 - Typical Photocathode Spectral Response Characteristics

BURLE Electron Tubes

Performance Data

With the voltage distribution of Table 1, an anode-to-cathode supply voltage of 1500 volts, a tungsten light source at a color temperature of 2856K, and an ambient operating temperature of 22 °C, unless otherwise indicated.

	C31034 Series			C31034A Series			
	Min.	Typ.	Max.	Min.	Typ.	Max.	Units
Anode Responsivity:							
Radiant at 860 nm	-	2.6×10^5	-	-	4.9×10^5	-	A/W
Luminous (1×10^{-6} lm)	100	210	-	100	430	-	A/m
Cathode Responsivity:							
Radiant at 860 nm[4]	25	52	-	70	81	-	mA/W
Luminous (1×10^{-6} lm, 200 volts between cathode and all other electrodes connected as anode)	210	440	-	560	720	-	uA/lm
Current Amplification (Gain)	-	0.5×10^6	-	-	0.6×10^6	-	
Rise Time	-	2.6	-	-	2.5	-	ns
Transit Time	-	33	-	-	33	-	ns
Anode Dark Current[5]	-	15	40	-	15	40	nA
Equivalent Anode Dark Current Input:							
at a Current Amplification of 10^6	-	3.4×10^{-11}	-	-	2.1×10^{-11}	-	lm
at 860 nm	-	2.9×10^{-10}	-	-	1.9×10^{-13}	-	W
Dark Pulse Summation:[6]							
C31034	-	-	50	-	-	-	cps
C31034-02	-	-	12	-	-	-	cps
C31034-06	-	-	25	-	-	-	cps
C31034A	-	-	-	-	-	60	cps
C31034A-02	-	-	-	-	-	12	cps
C31034A-05	-	-	-	-	-	25	cps

1. In accordance with the Absolute Maximum rating system as defined by the Electronic Industries Association Standard RS-239A, formulated by the JEDEC Electron Tube Council.
2. The gallium arsenide photocathode of the C31034 Series is highly susceptible to a permanent loss in photocathode responsivity when high levels of average anode current are used. Accordingly, this maximum rating must never be exceeded. To obtain best long-term stability and life, the tube should be operated at average anode current values well below the specified maximum of 100 nanoamperes.
3. The use or storage of these tubes with attached teflon sockets, such as the AJ2144A or AJ2145A, at temperatures below -50° C can destroy the tube. When tube operation below -50° C is desired, the teflon body of the socket should be removed and only socket-contact to circuit-element connections should be used.
4. Measured using a calibrated narrow bandpass filter having a center wavelength at 860 nanometers and a bandpass (FWHM) of approximately 10 nanometers. The light source is a tungsten-filament lamp having a lime-glass envelope. The lamp is operated at a color temperature of 2856 K and the value of light flux incident on the filter is 1×10^{-6} lumen. 200 volts are applied between cathode and all other electrodes connected as anode.
5. The supply voltage is adjusted so that the tube is operating at a current amplification of 10^6. The dark current is recorded. Dark current caused by thermionic emission may be reduced by cooling.
6. Maximum dark count summation made between 1/8 pe and 16 pe at -30° C and a Gain of 10^6 (See Figure 4)

   A Multichannel Pulse Height Analyzer having 256 channels is employed. The anode load is a 100 ± 5% kilohm resistor in parallel with a total capacitance of 100 ± 3% pF. The tube is measured in complete darkness. The pulse height for the single photoelectron equivalent is determined by using a light source operated at a low color temperature to assure the high probability of single photoelectron emission from the photocathode of the tube. The light is removed before the dark pulse summation is measured. The supply voltage is adjusted so that the peak of the single electron distribution lies in channel No.8. This corresponds to a tube gain of approximately 10^6.

Table 1

Voltage to be Provided by Divider	
Voltage (E) Between	8.06% o Supply Multiplied By:
Cathode and Dynode No.1	1.0
Dynode No.1 and Dynode No 2	1.4
Dynode No 2 and Dynode No 3	1 0
Dynode No.3 and Dynode No 4	1 0
Dynode No 4 and Dynode No 5	1 0
Dynode No.5 and Dynode No.6	1.0
Dynode No 6 and Dynode No 7	1 0
Dynode No 7 and Dynode No 8	1 0
Dynode No 8 and Dynode No 9	1 0
Dynode No.9 and Dynode No 10	1.0
Dynode No 10 and Dynode No.11	1 0
Dynode No.11 and Anode	1.0
Anode and Cathode	12 4
Aperture plate (pin No.17) is connected to photocathode (pin No.1).	

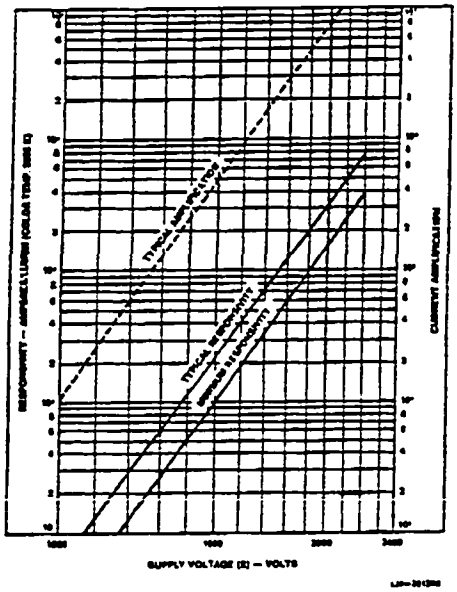


Figure 2 - Responsivity and Current Amplification Characteristics - C31034 Series

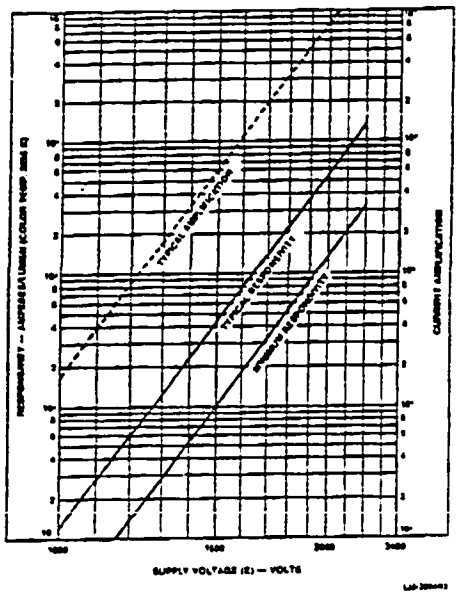


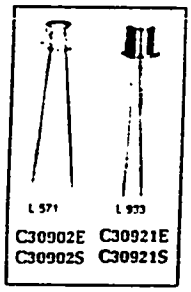
Figure 3 - Responsivity and Current Amplification Characteristics - C31034A Series

RCA Electro Optics

Silicon Avalanche Photodiodes
C30902E, C30902S, C30921E, C30921S

DATA SHEET

High Speed Solid State Detectors for Fiber Optic and Very Low Light-Level Applications



C30902E C30921E
C30902S C30921S

RCA Type C30902E avalanche photodiode utilizes a silicon detector chip fabricated with a double-diffused "reach-through" structure. This structure provides high responsivity between 400 and 1000 nanometers as well as extremely fast rise-and falltimes at all wavelengths. Because the fall-time characteristics have no "tail", the responsivity of the device is independent of modulation frequency up to about 800 MHz. The detector chip is hermetically-sealed behind a flat glass window in a modified TO-18 package. The useful diameter of the photosensitive surface is 0.5 mm.

RCA Type C30921E utilizes the same silicon detector chip as the C30902E, but in a package containing a lightpipe which allows efficient coupling of light to the detector from either a focussed spot or an optical fiber up to 0.25 mm in diameter. The internal end of the lightpipe is close enough to the detector surface to allow all of the illumination exiting the lightpipe to fall within the active-area of the detector. The hermetically-sealed TO-18 package allows fibers to be epoxied to the end of the lightpipe to minimize signal losses without fear of endangering detector stability.

The C30902E and C30921E are designed for a wide variety of uses including optical communications at data rates to 1 GBit/second, laser range-finding, and any other applications requiring high speed and/or high responsivity.

The C30902S and C30921S are selected C30902E and C30921E photodiodes having extremely low noise and low bulk dark-current. They are intended for ultra-low light level applications (optical power less than 1 pW) and can be used in either their normal linear mode ($V_R<V_{BR}$) at gains up to 250 or greater, or as photon counters in the "Geiger" mode ($V_R>V_{BR}$) where a single photoelectron may trigger an avalanche pulse of about 10^8 carriers. In this mode, no amplifiers are necessary and single photon detection probabilities of up to approximately 50% are possible.

Photon-counting is also advantageous where gating and coincidence techniques are employed for signal retrieval.

