ABSTRACT

Title of dissertation: NEW TECHNOLOGIES FOR BROADBAND QUANTUM KEY DISTRIBUTION: SOURCES, DETECTORS, AND SYSTEMS

Daniel J. Rogers, Doctor of Philosophy, 2008

Dissertation directed by: Dr. Julius Goldhar
Chemical Physics Program,
Institute for Physical Science and Technology

In this thesis I describe three independent projects that advance the development of broadband quantum cryptography. While each project pertains to a different part of the QKD chain, together they provide key developments in implementing QKD at bit rates that are practical for use in the modern telecommunications infrastructure.

The first project comprises the bulk of the thesis and involves developing a novel source of correlated photon pairs for use in free-space QKD. This source is based on a birefringent semiconductor optical waveguide as a Kerr medium. We demonstrate the feasibility of using birefringent phase-matched four-wave mixing to

generate correlated photon pairs. We further propose that, by reversing the process and pumping with conjugate wavelengths, one can use the same effect to produce entangled photon pairs with the same device. These pairs can then be used for QKD to realize the most secure and efficient quantum cryptographic data links.

The second project examines the implications of operating a BB84 QKD protocol at clock rates that are faster than the recovery time of the constituent detectors. We show that operating such systems under conventional protocols results in a security violation that allows an eavesdropper to learn significant information about the key and present a modification to the BB84 protocol that maintains key security at fast transmission rates. This modification to the protocol will become vital to QKD viability as links become faster and clock rates go into the tens of gigahertz. We also demonstrate, rather counterintuitively, that there exists an optimal transmission rate for a QKD system that exceeds the inverse of an individual detector's dead time.

The final project describes a new design for a free-space QKD link that centers around faster silicon detectors. These detectors have a peak quantum efficiency in the visible range, requiring that the system operate at a wavelength that is more susceptible to solar interference. To mitigate this effect, the link is designed around a Fraunhofer line in the solar spectrum where the background solar light levels are reduced by up to 90%. By implementing this system, we expect at least a two-fold increase in the secret key rate, coming ever closer to the goal of a 10 Mb/s QKD system compatible with first-generation ethernet technology.

NEW TECHNOLOGIES FOR BROADBAND QUANTUM KEY DISTRIBUTION: SOURCES, DETECTORS, AND SYSTEMS

by

Daniel J. Rogers

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2008

Advisory Committee:
Professor Julius Goldhar, Chair/Advisor
Dr. Charles W. Clark
Professor Chris Davis
Professor Michael Coplan
Professor Thomas Murphy

Dedication

To Mim, who somehow kept me smiling throughout this whole crazy ordeal and who will undoubtedly find every single spelling mistake in here.

And in memory of Nora, who I'm sure is reading this somewhere.

Acknowledgments

"One man alone can be pretty dumb sometimes, but for real bona fide stupidity, there ain't nothin' can beat teamwork." -Edward Abbey

There are so many people without whom this whole adventure of graduate school would not have been possible. First off, I must acknowledge the support and encouragement from my dynamic duo of advisors, Charles Clark and Julius Goldhar. They have pushed me to succeed while simultaneously holding me to the highest standards. I must also acknowledge the support of Michael Coplan at every step of the way, ensuring that I get the most out of my graduate career and indulging (even sometimes supporting) even my most whimsical choices as a student.

I should acknowledge the generosity of my sponsors through my graduate program, including the NIST/Chemical Physics Fellowship, the Department of Homeland Security Science and Technology Graduate Fellowship, and the Ann G. Wylie Dissertation Grant from the Graduate School, all of which allowed me to pursue the topics that were most interesting to me and focus entirely on my studies.

I also must thank those individuals who have generously helped me along the way, though they may have had no personal stake in my success. First off, I appreciate the generous assistance of Chris Richardson, Victor Yun, Yongzhang Leng, Joe DiPasquale, and Jon Sugrue at LPS. At NIST, I sincerely thank my partners in crime, Josh Bienfang, Alessandro Restelli, and Jemellie Galang. I also owe many thanks to Carl Williams, Alan Migdall, Jay Fan, Sergey Polyakov, Alex

Ling, Matt Eisaman, Aaron Pearlman, Xiao Tang, Hai Xu, Lijun Ma, Alan Mink, Tassos Nakassis, Barry Hershman, Gerry Fraser, Paul Lett, Trey Porto, and Carlos Lopez-Mariscal. At Maryland, I must acknowledge the help of Mario Dagenais, Chris Davis, and Edo Waks. I would also like to specifically mention all of the crucial contributions from Tom Murphy, whose class and personal interest in my work were critical to its success. In addition, I would like to thank Dr. Richard Gray of Appalachian State University for generously sharing his expertise and Dr. William Jeffrey for both his technical suggestions on the project and his personal advice.

Of course, none of this would have ever made it through the administrative straits without the tireless help of Helen Felrice and Debbie Jenkins. I suspect that it is truly they who keep the Earth spinning on a daily basis.

I must also mention the support of my graduate school comrades, Brooke, Mandy, Craig, Ching-Yee, Yi-Hsing, and Sophia, without whom I probably would have thrown in the towel early on.

Lastly, I must thank my family - Mom, Dad, Bethy, Jonny, Dan, Sarah, Ricky and Lucy. You have provided me with the inspiration and motivation to succeed even when faced with the toughest of challenges.

Table of Contents

List of Abbreviations

2PA	Two-Absorption
4WM	Four-Wave Mixing
APD	Avalanche Photo-Diode
BB84	Bennett and Brassard, 1984 QKD Protocol
B92	Bennett, 1992 QKD Protocol
BBS	Blum Blum Shub
CW	Continuous Wave
E91	Ekert 1991 Protocol
EA	Electro-Absorption
EC/PA	Error Correction and Privacy Amplification
EO	Electro-Optic
FEC	Forward Error Correction
FPGA	Field Programmable Gate Array
GEO	Geostationary Earth Orbit
GVD	Group Velocity Dispersion
LEO	Low Earth Orbit
MBE	Molecular Beam Epitaxy
MBPS	Megabits per second
MPD	Micro Photon Devices
PRNG	Pseudo-random number generator
QBER	Quantum Bit Error Rate
QKD	Quantum Key Distribution
RHEED	Reflection High Energy Electron Diffraction
RNG	Random Number Generator
SPAD	Single-Photon Avalanche Detector
SPCM	Single-Photon Counting Module
SSPD	Superconducting Single Photon Detector
TTL	Transistor-Transistor Logic
VCSEL	Vertical Cavity Surface Emitting Laser

Chapter 1

Introduction to Broadband Quantum Key Distribution

"The speed of communications is wondrous to behold. It is also true that speed can multiply the distribution of information that we know to be untrue." -Edward R. Murrow

1.1 QKD - Why?

The year is 47 BC. Four Roman Legions sit poised atop a hill overlooking the Turkish town of Zile. The generals await word from their brilliant commander, Julius Caesar, to end the coup led by Pharnaces, the well-known enemy of Rome. Suddenly, a messenger appears carrying a small parchment. At first glance, Caesar's orders look to be gibberish, but the generals quickly rewrite the note, replacing each letter with the one that occurs three before it in the alphabet. Suddenly the message becomes clear: *begin the attack*. They shout orders to their lieutenants and begin a five-day siege that lays waste to Pharnces' 20,000 warriors. Caesar, satisfied that he has crushed the coup, writes to Amantius in Rome, "I came, I saw, I conquered."[36]

Caesar's simple shift cipher that he used to communicate with his generals was surprisingly secure. Even in Caesar's time, people had yet to develop the art of

codebreaking, and most of Caesar's enemies were not even able to read Latin writing. Records indicate that it wasn't until the 9th century that Arab mathematicians developed the kind of cryptanalysis that could have compromised Caesar's messages [59]. However, once they did, the cat-and-mouse game between the codemakers and the codebreakers began.

In today's information-centric world, data encryption is ubiquitous. Anyone who buys a book online using their credit card understands the importance of data security. But we often take for granted that little padlock icon in the corner of our browser window. What does it mean, and will it always remain secure?

Most of today's encryption depends primarily on two steps: public key encryption to exchange a key between two parties, and a symmetric cipher that uses this key to encrypt the data. Both of these steps must be secure in order to maintain data integrity. Assuming the key is secret, we can say that the symmetric cipher step is relatively safe given modern technology. While it is by no means unbreakable, it does well against most publicly known computer hardware. The vulnerability comes when the key is stolen, allowing an eavesdropper to steal the symmetric ciphertext and decrypt it. Public-key encryption is based on the assumption that computers cannot efficiently factor a large prime number (or perform a logarithm modulo-n, which is an equivalent algorithmic problem). However, no one has yet proven that this is true. It is possible that we can wake up tomorrow morning and find that a clever mathematician or computer scientist somewhere has thought of an efficient way to do so, compromising all of our online financial transactions, not to mention various other sensitive government, military, and health care data. In fact, comput-

ers based on quantum mechanical bits are already being developed that have been shown to be able break public-key encryption [58].

So what can we do? Can we end this cat-and-mouse game that has been going on since the time of Caesar and keep our data secure beyond the first quantum computer? In some ways, we already have, and have done so since World War I. In 1917, a telephone engineer by the name of Gilbert Vernam devised a scheme to encrypt data using a string of random digits [68]. The fundamental idea is this: Consider your message to be a sequence of binary digits (as all messages can be thought of). Now write down a string of random digits (1's and 0's) that is exactly as long as the message itself; this will be the key. Now, bit by bit, perform an addition modulo-2 (also known as an exclusive OR, or XOR). The subsequent ciphertext now looks like a string of random digits, and without knowledge of the key, the encrypted message seems to contain no information. In fact, Claude Shannon later showed that this text does indeed contain zero information without the key and cannot be broken [54]. Of course, if you do know the key, you can simply subtract it (another bitwise XOR operation) and retrieve the original message. It is a truly unbreakable encryption scheme now known as the Vernam cipher or one-time pad.

So it seems that, by 1949, we had ended this cat-and-mouse game. But there remained a problem that has plagued the Vernam cipher since its invention. How does one transmit the key from the sender to the receiver without it being stolen? It turns out that this is just as difficult as sending the message itself; it is, after all, a string of digits exactly as long as original message. This problem essentially rendered the Vernam cipher useless until 1984, when Bennett and Brassard realized that one

can exploit the quirky principles of quantum mechanics to transmit a random key and guarantee that it remains secret for the duration of its trip[1] [6].

1.2 The Fundamentals of QKD

The fundamental procedure for what is commonly called BB84 (after Bennett and Brassard's 1984 paper) QKD goes as follows [6]: Suppose our sender (we'll call her Alice) wants to send a secret, binary message to her friend, whom we will call Bob. To do so, she would like to use a Vernam cipher, so before she encrypts her message, she will perform a joint quantum optics experiment with Bob in order to agree on a secret, random key. She begins by setting up a channel that can send single photons to Bob. These single photons will be sent in one of four quantum states, organized into two pairs of non-orthogonal bases, each containing two orthogonal states that represent the values 1 and 0, respectively. The simplest implementation uses polarization state encoding on the single photon. For example, the states $|H\rangle$, $|V\rangle$, $|+45\rangle$, and $|-45\rangle$ are commonly used, where $|H\rangle$ and $|+45\rangle$ represent a bit value of 0, while $|V\rangle$ and $|-45\rangle$ represent a bit value of 1. Alice then makes two random choices: First, she chooses a basis to use (either H-V or $\pm$ 45). She then chooses a random bit value and sends a single photon encoded in one of the four states corresponding to her random choice.

Bob now prepares to measure the incoming photon. Since we are using polarization state encoding, this involves choosing a basis in which to measure the

[1]The original inspiration for quantum key distribution came out of ideas on quantum coding expressed by Stephen Wiesner and published at the behest of Charles Bennett in [72]

photon's polarization (Bob and Alice agree beforehand which basis options they will choose between). Bob makes a completely random choice here. He will either measure the photon's polarization in the same basis in which Alice sent it and presumably obtain the same bit value that Alice sent, or he will measure it in the wrong basis and have a 50% probability of measuring the opposite bit value. Bob makes his measurement, and then, over an *open*, classical communication channel, he announces to Alice what basis he used in his measurement. If Alice responds (again over the open classical channel) that she used that same basis to transmit the photon, then Bob keeps his measured bit value and they repeat the process again. If they disagree on the bases used in transmission and measurement, Bob simply discards his measurement result and they repeat the process. This process is referred to as *sifting* and results in what is termed the *sifted key*.

Note that, even though they communicate basis choices on an open, classical channel, no information about the actual key bits is revealed thus far. Now suppose a nefarious eavesdropper (we will, of course, call her Eve), tries to intercept the photons on their way from Alice to Bob. The first line of defense against such an ignoble action is the simple fact that Alice only sent *one* photon. By definition, one photon cannot be split in half, so Eve would have to intercept the photon, measure it the same way Bob does, and send a *new* photon to Bob to replace the one she stole. In doing so, she must make the same random measurement choice that Bob does. This means that 50% of the time, she will make an incorrect basis choice and send Bob a photon in a different basis than Alice originally transmitted. 50% of *that* time, Bob will choose the *correct* basis, and finally, 50% of the final measurements

will have incorrect bit values. The final result is that there will be up to a 12.5% error rate per basis, resulting in an overall 25% maximum error rate in the sifted key in the presence of an eavesdropper.

The final step in the quantum key exchange occurs when Alice and Bob have an open conversation over the classical channel comparing their sifted key values. Under the most stringent assumption that all errors introduced in the sifted key are the result of an eavesdropper, Alice and Bob perform forward error correction to eliminate the discrepancies in their respective keys. While the details of FEC are irrelevant to this thesis (for a more in-depth discussion of error-correction codes, see [14], [45], or [28]), the important feature is that Alice and Bob can eliminate the portions of the sifted key that are in error *without* revealing any bits from the portions of the key that they keep. This process, known collectively as *error correction and privacy amplification* or *reconciliation*, results in a key that is guaranteed to be secure from eavesdroppers, also known as the *secret key*. This entire process is performed until the secret key is as long as the message itself. At that point, it can safely be used as the key for a one time pad encryption and secure transmission of the message over the open, classical channel.

It is apparent that this QKD concept seems to achieve the ultimate goal in data security - to end the cat-and-mouse game that has dogged cryptography for over two thousand years. However, there are a number of details that must be clarified when discussing the security of QKD. First of all, along with the postulate that a single quanta cannot be split by definition, the security proof depends on the *no-cloning theorem* of Wooters and Zurek[73]. This theorem states that an unknown quantum

state cannot be cloned without first measuring (and hence destroying) the original state. This result is critical to the security of QKD, as an unknown quantum state is precisely what Eve encounters in her attack on the QKD link. It also must be noted that QKD links are still susceptible to the so-called denial of service attacks. In fact, if the error rate increases beyond a certain point, the error-correcting codes can break down and reconciliation cannot occur [9]. Furthermore, when this system is implemented using real sources and detectors, further security threats come into play that must be dealt with if quantum cryptography is to be practical. These extensions will be discussed in Chapter 3.

It should also be noted that there are other schemes for implementing QKD. One is a variation on BB84, called B92 (after Bennett's 1992 paper describing it) that involves only two non-orthogonal quantum states [7]. While this protocol is simpler to implement in hardware, it does sacrifice efficiency. There is also a protocol devised by Artur Ekert in 1991 that uses the exotic properties of nonclassical light [17]. This protocol will be discussed further in Chapter 2.

1.3 Broadband QKD and its Limitations

While the underlying protocol behind quantum key distribution seems straightforward, actually implementing the process in a practical way proves to be a challenge that the community is still striving to achieve. Concerning practicality, I refer generally to creating an ultra-secure data link over out-of-earshot distances at speeds compatible with the modern telecommunications infrastructure. A good

speed benchmark for a typical QKD link might be a secret key rate in excess of 10 Mb/s in order to be compatible with first-generation ethernet hardware. A number of links, including ones at NIST, have begun to approach this secret key rate [9], [75], [63].

The commonly accepted distance benchmark is the range from the ground to a low-earth orbit (LEO) satellite. This metric derives from the concept of operations involving secure quantum key transfer between a ground station, a LEO satellite, and another ground station somewhere else on the globe. Assuming the satellite can store the key securely, this arrangement, along with its cousin involving multiple GEO satellites, is viewed as the most promising model for a global, secure, quantum communications network [48].

A number of advances have begun to demonstrate the feasibility of performing such quantum data transfers [69]. Most of these efforts have focused on link distance as the most obvious metric, including a recent demonstration of quantum key exchange over the distance of 144 km between two of the Canary Islands [53]. While the link range is rather impressive, the experiment in [53] only operated at a secret key rate of about 13 bits per second. Considering the limited time window in which a LEO satellite is overhead, this is not yet a practical demonstration. Fast key rates on the order of 10 Mb/s are required for any meaningful key exchange between a ground station and a LEO satellite [10], [50].

Hence our focus in the NIST QKD effort has been to achieve truly *broadband* quantum key distribution. We demonstrated in 2004 a system that exchanged secret key at $>$ 1 Mb/s, fast enough to transmit quantum-encrypted streaming video

over a distance of 700 m. To achieve this, we incorporated a number techniques from classical telecommunications engineering to more efficiently perform the steps involved in the BB84 protocol. For example, we clocked the transmitter at 1.25 GHz, using as our source high-speed, gain-switched VCSELs to create very short, weak laser pulses. We implemented all of the control and sifting processes on a dedicated FPGA and custom PCI card, and we optimized and multithreaded the FEC and reconciliation procedures to make them more efficient [9]. However, in building this link around these cutting-edge technologies, we encountered a fundamental speed limit in the system: Timing jitter in the single photon APDs.

Avalanche photodiodes biased to operate in Geiger mode, transmit an electrical pulse for each photon absorbed in the active semiconductor region (with some efficiency, of course). Where in the absorption region of the device the photon is absorbed creates fundamental uncertainty between the clocked photon transmission and the time at which the SPAD sends its TTL pulse. This uncertainty, known as *timing jitter* is about 550 ps at 3 dB for the most common commercial SPADs [2]. Because one cannot transmit photons faster than the inverse of the jitter, the transmission speed is limited. More detail on techniques to overcome this limitation will be discussed in Chapter 4.

This thesis addresses a number of hurdles associated with making faster QKD links. As mentioned above, Chapter 2 addresses sources of entanglement for QKD

[2]Because error rates can significantly increase when pulses are miscounted in adjacent clock bins, a better measure of the timing jitter for this application is the 1/100 point. For the commercial SPAD units, this can be as long as 2.5 ns [9].

schemes that exploit the often-counterintuitive properties of quantum states of light. QKD systems of this nature are fundamentally more secure than their counterparts that use attenuated lasers. Chapter 2 will delve into the reasoning behind this assertion and describe a project that examines a semiconductor source of correlated and ultimately entangled photons. This effort also provides insight into some fundamental questions in nonlinear optics that still remained open. Chapter 3 then examines what happens to the security of broadband QKD when implemented using real detectors. In addition to an intrinsic timing jitter as described above, SPADs also exhibit a recovery time that is often much longer than the transmission period. We discuss the security implications of operating in this regime, propose a modification to the BB84 protocol to mitigate these effects and provide results that show the existence of an optimal transmission rate in excess of the inverse of the detector's recovery time. Finally, in Chapter 4 we put forth a design for a new free-space QKD system at a wavelength that overlaps with a Fraunhofer line in the solar spectrum. Based on newly available SPADs with reduced timing jitter, this system promises to have an increased secret key rate of at least twice the previous demonstration, approaching the milestone of compatibility with first-generation ethernet connections.

Chapter 2

Birefringent phase-matched four wave mixing in a semiconductor waveguide for correlated photon generation

"God does not play dice with the universe; He plays an ineffable game of his own devising, which might be compared, from the perspective of any of the other players, to being involved in an obscure and complex version of poker in a pitch dark room, with blank cards, for infinite stakes, with a dealer who won't tell you the rules, and who smiles all the time." -Terry Pratchett

2.1 Sources of correlated and entangled photons

The BB84 QKD protocol described in Chapter 1 is only a semi-classical approximation of the absolute security that truly quantum communication can provide. Often called a *prepare and send* method for obvious reasons, it requires that Alice makes two initial random choices before even introducing any quanta into the scheme. Thus the cryptosystem, like nearly all others, is limited by the quality of Alice's random number generator. The NIST system, for example, relies on a Mersenne Twister [40] pseudo-random number generator to create Alice's random bits. The Mersenne Twister is computationally fast and is often used for Monte-

Carlo simulations, but from a security perspective it is considered unsuitable for cryptography because it only requires observation of 624 iterates to determine all of the parameters needed to predict its next output bit [40]. While the EC/PA process mitigates the problem to some degree, the PRNG is certainly the weakest link in the NIST QKD system. One obvious solution would be to use a more secure PRNG. However, the current standard in cryptographically secure random bits is the Blum Blum Shub algorithm [11]. Like all modern classical cryptography, BBS relies on the product of two large prime numbers to produce random bits. Because QKD is meant to protect against attackers equipped with potential fast factoring capabilities, it would be quite counterproductive to base the random bit choices on a PRNG susceptible to a factoring attack.