- **High Quantum Efficiency**
  77% Typical at 830 nm
- **C30902S and C30921S in Geiger Mode:**
  - Single-Photon Detection Probability to 50%
  - Low Dark-Count Rate at 5% Detection Probability – Typically
    15,000/second at +22° C
    350/second at -25° C
  - Count Rates to 2 x 10^6/second
- **Hermetically Sealed Package**
- **Low Noise at Room Temperature**
  - C30902E, C30921E –
    2.3 x 10^{-13} A/Hz$^{1/2}$
  - C30902S, C30921S –
    1.1 x 10^{-13} A/Hz$^{1/2}$
- **High Responsivity –**
  Internal Avalanche Gains in Excess of 150
- **Spectral Response Range – (10% Points)**
  400 to 1000 nm
- **Time Response – Typically 0.5 ns**
- **Wide Operating Temperature Range –**
  -40° C to +70° C

Optical Characteristics

C30902E, C30902S (Figure 13)

Photosensitive Surface:

Shape	Circular	
Useful area	0.2	mm²
Useful diameter	0.5	mm

Field of View:

Approximate full angle for totally illuminated photosensitive surface	100	deg

C30921E, C30921S (Figure 14)

Numerical Aperture of Light Pipe	0.55	
Refractive Index (n) of Core	1.61	
Light Pipe Core Diameter	0.25	mm

Maximum Ratings, Absolute-Maximum Values

Reverse Current at 22° C:		
Average value, continuous operation	200	μA
Peak value (For 1 second duration, non-repetitive	1	mA
Forward Current, I_F at 22°C:		
Average value, continuous operation	5	mA
Peak value (For 1 second duration, non-repetitive)	50	mA
Maximum Total Power Dissipation at 22° C	60	mW
Ambient Temperature —		
Storage, T_{stg}	-60 to + 100	°C
Operating, T_A	-40 to +70	°C
Soldering:		
For 5 seconds	200	°C

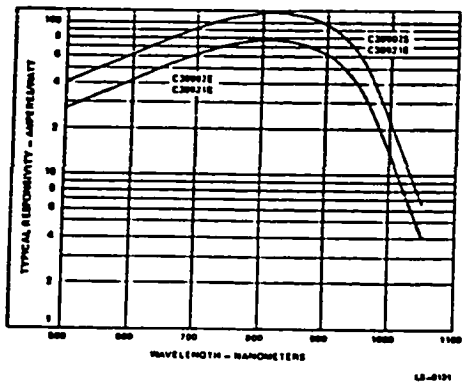


Fig. 1 Typical Spectral Responsivity at 22° C

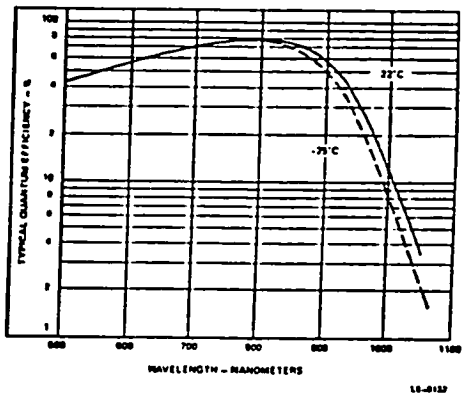


Fig. 2 Typical Quantum Efficiency vs Wavelength

Electrical Characteristics[1] at T_A = 22° C

	C30902E, C30921E			C30902S, C30921S			Units
	Min.	Typ.	Max.	Min.	Typ.	Max.	
Breakdown voltage, V_{BR}	–	225	–	–	225	–	V
Temperature Coefficient of V_R for Constant Gain	0.5	0.7	0.8	0.5	0.7	0.8	V/° C
Gain	–	150	–	–	250	–	
Responsivity:							
At 900 nm	55	65	–	92	108	–	A/W
At 830 nm	70	77	–	117	128	–	A/W
Quantum Efficiency:							
At 900 nm	–	60	–	–	60	–	%
At 830 nm	–	77	–	–	77	–	%
Dark Current, I_d	–	1.5x10⁻⁹ (Figure 6)	3x10⁻⁹	–	1x10⁻⁹ (Figure 6)	3x10⁻⁹	A
Noise Current, i_n:[2] f = 10 kHz, Δf = 1.0 Hz	–	2.3x10⁻¹² (Figure 3)	5x10⁻¹²	–	1.1x10⁻¹² (Figure 3)	2x10⁻¹²	A/Hz$^{1/2}$
Capacitance, C_d	–	1.6	2	–	1.6	2	pF
Rise Time, t_r: R_L = 50 Ω, λ = 830 nm, 10% to 90% points	–	0.5	0.75	–	0.5	0.75	ns
Fall Time: R_L = 50 Ω, λ = 830 nm, 90% to 10% points	–	0.5	0.75	–	0.5	0.75	ns
Geiger Mode (See Appendix)							
Dark Count Rate at 5% Photon Detection Probability[3] (830 nm):							
22° C	–	–	–	–	15,000	30,000	cps
-25° C	–	–	–	–	350	700	cps
Voltage Above V_{BR} for 5% Photon Detection Probability[3] (830 nm) (see Figure 8)	–	–	–	–	2	–	V
Dead-Time Per Event (See Appendix)	–	–	–	–	300	–	ns
After-Pulse Ratio at 5% Photon Detection Probability (830 nm) 22° C[4]	–	–	–	–	2	15	%

1 At the DC reverse operating voltage V_R supplied with the device and a light spot diameter of 0.25 mm (C30902E, S) or 0.10 mm (C30921E, S). Note that a specific value of V_R is supplied with each device. When the photodiode is operated at this voltage, the device will meet the electrical characteristic limits shown above. The voltage value will be within the range of 180 to 250 volts.

2 The theoretical expression for shot noise current in an avalanche photodiode is $i_n = (2q\,(I_{ds} + (I_{db}M^2 + P_oRM)\,F)\,B_w)^{1/2}$ where q is the electronic charge, I_{ds} is the dark surface current, I_{db} is the dark bulk current, F is the excess noise factor, M is the gain, P_o is the optical power on the device, and B_w is the noise bandwidth. For these devices F = 0.98 (2-1/M) + 0.02 M. (Reference: PP Webb, RJ McIntyre, JJ Conradi, "RCA Review", Vol. 35, p. 234, (1974).

3 The C30902S and C30921S can be operated at a substantially higher Dectection Probabilities. See Appendix.

4 After-Pulse occurring 1 microsecond to 60 seconds after main pulse.

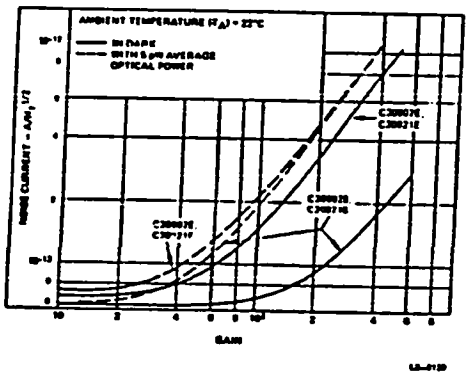


Fig. 3 Typical Noise Current vs Gain

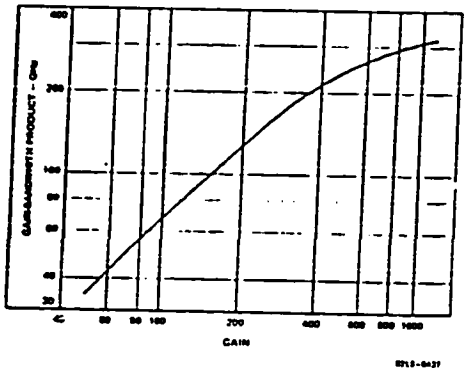


Fig. 5 Typical Gain-Bandwidth Product vs Gain

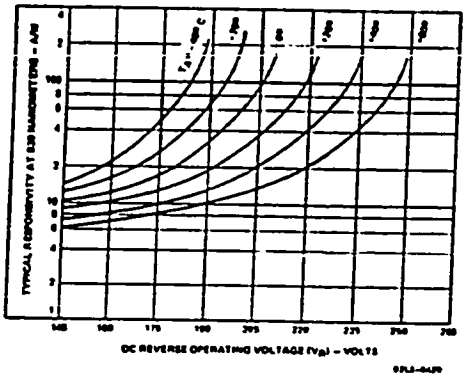


Fig. 4 Typical Responsivity at 830 nm vs Operating Voltage

Note: Operation below 145 volts is not recommended, since the device is not fully depleted below this value.

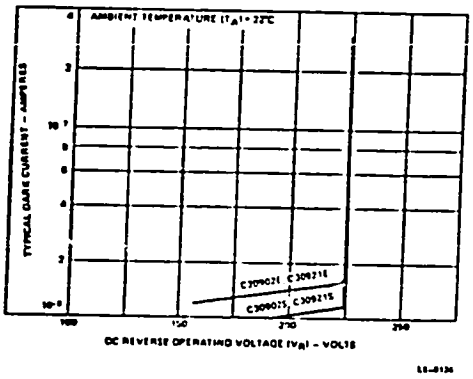


Fig. 6 Typical Dark Current vs Operating Voltage ($V<V_{BR}$)

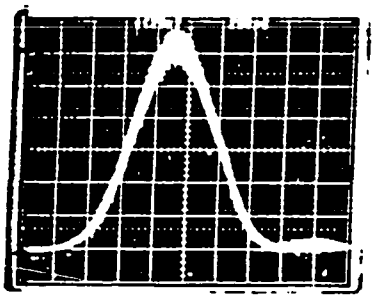


Fig. 7 Avalanche Photodiode Response to a 100 ps Laser Pulse as Measured With a 350 ps Sampling Head. (Horizontal Axis: 200 ps /Division)
Normal Linear Mode $V_R<V_{BR}$