As mentioned briefly in Chapter 1, there is an alternative to the prepare and send QKD protocols. A protocol based on entanglement, first proposed by Artur Ekert in 1991[17] and called the E91 protocol, involves creating a pair of polarization-entangled photons in a Bell state and sending one photon of the pair to Alice and one to Bob. Alice and Bob then simultaneously measure their respective photons using the same procedure that Bob uses in the BB84 protocol. Since the photons are polarization entangled, the measurement outcomes are correlated and can be used to generate random keys in two different places. This method (and others based on entanglement of properties other than polarization, like time or frequency) has two distinct advantages over prepare and send methods. First, the randomness between the ones and zeros no longer depends on a quality PRNG. Instead, it is intrinsic to the physical state of the photon pair and is as truly random as possible in nature.

Second, the security of the information, as far as the laws of quantum mechanics are understood, is unequivocally guaranteed. That is, in prepare and send schemes, one must always worry about side channels of information. For example, information about which photon state Alice sent might be gleaned by analyzing the timing information, looking at RF leakage from the drive electronics, or monitoring some other overlooked detail of the system. However, because of the nature of quantum entanglement, one can guarantee that, if the photons violate the Bell inequality, there can be no information leaked through *any* side channel, even ones that are unknown to Alice and Bob (or anyone else, for that matter). For more details about this concept, see [3], [5], and [4].

While entangled photons offer the path to truly unbreakable encryption, they still present challenges, namely in ways to generate them quickly and efficiently. Currently the most ubiquitous method for generating entanglement involves type II parametric downconversion in a BBO crystal [32]. This method has been demonstrated as a source for E91 QKD up to the kb/s range [39]. However, this source is limited to these slow speeds and requires bulk optical crystals pumped with free-space beams. Thus there is a strong interest in a source of entanglement that is compact, integrated, and able to operate at GHz pair generation rates.

There are a number of approaches to integrated sources of entangled pairs. For example, many groups are investigating third order nonlinearities in silicon nanowires [22], [55]. In these devices, the sub-wavelength dimensions of the waveguide allow the modal dispersion to be tailored to achieve phase matching [65]. However, because they are so small, they can potentially have high linear loss and be

very difficult to couple to efficiently. Another approach uses four wave mixing in a microstructure optical fiber to generate correlated and entangled pairs [19], [46], [70]. In this scheme, the small mode confinement and long propagation length compensates for the relatively small third-order susceptibility of glass. Unlike Si nanowires, this source has an additional advantage of having very low linear loss. Other exotic schemes involve second order processes in highly birefringent III-V semiconductor waveguides [20], as well as a host of other approaches.

Each approach has its respective advantages and disadvantages. For example, in certain materials, the second order nonlinearity is quite strong, promising good pair generation efficiency. However, in second order processes, the signal-idler detuning is on the order of the pump wavelength, so creating integrated devices that guide well across such a broad spectral range is rather difficult. Hence third order processes are preferred for integrated devices. However, correlated pair generation from four photon mixing is often masked by strong, uncorrelated Raman scattering, and groups working with these devices have gone to great lengths to eliminate this source of noise [34].

It should be noted that compact, integrated sources of entangled photon pairs have utility beyond quantum communications. For example, these devices can enable compact two photon interferometers for better optical sensing applications [57], comprise sources for linear optical quantum computing, or be used in correlated photon metrology [42]. Of course, as with many quantum information technologies, the most likely 'killer' applications are the ones that have not yet been imagined.

2.2 Birefringent phase-matched four wave mixing

As mentioned above, one of the main drawbacks to third-order processes for entanglement generation is noise generated by Raman scattering. As Lin, *et al.*, point out in their 2006 paper, one way to mitigate this effect is to generate signal and idler pairs at the opposite polarization to the pump [35]. They show that the Raman gain is insignificant at the orthogonal polarization. This polarization diversity scheme has the added benefit of making it simpler to separate the signal-idler pair from the pump with only a polarizer on the output. Per the proposal of Lin, *et al.*, we examine and demonstrate third-order nonlinearity in a semiconductor waveguide using birefringent phase matching in order to generate correlated photon pairs without the Raman noise. Furthermore, because the material that we investigate is a crystalline semiconductor, the Raman noise generated is narrow band, as opposed to the broadband Raman noise produced in glass devices, further reducing the interference from uncorrelated processes.

2.2.1 Linear and nonlinear optics

Before we present the details of the device, it behooves us to briefly review the theory of nonlinear optics in general and four wave mixing in particular[1]. We begin by considering a simple atom, comprised of a nucleus and an electron cloud, in the presence of a time-dependent electric field. This system can be approximated

[1]This treatment is very cursory. For a much more detailed treatment of the nonlinear susceptibility tensors, coupled wave equations, etc., see [13].

by a mass on a spring, where the restoring force of the spring is analogous to the Coulomb force that the electron experiences in the presence of the field. At first, we will assume that this force is linear in the displacement. Like any good physics derivation, we begin with Newton's second law [43],

$$F = ma = m\ddot{x}(t) = -eE(t) - kx(t) - 2m\gamma\dot{x}(t) \tag{2.1}$$

where the first term is the driving force, the second term is the restoring force, the third term represents a damping loss term. Rearranging terms, we obtain the linear differential equation

$$\ddot{x}(t) + 2\gamma\dot{x}(t) + \Omega^2 x(t) = -\frac{e}{m}E(t) \tag{2.2}$$

where $\Omega = \sqrt{\frac{k}{m}}$ is defined as the natural oscillation frequency. If we assume that the incident electric field is a monochromatic plane wave of the form $E(t) = E_0\cos(\omega t) = \frac{1}{2}E_0 e^{-i\omega t} + c.c.$, then the AC steady state response of the atom is

$$x(t) = \frac{eE_0}{2m\left(\Omega^2 - 2i\gamma\omega - \omega^2\right)}e^{-i\omega t} + c.c. \tag{2.3}$$

Now that we know the response, consider a medium consisting of a large number of these atoms in an isotropic, uniform distribution (for example, an atomic vapor), with number density N. In such a medium, the macroscopic *polarization* of the material in one dimension is defined as

$$P(t) = -Nex(t) \tag{2.4}$$

where the factor of $ex(t)$ represents the dipole moment of each individual atom. Substituting the solution in 2.3 into 2.4, we obtain

$$P(t) \quad = \quad \frac{Ne^2E_0}{2m\left(\Omega^2 - 2i\gamma\omega - \omega^2\right)}e^{-i\omega t} + c.c. \tag{2.5}$$

$$= \frac{1}{2}P_0 e^{-i\omega t} + c.c.$$

where $P_0 = \frac{Ne^2E_0}{m(\Omega^2 - 2i\gamma\omega - \omega^2)}$. In the frequency domain, we can write this as

$$\hat{P}(\omega) = \epsilon_0 \chi^{(1)}(\omega)\hat{E}(\omega) \tag{2.6}$$

where $\hat{P}(\omega)$ and $\hat{E}(\omega)$ are the Fourier transforms of $P(t)$ and $E(t)$ respectively, and $\chi^{(1)}$ is defined as the *linear susceptibility* of the material. Since $E(t)$ is monochromatic, we know from 2.6 that

$$P_0 = \epsilon_0 \chi^{(1)}(\omega)E_0 \tag{2.7}$$

Comparing 2.7 to 2.5, we see that the linear susceptibility of the material is given by

$$\chi^{(1)} = \frac{Ne^2}{m\left(\Omega^2 - 2i\gamma\omega - \omega^2\right)} \tag{2.8}$$

Now consider the case where the restoring force is no longer linear in the displacement (as we know to be the case), but instead some arbitrary function. We can expand this function into a Taylor series and incorporate it into the differential equation in 2.2, giving, again in one dimension,

$$m\ddot{x}(t) = -eE(t) - \left[k_1 x(t) + k_2 x^2(t) + k_3 x^3(t) + \ldots\right] - 2m\gamma\dot{x}(t) \tag{2.9}$$

Solving this nonlinear differential equation, we can write down the total polarization of the material as $P_0 = P_0^{(1)} + P_0^{(2)} + P_0^{(3)} + \ldots$, a sum of polarizations from each term in the expansion, giving us

$$P_0^{(1)} = \epsilon_0 \chi^{(1)} E_0 \tag{2.10}$$

$$
\begin{aligned}
P_0^{(2)} &= \epsilon_0 \chi^{(2)} E_0^2 \\
P_0^{(3)} &= \epsilon_0 \chi^{(3)} E_0^3 \\
&\vdots
\end{aligned}
$$

Until now our analysis has been entirely one-dimensional. Consider now the case where the electric field and the polarization are full three-dimensional vectors. In the linear case, 2.6 becomes

$$\vec{P} = \epsilon_0 \chi^{(1)} \vec{E} \tag{2.11}$$

where $\chi^{(1)}$ is now a rank 2 tensor with nine elements. In full matrix notation, we can write this as

$$
\begin{pmatrix} P_x \\ P_y \\ P_z \end{pmatrix} = \epsilon_0 \begin{pmatrix} \chi_{xx}^{(1)} & \chi_{xy}^{(1)} & \chi_{xz}^{(1)} \\ \chi_{yx}^{(1)} & \chi_{yy}^{(1)} & \chi_{yz}^{(1)} \\ \chi_{zx}^{(1)} & \chi_{zy}^{(1)} & \chi_{zz}^{(1)} \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix}
$$

Using the Einstein notation, where sums over repeating subscripts are implied, we can abbreviate this expression by

$$P_j = \epsilon_0 \chi_{jk}^{(1)} E_k \tag{2.12}$$

Similarly, the *nonlinear susceptibilities* now become

$$
\begin{aligned}
P_j^{(1)} &= \epsilon_0 \chi_{jk}^{(1)} E_k \\
P_j^{(2)} &= \epsilon_0 \chi_{jkl}^{(2)} E_k E_l \\
P_j^{(3)} &= \epsilon_0 \chi_{jklm}^{(3)} E_k E_l E_m \\
&\vdots
\end{aligned}
\tag{2.13}
$$

where $\chi^{(2)}$ and $\chi^{(3)}$ are third and fourth rank tensors with 27 and 81 elements, respectively. Also note that, for the purposes of this derivation, all waves were assumed to be at one frequency, ω. When this is not valid (which is most of the time), the susceptibilities would be functions of the various frequencies involved.

2.2.2 Third-order processes and four wave mixing

The second-order susceptibility is responsible for processes such as second harmonic generation, parametric downconversion, and sum frequency generation, among others. Because it involves three different wavelengths (hence the name *three-photon processes*), and because conservation of energy dictates that the total photon energy incident on the medium must equal the photon energy transmitted out of the medium, the detuning between input and output wavelengths are often very large, on the order of the wavelength itself. Additionally, because the second-order susceptibility is related to the symmetric part of the restoring force (proportional to x^2), only materials that are non-centrosymmetric will exhibit second-order nonlinear optical phenomena. Of the 32 crystal classes, only 21 fall into this category and will have non-zero $\chi^{(2)}$ elements. All others do not exhibit second-order or, in fact, any even order behavior.

In this experiment we choose to focus on third-order processes for two reasons. First, because of the nature of these *four* photon processes, the detuning between input and output wavelengths can be much smaller. This helps in designing optical devices both because they do not have to operate over such a large spectrum

and because achieving phase matching is significantly easier (this will be discussed in more detail in the next section). In addition, *all* materials exhibit third-order characteristics, opening up a larger field of possible materials with which to work. One drawback to $\chi^{(3)}$ processes, however, is that they are typically weaker than second-order effects. Still, because phase matching is easier, it is possible to achieve adequate interaction strength with simpler designs.

We will focus on the specific process of *partially-degenerate four wave mixing*, where two photons at a pump wavelength are converted via the third-order nonlinear susceptibility of the medium into two photons, generally called *signal* and *idler* photons, at frequencies slightly detuned longer and shorter than the pump wavelength. To understand how this process occurs, consider again a centro-symmetric material with an anharmonic response. Suppose, for simplicity, that a monochromatic plane wave is incident on this material. Equation 2.9 still applies, albeit with $k_2 = 0$. If we take the perturbation approach, expanding $x(t) = x^{(1)}(t) + x^{(3)}(t) + \ldots$, with $x^{(1)} \gg x^{(3)}$, then 2.9 becomes

$$\begin{aligned} \frac{d^2}{dt^2} \quad \left[x^{(1)}(t) + x^{(3)}(t)\right] + 2\gamma \tfrac{d}{dt}\left[x^{(1)}(t) + x^{(3)}(t)\right] + \Omega^2\left[x^{(1)}(t) + x^{(3)}(t)\right] \quad (2.14) \\ = -\tfrac{e}{m}E(t) - \tfrac{k_3}{m}\left[x^{(1)}(t) + x^{(3)}(t)\right]^3 \end{aligned}$$

Because $x^{(3)}$ is small, we can approximate the last term on the right hand side by $-\frac{k_3}{m}\left[x^{(1)}(t)\right]^3$. Subtracting 2.2 from 2.14, we obtain

$$\ddot{x}^{(3)}(t) + 2\gamma\dot{x}^{(3)}(t) + \Omega^2 x^{(3)}(t) = -\frac{k_3}{m}\left[x^{(1)}(t)\right]^3 \quad (2.15)$$

We can expand the right hand side, using the linear solution from 2.3, into

$$\left[x^{(1)}(t)\right]^3 = \quad \left[\tfrac{1}{8}x_0^{(1)}x_0^{(1)}x_0^{(1)}e^{-i3\omega t} + c.c.\right] + \left[\tfrac{1}{8}x_0^{(1)^*}x_0^{(1)}x_0^{(1)}e^{-i\omega t} + c.c.\right] \quad (2.16)$$

$$+ \left[\tfrac{1}{4} x_0^{(1)} x_0^{(1)} x_0^{(1)^*} e^{-i\omega t} + c.c.\right]$$

The first term oscillates at three times the input frequency and is responsible for the effect of *third harmonic generation*, while the second and third term oscillate at the original input frequency. We can treat these two frequency components separately, splitting the total polarization of the material into two parts, $\vec{P}^{(3)}(3\omega)$ and $\vec{P}^{(3)}(\omega)$. Using the abbreviated Einstein notation, we can write these in terms of the third-order nonlinear susceptibilities,

$$P_j^{(3)}(3\omega) = \frac{3}{2}\epsilon_0 \chi_{jklm}^{(3)}(-3\omega;\omega,\omega,\omega) E_k(\omega) E_l(\omega) E_m(\omega) \tag{2.17}$$

$$P_j^{(3)}(\omega) = \frac{3}{4}\epsilon_0 \chi_{jklm}^{(3)}(-\omega;\omega,\omega,-\omega) E_k(\omega) E_l(\omega) E_m^*(\omega) \tag{2.18}$$

where the factors of $\frac{3}{2}$ and $\frac{3}{4}$ are related to the degeneracies and symmetries of the susceptibility tensor [13]. The second expression represents the third-order contribution to the material polarization at the input frequency ω. As there is also a contribution from the linear polarization, this perturbation results in an effective index of refraction that is proportional to the input field intensity, resulting in the optical Kerr effect [13].

So far we have only examined the simple case where the input field is monochromatic. However, consider the case where the input field is composed of two closely spaced frequencies, a pump at ω_p and a signal at $\omega_s = \omega_p + \delta$, with the component at the pump frequency significantly stronger than the component at the signal frequency. It is not difficult to see that the resulting third-order polarization will have components oscillating at a number of different frequencies, including the pump frequency, the signal frequency, the third harmonics of both frequencies, and a new

frequency, called the *idler* frequency, at $\omega_i = \omega_p - \delta$. This polarization component can be written as

$$P_j^{(3)}(\omega_i) = \frac{3}{2}\epsilon_0 \chi_{jklm}^{(3)}(-\omega_i; \omega_p, \omega_p, -\omega_s) E_k(\omega_p) E_l(\omega_p) E_m^*(\omega_s) \tag{2.19}$$

It should be noted that, nominally, the fourth-rank tensor $\chi^{(3)}$ has 81 elements. However, under certain reasonable assumptions about the material properties, as well as the symmetry properties of nonlinear susceptibility tensors, we can describe the third-order susceptibility by only two numbers, $\chi_{xxxx}^{(3)}$ and $\chi_{xxyy}^{(3)}$, where the first number represents the strength of processes where the pump and signal are co-polarized, and the second number represents the case where the two are cross-polarized. Note that these assumptions are only valid for semiconductor materials with cubic symmetry in spectral regions where they are essentially lossless. In fact, in isotropic materials such as glass, we can reduce the tensor even further with the relation $\chi_{xxyy}^{(3)} = \frac{1}{3}\chi_{xxxx}^{(3)}$. However, for the cubic semiconductors in which we are interested for this experiment, there is no analytical relationship between the two and we must rely on experimental measurements to determine the interaction strength of cross-polarized processes.

In order to examine the behavior of the idler with respect to the other signals, we must derive a set of coupled wave equations describing the joint evolution of the electromagnetic fields as they propagate through the medium[70]. We begin with Maxwell's equations for a non-conducting, non-magnetic material,

$$\begin{aligned} \bigtriangledown \times \vec{E} &= -\frac{\partial \vec{B}}{\partial t} \\ \bigtriangledown \times \vec{H} &= \frac{\partial \vec{D}}{\partial t} \end{aligned} \tag{2.20}$$

$$\begin{aligned} \bigtriangledown \cdot \vec{D} &= 0 \\ \bigtriangledown \cdot \vec{B} &= 0 \end{aligned}$$

Combining the first two, we obtain

$$\begin{aligned} \bigtriangledown \times \bigtriangledown \times \vec{E} &= -\mu_0 \frac{\partial}{\partial t}\left(\bigtriangledown \times \vec{H}\right) \\ &= -\mu_0 \frac{\partial^2 \vec{D}}{\partial t^2} \end{aligned} \tag{2.21}$$

We know that, by definition, $\vec{D} = \epsilon_0 \vec{E} + \vec{P}$. Substituting, we obtain

$$\bigtriangledown \times \bigtriangledown \times \vec{E} + \mu_0 \epsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2} = -\mu_0 \frac{\partial^2 \vec{P}}{\partial t^2} \tag{2.22}$$

The polarization, $\vec{P} = \vec{P}^{(\text{linear})} + \vec{P}^{(NL)}$, can be written as separate linear and nonlinear parts. If we consider only the nonlinear part of the polarization, then 2.22 simplifies to

$$\bigtriangledown^2 \vec{E} - \frac{n^2}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} = \mu_0 \frac{\partial^2 \vec{P}^{(NL)}}{\partial t^2} \tag{2.23}$$

With the nonlinear polarization described above, we can use this wave equation to examine the evolution of the electromagnetic wave as it propagates through the material. Consider that the wave has the form

$$E_j(\vec{r}, t) = \frac{1}{2} A_j(z) U(x, y) e^{i k_j z - \omega_j t} + c.c. \tag{2.24}$$

where $j = p, s, i$ is the index of each component (pump, signal, and idler). Note that the wave is factored into a slowly-varying envelope $A_j(z)$, a transverse mode $U(x, y)$, and complex phasor. We make the following assumptions: First, that all of the waves are co-polarized, allowing us to treat the field in one dimension; we will revisit this assumption later. Second, we will assume that all $U(x, y)$ are the

same for each component of the field, which is a valid assumption for closely spaced frequencies in a single-mode waveguide. Finally, we will assume that $A_s, A_i \ll A_p$, that is, the pump field is much stronger than the signal and idler fields. This allow for two simplifications - first, we can assume that the pump is *undepleted* by the parametric process. Second, when considering the nonlinear polarization, we can ignore any terms that result from the product of the signal and/or idler fields, only keeping terms that involve products with the pump field. This results in three expressions for the nonlinear polarization.

$$\begin{aligned} P(\omega_p) &= \xi_p(t)\left|A_p\right|^2 A_p \\ P(\omega_s) &= \sigma_s\left[2\left|A_p\right|^2 A_s + A_p^2 A_i^* \xi_s(t)\right] \\ P(\omega_i) &= \sigma_i\left[2\left|A_p\right|^2 A_i + A_p^2 A_s^* \xi_s(t)\right] \end{aligned} \tag{2.25}$$

where we have simplified the notation[2] by incorporating all of the constants such as the susceptibility, c, ϵ_0, etc., into the respective $\sigma_{p,s,i}$ and rewriting all of the time-dependent phasors as the various $\xi_{p,s,i}(t)$. Since we assume that the pump remains undepleted throughout the process, the only term that remains is the one responsible for self-phase modulation. In the polarizations of the signal and idler, two terms remain significant - the cross-phase modulation from the pump signal, represented by the first term, and the four-wave mixing term that is the subject of this analysis.