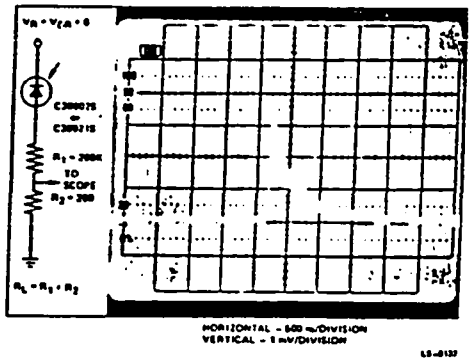


Fig. 9 Passively Quenched Circuit and Resulting Pulse Shape

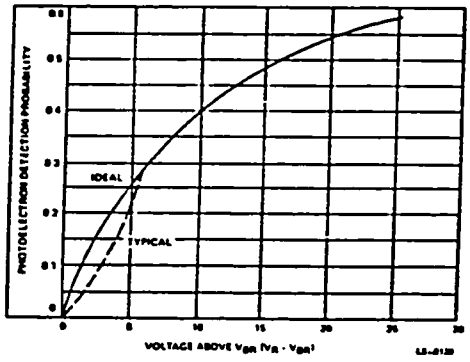


Fig. 8 Gelger Mode, Photoelectron Detection Probability vs Voltage Above V_{BR} ($V_R>V_{BR}$)

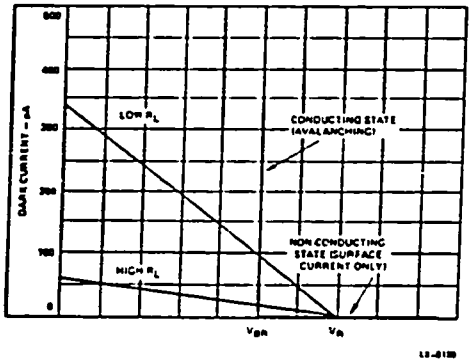


Fig. 10 Load Line for C30921S in the Geiger Mode

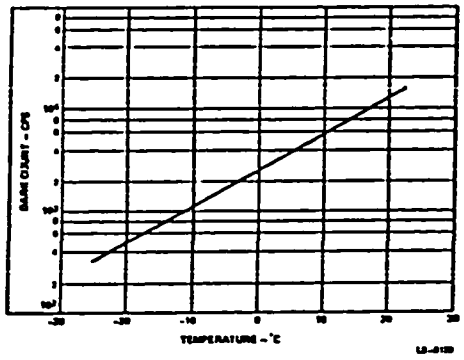


Fig. 11 Typical Dark Count vs Temperature at 5% Photon (830 nm) Detection Efficiency

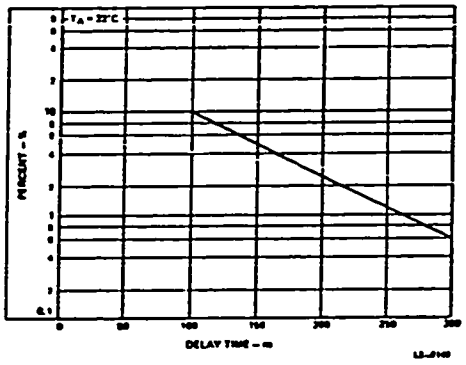
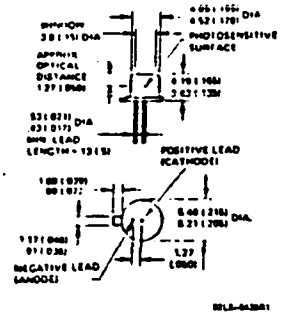


Fig. 12 Chance of an After-Pulse Within the Next 100 ns vs Delay-Time in an Actively Quenched Circuit.
(Typical for C30902S, C30921S at V_{BR} + 25)

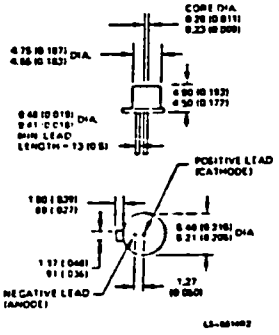


Modified TO-18 Package

Dimensions in millimeters. Dimensions in parentheses are in inches.

Note: Optical distance is defined as the distance from the surface of the silicon chip to the front surface of the window.

Fig. 13 Dimensional Outline - C30902E, C30902S



TO-18 Package

Dimensions in millimeters. Dimensions in parentheses are in inches.

Fig. 14 Dimensional Outline - C30921E, C30921S

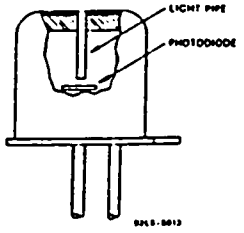


Fig. 15 Cutaway of the RCA C30921E, C30921S

Appendix

Operation of the C30902S & C30921S in the Geiger Mode

Introduction

When biased above the breakdown voltage, an avalanche photodiode will normally conduct a large current. However, if the current is such that the current is limited to less than a particular value (about 50 µA for these diodes), the current is unstable and *can switch off by itself. The explanation of this behaviour is that* the number of carriers in the avalanche region at any one time is small and fluctuating wildly. If the number happens to fluctuate to zero, the current must stop. It subsequently remains off until the avalanche pulse is retriggered by a bulk-or photo-generated carrier.

The C30902S and C30921S are selected to have small bulk-generated dark-current. This makes them suitable for low-noise operation below V_{BR} or of photon-counting above V_{BR} in the Geiger mode. In this so-called Geiger mode, a single photoelectron (or thermally-generated electron) may trigger an avalanche pulse which discharges the photodiode from its reverse voltage V_R to a voltage slightly below V_{BR}. The probability of this avalanche occuring is shown in Figure 8 as the "Photoelectron Detection Probability" and as can be seen, it increases with reverse voltage V_R. For a given value of V_R-V_{BR}, the Photoelectron Detection Probability is independent of temperature. To determine the Photon Detection Probability, it is necessary to multiply the Photoelectron Detection Probability by the Quantum Efficiency, which is shown in Figure 2. the Quantum Efficiency also is relatively independent of temperature, except near the 1000 nm cut-off.

The C30902S and C30921S can be used in the Geiger mode using either "passive" or "active" pulse quenching circuits. The advantages and disadvantages of each are discussed below.

Passive-Quenching Circuit

The simplest, and in many cases a perfectly adequate method of quenching a breakdown pulse, is through the use of a current-limiting load resistor. An example of such a "passive" quenching circuit is shown in Figure 9. The load line of the circuit is shown in Figure 10. To be in the conducting state at V_{BR} two conditions must be met:

1. The avalanche must have been triggered by either a photoelectron or a bulk-generated electron entering the avalanche region of the diode. (Note: holes are inefficient at starting avalanches in silicon.) The probability of an avalanche being initiated is discussed above.
2. To continue to be in the conducting state, a sufficiently large current, called the latching current I_{LATCH}, must be passing through the device so that there is always an electron or hole in the avalanche region. Typically in the C30902S and C30921S, I_{LATCH} = 50 µA, for currents $(V_R-V_{BR})/R_L$ much greater than I_{LATCH}, the diode remains conducting. If the current $(V_R-V_{BR})/R_L$ is much less than I_{LATCH}, the diode switches almost immediately to the non-conducting state. If $(V_R-V_{BR})/R_L$ is approximately equal to I_{LATCH}, then the diode will switch at an arbitrary time from the conducting to the non-conducting state depending on when the number of electrons and holes in the avalanche region statistically fluctuates to zero.

When R_L is large, the photodiode is normally nonconducting, and the operating point is at $V_R - I_xR_L$ in the non-conducting state. Following an avalanche breakdown, the device recharges to the voltage $V_R - I_xR_L$ with the time constant CR_L where C is the total device capacitance including stray capacitance. Using C = 1.6 pF and R_L = 200 2 KΩ a recharge time constant of .32 microseconds is calculated. In reasonable agreement with observation as shown in Figure 9. As is also evident from Figure 9, the rise-time is fast, 5 to 50 nanoseconds, decreases as V_R-V_{BR} increases, and is very dependent on the capacitances of the load resistors, leads, etc. The jitter at the half-voltage point is typically the same order of magnitude as the rise time. For timing purposes where it is important to have minimum jitter, the lowest possible threshold of the rising pulse should be used.

Active-Quenching Circuit

Until the C30902S or C30921S is recharged, the probability of detecting another incoming photoelectron is relatively low. To *avoid an excessive dead time when operating at a large voltage* above V_{BR}, an "actively quenched" circuit can be used. The circuit temporarily drops the bias voltage for a fraction of a microsecond following the detection of an avalanche discharge. This delay time allows all electrons and holes to be collected, including most of those temporarily "trapped" at various impurity sites in the silicon. When the higher voltage is reapplied, there are no electrons in the depletion region to trigger another avalanche or latch the diode. Recharging can now be very rapid through a small load resistor. Alternatively, the bias voltage can be maintained but the load resistor is replaced by a transistor which is kept off for a short time after an avalanche, and then turned on for a period sufficient to recharge the photodiode.

After-Pulsing

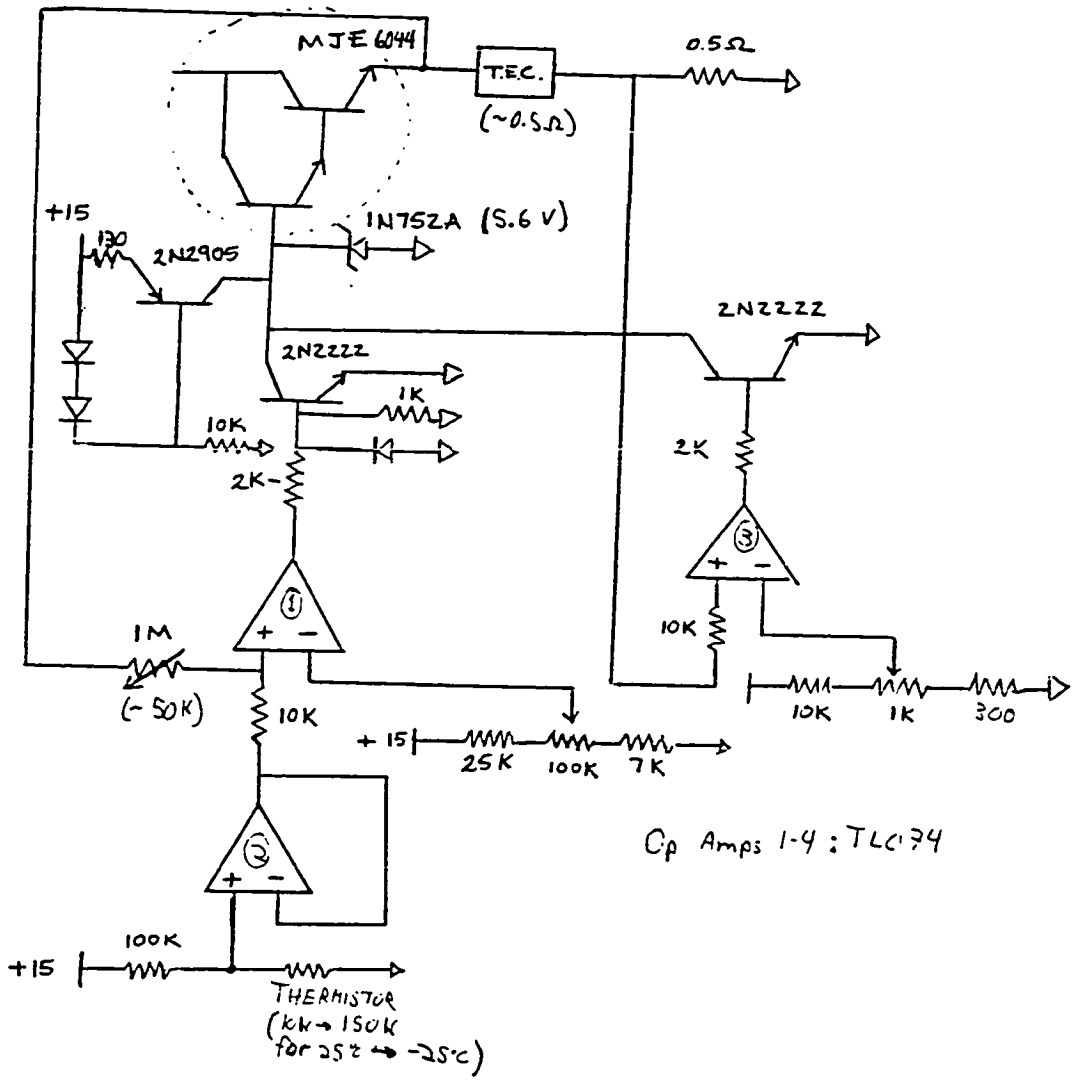
An after pulse is an avalanche breakdown pulse which follows a photon generated pulse and is induced by it. An after pulse is usually caused by one of the approximately 10^6 carriers which pass through the diode because of the first avalanche. This electron or hole is captured and trapped at some impurity site in the silicon, as previously described. When this charge carrier is liberated, usually in less than 100 nanoseconds but sometimes several milliseconds later, it may start another avalanche. The probability of an after pulse occurring more than one microsecond later is typically less than 2% at 2 volts above V_{BR}, using the circuit shown in Figure 9. After-pulsing increases with bias voltage. If it is necessary to reduce after-pulses, it is recommended that one keep V_R-V_{BR} low, use an actively quenched circuit with a long delay-time (see Figure 12), or a passively-quenched circuit with a long R_LC constant. Stray capacitances must also be minimized. Electronic gating of the signal can be performed in certain situations. Should after-pulses be a serious complication in a particular application, operation below V_{BR} with a good amplifier might be considered.

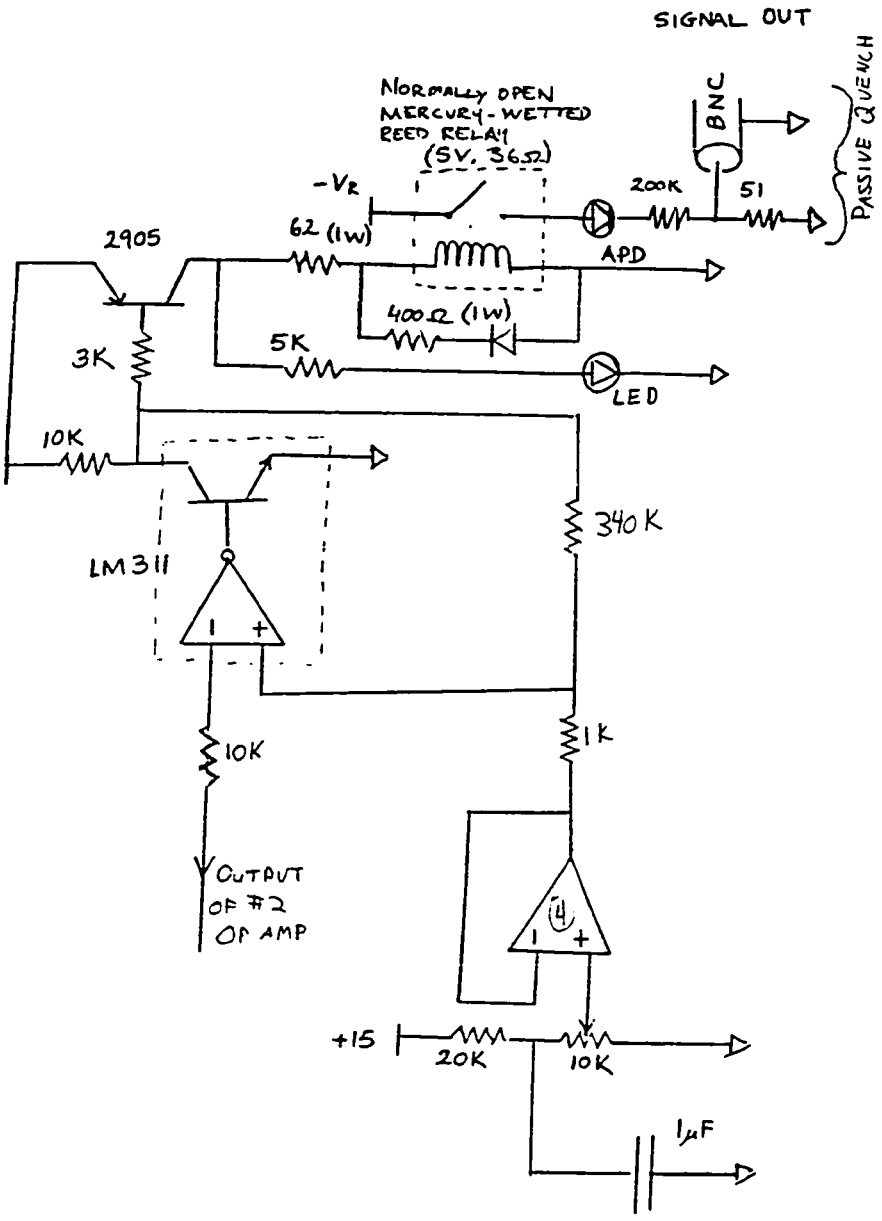
Dark Current

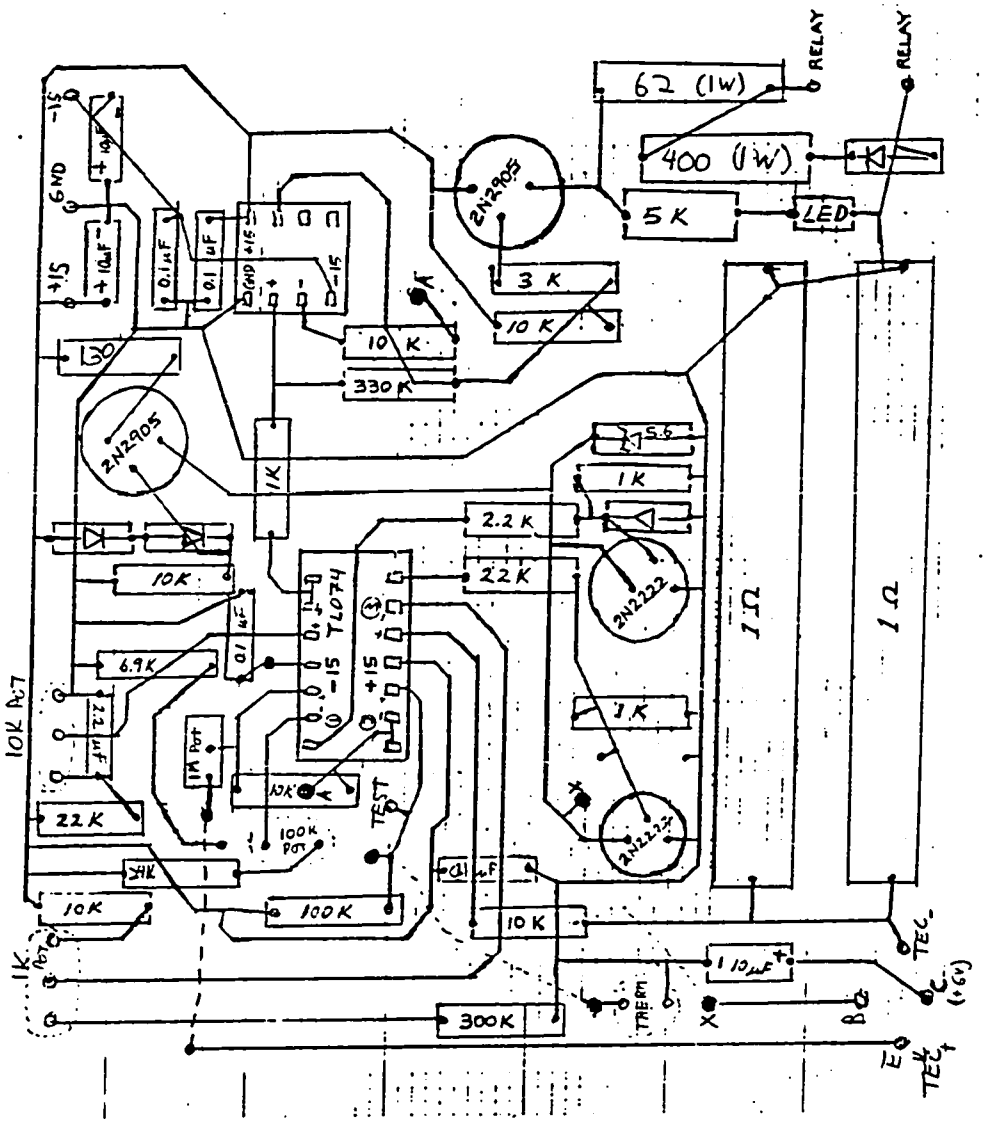
Both the C30902S and C30921S have been selected to have a low dark count rate. Cooling to -25° C can reduce this by a factor of 50, since the dependence of dark-count rate on temperature is exponential.