We can now substitute 2.24 and 2.25 into 2.23 to obtain three differential equations describing the coupled evolution of the electromagnetic field as it propagates

[2]This notation is similar to that used in [70].

through the medium. Under the assumption that the envelopes $A_{p,s,i}$ are slowly varying compared to the frequencies $\omega_{p,s,i}$, we can say that

$$\left|\frac{\partial^2}{\partial z^2}A_{p,s,i}\right| \ll \left|k_{p,s,i}\frac{\partial}{\partial z}A_{p,s,i}\right| \tag{2.26}$$

This allows us to ignore all of the second-order derivatives. If we further assume that the medium is lossless and that the phase-matching conditions are met ($2k_p - k_s - k_i = 2\omega_p - \omega_s - \omega_i = 0$), this gives us

$$\begin{aligned}
\frac{dA_p}{dz} &= \frac{3i\omega_p^2\chi^{(3)}_{xxxx}}{2k_pc^2}\left|A_p\right|^2 A_p \\
\frac{dA_s}{dz} &= \frac{3i\omega_s^2\chi^{(3)}_{xxxx}}{2k_sc^2}\left(2\left|A_p\right|^2 A_s + A_p^2A_i^*\right) \\
\frac{dA_i}{dz} &= \frac{3i\omega_i^2\chi^{(3)}_{xxxx}}{2k_ic^2}\left(2\left|A_p\right|^2 A_i + A_p^2A_s^*\right)
\end{aligned} \tag{2.27}$$

We can readily solve the first equation in 2.27, giving

$$\begin{aligned}
A_p(z) &= A_p(0)e^{i\phi(z)} \\
\phi(z) &= \frac{3i\omega_p^2\chi^{(3)}_{xxxx}}{2k_pc^2}\left|A_p\right|^2 z
\end{aligned} \tag{2.28}$$

With the proper substitutions for the nonlinear index $n_2 = \frac{3\chi^{(3)}_{xxxx}}{4n_0^2\epsilon_0 c}$ and the various factors relating A_p to the pump power P_0, we can show that the phase function $\phi(z)$ in 2.28 can be written as

$$\phi(z) = \gamma P_0 z \tag{2.29}$$

where $\gamma = \frac{2\pi n_2}{\lambda A_{\text{eff}}}$ is the *nonlinear parameter* indicating the strength of the nonlinear interaction and A_{eff} is the effective mode area. This expression $\gamma P_0 z$ appears often in parametric nonlinear processes. The parametric gain in four-wave mixing, as we will show, is given by $G = |\gamma P_0 z|^2$. Since we are keen to deal in the regime where

both the signal and idler signals are composed of single photons, we can make the valid assumption that $|\gamma P_0 z|^2 \ll 1$ and that both the self-phase and cross-phase modulation are small.

Using this solution in 2.28 and 2.29, we can reduce the number of coupled equations from three to two, allowing us to write, to first-order approximation, the following coupled equations describing the evolution of the signal and idler beams:

$$\begin{aligned} \frac{dA_s}{dz} &= i\gamma P_0 \left(A_s + A_i^*\right) \\ \frac{dA_i^*}{dz} &= -i\gamma P_0 \left(A_s^* + A_i\right) \end{aligned} \tag{2.30}$$

Collecting terms, we can show that

$$\begin{aligned} \frac{d}{dz}\left(A_s + A_i^*\right) &= 0 \\ \frac{d}{dz}\left(A_s - A_i^*\right) &= 2i\gamma P_0 \left(A_s + A_i^*\right) \end{aligned} \tag{2.31}$$

We can solve these differential equations to obtain

$$\begin{aligned} A_s(z) + A_i^*(z) &= A_s(0) + A_i^*(0) \\ A_s(z) - A_i^*(z) &= \left(A_s(0) - A_i^*(0)\right) + 2i\gamma P_0 z \left(A_s(0) + A_i^*(0)\right) \end{aligned} \tag{2.32}$$

Solving for the signal and idler envelopes under the intitial conditions of $A_s(0) = A_0$ and $A_i^*(0) = 0$ gives us

$$\begin{aligned} A_s(z) &= (1 + \gamma P_0 z) A_0 \\ A_i^*(z) &= \gamma P_0 z A_0 \end{aligned} \tag{2.33}$$

Note that these expressions assume perfect phase matching, as well as co-polarization of all of the beams. These assumptions will be revisited in later sections, but suffice to say they are valid in understanding what is meant by *correlated photons*.

2.2.3 Parametric gain, photon correlation, and spontaneous pair generation

From the coupled equations in 2.33, we can see, as indicated above, that the parametric gain at the signal and idler wavelengths is the same. That is, for a device of length L with an initial signal input of $A_s(0)$ and a pump power of P_0, the parametric gain is given by

$$G = \frac{|A_i^*(L)|^2}{|A_s(0)|^2} = |\gamma P_0 L|^2 \tag{2.34}$$

The fact that these intensities increase by the same amount is a first indication of what is meant by *correlated photons*. More deeply, however, we can consider what happens to the coupled wave equations when we quantize the signal and idler fields and consider them on the single-photon level. We define the signal and idler field operators in the usual way [37]:

$$\begin{aligned} \hat{E}_s(z,t) &= \sqrt{\frac{\hbar\omega_s}{2\epsilon_0 V}}\left(\hat{a}_s(t)e^{i(k_s z-\omega_s t)} + \hat{a}_s^\dagger(t)e^{-i(k_s z-\omega_s t)}\right) \\ \hat{E}_i(z,t) &= \sqrt{\frac{\hbar\omega_i}{2\epsilon_0 V}}\left(\hat{a}_i(t)e^{i(k_i z-\omega_i t)} + \hat{a}_i^\dagger(t)e^{-i(k_i z-\omega_i t)}\right) \end{aligned} \tag{2.35}$$

where V is the mode volume. Note that the creation and annihilation operators satisfy the commutation relations

$$\begin{aligned} \left[\hat{a}_i(t), \hat{a}_j^\dagger(t)\right] &= \delta_{ij} \\ \left[\hat{a}_i(t), \hat{a}_j(t)\right] &= 0 \end{aligned} \tag{2.36}$$

for $i, j = s, i$. Since all of the waves are close together, we can make the assumption that $\omega_s \approx \omega_i \approx \omega_p = \omega$. Using this assumption, we can write the interaction

Hamiltonian for the parametric four-wave mixing process as

$$\hat{H}_{int} = \frac{\hbar\omega\gamma P_0 c}{n}\left(\hat{a}_s^\dagger\hat{a}_s + \hat{a}_i^\dagger\hat{a}_i + \hat{a}_s^\dagger\hat{a}_i^\dagger + \hat{a}_s\hat{a}_i\right) \tag{2.37}$$

Since we want to find the time-evolution of the field operators, we can use the Heisenberg equation of motion to show that

$$\begin{aligned} i\hbar\frac{d}{dt}\hat{a}_s &= \left[\hat{a}_s, \hat{H}_{int}\right] \\ &= \frac{\hbar\omega\gamma P_0 c}{n}\left(\left[\hat{a}_s, \hat{a}_s^\dagger\right]\hat{a}_s + \left[\hat{a}_s, \hat{a}_s^\dagger\right]\hat{a}_i^\dagger\right) \\ i\hbar\frac{d}{dt}\hat{a}_i^\dagger &= \left[\hat{a}_i^\dagger, \hat{H}_{int}\right] \\ &= \frac{\hbar\omega\gamma P_0 c}{n}\left(\left[\hat{a}_i^\dagger, \hat{a}_i\right]\hat{a}_s + \left[\hat{a}_i^\dagger, \hat{a}_i\right]\hat{a}_i^\dagger\right) \end{aligned} \tag{2.38}$$

These expressions give us the coupled equations for time evolution of the signal and idler field operators:

$$\begin{aligned} \frac{d}{dt}\hat{a}_s &= \frac{-i\gamma P_0 c}{n}\left(\hat{a}_s + \hat{a}_i^\dagger\right) \\ \frac{d}{dt}\hat{a}_i^\dagger &= \frac{i\gamma P_0 c}{n}\left(\hat{a}_s + \hat{a}_i^\dagger\right) \end{aligned} \tag{2.39}$$

Solving these in a similar manner to the classical case above, we obtain

$$\begin{aligned} \hat{a}_s(t) &= \mu\hat{a}_s(0) + \nu\hat{a}_i^\dagger(0) \\ \hat{a}_i^\dagger(t) &= \nu^*\hat{a}_s(0) + \mu^*\hat{a}_i^\dagger(0) \end{aligned} \tag{2.40}$$

where $\mu = 1 - \frac{i\gamma P_0 ct}{n}$ and $\nu = \frac{-i\gamma P_0 ct}{n}$. From 2.39 and 2.36, we find that

$$\begin{aligned} \frac{\partial}{\partial t}\left(\hat{n}_s(t) - \hat{n}_i(t)\right) &= \frac{\partial}{\partial t}\left(\hat{a}_s^\dagger(t)\hat{a}_s(t) - a_i^\dagger(t)\hat{a}_i(t)\right) \\ &= a_s^\dagger(t)\frac{\partial}{\partial t}\hat{a}_s(t) + \left(\frac{\partial}{\partial t}\hat{a}_s^\dagger(t)\right)\hat{a}_s(t) \\ &\quad -a_i^\dagger(t)\frac{\partial}{\partial t}\hat{a}_i(t) - \left(\frac{\partial}{\partial t}\hat{a}_i^\dagger(t)\right)\hat{a}_i(t) = 0 \end{aligned} \tag{2.41}$$

This implies that the difference in the number operators (defined as $\hat{n}_j = \hat{a}_j^\dagger \hat{a}_j$ for $j = s, i$) of the signal and idler fields $\hat{n}_s(k,t) - \hat{n}_i(k,t)$ is a *constant of motion.* More plainly, it indicates that whenever a signal photon is created, an idler photon will be created as well. This is the fundamental meaning of *correlated photon pairs.*

In addition to their correlation, we must examine how these pairs can be generated *spontaneously.* Using the solutions in 2.40, let us examine the case when there is nothing but vacuum fluctuations present at signal and idler modes at $z = 0$. Noting that $\hat{a}_j \hat{a}_j^\dagger = \hat{1}$, the vacuum matrix element of the signal number operator at time $t = \frac{nL}{c}$ becomes (after some algebra)

$$\begin{aligned} \left\langle 0_s, 0_i \left| \hat{n}_s \left(\frac{nL}{c} \right) \right| 0_s, 0_i \right\rangle &= \left\langle 0_s, 0_i \left| \hat{a}_s^\dagger \left(\frac{nL}{c} \right) \hat{a}_s \left(\frac{nL}{c} \right) \right| 0_s, 0_i \right\rangle \\ &= |\nu|^2 = |\gamma P_0 L|^2 \end{aligned} \tag{2.42}$$

Thus we can see that, even in the presence of nothing but vacuum fluctuations on the signal input, signal photons are generated spontaneously. Additionally, since we have shown that signal and idler photons are always generated in pairs, we have established a mathematical basis for the spontaneous generation of correlated photon pairs.[3]

[3] Classically, we can show from the Manley-Rowe relations [38] that the photon flux in each beam is correlated, i.e., $\frac{d}{dz}\left(\frac{I_s}{\omega_s}\right) = \frac{d}{dz}\left(\frac{I_i}{\omega_i}\right) = -2\frac{d}{dz}\left(\frac{I_p}{\omega_p}\right)$, where $I_{p,s,i}$ are the intensities of the pump, signal, and idler beams respectively. However, by treating the beams quantum mechanically, we can quantify a measure of the correlation, as we will do in later sections.

2.2.4 Birefringent phase matching and the cross-polarized susceptibility tensor element

In the derivation above, we have established that, under certain assumed conditions, spontaneous pairs of correlated photons are generated via four-wave mixing. However, a number of the assumptions made above are non-trivial. First, we assumed that all of the waves are co-polarized. While violating this assumption does not necessarily void the above result, it does change some of the interaction coefficients. Specifically, the definitions of n_2 and γ are based on the real part of the tensor element $\chi^{(3)}_{xxxx}$. If, for example, we used a process that involved pump photons at one polarization interacting with signal and idler photons at another wavelength, these parameters would have to be defined in terms of the $\chi^{(3)}_{xxyy}$ tensor element. In an isotropic medium, such as a glass optical fiber, the relationship is well-known to be $\chi^{(3)}_{xxyy} = \frac{1}{3}\chi^{(3)}_{xxxx}$. However, for cubic semiconductors such as Si, GaAs, and AlGaAs, there is no analytical relationship. There are some experimental results that provide estimates of the ratio $\frac{\chi^{(3)}_{xxyy}}{\chi^{(3)}_{xxxx}}$, but they range from 0.25 to over 0.9 and are inconclusive [16], [21]. An estimate of this ratio is one of the immediate goals of this project. For comparison, Table 2.1 compares the published n_2 values for various common optical materials.

While using interacting fields at orthogonal polarizations leads to uncertainty in the strength of the interaction, it does offer a number of advantages in a practical nonlinear device. From an experimental point of view, generating the signal and idler orthogonal to the pump allows for easier elimination of the pump beam, starting

Table 2.1: A comparison of the nonlinear refractive indices of different materials [13], [19], [33]. Co-polarized processes are strongest in AlGaAs, but the strength of cross-polarized processes is unknown.

Material	n_2 in m^2/W
Silica	2.8×10^{-20}
Silicon	5×10^{-18}
AlGaAs	3.4×10^{-17}

with a simple polarizer on the output of the device. More importantly, however, generating the signal and idler orthogonally polarized to the pump allows for better elimination of Raman noise. As Lin, et al., point out in [35], the limiting factor in most fiber-based entanglement generation systems is noise from broadband Raman scattering in the medium. The Raman gain is mainly polarized in the direction of the pump beam, with negligible Raman noise generated at the orthogonal polarization. Hence they propose using birefringent phase matched four wave mixing as a way to eliminate Raman noise in entanglement generation systems. For this reason, we choose to focus our experiment on using birefringent phase-matched four-wave mixing in a semiconductor waveguide as a method to generate entanglement without interference from Raman noise.

Birefringent phase matching is certainly not a new idea [61]. However, it has yet to be examined as a method to generate correlated pairs, especially in a

semiconductor waveguide. It is not well established that birefringence is something that is easily designed into a waveguide, as the typical goal of integrated photonics design is to eliminate birefringence altogether. In addition, it is not obvious how to use linearly co-polarized correlated photon pairs to generate entanglement. Though ways do exist, whether they are practical and efficient has yet to be determined.

How does the approach of birefringent phase matching relate to the analysis above? Recall that we assumed that the phase mismatch, $\Delta k = 2k_p - k_s - k_i$, was zero, giving us a parametric gain of $G = |\gamma P_0 L|^2$. If $\Delta k \neq 0$, then one can show that the expression for parametric gain becomes

$$G = |\gamma P_0 L|^2 \cdot \frac{\sin^2\left(\frac{1}{2}\Delta k L\right)}{\left(\frac{1}{2}\Delta k L\right)^2} \tag{2.43}$$

Because this additional factor approaches zero very rapidly for any $\Delta k \neq 0$, it becomes imperative that the system be designed such that the phase mismatch is truly zero. This becomes the challenge in most nonlinear optics experiments involving dispersive media, and many diverse approaches exist[4]. In the case of birefringent phase matching, we can write the phase mismatch as

$$\Delta k = 2\pi \left[\frac{2}{\lambda_p} n_x(\lambda_p) - \frac{1}{\lambda_s} n_y(\lambda_s) - \frac{1}{\lambda_i} n_y(\lambda_i)\right] \tag{2.44}$$

The phase matching condition is tantamount to setting $\Delta k = 0$ in this expression. Note that the index of refraction at the x and y polarization are different and wavelength-dependent.

[4]For a discussion of the various phase-matching techniques, including angle tuning, quasi-phase matching, and dispersion tailoring, see [13] and [65].

2.2.5 Figure of merit and the effects of linear and nonlinear loss

The above analysis assumes negligible linear loss and two-photon absorption. The presence of significant absorption can strongly affect the correlated pair generation efficiency. To understand how these effects enter into the analysis, we can introduce a measure of the correlation of changes in the signal and idler fields [24], defined as

$$C = \frac{\langle \hat{n}_s \hat{n}_i \rangle - \langle \hat{n}_s \rangle \langle \hat{n}_i \rangle}{\langle \hat{n}_s^2 \rangle - \langle \hat{n}_i^2 \rangle} \tag{2.45}$$

In the absence of loss, this expression becomes $C = \frac{|\mu|^2}{|\nu|^2+1} = 1$. We can also compute the value of C in the presence of linear loss. First, we redefine the creation and annihilation operators to include linear loss:

$$\hat{b}_{s,i} = \sqrt{\eta}\hat{a}_{s,i} + \hat{\xi} \tag{2.46}$$

where η is a parameter representing linear loss (close to unity) and $\hat{\xi}$ is the noise operator required to maintain the commutation relation $\left[\hat{b}_i, \hat{b}_j^\dagger\right] = \delta_{i,j}$ for $i, j = s, i$. ξ itself satisfies the commutator $\left[\hat{\xi}, \hat{\xi}^\dagger\right] = 1 - \eta$. Using these new operators, we can show that C becomes

$$C = \frac{\eta^2 |\mu|^2}{\eta^2 |\nu|^2 + 1} = \frac{\eta^2 \left(|\nu|^2 + 1\right)}{\eta^2 |\nu|^2 + 1} \approx \eta^2 \tag{2.47}$$

since we assume $|\nu|^2 << 1$. We see that the correlation scales with η^2, and η is close to but less than unity, so the photons become decorrelated quite quickly as the linear loss increases.

For a material with appreciable two-photon absorption, we can show that the

correlation function becomes

$$C = \frac{|\nu|^2 + e^{-2\beta P_0 z}}{|\nu|^2 + 1} \tag{2.48}$$

where β is the 2PA coefficient. In this case, as the mean photon number increases with the presence of significant 2PA, the correlation rapidly decreases.

To estimate the impact of 2PA on correlated photon generation, we can define a figure of merit to compare the relative strengths of 4WM and 2PA:

$$f = \left(\frac{2\pi n_2}{\lambda\beta}\right)^2 \tag{2.49}$$

If f is greater than unity, then we can assume that 4WM will dominate 2PA and the device will potentially be useful as a source of correlated photon pairs.

2.3 Device design and fabrication

2.3.1 Wavelength and material selection

In order to fabricate an actual device to perform 4WM, we must first select a wavelength at which to operate. We determine the operating wavelength based on the following criteria: First, that the system be compatible with Si SPADs. We demand this simply because single-photon measurements are much more straightforward using Si SPADs than measurements made, for example, at C-band wavelengths with cooled InGaAs detectors. Second, the operating wavelength must be compatible with free-space quantum key distribution; that is, it must not fall within a water absorption band in the atmosphere. Finally, we must select a wavelength for which optics are available. Of the possibilities, the wavelength that best satisfies these

criteria is 780 nm. It lies well within the efficient absorption spectrum of Si SPADs, it propagates readily through the atmosphere (as shown in Figure 2.1), and it also happens to be a rubidium absorption line, so optics are readily available at that wavelength.

Having chosen 780 nm as the center wavelength for the device, we must choose a semiconductor material that is transparent in that region. The obvious candidate is aluminum gallium arsenide, the preferred photonic material for that wavelength. The empirical relationship of the bandgap energy in eV to the aluminum content for $Al_xGa_{1-x}As$ is given by[1]

$$E_g = 1.424 + 1.266x + 0.26x^2 \tag{2.50}$$

Using this expression, we can choose a material composition that is transparent at 780 nm, or a photon energy of 1.59 eV. We want this photon energy to be roughly 20% lower than the bandgap energy of the material. Solving backwards indicates that a waveguide with roughly 30% aluminum should suffice, with a bandgap energy of 1.798 eV. This choice of material serves as our starting point for designing the device.