The Dark-Count increases with voltage following the same curve as the Photoelectron Detection Probability until a voltage where after pulsing is responsible for a feedback mechanism which dramatically increases the dark-count rate. This maximum voltage is circuit dependent, and is not warranted other than the values listed on page 3. In most cases, with a delay time of 300 ns, the diode can be used effectively at V_R up to V_{BR} + 25 V.

The C30902S and C30921S should not be forward biased or, when unbiased, exposed to strong illumination. These conditions result in a greatly enchanced dark-count which requires up to 24 hours to return to its nominal value.

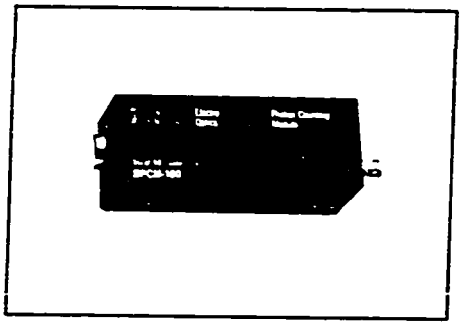

MJE 6044
T.E.C.
(~0.5Ω)
0.5Ω
+15
130
2N2905
1N752A (5.6 V)
2N2222
2N2222
1K
10K
2K
2K
1M
(~50K)
10K
10K
+15
25K
100K
7K
10K
1K
300
100K
+15
THERMISTOR
(kΩ → 150K
for 25°C → -25°C)
Op Amps 1-4 : TLC274


SIGNAL OUT
NORMALLY OPEN
MERCURY-WETTED
REED RELAY
(5V, 3652)
BNC
PASSIVE QUENCH
$-V_R$
200K
51
2905
62 (1w)
APD
400Ω (1w)
5K
3K
LED
10K
340K
LM311
10K
1K
OUTPUT
OF #2
OP AMP
+15
20K
10K
1μF


62 (1W)
RELAY
RELAY
400 (1W)
5K
LED
2N2905
3 K
10 K
10 K
330 K
2N2905
1K
5.6
1K
2.2 K
22K
2N2222
10K
6.9K
TL074
1 Ω
1 Ω
1K
2N2222
22K
1M Pot
10K
100K pot
TEST
100K
10 K
1 10μF
TEC
300K
TREM
10K Adj
1K Adj
22K
0.1μF
0.1μF
10μF
+15
GND
-15

Electro Optics

Single Photon Counting Module
SPCM-100-PQ
DATA SHEET



SPCM-100-PQ - General Purpose
SPCM-100-PQ-FC - *FC Fiber Optic Receptacle*

- **Peak Photon Detection Efficiency:**
  >40% at 633 nm
  >25% at 830 nm
- **Active Area: 150 microns Diameter for Good Sensitivity; 100 microns Diameter for Peak Photon Detection Efficiency**
- **Low Voltage Inputs: +12, +5, -5**
- **Digital Output Pulses**
- **VERY EASY TO USE**

The SPCM-100-PQ is a self-contained module which detects single photons of light over the wavelength range from 400 nm to 1060 nm ... a range and sensitivity which often outperforms photomultiplier tubes.

Applications include Lidar, Photon Correlation Spectroscopy, Astonomical Observations, Optical Rangefinding, Optical Fiber Test and Fault Location, Ultrasensitive Spectroscopy etc., where the light being measured can be focused to a small spot.

The module utilizes a unique silicon avalanche photodiode which has a circular active area whose peak photon detection efficiency, within 50 microns of the center, exceeds 40% at 633 nm. At larger distances from the center, the response decreases, so that at 75 microns from the center the photon detection efficiency is only 20% (see figure 2).

The module has a so-called "passive quench" circuit, which can count at speeds up to 1,000,000 counts per second. There is a "dead time" of 200 ns between pulses. The avalanche photodiode sensing element is temperature controlled, ensuring stablized performance despite changes in the ambient temperature.

Single photon arrival times can be measured with an accuracy of 3 ns rms.

The SPCM-100-PQ requires +12, +5, -5 volt power supplies (a mating cable is provided with each module). A digital pulse, 2 volts high and approximately 200 ns wide, is output at the rear BNC connector as each photon is detected. The case temperature should be kept between 5 and 40° C during operation.

Electrical Characteristics	Minimum	Typical	Maximum	Units
Supply Voltages Required:				
+12 @ .08 A max	11.9	12.0	12.1	V
+5 @ 1.3 A max	4.9	5.0	5.1	V
-5 @ .02 A max	-4.9	-5.0	-5.1	V
Supply Voltage Ripple (all supplies)	—	—	50	mV rms
Case Operating Temperature	5	—	40	°C
Humidity (Tests in Progress)	—	—	70 (estimate)	% R.H
Photon Detection Efficiency (P_d) Within 50 microns of Center:				
λ = 1060 nm	—	0.8	—	%
λ = 830 nm	25	32	—	%
λ = 633 nm	40	43	—	%
λ = 400 nm	—	20	—	%
Photon Detection Efficiency Variation at Constant Case Temperature (12h)	—	± 1	± 3	% Relative
Photon Detection Efficiency Variation from 5° to 40°C compared with 25°C Case temperature	—	± 5	± 10	% Relative
Dark Count	—	500	1000	counts/s
Dark Count Variation at Constant Case Temperature (12h, no illumination)	—	± 2	± 10	% Relative
Dark Count Variation from 5° to 40°C compared with 25°C Case Temperature	—	± 5	± 15	% Relative
Dead Time	—	200	—	ns
Output Count Rate Before Saturation	1.3×10^6	1.8×10^6	—	counts/s
Single Photon Timing Resolution (rms)	—	3	—	ns
Settling Time Following Power Up	—	20	100	s
Threshold Setting Required for Digital Output Pulse (Terminate in 50 ohms)	0.5	1.0	1.5	V

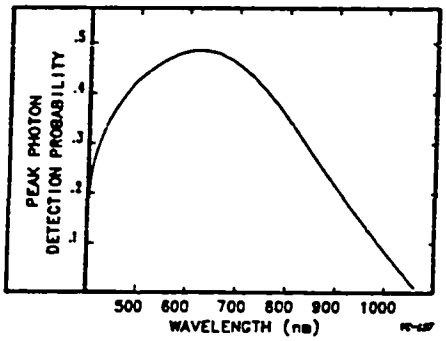


Figure 1 — Typical Peak Photon Detection Efficiency vs Wavelength

Options

The SPCM-100-PQ-FC has an "FC" fiber optic receptacle prealigned to the optical detector for ease in measuring the signals from optical fibers. Optical fibers with an FC connector on one end are available. Consult Customer Service at the factory for any fibers not listed in Table 2.

Safety

The SPCM-100-PQ contains a high voltage power supply. All internal settings have been preset, and there are no user adjustments.

Units which appear defective or have suffered mechanical damage should not be used because of possible electrical shorting of the high voltage power supply.

Warranty

Any RCA warranty is null and void if the module case is opened.

Light Emission During Photon Detection

One peculiarity of the avalanche photodiode sensing element is that as an incoming photon is detected, a small amount of light is emitted by the avalanche photodiode itself. In most cases this is not a problem. However it can cause some confusion if another detector is monitoring light, or if the optical system is such that light emitted from the SPCM-100-PQ is reflected back on itself. (An example of the latter case would be where a long optical fiber is used to bring light to the module. Some of the photons from the avalanche photodiode could be captured by the fiber, guided down its length, reflected from the far end of the fiber, and travel back to the module. If these photons return more than 200 ns after the initial event, then they will be detected.)

The light emitted from the module has a broad spectral distribution. With reasonable optical design, the light emission during photon detection is rarely a problem.

Saturation

At higher incoming light levels the count rate actually decreases. As an extreme example, if the module is exposed to room light, the count rate will fall to zero (if this occurs for a short time, there is no deterioration in the performance of the SPCM-100-PQ). Consequently, in certain applications, some test should be performed by the operator to ensure that a low count rate is not caused by detector saturation. One simple test is to deliberately increase the photon flux, eg, turn on an extremely weak light source during the experiment, and thereby verify that the count rate goes up, rather than down.