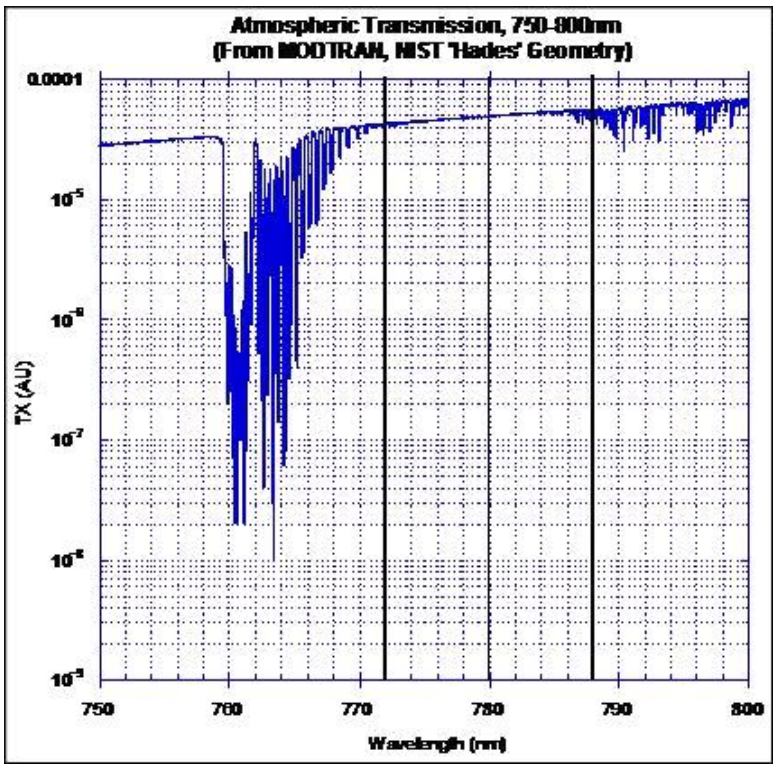

Figure 2.1: A MODTRAN plot of the atmospheric transmission around 780 nm. Note that the wavelength range of interest does not contain any water absorption lines. For more details on MODTRAN, see chapter 4.

2.3.2 Birefringence and detuning estimation

The next step in the design process involves choosing a range in which we want to generate photon pairs, and determining the birefringence required to achieve such pump-signal detuning. In order to make this determination, we can use the temperature-dependent Sellmeier equations of Gehrsitz, et al.[23], for AlGaAs with arbitrary aluminum content. The full calculation of the index of refraction as a function of the wavelength, aluminum content, and temperature is not trivial. An implementation of the calculation in MATLAB compatible C code is included in Appendix A. We can use this code, along with the expression for the phase mismatch in 2.44, to estimate the birefringence required to achieve phase matching. Setting $\Delta k = 0$ in 2.44 and solving for the birefringence, $\delta n = n_x - n_y$ gives us

$$\delta n = \frac{\lambda_s \lambda_i}{\lambda_s + \lambda_i} \left(\frac{2n(\lambda_p)}{\lambda_p} - \frac{n(\lambda_s)}{\lambda_s} - \frac{n(\lambda_i)}{\lambda_i} \right) \tag{2.51}$$

where $n(\lambda)$ is the Sellemeir equation for bulk AlGaAs from [23], assuming 30% aluminum at a temperature of 300 K. A plot of this curve is shown in Figure 2.2. Of course, this calculation makes a number of assumptions, the most cavalier (though not unreasonable) assumes that the birefringence is constant with wavelength. Still, the plot in Figure 2.2 provides us with at least an estimate of the range of birefringence to which we must design the device.

When choosing the pair generation wavelengths, we must consider both experimental practicalities and proper operation device. As the detuning decreases, we are limited by the ability to experimentally separate the pump, signal, and idler photons in the spectral domain. This ostensibly limits the detuning to at least 3

nm, implying the device should have a birefringence of at least $\delta n = 10^{-4}$. In the other direction, we wish to avoid noise due to Raman interference, one of the motivations behind pursuing a birefringent process. One advantage of using a crystalline semiconductor over an amorphous glass is the strong localization of the Raman spectrum. Data outlined in Holtz, et al.[27], suggests that AlGaAs has LO phonon peaks at approximately 300 cm^{-1}, which corresponds to peaks between 10 and 20 nm detuned from a 780 nm pump. Hence when designing our device, we should limit the pump-signal detuning to less than the detuning of the Raman peaks[5]. From the plot in Figure 2.2, we see that the upper bound on our detuning implies that the device birefringence should remain less than $\delta n = 10^{-3}$.

[5]It is preferable, at least for optical fibers, to operate at detunings smaller than the Raman scattering peaks, as outside the Raman band the nonlinear gain for four photon mixing is lower [35]. It is also desirable to have smaller detunings since designing a waveguide for closely-spaced wavelengths is much more straightforward.

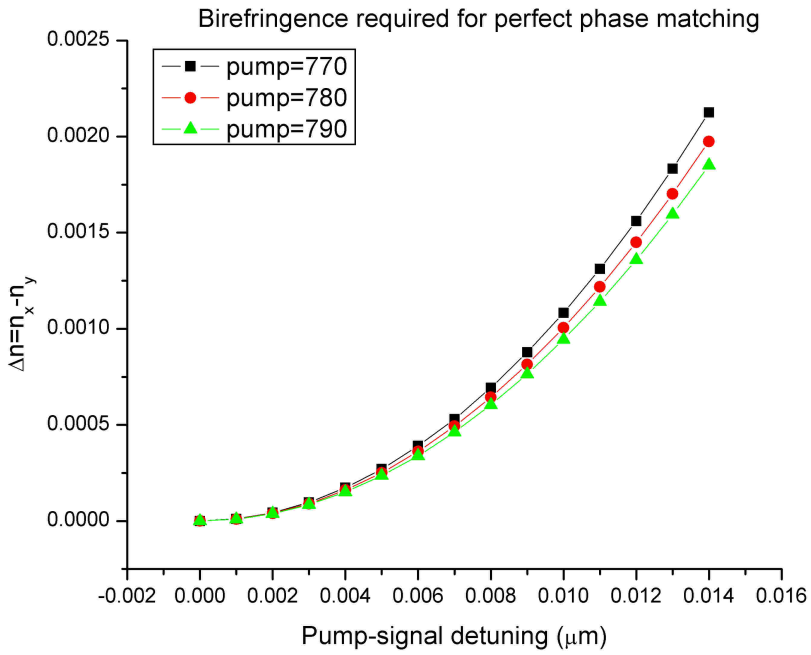

Figure 2.2: A plot of the required birefringence to achieve perfect phase matching as a function of the desired pump-signal detuning. The calculations show that the phasematching is sensitive primarily to the birefringence - the precise pump wavelength has little effect on the curve.

2.3.3 Form birefringence, layer structure, and feature size determination

The design requirements dictate that the detuning fall within the range of $10^{-4} < \delta n < 10^{-3}$. This range of values is rather small in the context of integrated photonics and is on the order of the birefringence observed in AlGaAs waveguides due simply to strain in the material[66]. It is not trivial to design a waveguide with such a small but precisely defined birefringence. In fact, often the goal of integrated photonic designs is to eliminate birefringence altogether. Hence one of the primary goals of this project is to demonstrate the ability to design a waveguide with a small but predictable and controllable birefringence. To do this, the most obvious approach is to use the property of *form birefringence*. This phenomenon occurs when a material consists of layers of alternating index of refraction. Consider a structure consisting of alternating layers of different bulk isotropic materials A and B, where material A has layer thickness t_A and index n_A, and material B has layer thickness t_B and index n_B. The structure is depicted in Figure 2.3. Given this composite material, the induced birefringence from the alternating layers is given by

$$n_e^2 - n_o^2 = -\frac{f_A f_B \left(n_A^2 - n_B^2\right)^2}{f_A n_B^2 + f_B n_A^2} \tag{2.52}$$

where n_o and n_e are the ordinary and extraordinary indices of refraction of the composite material, and $f_A = \frac{t_A}{t_A+t_B}$ and $f_B = \frac{t_B}{t_A+t_B}$ are the fractional thicknesses of the A and B layers[12].

Using this concept, we can create a device with a core region that consists

of alternating layers with slightly different aluminum content, creating an artificial birefringence that is controllable by the material properties. In addition, if we use a rib waveguide (which is the simplest waveguide design and the one that we will use), we can have further control over the birefringence by varying the width of the rib and thus controlling the aspect ratio of the guided mode. However, this results in a tradeoff between higher birefringence and less efficient input and output coupling; the efficiency is controlled by the overlap integral of the mode with a presumably gaussian beam, and a high aspect ratio mode results in a smaller coupling efficiency. Hence the goal is to create a birefringent mode with an aspect ratio as close to unity as possible.

There is another limitation on the feature size in addition to the mode properties. Because of the sensitivity of correlated pair generation to linear loss, the device must be designed to have as low a linear loss as possible while still remaining single mode. The lower limit on feature size is a product of limitations in the etching process during fabrication; exposure of the mask and the chemical etch process create features with roughness on the sidewalls. Hence it is good practice to limit the waveguide to a 2 μm rib width in order to maintain good guiding properties. In addition, we choose to design the device so that the mode is guided beneath the device surface in order to further mitigate losses due to surface roughness.

The last design consideration when laying out the waveguides is the very practical one of cost. As mentioned above, rib waveguides are the simplest designs to fabricate. Thus we choose to pursue designs based on this waveguide structure. The more important consideration, however, is the cost of the mask. Masks are one of

the most expensive components in semiconductor fabrication, so the ability to use an existing mask significantly reduces the cost and time of fabrication. Because of the risky nature of this project, we chose to base our design upon an existing mask, shown in Figure 2.4. This mask consists of ribs that vary from 2.0 to 7.0 μm in 0.5 μm steps, allowing us to test a range of devices until we find one that has the proper birefringence.

All of these considerations were taken into account and put into the design model. We simulated a myriad of various layer structures and finally settled on the device design shown in Figure 2.5. The structure is relatively simple to grow via molecular beam epitaxy, involving only various alloys of AlGaAs. The layer thicknesses are controllable in MBE to within monolayers, providing much finer precision than necessary. The aluminum concentrations can vary during the process and must be calibrated but are nominally controllable to within a few percent. More important, the *relative* aluminum concentrations that are critical to determining the form birefringence can be controlled to greater accuracy than the absolute aluminum concentrations.

This layer structure was verified by computional modeling to have the proper birefringence for devices ranging from 2 to 7 μm. A solution for the lowest order mode, from the OWMS commercial mode solving software, is shown in Figure 2.6. Plots of the birefringence versus rib width are shown in Figure 2.7. The plots show that the devices should have a birefringence within the desired range, while guiding only a single mode with a reasonably small aspect ratio.

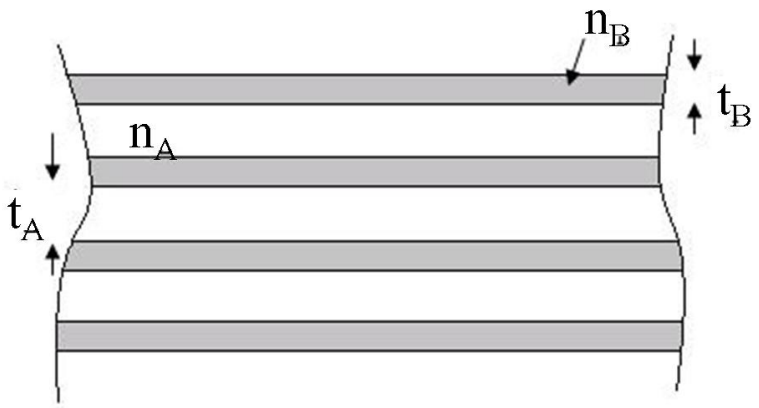

Figure 2.3: A cartoon of a material with form birefringence, consisting of alternating layers of isotropic materials with different indices and thicknesses. It is assumed that $\frac{\lambda}{n_{A,B}} >> t_{A,B}$

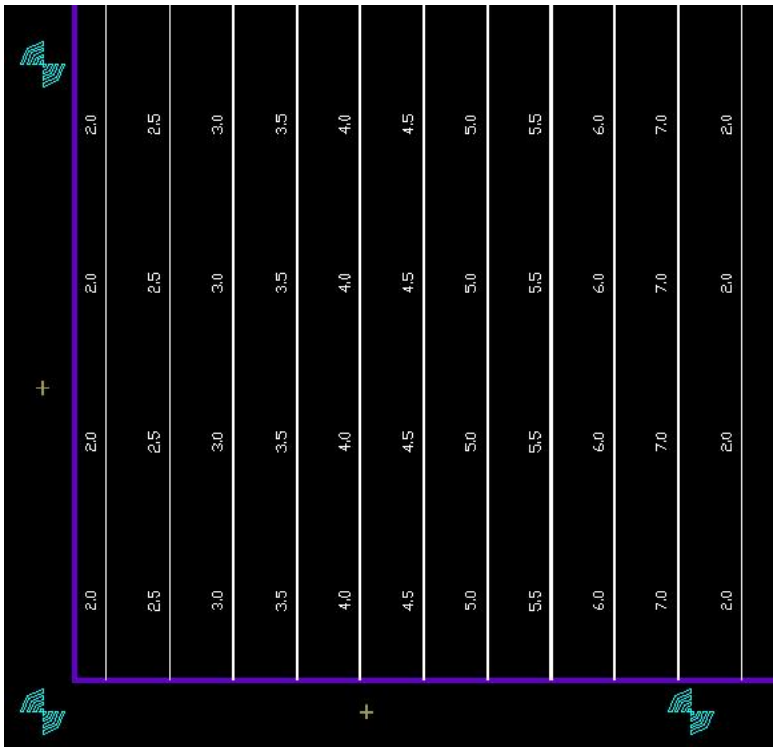

Figure 2.4: A partial schematic of the mask used to fabricate the chip. The chip consists of a series of waveguides of different rib widths (indicated by the numbers, in microns).

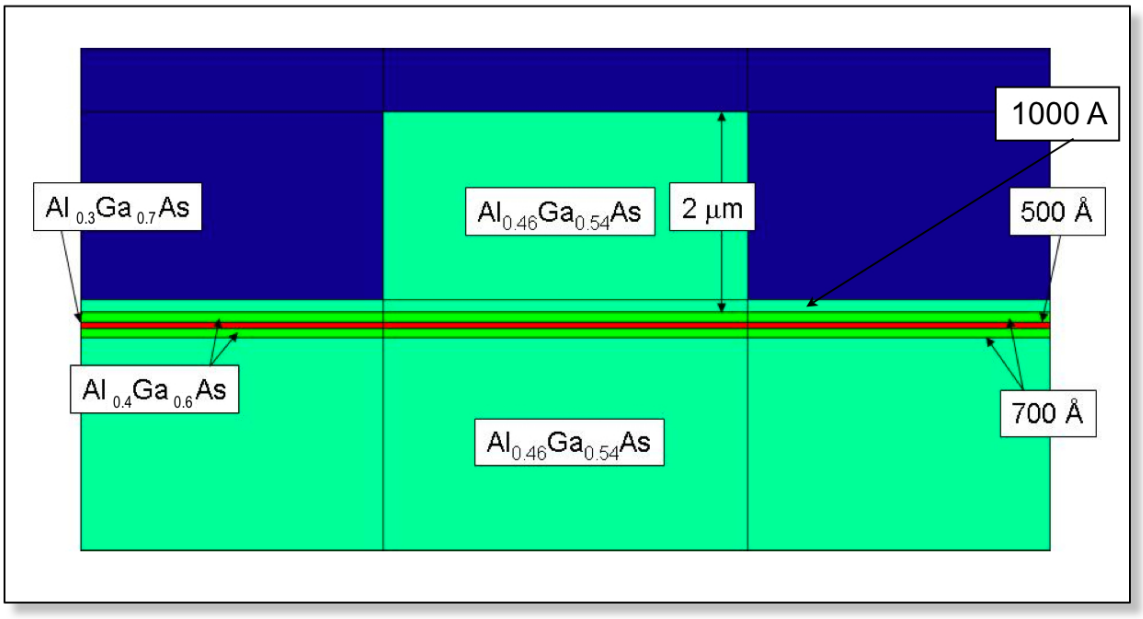

Figure 2.5: The layer structure settled on during modeling, showing a multi-layer birefringent core buried 100 nm beneath the surface. The lateral confinement is achieved by a rib above the core. Varying this rib width changes the overall birefringence of the mode.

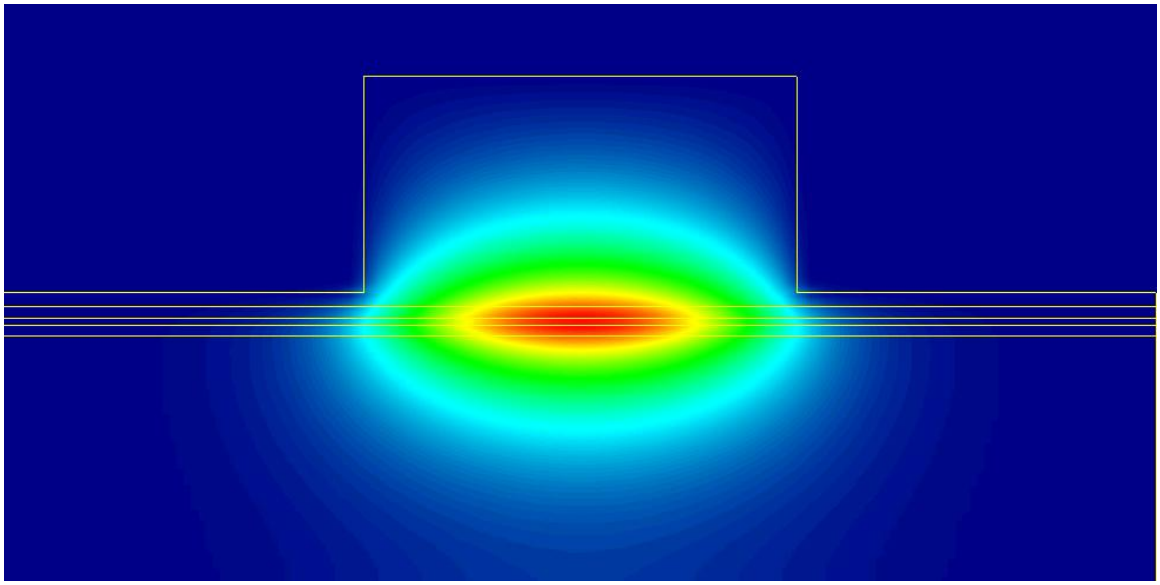

Figure 2.6: An image of the computed lowest order mode generated by OWMS.

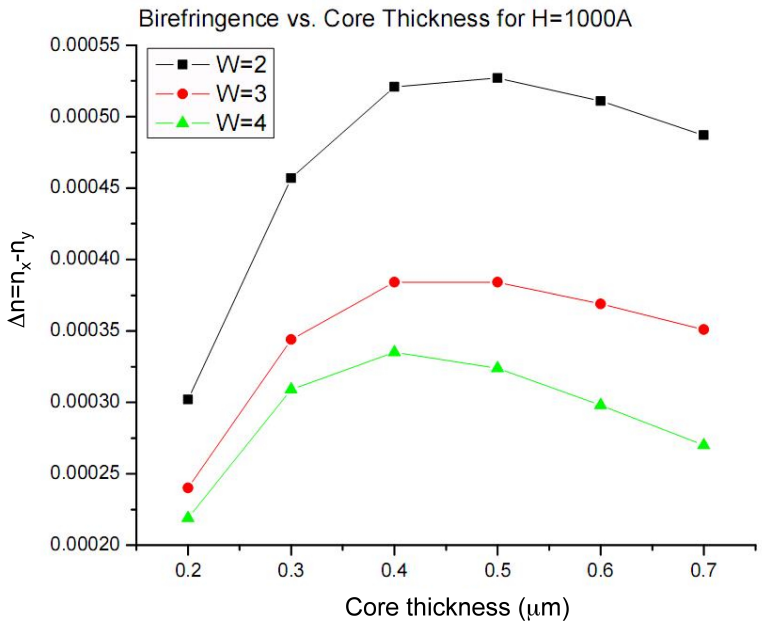

Figure 2.7: A plot of a sample of OWMS modeling results, showing the computed birefringence for various rib widths and core thicknesses. The modeling indicates that we can expect a birefringence in the target range using this structure.

2.3.4 Device fabrication and X-ray analysis

Given the design outlined above, the first step in the fabrication process involved growing the material on a GaAs wafer using MBE[6]. Two superlattice layers were first grown on the GaAs substrate and were used for calibrating the aluminum content and growth rate using Reflection High Energy Electron Diffraction. RHEED is a technique that uses electron diffraction from surfaces to calibrate the number of monolayers present during growth; diffraction energy peaks when the surface is covered with 50% of one monolayer and is at a minimum when the monolayer is completely filled. By monitoring the oscillations in the RHEED energy, one can determine the growth time required for one monolayer of material growth. The defect count in the final growth was approximately 100 per sq. cm.

The material composition was verified using X-ray diffraction. As indicated above, there is some uncertainty in the actual aluminum content achieved during growth, but the *relative* percentages are very precisely controlled. Thus by verifying the aluminum percentage of one layer, one can be reasonably sure that the aluminum content of the other layers is shifted by the same amount. The X-ray data are shown in Figure 2.8. They indicate (as explained in the caption) that the final aluminum content is only 3% higher than intended in the recipe.

The device patterning was performed using a photoresist pattern and a chlorine-

[6]The fabrication of the device was perform by the very adept Dr. Chris Richardson in the state-of-the-art III-V semiconductor fabrication facilities located at the Laboratory for Physical Sciences in College Park, MD. The information in this section is adapted from his report, located at http://www.quantumcollective.org.

based etch in an inductively coupled plasma etch tool. The final devices have widths of 2.15, 2.62, 3.11, 4.7, 5.1, 5.7, and 6.2 μm, as measured by scanning electron microscopy. An SEM image of the 2.5 μm device is shown in Figure 2.9.

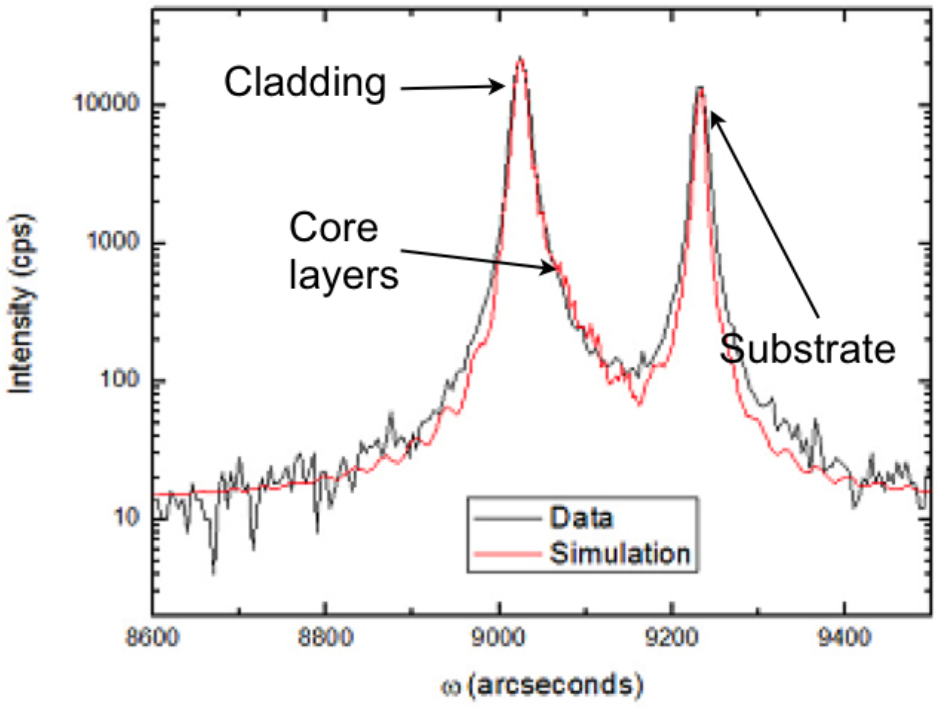

Figure 2.8: X-ray diffraction data from measurements of the final growth. The technique involves measuring the diffracted X-ray power as a function of the angle of incidence on the wafer and essentially determines the lattice constant. The data are interpreted by computing an expected curve based on the intended growth recipe, then modifying the layer contents in the model until it fits reasonably well with the experimental data. The peak on the left is from the cladding layers, which dominate the diffraction due to their relatively large thickness. The location of this peak corresponds to an aluminum content of 49%, 3% higher than was intended. From this we can infer that the core layers are at 33% and 43%.

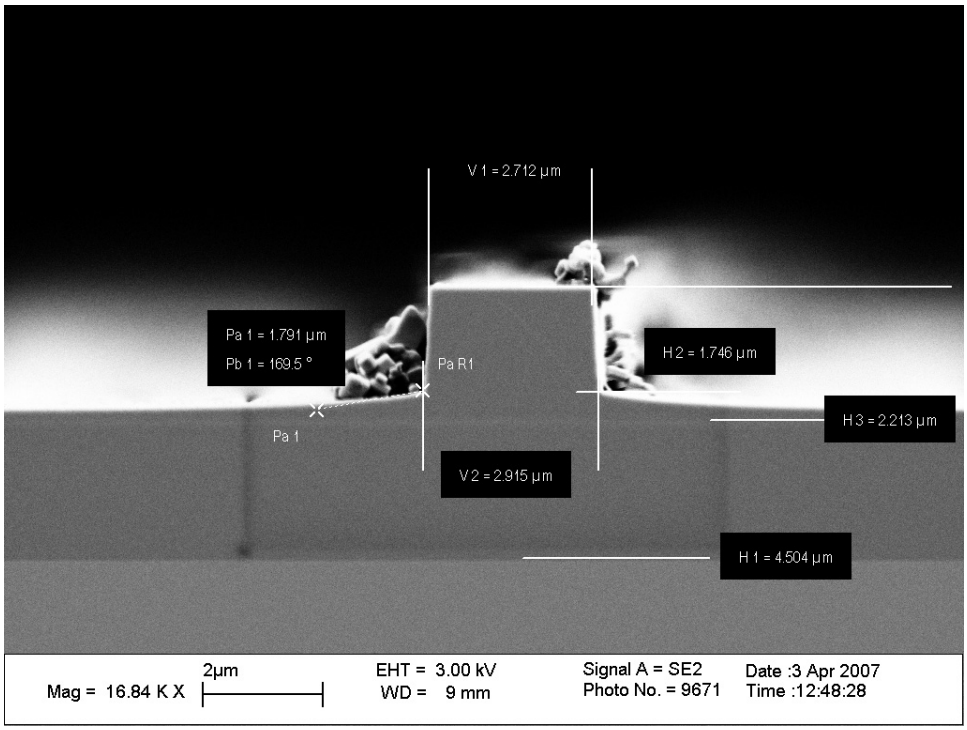

Figure 2.9: An SEM cross-section of one of the devices, showing the measured dimensions of each feature. The core layers and GaAs substrate are visible below the surface. The halo is due to the non-conducting properties of the uncoated material and the dirt on the device can safely be ignored.

2.4 Tests of birefringence, loss, and nonlinearity

2.4.1 Birefringence measurements and further modeling

The first experiment required to validate the operation of the device is a measurement of its birefringence. There are a number of ways to measure birefringence. One involves measuring the Fabry-Perot fringes of the waveguide with reflective facets at both polarizations, measuring the difference in fringe spacing between the two to determine the birefringence [62]. However, because the devices are at most 8 mm long, the fringe spacing is beyond the resolution of most optical spectrum analyzers. Alternatively, one can use a setup similar to a Lyot filter, where the device under test is placed between two crossed polarizers oriented 45^o with respect to the birefringent axes of the device and pumped with a broadband light source. In this configuration, the birefringent waveguide rotates the 45^o polarized light by an amount that depends on the wavelength, allowing more or less light to pass through an analyzer oriented at -45^o. The result is a series of spectral fringes whose spacing depends on the birefringence of the device. If the analyzer is rotated by 90^o, the fringes invert, allowing us to accurately determine the spacing using the crossing points of the two spectra. There is one problem with this method; for such a small birefringence, we only expect to see part of one fringe, making it very difficult to measure the fringe spacing. In order to compensate, we insert a birefringent potassium diphosphide, or KDP, crystal with the same orientation as the waveguide. By measuring the crystal's birefringence independently first, we can then add the waveguide and measure the change. Because the two axes are aligned, the birefringence

is purely additive, and the magnitude of the change in fringe spacing corresponds directly to the birefringence. To understand how this occurs, consider the intensity as a function of the wavelength, determined from the product of the Jones matrices for the setup as shown in Figure 2.10:

$$\begin{aligned} I(\lambda) &= \left| \frac{1}{2} \begin{pmatrix} 1 \\ -1 \end{pmatrix}^T \begin{pmatrix} e^{\frac{i\pi\delta n_0 L_0}{\lambda}} & 0 \\ 0 & e^{\frac{-i\pi\delta n_0 L_0}{\lambda}} \end{pmatrix} \begin{pmatrix} e^{\frac{i\pi\delta nL}{\lambda}} & 0 \\ 0 & e^{\frac{-i\pi\delta nL}{\lambda}} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right|^2 \\ &= 2\sin^2\left(\frac{\pi}{\lambda}(\delta n_0 L_0 + \delta nL)\right) \\ &= 2\sin^2\left(\frac{k}{2}(\delta n_0 L_0 + \delta nL)\right) \end{aligned} \tag{2.53}$$

where δn_0 and L_0 are the birefringence and length of the KDP crystal, δn and L are the birefringence and length of the waveguide, and k is the wavenumber corresponding to the wavelength λ. The resulting spectrum should show fringes that are spaced linearly in the wavenumber, whose period is a sum of the birefringence of the crystal and the waveguide. If we measure the birefringence of the crystal independently, we can subtract it out to determine the birefringence only due to the waveguide using the following expressions:

$$\begin{aligned} \delta n_0 &= \frac{\pi}{L_0 \Delta K_0} \\ \delta n &= \left| \frac{\pi}{L\Delta K} - \frac{L_0}{L}\delta n_0 \right| \end{aligned} \tag{2.54}$$

where Δk_0 is the fringe spacing without the waveguide and Δk is the fringe spacing with waveguide added.

A sample scan of one device is shown in Figure 2.11. Also shown is a scan of the KDP alone. Determining the birefringence from these data is nontrivial. First,

one must replot the data in terms of wavenumber, then read off the position of each point where the crossed- and co-polarized spectra coincide. This gives an accurate determination of the position of each fringe. Plotting that position as a function of fringe number gives a nominally straight line whose slope is equal to the fringe spacing. We can then plot the lines of the various devices, measuring the difference in slope between the the line for the device and the line for the KDP only. The birefringence measurement of the KDP alone agreed to within 10% of the published value in [44]. This is a good validation of the measurement technique. The plots for the various devices under test are shown in Figure 2.12, along with a plot of the determined birefringence of each device. Note that all of the devices 3 μm and wider appear to be multimode, and the measured birefringence matches reasonably well with simulations of higher order modes. The fact that the devices are multimode was confirmed with microscope images of the output facets taken while the devices were illuminated. However, the 2.5 μm device appears to be single mode and has a birefringence of 4.4×10^{-4}, well within the target range of the design. Thus this device appears to be a good candidate with which to proceed.

In order to determine precisely where this device should be phase matched and how well we can control the birefringence, we ran more accurate models of the 2.5 μm device. Instead of using the previous commercial modesolving software, we employed the MATLAB waveguide modesolver of Professor Tom Murphy[18]. In addition to being open source, this code provides scripting ability, enabling us to automate its execution for a series of wavelegnths to generate full dispersion curves. We can also incorporate the temperature-dependent Sellmeier equations discussed

above into the model to obtain a more accurate description of the material. We implemented this code, compiled into an executable, and ran it on a Dell PowerEdge workstation with 12 GB of RAM in order to determine the full dispersion curves shown in Figure 2.15. We modeled two separate devices: The first, as we originally designed it, with the target aluminum contents and feature sizes. The second model was run using the measured aluminum content and feature size from the final 2.5 μm device. As the data shows, the change in aluminum content results in a significant shift in the overall index of the device. However, taking the difference of the two curves shows that the *birefringence* of the device is not significantly different. In fact, if we extrapolate the red curve to the wavelength where the birefringence measurement was made, we see that it matches to within 10%. This indicates that the modeling is a very accurate way to predict device birefringence and shows that one can design a waveguide to a specific birefringence without uncertainties due to growth effects such as strain. This is a very promising result for future work on similar devices and for the modeling capabilities as well.

We can further use these dispersion curves to calculate the phase-matched detuning that we expect. The results of this calculation are shown in Figure 2.16. The two curves correspond to the device as designed and as fabricated and show only a shift of less than 0.5 nm in the phase matched wavelength. This is also a promising result, as it shows that the phase matching wavelength is not terribly sensitive to fabrication tolerances. Judging from this data, we expect to see correlated pairs detuned between 9 and 9.5 nm from the pump wavelength with a bandwidth of 0.4 nm.

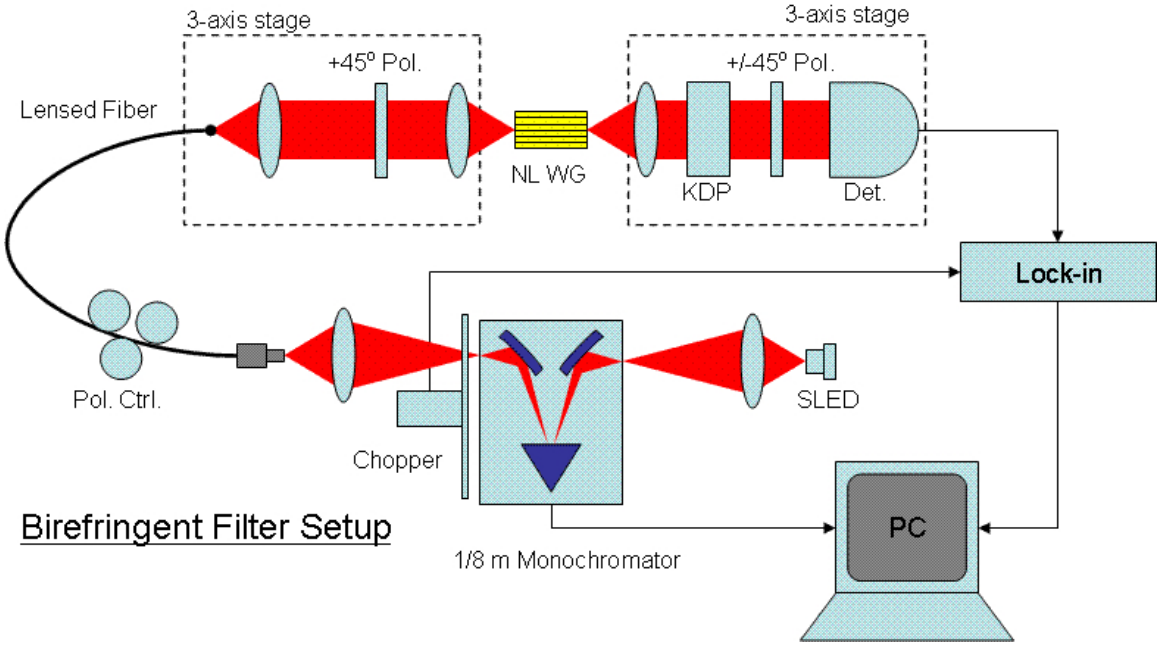

Figure 2.10: The setup used to measure birefringence. The birefringence of the device is determined by observing the change in the fringe spacing between scans with only the KDP crystal and with both the crystal and the waveguide.

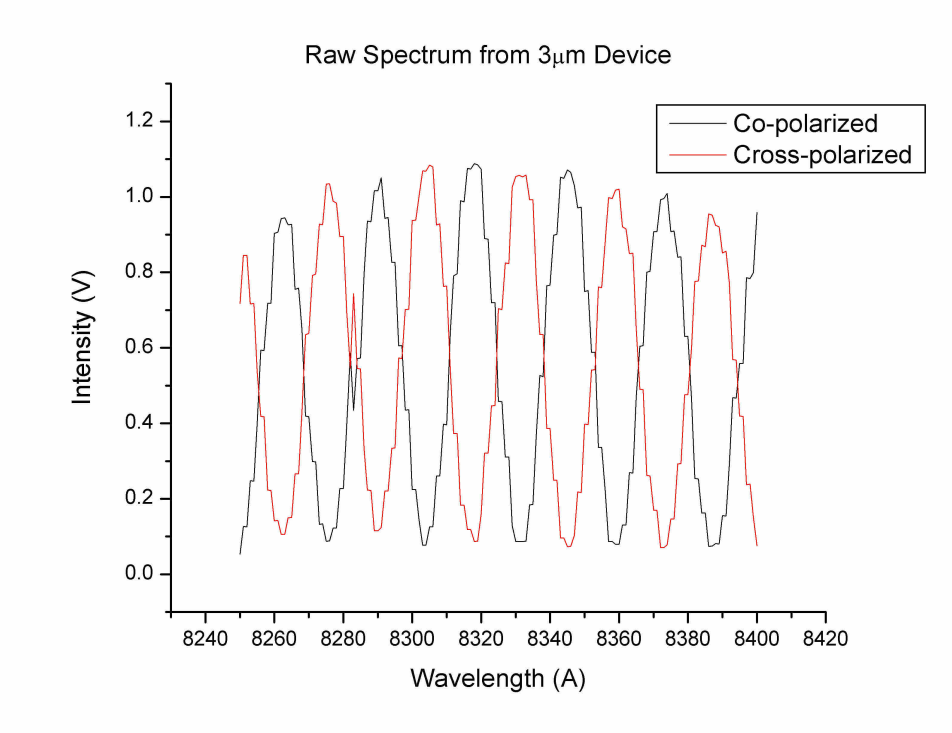

Figure 2.11: A sample of the spectrum of a 3 μm device, showing the fringes with the output analyzer both cross- and co-polarized to the input polarizer.

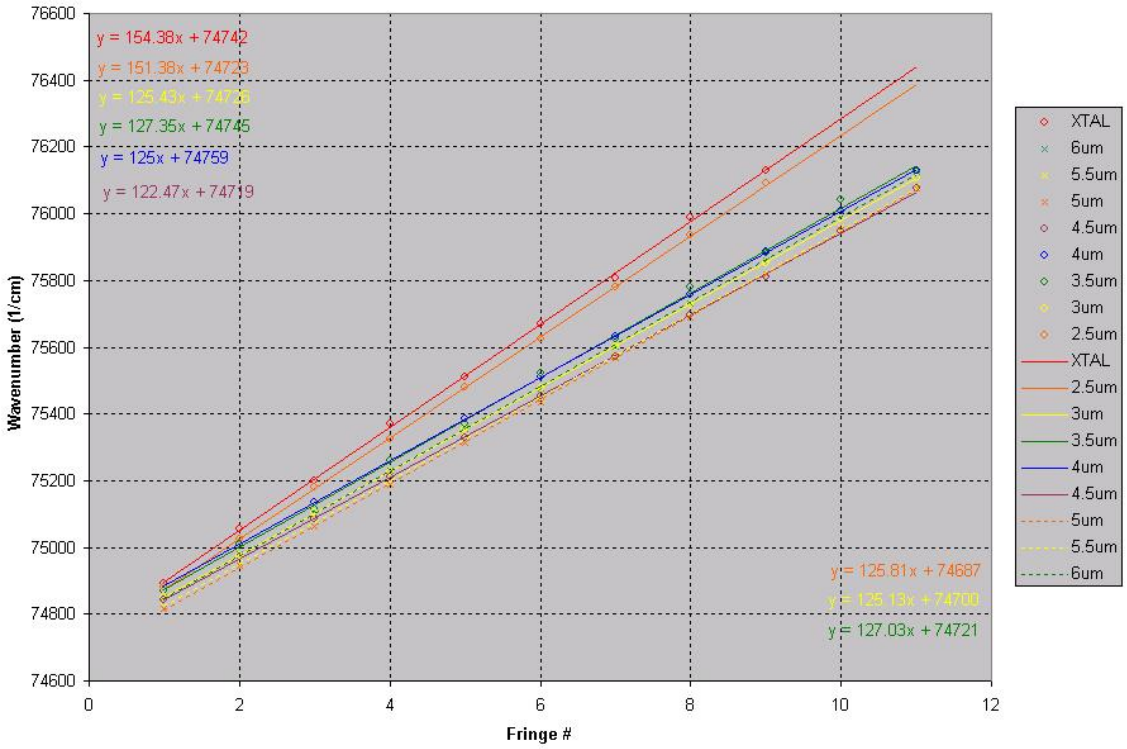

Figure 2.12: Plots of the fringe spacing for a series of devices. Verifying that these results are statistically significant, the slope of the line for the KDP only was measured to be 154.3 ± 1.2, while the $2.5\mu m$ device has a slope of 151.38 ± 0.67, or approximately six standard deviations lower.