Table 2 — Standard fiber pigtail options. Standard length is 1 meter.

SPCM-QCX	FIBER TYPE	FIBER MANUFACTURER	DIAMETERS			NUMERICAL APERTURE	CUT-OFF WAVELENGTH
			CORE	CLADDING	OUTER		
-QC1	SINGLE MODE	YORK SM600	3.5 μm	125 μm	3 mm	0.13	590 nm
-QC2	SINGLE MODE	YORK SM750	4.5 μm	125 μm	3 mm	0.13	635 nm
-QC3	MULTIMODE	CANSTAR	50 μm	125 μm	3 mm	0.20	Not Applicable
-QC4	MULTIMODE	CANSTAR	62.5 μm	125 μm	3 mm	0.27	Not Applicable
-QC5	MULTIMODE	CANSTAR	85 μm	125 μm	3 mm	0.26	Not Applicable
-QC6	MULTIMODE	CANSTAR	100 μm	140 μm	3 mm	0.29	Not Applicable

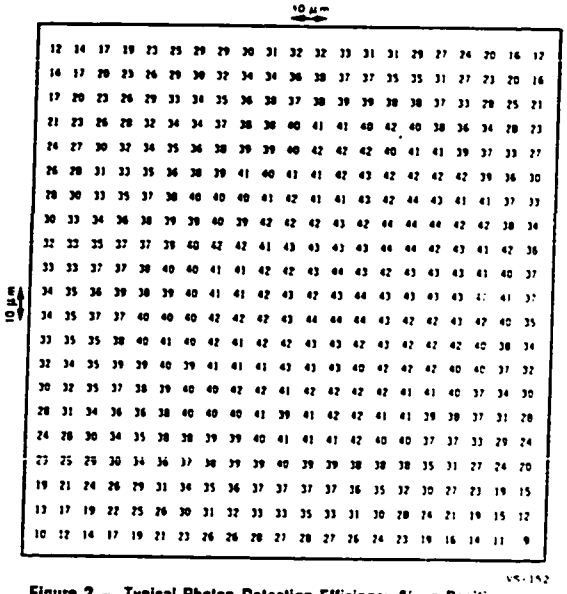


Figure 2 — Typical Photon Detection Efficiency % vs Position. The device was measured with a spot 10 microns in diameter, wavelength λ = 633 nm

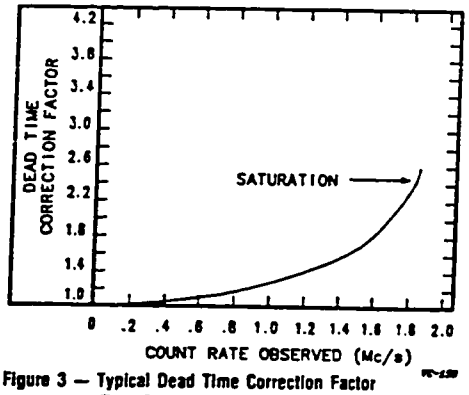


Figure 3 — Typical Dead Time Correction Factor (True Counts/Observed Counts) vs Observed Count Rate.

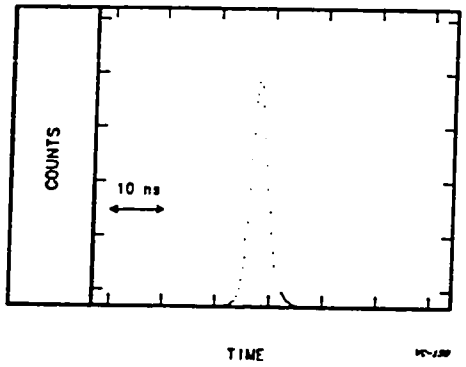


Figure 4 — Typical Histogram of Single Photon Timing Resolution (Spot Focussed at Center).

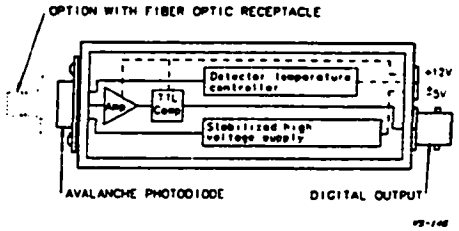


Figure 5 — Block Diagram of Module

LEMO SIZE 1, 4 PIN CONNECTOR FOR SUPPLY VOLTAGE INPUT

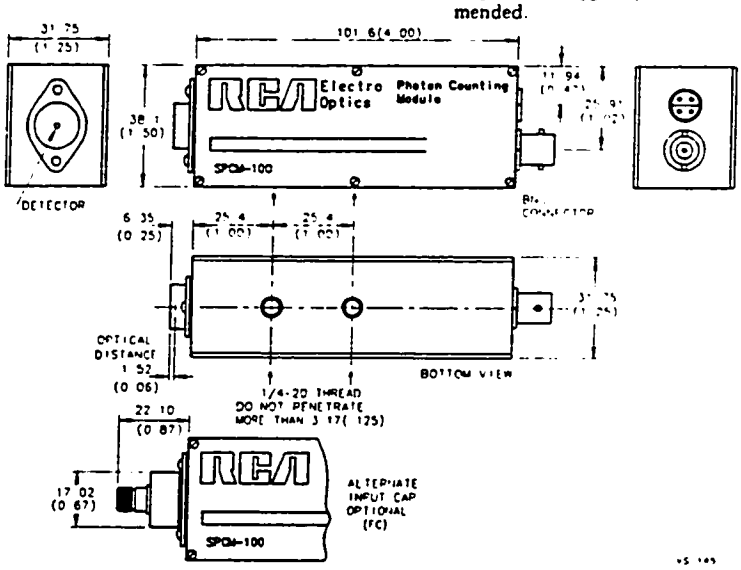
Case Pin	Supply	Wire*
1	+12V	WHITE
2	+5V	RED
3	GND	GREEN
4	−5V	BLACK
SHIELD	GND	BRAID

*MATING CABLE PROVIDED

STANDARD BNC CONNECTOR FOR DIGITAL OUTPUT PULSE

Figure 6 — Electrical Connections.

The shield must be connected to the ground wire at the power supply end. This has been done in the 1 meter cable supplied. The digital output pulse has a minimum height of 2.0 volts, and it should be terminated with a 50 ohm load to avoid distortion and ringing. A triggering level of 1.0 volts is recommended.



Dimensions in millimeters. Dimensions in parentheses are in inches.

Figure 7 — Mechanical Drawing. Care should be taken not to warp the side walls of the module during mounting since electronic components are attached to them. The module is rated for a case temperature of 5° to 40°C, so it may be necessary to ensure good heatsinking or ventilation. In a room temperature laboratory environment (not a closed equipment box) natural convection is normally sufficient.

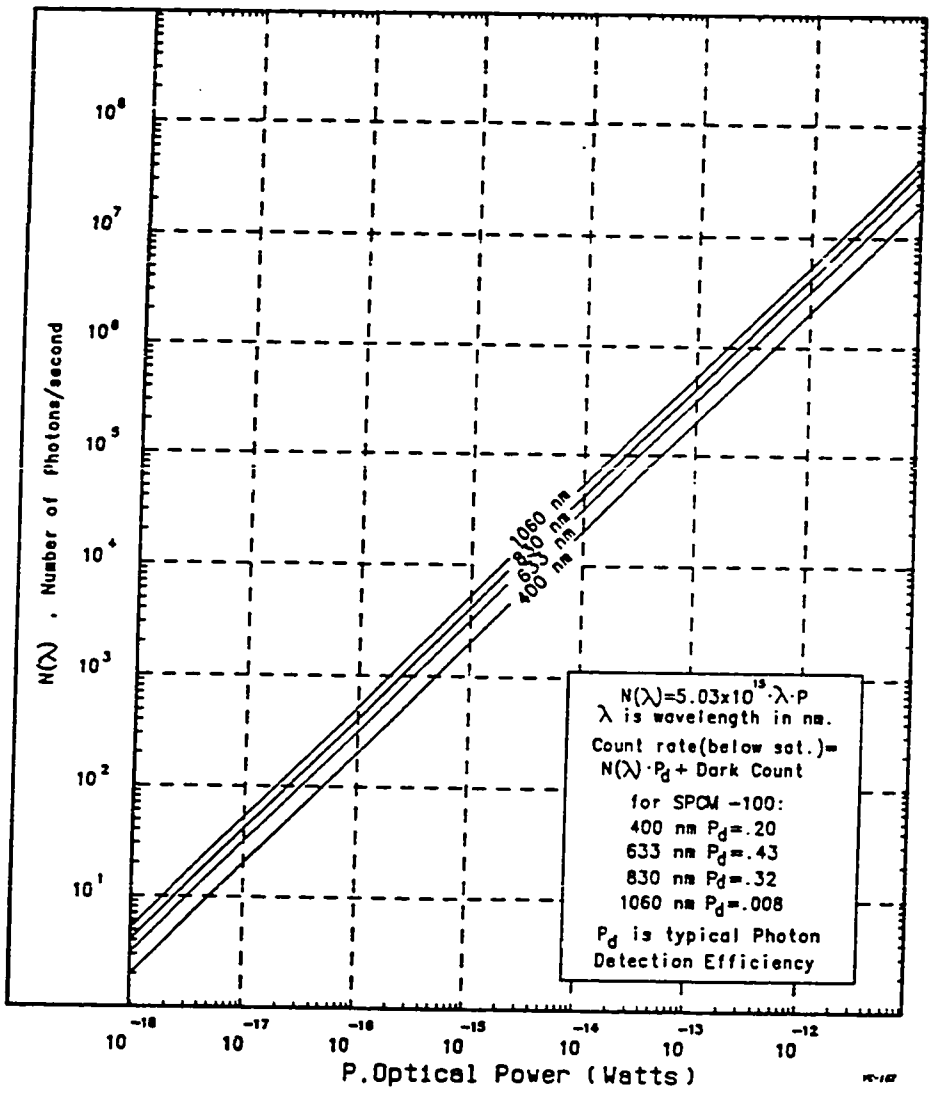


Figure 8 — Number of Photons / second vs Optical Power at various wavelengths.