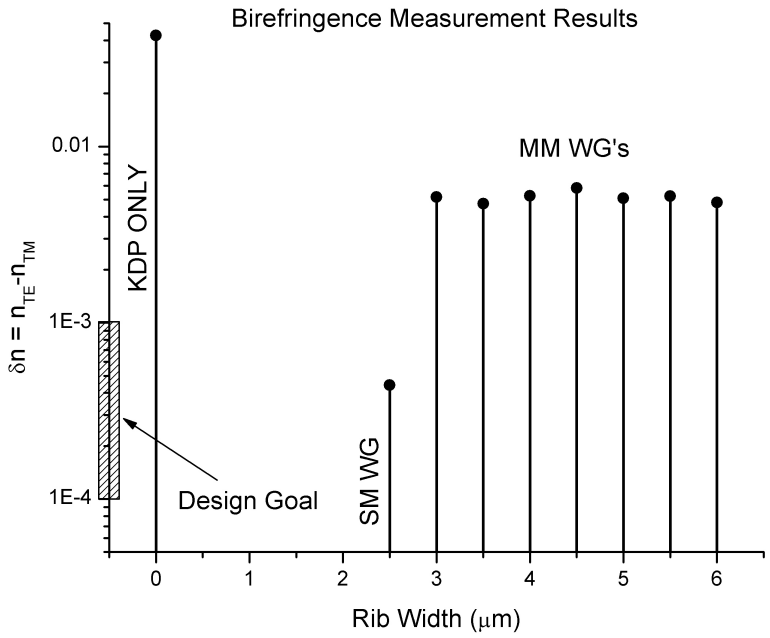

Figure 2.13: The final results of the birefringence measurement, showing that most of the devices are actually multimode with an order of magnitude greater birefringence than the design goal. The 2.5 μm device, however, exhibited a birefringence in the design target range.

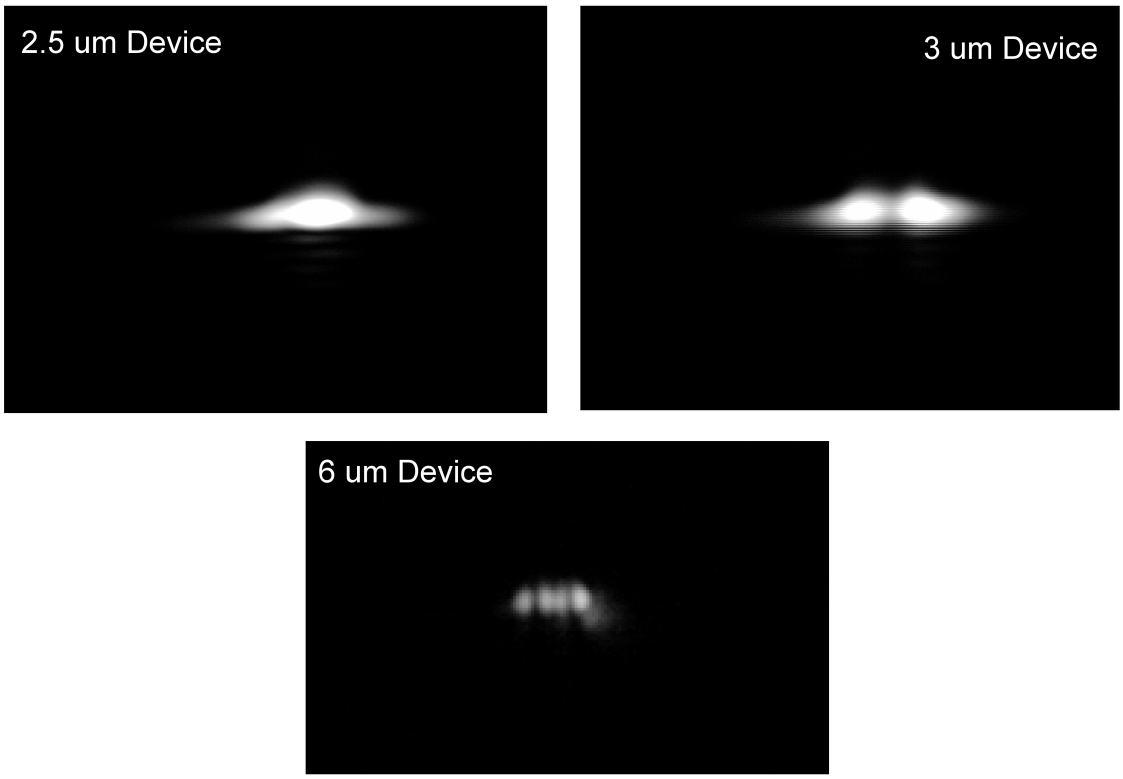

Figure 2.14: Microscope images of the mode outputs of various devices, showing that the devices wider than 2.5 μm are indeed multimode.

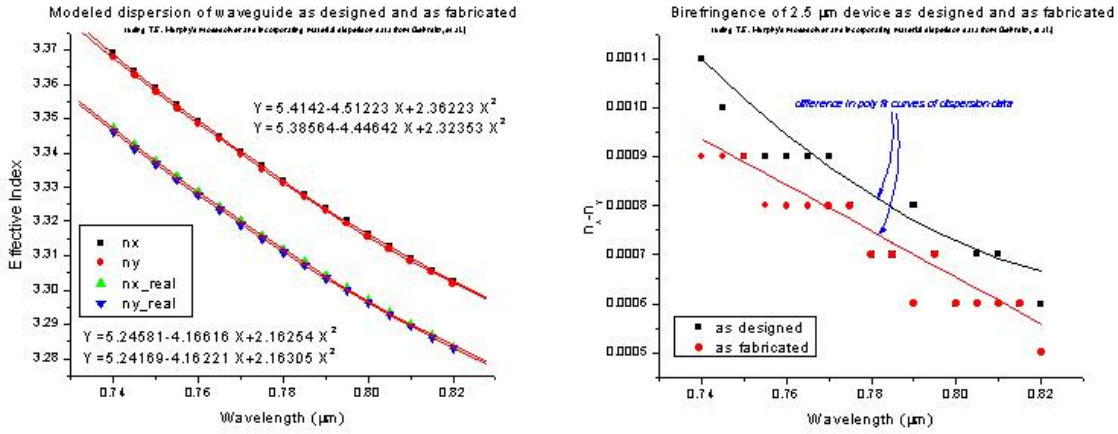

Figure 2.15: A more in-depth look at the modeling of the 2.5 μm device using Professor Tom Murphy's MATLAB modesolver [18] with material dispersion modeling included (see Appendix A). While the variation in aluminum content during growth shifted the dispersion curve, the actual birefringence seemed to be unaffected

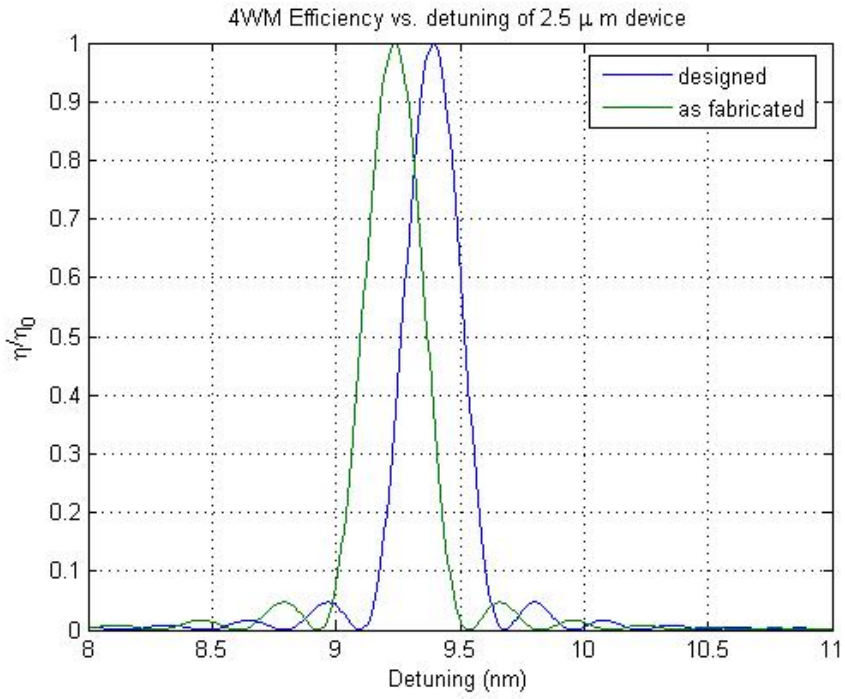

Figure 2.16: A plot of the predicted phase-matching peaks for both the device as designed and as fabricated. We expect to see 4WM phase matched at between 9 and 9.5 nm detuned from the pump, with little effect from fabrication tolerances.

2.4.2 The PPLN pump source

In order to probe the nonlinear properties of this device, we need a high power pump source at 780 nm. While a number of choices exist, we chose a bulk, periodically poled lithium niobate crystal to frequency double a 1560 nm fiber laser[7]. This offers some practical advantages over a titanium sapphire laser mainly in cost and compactness. We obtained a 35 mm long MgO:PPLN bulk crystal with 5 poling periods ranging from 18.0 to 19.0 μm. We chose 19 μm as the poling period primarily because it is the largest and most easily identified track on the crystal. To achieve phase matching at 780 nm and to avoid photorefractive damage of the crystal, we determined using the SNLO software[60] that the crystal must be heated to 176^o C. The first version of the PPLN pump source is shown in Figure 2.17. Data showing measurements of the output are shown in Figure 2.18. The data shows that, when pumped at the optimal wavelength of 1557.03 nm, it is possible to achieve almost 20% conversion efficiency. We further verified that the acceptance bandwidth of the PPLN is 0.345 nm, closely matching our prediction for a 35 mm crystal.

[7]The details of quasi-phase-matched second harmonic generation in PPLN are well known and need not be reproduced here. For details, see [31] or [13]

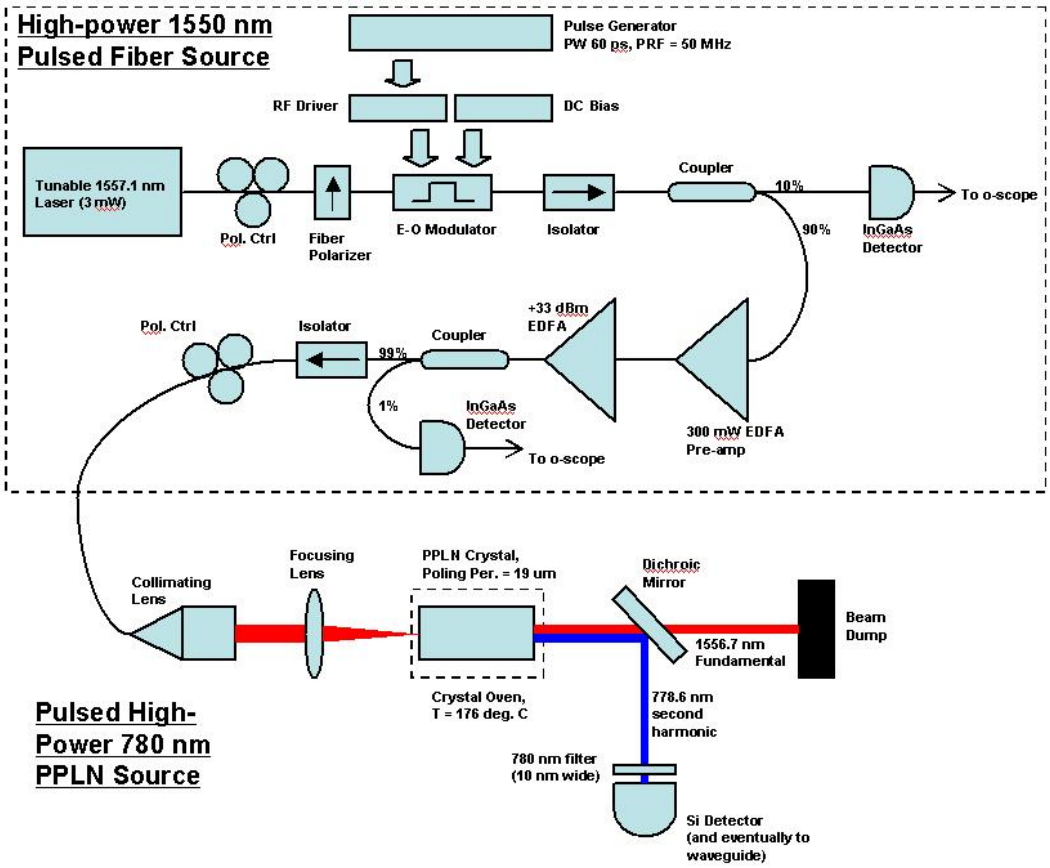

Figure 2.17: A schematic of the first version of the PPLN pump source.

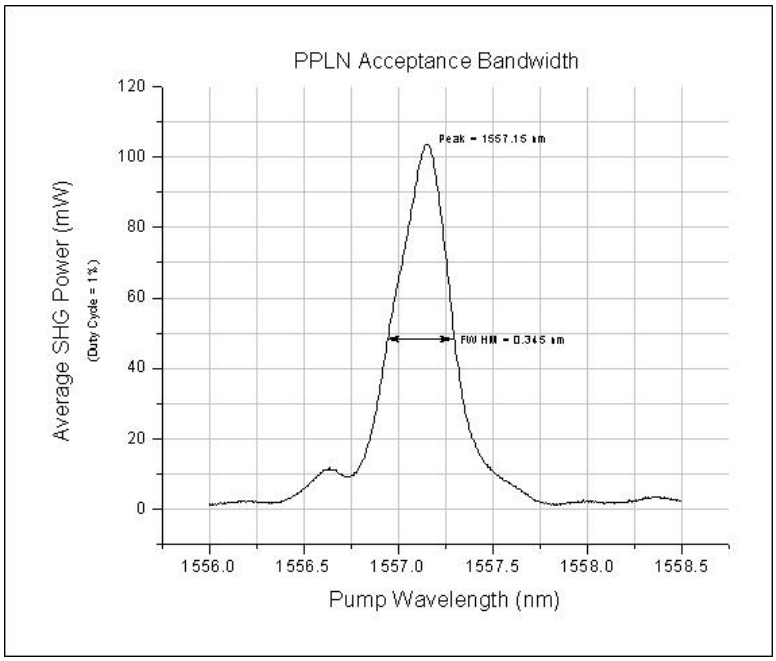

Figure 2.18: A scan of the SHG power vs. the input wavelength, showing secondary phase matching peaks characteristic of nonlinear processes. The peak output power of the PPLN is in the tens of watts, approaching the theoretical limit of approximately 20% conversion efficiency. The measured bandwidth matches to within 1.5% of the calculated bandwidth for a 35 mm long crystal.

2.4.3 Detector calibration procedure

The detector in each measurement, whether a low-noise PIN diode, an amplified APD, or a PMT, was calibrated using the following procedure. Pump light was launched into the device and maximum coupling was achieved. The monochromator was set to the center wavelength of the pump, where nominally the detector and lock-in amplifier were saturated. Four stages of neutral density filters were added to achieve up to 60 dB of attenuation, until the lock-in was no longer saturated. This signal voltage was then compared to a power measurement made directly after the output coupling objective by a calibrated power meter. Comparing the average power (and the calculated peak power using the duty cycle) to the lock-in voltage reading resulted in a calibration curve used to estimate the photon number detected at each wavelength.

2.4.4 Preliminary four wave mixing measurements

Using this high power pump source, we were able to make initial measurements of spontaneous sideband generation. This involved launching the pump beam using free-space optics, as shown in the schematic in Figure 2.19. One of the key components of the coupling system is the free-space coupling stage. Based on a design courtesy of Dr. Richard Mirin at NIST in Boulder, CO, and of staff members at JILA, this configuration allows a free-space beam to be coupled to the device via a moveable objective lens; each mirror is translated along the beam axis independently, letting the input beam alignment to stay fixed to the rest of the setup while

still allowing the objective lens to translate.

Using this setup, we were able to take the spectral scan shown in Figure 2.20. This scan indicates spontaneous sideband generation 8.5 nm detuned from the pump, within about 10% of the expected detuning. However, before we could perform any further tests to validate these results, we caused irreparable damage to the input facet of the device due to too much input average power. Microscope images of this damage are shown in Figure 2.21.

It was initially uncertain whether the damage was being caused by thermal heating of the device, an effect that is sensitive to average power, or by extremely high field strengths due to high peak power. Further tests later confirmed that the damage was indeed due to thermal effects resulting from excessive average pump power.

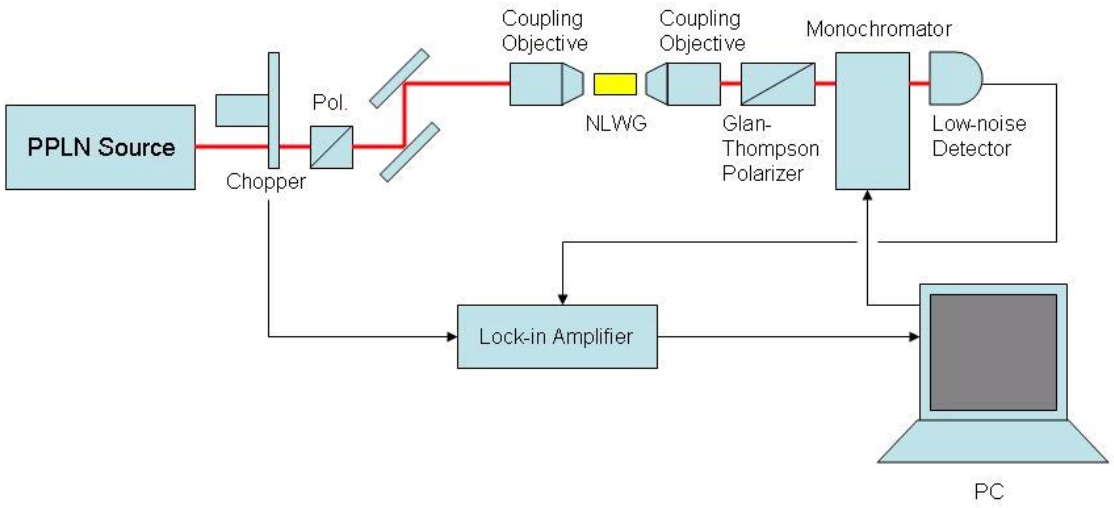

Figure 2.19: A schematic of the setup used to look for spontaneous 4WM. The free-space coupling stage is not shown, but allows for a free-space beam to stay aligned to the source while the coupling objective is moved to align to the waveguide.

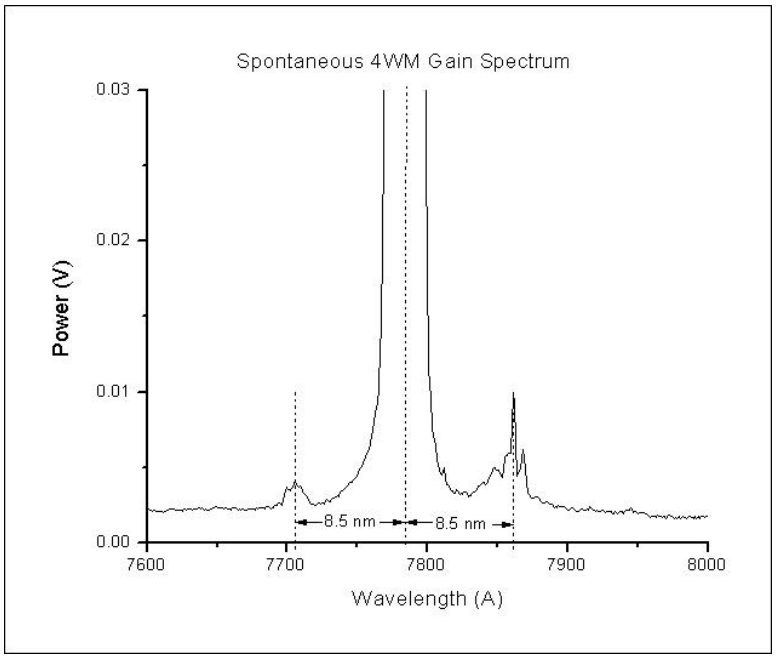

Figure 2.20: A preliminary scan of the waveguide output, showing spontaneous sideband generation at 8.5 nm detuned from the pump. Before we could further investigate these peaks, the input coupling face of the waveguide was destroyed by the high-power pump pulses.

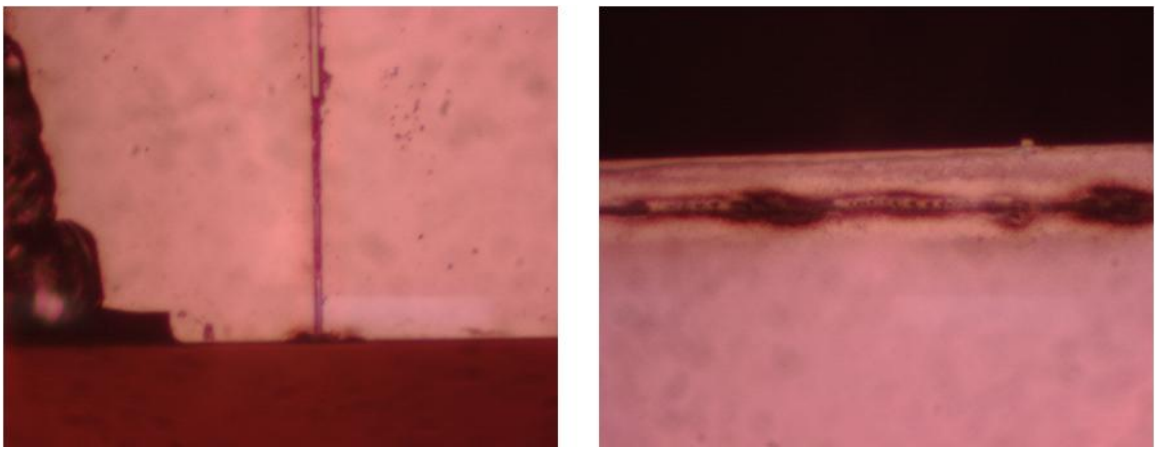

Figure 2.21: Microscope images of the damaged device. High average power caused thermal heating in the chip, resulting in damage to both the input facet and the rib.

2.4.5 Modified source and subsequent four wave mixing measurements

In order to avoid damaging the device, we were forced to modify the pump source in order to achieve nominally the same peak pulse power while reducing the overall average power. One attempt to do this involved using a mode-locked laser in place of the CW tunable laser and modulator to reduce the duty cycle by an order of magnitude. While this did manage to decrease the average power and avoid damaging the chip, it did have two undesirable side effects. First, the mode-locked laser produced 200 fs pulses, resulting in a bandwidth that was significantly wider than the 0.345 nm acceptance bandwidth of the PPLN. This caused a significant loss in conversion efficiency but, more importantly, resulted in an extremely noisy pump signal. A plot of the output spectrum of the pump is shown in Figure 2.22. Even after using a grating to filter the output, there was still significant noise on the single-photon level where our detector is sensitive. The mode locked laser was also significantly more unstable than the pulse-carved CW source, At low conversion efficiency, the PPLN output is sensitive to the square of the input power (the efficiency is dependent on the pump power itself), so the SHG process magnified the instabilities in the mode-locked laser, resulting in a source that was not stable even long enough to make a single spectral scan.

The final configuration of the source involved returning the the CW pulse-carved scheme, but driven with an external oscillator at a lower duty cycle, as shown in Figure 2.23. Because the duty cycle was significantly lower, we required

a more sensitive detection system. After trying a fast, amplfied APD from Menlo Systems that allowed us to lock in at the fast pulse frequency rather than the slow chopper frequency, we settled on a photo-multiplier tube that is sensitive down to the single-photon level. Because the detector still had a bandwidth of 12 kHz, it still required that we chop the beam at low frequency. However, this was compensated by the high sensitivity of the detector; upon calibrating the detector, we found it to be sensitive on the single photon level with integration times on the order of 100 ms, as shown in Figure 2.24.

The spectral scans shown in Figure 2.27 depict data acquired with this modified pump source and a second 2.5 μm device. They show that there are features of interest at exactly 9.5 nm detuned from the pump, the shift predicted by the simulation data. While there are other spectral features of unknown origin, the plot in Figure 2.27 shows that only the features detuned 9.5 nm from the pump exhibit nonlinear power dependence. Furthermore, the features exhibit the expected polarization behavior - namely when a scan is taken with an extra polarizer oriented at the pump polarization, the features are less prominent. When the extra polarizer is rotated by 90^o, the features become stronger in power.

One feature to note in the nonlinear spectral scan data is that the red-shifted peak appears to be weaker than the blue-shifted peak. One possible explanation for this is the wavelength dependence of the 2PA coefficient (discussed in more detail in a later section). Because of the thermal distribution of the density of states in the conduction band, the 2PA coefficient actually peaks when the photon energy is equal to approximately 60% of the gap energy [56] (the 2PA coefficient is necessarily

zero when the photon energy is less than half of the gap energy). Because we are pumping the device at roughly 80% of the gap energy of the core material, we expect that the 2PA coefficient should be smaller when detuned blue from the pump and larger when detuned red from the pump. This is consistent with our observations in Figure 2.27.

2.4.6 Expected 4WM efficiency and corresponding observations

Using the nonlinear index from [33], we are able to make a zeroth order prediction regarding the efficiency of the nonlinear process. Figure 2.31 shows a plot of the expected parametric gain versus the input power for a 2.5 μm, 8 mm long device. This prediction is based on the co-polarized value of the nonlinear susceptibility tensor. We can use this plot to calculate the expected pair generation rate using the following expression [70]:

$$\frac{dN}{dt} = \eta \left|\gamma P_0 L\right|^2 \Delta\nu\tau R \tag{2.55}$$

where η is the detection efficiency, $\Delta\nu$ is the phase-matched bandwidth, τ is the pulse duration, and R is the pulse repetition rate. The calibration procedure accounts for the detector efficiency, so we can safely ignore the factor of η. We can then divide by R to obtain the predicted number of photons generated per pulse:

$$\rho = \left|\gamma P_0 L\right|^2 \Delta\nu\tau \tag{2.56}$$

Figure 2.28 shows a calibrated scan with what seems to be spontaneously generated sidebands with 400 mW of input pump power. At this pump power, we expect from

Figure 2.31 that the parametric gain is greater than 0.1. Since the pulse bandwidth is approximately one tenth of the phase-matched bandwidth of the device, the product $\Delta\nu\tau$ is approximately 10. Hence we would expect to see on the order of $\rho = 1$ photon/pulse in the spontaneous sidebands. However, we observe less than 0.1, indicating that the observed gain is reduced by an order of magnitude. Since the gain is proportional to the square of the susceptibility, we can conclude that the ratio $\frac{\chi^{(3)}_{xxyy}}{\chi^{(3)}_{xxxx}} < \sqrt{0.1} \approx 0.3$. This poses a significant problem, as it implies that the power required for efficient correlated photon generation could be too great for practical operation. It is, however, consistent with observations for other zincblende semiconductors, as outlined in [21].

2.4.7 Linear and nonlinear loss measurements

For the sake of thoroughness, it is important that we characterize both the linear and nonlinear loss of the device. Because we are operating at greater than half of the bandgap energy, we expect to see significant two-photon absorption. In addition, the correlation of the generated photon pairs is very sensitive to linear loss in the device, so it is important to verify that the linear loss is relatively low.

In order to estimate the linear loss, we chose to perform a simple yet rather crude experiment. We pumped the device with rather high power and measured the change in the out-of-plane scattered light levels using digitized microscope images. A sample of the data is shown in Figure 2.29. By measuring the rate of decay, and assuming the impurities from which the light is scattered are uniformly distributed,

we estimate the device to have a linear loss of between 3 and 6 dB/cm. The overall insertion loss, including coupling efficiency, was measured to be 13 dB.

In order to measure nonlinear loss, we performed another simple experiment where we measured the output power versus the input power. The plot in Figure 2.30 shows that the transmission is dependent on the input power, indicating measurable nonlinear absorption. We can use these data to calculate the 2PA coefficient in the following way: Consider the differential equation governing the intensity of the pump as it propagates through the medium,

$$\frac{dI}{dx} = -\alpha I - \sqrt{2}\beta I^2 \tag{2.57}$$

where the first term represents the linear loss and the second term represents the intensity-dependent nonlinear loss resulting from 2PA given a gaussian pulse. Directly integrating both sides from $z = 0 \cdots L$ gives us

$$I(L) = \frac{I(0)}{e^{\alpha L} + \sqrt{2}\beta I(0)\frac{e^{\alpha L}-1}{\alpha}} \tag{2.58}$$

If we define an effective length as $L_{\text{eff}} = \frac{e^{\alpha L}}{\alpha}$, where $\lim_{\alpha \to 0} L_{\text{eff}} = L$, then we can rewrite the expression in 2.58 as

$$T^{-1} = \frac{I(0)}{I(L)} = e^{\alpha L} + \sqrt{2}\beta L_{\text{eff}} I(0) \tag{2.59}$$

Assuming an effective mode area of 2.5 μm^2, we can replot the data in Figure 2.30 in terms of the inverse of the transmission versus the input intensity. This gives us a line, as shown in Figure 2.30, where the y-intercept represents the exponential of the linear loss times the device length and the slope represents the nonlinear loss times the effective length. Since we do not have an accurate determination of the

effective length of the device, we can only estimate the range of values of the 2PA coefficient. Given the results of both loss experiments, we can estimate that the 2PA coefficient of the device is between 3 and 25 cm/GW. This range of values is of the same order as measured for similar materials [67].

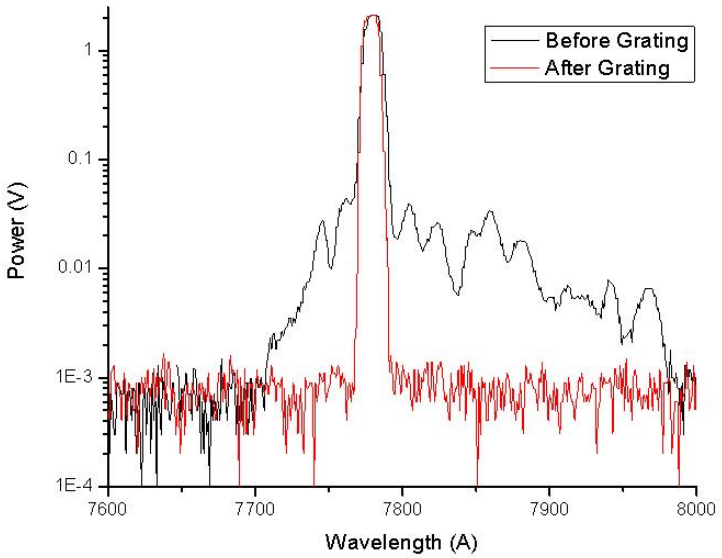

Figure 2.22: A plot of the output spectrum of the PPLN before and after a grating. The broadband nature of the mode-locked laser caused significant light to be generated even outside the quasi-phasematched bandwidth. Even after the grating, significant structure was observed at the single-photon level.

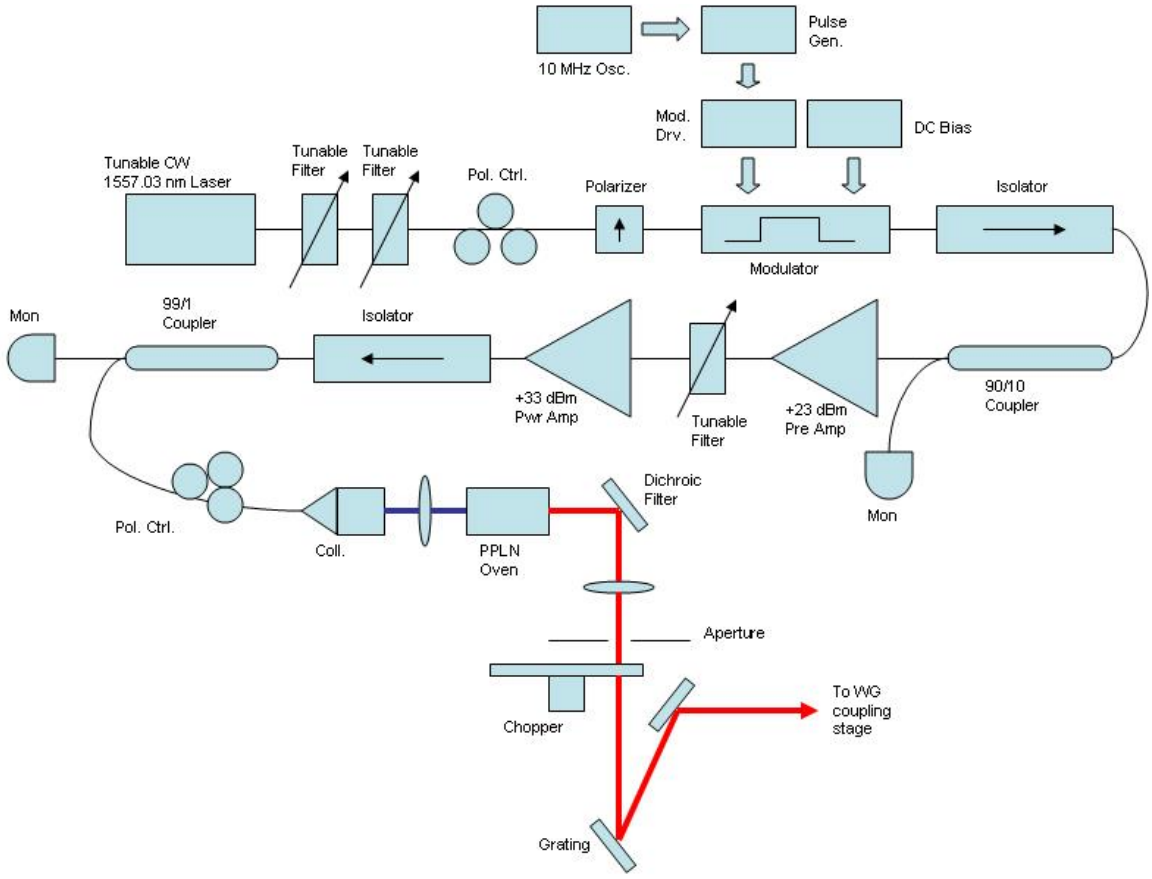

Figure 2.23: A schematic of the final waveguide pump source. The seed is still a CW tunable laser with an external Mach-Zender modulator. The duty cycle is reduced to 0.8% using an external oscillator.

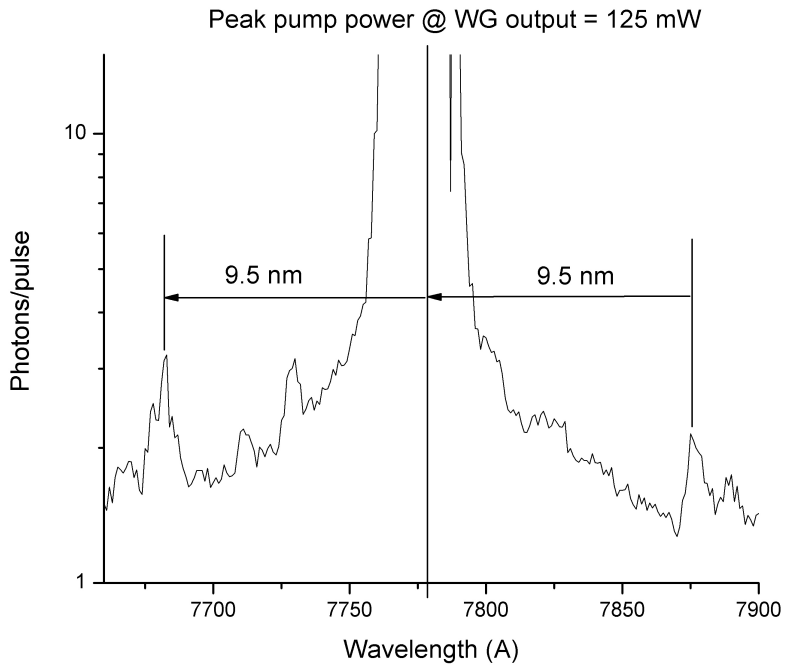

Figure 2.24: A spectral scan of the waveguide output showing features at 9.5 nm detuning. Note the near-single-photon sensitivity.

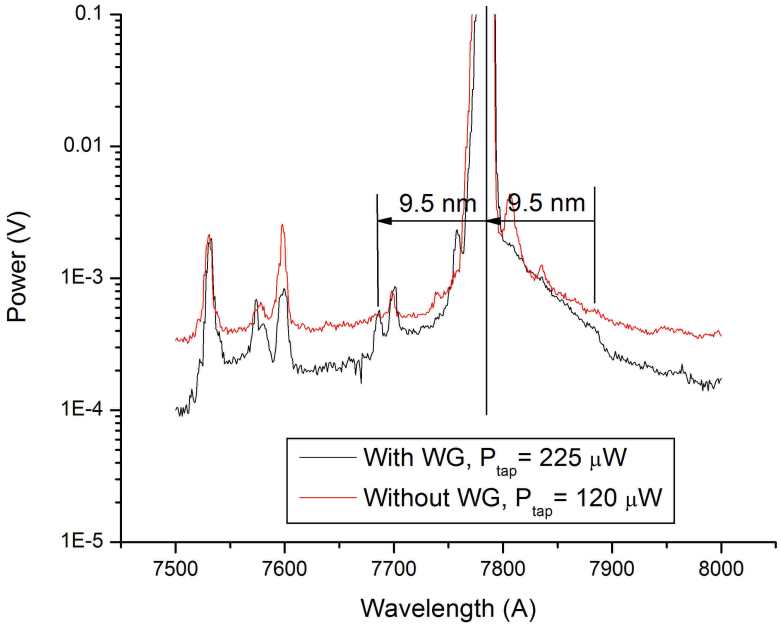

Figure 2.25: Spectral scans of the pump light without the waveguide and of the waveguide output overlaid. Most of the structure is on the pump beam itself, but note that there are significant differences in the spectra around 9.5 nm detuned from the center.

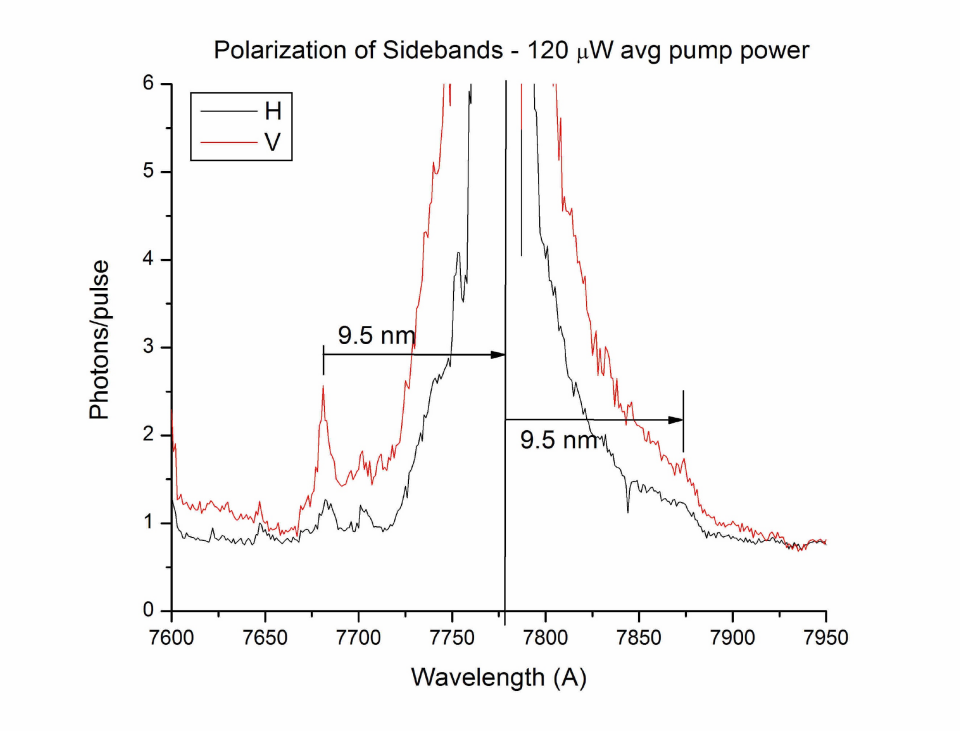

Figure 2.26: Two spectral scans taken with an extra sheet polarizer on the output, showing the proper behavior for birefringent phase matching. The sidebands are more visible when the sheet polarizer is aligned to the fast axis of the device, while the pump remains polarized on the slow axis. Note that the pump beam is so broadened that its wings overlap with the expected sidebands.