For further information, please contact your local GE Canada Electro Optics representative or
General Electric Canada Inc., Electro Optics, P.O. Box 900, Vaudreuil, Canada J7V 7X3
Tel.: (514) 424-3300 • Fax: (514) 424-3411 • Tel. USA: 717-653-9122 • Tel. Europe: (011-33) 1-30547021

General Electric Canada Inc., Electro Optics formerly RCA Inc., Electro Optics

GENERAL ELECTRIC CANADA INC., Electro Optics Printed in Canada

FD-0043/12 89

REFERENCES

Y. Aharonov and J. Anandan, Phys. Rev. Lett. **58**, 1593 (1987).

A. Aspect, J. Dalibard and G. Roger, Phys. Rev. Lett. **49**, 1804 (1982).

A. Aspect, P. Grangier and G. Roger, Phys. Rev. Lett. **47**, 460 (1981).

A. Aspect, P. Grangier and G. Roger, Phys. Rev. Lett. **49**, 91 (1982).

S. M. Barnett and S. J. D. Phoenix, submitted to Phys. Rev. Lett. (1993).

J. S. Bell, Physics **1**, 195 (1964).

J. S. Bell, Ann. New York Acad. Sci. **480**, 263 (1986).

J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge Univ. Press, Cambridge, 1987).

C. H. Bennett, G. Brassard and N. D. Mermin, Phys. Rev. Lett. **68**, 557 (1992).

C. H. Bennett, Phys. Rev. Lett. **68**, 3121 (1992).

M. V. Berry, Proc. Roy. Soc. (London) **A392**, 45 (1984).

R. Bhandari and J. Samuel, Phys. Rev. Lett. **60**, 1210 (1988).

D. Bohm, Phys.Rev. **85**, 166 (1952).

D. Bohm and B. J. Hiley, Found. Phys. **14**, 255 (1984).

D. Bohm, B. J. Hiley and P. N. Kaloyerou, Phys. Rep. **144**, 323 (1987).

N. Bohr, in *Essays 1958-1962 on Atomic Physics and Human Knowledge*, (Wiley, New York, 1963), p. 60.

G. Brassard, *Modern Cryptology: A Tutorial*, in *Lecture Notes in Computer Science*, Vol. 325, edited by G. Goos and J. Hartmanis (Springer-Verlag, New York, 1988).

S. L. Braunstein and A. Mann, Phys. Rev. A **47**, R2427 (1993).

J. Brendel, E. Mohler and W. Martienssen, Phys. Rev. Lett. **66**, 1142 (1991).

J. Brendel, E. Mohler and W. Martienssen, Europhys. Lett. **20**, 575 (1992).

D. C. Burnham and D. L. Weinberg, Phys. Rev. Lett. **25**, 84 (1970).

Y. Q. Cai, G. Papini, W. R. Wood and S. R. Valluri, Quantum Opt. **1**, 49 (1989).

R. Campos, B. Saleh and M. Teich, Phys. Rev. **A 42**, 4127 (1990).

C. Chen, Laser Focus World 129 (November 1989).

R. Y. Chiao, in *Analogies in Optics and Microelectronics*, edited by W.V. Haeringen

and D. Lenstra, (Kluwer Academic Publishers, Dordrecht, the Netherlands, 1990), p. 151.

R. Y. Chiao and T. F. Jordan, Phys. Lett. **A132**, 77 (1988).

R. Y. Chiao, P. G. Kwiat and A. M. Steinberg, Physica **B175**, 257 (1991).

R. Y. Chiao, P. G. Kwiat and A. M. Steinberg, in *Workshop on Squeezed States and Uncertainty Relations*, edited by D. Han, Y.S. Kim and W.W. Zachary NASA Conference Publication 3135, 1991), p. 61.

R. Y. Chiao, P. G. Kwiat, A. M. Steinberg and I. H. Deutsch, in *Recent Developments in Quantum Optics*, edited by R. Inguva, (Plenum Press, New York, 1993), p. 145.

R. Y. Chiao, P. G. Kwiat, W. A. Vareka, C. K. Hong and H. Nathel, in *Coherence and Quantum Optics VI*, edited by J.H. Eberly, L. Mandel and E. Wolf, (Plenum Press, New York, 1990), p. 155.

R. Y. Chiao and A. Tomita, Phys. Rev. Lett. **59**, 1789 (1987).

R. Y. Chiao and Y. S. Wu, Phys. Rev. Lett. **57**, 933 (1986).

T. H. Chyba, L. J. Wang, L. Mandel and R. Simon, Appl. Opt. **13**, 562 (1988).

J. F. Clauser, Phys. Rev. D **9**, 853 (1974).

J. F. Clauser, R. A. Holt, M. A. Horne and A. Shimony, Phys. Rev. Lett. **23**, 880 (1969).

J. F. Clauser and M. A. Horne, Phys. Rev. D **10**, 526 (1974).

J. F. Clauser and A. Shimony, Rep. Prog. Phys. **41**, 1881 (1978).

H. Z. Cummins and E. R. Pike, *Photon Correlation and Light Beating Spectroscopy* (Plenum, London, 1974).

H. Z. Cummins and E. R. Pike, *Photon Correlation Spectroscopy and Velocimetry* (Plenum, New York, 1977).

L. M. Davis, Phys. Lett. A **140**, 275 (1989).

L. deBroglie, *Non-Linear Wave Mechanics* (Elsevier, Amsterdam, 1960).

I. Deutsch, Ph. D. Thesis, University of California at Berkeley (1992).

B. S. DeWitt and N. Graham, eds., *The Many-Worlds Interpretation of Quantum Mechanics*, (Princeton Univ. Press, Princeton, 1973).

V. G. Dmitriev, G. G. Gurzadyan and D. N. Nikogosyan, *Handbook of Nonlinear*

Optical Crystals, in *Springer Series in Optical Sciences*, Vol. 64, edited by A.E. Siegman (Springer-Verlag, New York, 1991).

P. H. Eberhard, Phys. Rev. A **47**, R747 (1993).

D. Eimerl, L. Davis, S. Velsko, E. K. Graham and A. Zalkin, J. Appl. Phys. **62**, 1968 (1987).

A. Einstein, B. Podolsky and N. Rosen, Phys. Rev. **47**, 777 (1935).

A. K. Ekert, Phys. Rev. Lett. **67**, 661 (1991).

A. K. Ekert, J. G. Rarity, P. R. Tapster and G. M. Palma, Phys. Rev. Lett. **69**, 1293 (1992).

B. -G. Englert, J. Schwinger and M. O. Scully, Found. of Phys. **18**, 1045 (1988).

B. -G. Englert, H. Walther and M. O. Scully, Appl. Phys. B **54**, 366 (1992).

H. Everett, Rev. Mod. Phys. **29**, 454 (1957).

R. Feynman, R. Leighton and M. Sands, *The Feynman Lectures on Physics* (Addison Wesley, Reading, 1965).

J. D. Franson, Phys. Rev. D **31**, 2529 (1985).

J. D. Franson, Phys. Rev. Lett. **62**, 2205 (1989).

J. D. Franson, Phys. Rev. Lett. **67**, 290 (1991).

J. D. Franson, submitted to Phys. Rev. A (1993).

E. S. Fry and S. Li, in *Second Santa Fe Meeting on the Foundations of Quantum Mechanics*, edited by T.D. Black, *et al.* (World Scientific, Singapore, 1991), p. 175.

A. Garuccio and V. A. Rapisarda, Nuovo Cim. **65A**, 269 (1981).

R. J. Glauber, Phys. Rev. **130**, 2529 (1963).

P. Grangier, G. Roger and A. Aspect, Europhys. Lett. **1**, 173 (1986).

D. M. Greenberger, M. A. Horne and A. Zeilinger, in *Bell's Theorem, Quantum Theory and Conceptions of the Universe*, edited by M. Kafatos, (Kluwer, Dordrecht, 1989), p. 73.

F. D. M. Haldane, Opt. Lett. **11**, 730 (1986).

F. D. M. Haldane, Phys. Rev. Lett. **59**, 1788 (1987).

L. Hardy, Phys. Rev. Lett. **68**, 2981 (1992a).

L. Hardy, Phys. Lett. A **161**, 326 (1992b).

L. Hardy, submitted to Phys. Rev. Lett. (1993).

M. Hillery and M. O. Scully, in *Quantum Optics, Experimental Gravitation, and Measurement Theory*, edited by P. Meystre and M.O. Scully, (Plenum Press, New York, 1983), p. 65.

C. K. Hong, Ph. D. Thesis, University of Rochester (1988).

C. K. Hong, S. Friberg and L. Mandel, Appl. Opt. **24**, 3877 (1985).

C. K. Hong and L. Mandel, Phys. Rev. A **31**, 2409 (1985).

C. K. Hong and L. Mandel, Phys. Rev. Lett. **56**, 58 (1986).

C. K. Hong, Z. Y. Ou and L. Mandel, Phys. Rev. Lett. **59**, 2044 (1987).

M. A. Horne, A. Shimony and A. Zeilinger, Phys. Rev. Lett. **62**, 2209 (1989).

J. Jarrett, Nôus **18**, 569 (1984).

H. Jiao, S. R. Wilkinson, R. Y. Chiao and H. Nathel, Phys. Rev. A **39**, 3475 (1989).

M. Kitano and T. Yabuzaki, Phys. Lett. A **142**, 321 (1989).

D. N. Klyshko, Sov. J. Quantum Electron. **10**, 1112 (1980).

D. N. Klyshko, Soviet Physics-Uspekhi **32**, 555 (1989).

M. Kugler and S. Shtrikman, Phys. Rev. D **37**, 934 (1988).

P. G. Kwiat and R. Y. Chiao, Phys. Rev. Lett. **66**, 588 (1991).

P. G. Kwiat, A. M. Steinberg and R. Y. Chiao, in *Second Santa Fe Meeting on the Foundations of Quantum Mechanics*, edited by T.D. Black, *et al.* (World Scientific, Singapore, 1991), p. 193.

P. G. Kwiat, A. M. Steinberg and R. Y. Chiao, Phys. Rev. **A45**, 7729 (1992).

P. G. Kwiat, A. M. Steinberg and R. Y. Chiao, Phys. Rev. **A 47**, R2472 (1993a).

P. G. Kwiat, A. M. Steinberg, R. Y. Chiao, P. H. Eberhard and M. D. Petroff, to appear in Appl. Opt. (1993b).

P. G. Kwiat, A. M. Steinberg, R. Y. Chiao, P. Eberhard and M. Petroff, Phys. Rev. A **48**, R867 (1993c).

P. G. Kwiat, A. M. Steinberg and R. Y. Chiao, to appear in Phys. Rev. A (1993d).

P. G. Kwiat, W. A. Vareka, C. K. Hong, H. Nathel and R. Y. Chiao, Phys. Rev. **A41**, 2910 (1990).

C. T. Lee, Phys. Rev. A **41**, 1721 (1990).

L. Q. Li, L. M. Davis, S. I. Soltesz and C. J. Trottier, in *1992 OSA Annual Meeting*, 1992).

A. W. Lightstone, A. D. MacGregor, D. E. MacSween, R. J. McIntyre, C. Trottier and P. P. Webb, Electron. Eng. **61**, 37 (1989).

A. W. Lightstone and R. J. McIntyre, in *Photon Correlation Techniques and Applications*, (OSA, Washington, D.C., May 31 - June 2, 1988), p. 183.

L. Mandel, J. Opt. Soc. Am. B **1**, 108 (1984).

L. Mandel, Opt. Lett. **16**, 1882 (1991).

R. J. McIntyre, IEEE Trans. Electron Devices **20**, 637 (1973).

N. D. Mermin, Ann. New York Acad. Sci. **480**, 422 (1986).

N. D. Mermin, Phys. Rev. Lett. **65**, 1838 (1990).

W. J. Munro and M. D. Reid, Phys. Rev. A **47**, 4412 (1993).

A. Niemi and G. Semenoff, Phys. Rev. Lett. **55**, 927 (1985).

Z. Y. Ou and L. Mandel, Phys. Rev. Lett. **61**, 50 (1988).

Z. Y. Ou, L. J. Wang, X. Y. Zou and L. Mandel, Phys. Rev. A **41**, 566 (1990a).

Z. Y. Ou, X. Y. Zou, L. J. Wang and L. Mandel, Phys. Rev. Lett. **65**, 321 (1990b).

S. Pancharatnam, Proc. Ind. Acad. Sci. **A44**, 247 (1956).

A. Peres, Phys. Rev. **D 22**, 879 (1980).

A. Peres, Ann. New York Acad. Sci. **480**, 438 (1986).

M. D. Petroff, M. G. Stapelbroek and W. A. Kleinhans, Appl. Phys. Lett. **51**, 406 (1987).

J. G. Rarity, K. D. Ridley and P. R. Tapster, Appl. Opt. **26**, 4616 (1987).

J. G. Rarity and P. R. Tapster, Phys. Rev. Lett. **64**, 2495 (1990).

J. G. Rarity and P. R. Tapster, Phys. Rev **A45**, 2052 (1992).

J. G. Rarity, P. R. Tapster, J. A. Levenson, J. C. Garreau, I. Abram, J. Mertz, T. Debuisschert, A. Heidmann, C. Fabre, *et al.*, Appl. Phys. B **55**, 250 (1992).

M. G. Raymer and S. Yang, *Information and complementarity in a proposed which-path experiment using photons* (1991).

M. L. G. Redhead, Ann. New York Acad. Sci. **480**, 14 (1986).

G. Ripamonti, M. Ghioni and A. Lacaita, IEEE J. Lightwave Tech. **8**, 1278 (1990).

J. F. Roch, G. Roger, P. Grangier, J. Courty and S. Reynaud, Appl. Phys. B **55**, 291 (1992), and references therein.

J. N. Ross, Opt. Quant. Elect. **16**, 445 (1984).

M. H. Rubin and Y. H. Shih, Phys. Rev. A **45**, 8138 (1992).

E. Santos, Phys. Rev. Lett. **66**, 1388 (1991).

E. Santos, Phys. Rev. A **46**, 3646 (1992).

J. Schwinger, M. O. Scully and B.-G. Englert, Z. Phys. D. **10**, 135 (1988).

M. O. Scully, B.-G. Englert and H. Walther, Nature **351**, 111 (1991).

M. O. Scully, B.-G. Englert and H. Walther, Nature **351**, 111 (1991).

M. O. Scully, R. Shea and J. D. McCullen, Phys. Rep. **13**, 485 (1978).

M. O. Scully and H. Walther, Phys. Rev. A **39**, 5229 (1989).

J. Segert, Phys. Rev. A **36**, 10 (1987).

F. Selleri, *Quantum Mechanics Versus Local Realism: The Einstein-Podolsky-Rosen Paradox*, in *Physics of Atoms and Molecules*, edited by P.G. Burke and H. Kleinpoppen (Plenum, New York, 1988).

Y. R. Shen, *Nonlinear Optics* (Wiley, New York, 1984).

Y. H. Shih and C. O. Alley, Phys. Rev. Lett. **61**, 2921 (1988).

Y. H. Shih, A. V. Sergienko and M. H. Rubin, Phys. Rev. A **47**, 1288 (1993).

A. Shimony, *An Exposition of Bell's Theorem* (Plenum Press, New York, 1990).

R. Simon, H. J. Kimble and E. C. G. Sudarshan, Phys. Rev. Lett. **61**, 19 (1988).

H. P. Stapp, Phys. Rev. D **3**, 1303 (1971).

H. P. Stapp, Ann. New York Acad. Sci. **480**, 326 (1986).

H. P. Stapp, Found. of Phys. **21**, 1 (1991).

A. M. Steinberg, P. G. Kwiat and R. Y. Chiao, Phys. Rev. **A45**, 6659 (1992).

A. M. Steinberg, P. G. Kwiat and R. Y. Chiao, Phys. Rev. Lett. **68**, 2421 (1992).

A. M. Steinberg, P. G. Kwiat and R. Y. Chiao, Phys. Rev. Lett. **71**, 708 (1993).

A. M. Steinberg, P. G. Kwiat and R. Y. Chiao, in *XXVIIIe Rencontre de Moriond*, edited by J.T. Van (Editions Frontières, France, 1993).

A. Stern, Y. Aharonov and Y. Imry, Phys. Rev. **A41**, 3436 (1990).

C. Su and K. Wódkiewicz, Phys. Rev. **A44**, 6097 (1990).

S. M. Tan and D. F. Walls, Phys. Rev. A **47**, 4663 (1993).

A. Tomita and R. Y. Chiao, Phys. Rev. Lett. **57**, 937 (1986).

W. R. Tompkin, M. S. Malcuit, R. W. Boyd and R. Y. Chiao, J. Opt. Soc. Am. B **7**, 230 (1990).

J. von Neumann, in *Quantum Theory and Measurement*, edited by J.A. Wheeler and W.H. Zurek, (Princeton University Press, 1983), p. 549.

J. A. Wheeler, in *Problems in the Formulation of Physics*, edited by G.T. diFrancia, (North-Holland, Amsterdam, 1979).

J. A. Wheeler, in *Quantum Theory and Measurement*, edited by J.A. Wheeler and W.H. Zurek, (Princeton Univ. Press, Princeton, 1983), p. 182.

J. A. Wheeler and W. H. Zurek, *Quantum Theory and Measurement*, in (Princeton Univ. Press, Princeton, 1983).

E. P. Wigner, Phys. Rev. **40**, 749 (1932).

K. Wódkiewicz, Phys. Lett. **A129**, 1 (1988).

W. K. Wootters and W. H. Zurek, Nature **299**, 802 (1982).

A. Ya'sin and D. M. Greenberger, Ann. New York Acad. Sci. **480**, 622 (1986).

A. Yariv, *Quantum Electronics* (Wiley, New York, 1988).

A. Zajonc, Phys. Lett. **96A**, 61 (1983).

A. G. Zajonc, L. J. Wang, X. Y. Zou and L. Mandel, Nature **353**, 507 (1991).

A. Zeilinger, Phys. Lett. A **118**, 1 (1986).

X. Y. Zou, L. J. Wang and L. Mandel, Phys. Rev. Lett. **67**, 318 (1991).

W. H. Zurek, Ann. New York Acad. Sci. **480**, 89 (1986).

W. H. Zurek, Physics Today **44**, 36 (1991).