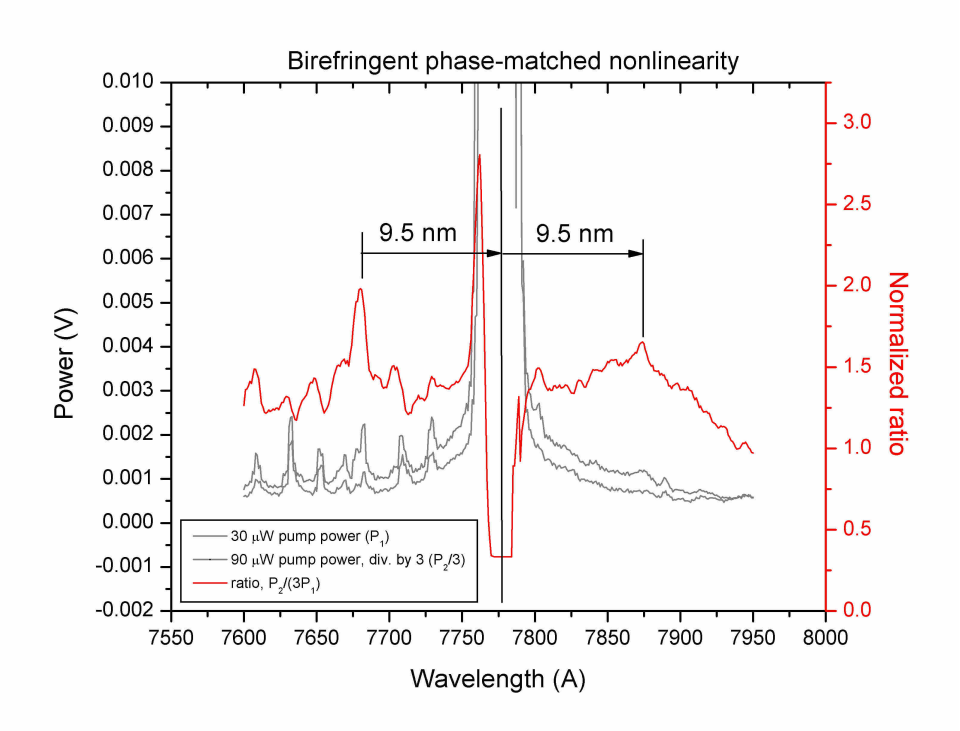

Figure 2.27: A plot of data indicating phase-matched nonlinearity at the expected detuning. Two scans were taken, one at high pump power and one at lower pump power (reduced by one third via a wire mesh placed before the input facet). The ratio between the two shows nonlinear power dependence at 9.5 nm detuned from the pump beam

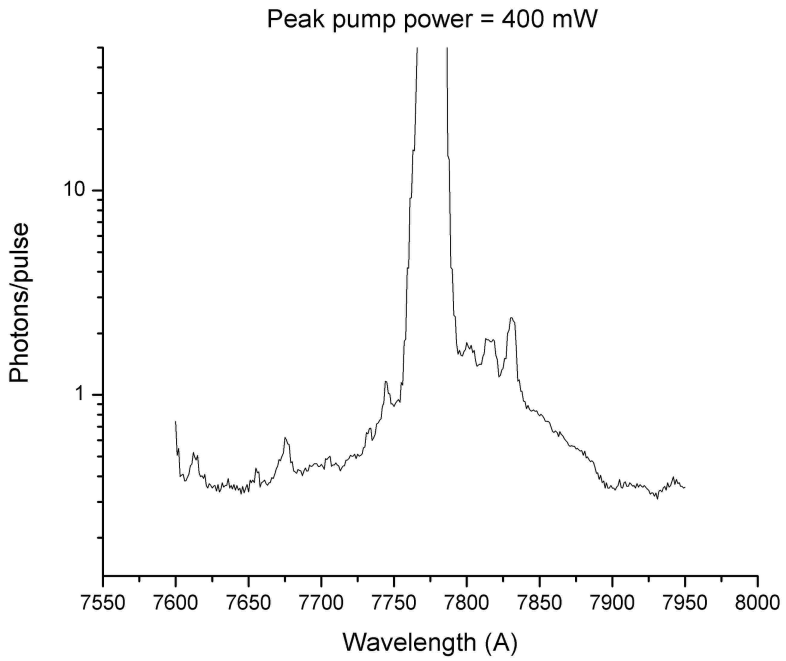

Figure 2.28: A spectral scan taken at higher pump power. There are still features at 9.5 nm detuning but their strengths are much smaller than expected. Using the published value for n_2 and the fact that the pulse bandwidth is one tenth of the phase-matched bandwidth, we would expect to observe approximately 1 photon/pulse on average. The plot shows an order of magnitude less.

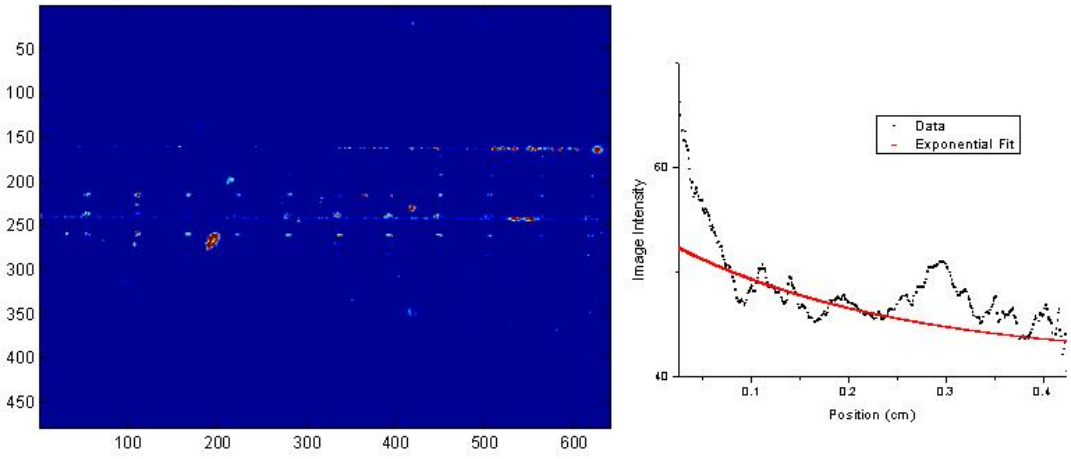

Figure 2.29: To estmate the linear loss, we fit an approximate exponential curve (left) to a digitized image of the out-of-plane scattering from the waveguide (right). While crude, this technique is minimally invasive and provides some indication of the order of magnitude of the device loss.

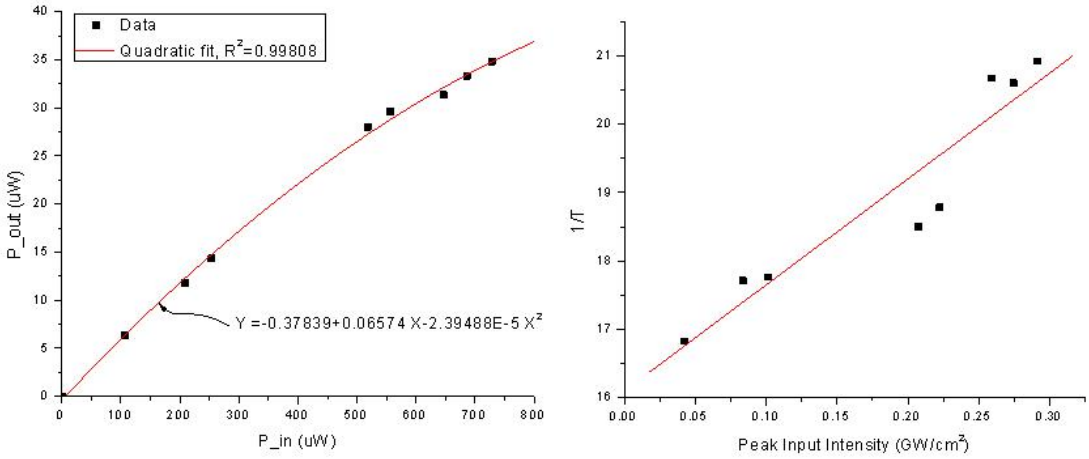

Figure 2.30: Plots of output power vs. input power, showing appreciable 2PA. By replotting the data in terms of the inverse of the transmission, we can estimate the 2PA coefficient.

2.4.8 Practical implications of observed inefficiency

As stated above, the observed nonlinear gain appears to be an order of magnitude smaller than expected. While the origin of the inefficiency is unknown, we can estimate its implications. For example, consider the figure of merit outlined above. If we assume for the moment the best case scenario, where the nonlinear index is equal to the expected value of 3×10^{-17} m^2/W (rather than 70% less, as was observed), and make a reasonable assumption that the 2PA coefficient is $\beta = 20$ cm/GW, the figure of merit defined above comes out to be 1.1, only slightly greater than unity. If the 4WM gain is less than expected, especially by an order of magnitude, then the figure of merit will be reduced significantly below 1 and 2PA will dominate.

In addition, the reduced 4WM gain indicates that pumping the device hard enough to generate appreciable pairs will require operating the device in the regime where damage previously occurred. Previously, damage was observed when the duty cycle was 1% and the peak power was around 10 W at 780 nm. Assuming the gain is an order of magnitude lower, one would need approximately three times the pump power to achieve the pair generation rate initially expected. Even assuming one shortens the pulse length by 90% to match (but not exceed) the acceptance bandwidth of the device, the duty cycle constraints would limit the pulse repetition rate to less than 300 MHz. While this is not necessarily slow, it offers no improvement over any of the other pair generation schemes and does not approach the GHz rate that broadband QKD applications demand.

2.4.9 Pump probe measurements and thermal effects

A conclusive test to measure the parametric gain of the nonlinear interaction would be to introduce a probe beam at the signal wavelength, cross-polarized to and simultaneous with the pump beam. This would allow one to determine the gain, given the power of both the pump and the probe. We implemented such an experiment using a CW DFB laser at 770 nm that was tunable via temperature control of the diode. This experiment yielded inconclusive results pertaining to 4WM, but did indicate that there are thermal effects that may be interfering with the nonlinear process of interest.

The data in Figure 2.32 shows both the magnitude and phase of the lock-in signal used to create the spectral scan. As is evident from the phase data in the spectral region of the probe beam, the probe is *demodulated* by the pump. That is, the presence of a burst of pulses from the pump causes a *decrease* in the strength of the probe beam. The cause of this is unknown, but there are at least two possibilities. First, it may be a slow, thermal effect, where heating in the waveguide due to the strong pump causes the guiding properties of the waveguide to change and the subsequent leakage of the mode carrying the CW probe signal. Alternatively, it could be a result of two photon absorption between the pump and probe. There is some indication that the cause is the former; a decrease in the strength of this interaction was noted when using the fast APD with an RF lock-in amplifier oscillating at the PRF of 20 MHz when compared to the scans with the slower detector and the 200 Hz mechanical chopper. Hence it seems that there

is some unknown thermal effect causing a change in the guiding properties of the waveguide in the presence of strong pump pulses. This is disconcerting, as it could easily imply that the phase-matching condition or even the nonlinearity itself is modified in the presence of a strong pump beam.

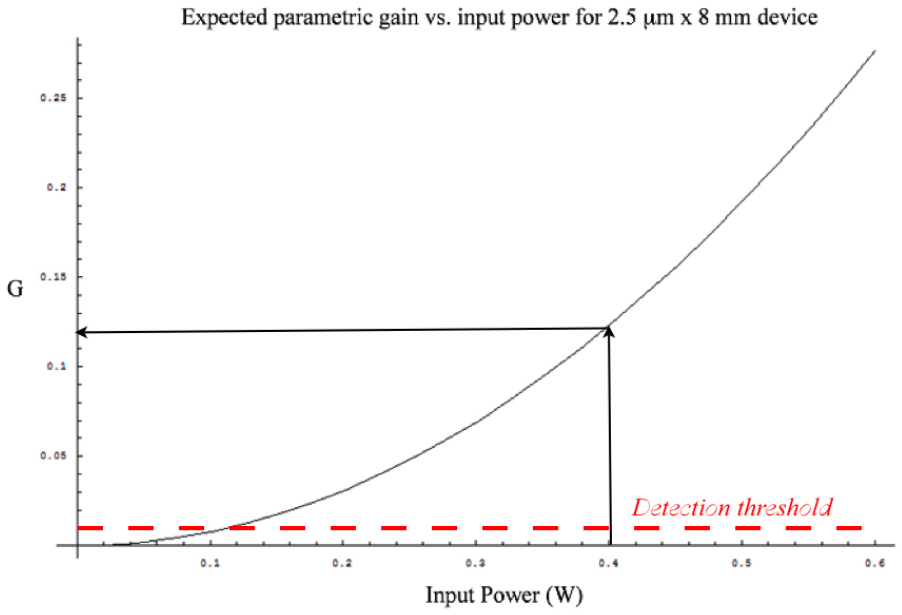

Figure 2.31: A plot of the expected parametric gain using published values of n_2. The actual gain curve depends on the other nonlinear susceptibility tensor element, the relative strength of which is unknown. The arrows indicate the expected gain if the cross-polarized susceptibility were as strong as the co-polarized one. Since the observed gain was barely above the detection threshold of 0.01 photons/mode, we can conclude that the cross-polarized susceptibility tensor element is reduced by 70% of the expected value.

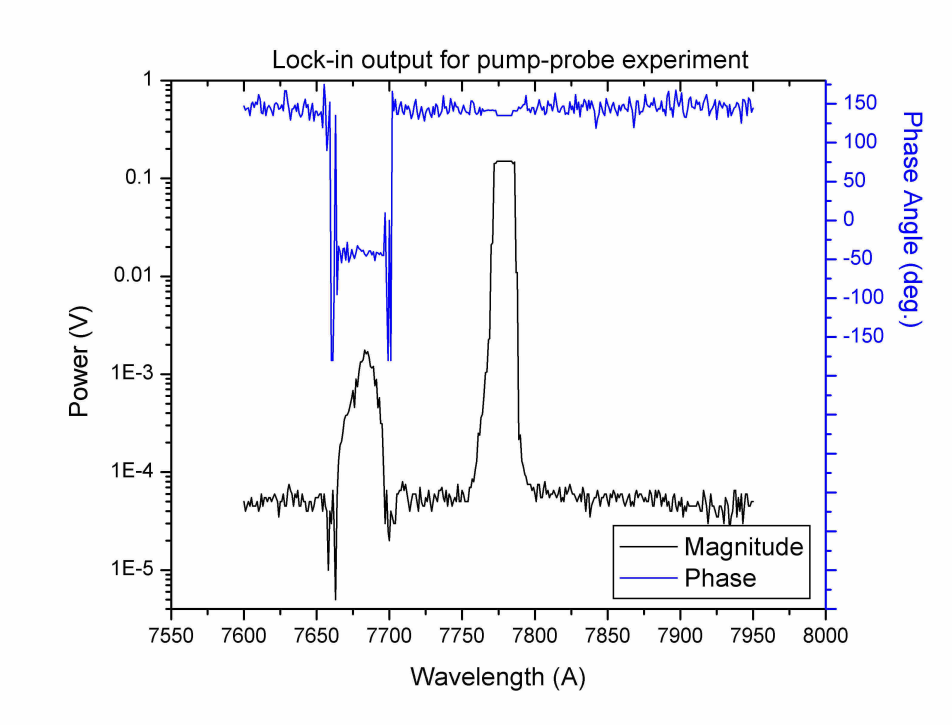

Figure 2.32: A spectral scan of the pump probe, showing both the magnitude and phase of the lock-in signal. The 180^o phase-shift at the lock-in detected signal indicates that the pump is causing the signal to decrease rather than increase. The origin of this effect requires further investigation.

2.5 Results and further work

2.5.1 Current results and open questions

The goal of this project is ultimately to demonstrate the feasibility of using four-wave mixing in a birefringent waveguide as a source of entanglement for applications such as QKD. In order to do so, a number of questions must be answered. First and foremost, it must be demonstrated that one can fabricate a waveguide with a controllable and predictable birefringence. We have shown the ability to accurately predict the waveguide birefringence, and we have demonstrated that one can fabricate a waveguide with a specific birefringence, even in the presence of uncertainties due to strain during material growth. We have fabricated the first set of devices and used them to measure such basic waveguide properties such as insertion loss, device loss, and nonlinear loss. We have also devised a straightforward way to accurately measure small birefringence. Finally, we have shown that there is indeed birefringent phase matched nonlinearity occurring at the detuning predicted by the model. However, we have shown that the strength of the cross-polarized nonlinear interaction is significantly weaker than we had originally expected, requiring potentially unreasonable pump power levels for efficient correlated photon generation.

These preliminary results beg some very fundamental questions. First, it is still unclear what role the group velocity dispersion plays in the phase matching. We know that the GVD plays a key role in the birefringent phase matching and that is is dominated by the material dispersion in devices whose feature size is larger than the wavelength[65]. However, we have no clear measurement of the

GVD of the device, forcing us to make an educated guess at the phase matched wavelength. Our measurements only predict the average birefringence of the device over a small spectral range. For a more thorough characterization of the ability to control phase matching, it is important to come up with a way to measure the GVD. One possibility is to use a phase shift technique similar to that used in Costa, et al., outlined in [15]. In this setup, a vector voltmeter is used to measure the phase shift of slowly-modulated, broadband optical signal at different wavelengths. The only uncertainty in the application of this technique is whether the device is long enough to introduce a phase shift larger than the vector voltmeter's phase precision. However, the technique is simple enough that it warrents further investigation and is compatible with the same setup used to measure the birefringence of the device.

Another issue related to the birefringence is a lack of understanding of the uniformity of the birefringence of the waveguide. While the measurement described above measured the overall birefringence, there was no information about the variation in the birefringence over the length of the device. Local strain or defects in the material may contribute to a variation in the birefringence that may be masked in the measurement but result in very poor phase matching for the nonlinear process. Currently we cannot determine the uniformity of the birefringence along the length of the device. One possible way to measure this would be to employ a technique similar to one used to measure the linear loss. That is, we can fabricate a chip with devices of the same dimensions as the one currently used, but with a series of different lengths, as shown in Figure 2.33. A chip with this layout would be useful for both measuring the uniformity of the birefringence and for directly measuring

the linear loss.

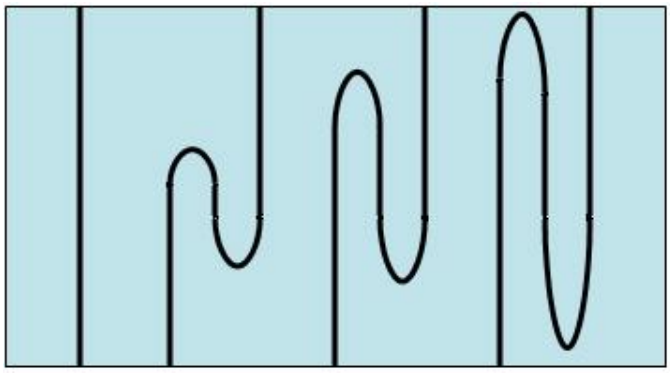

Figure 2.33: A concept of a chip with devices of various lengths. Because the U bends contribute constant loss, this chip is useful in measuring the linear loss of the device more precisely. In addition, by measuring the birefringence of the various length devices, we can glean some idea of the uniformity of the birefringence over the length of the device.

2.5.2 Further nonlinearity and correlation measurements

The next task is to quantify the strength of the nonlinear interaction. While the current pump setup is rather limited, with some modification we can create a pump-probe setup that can directly probe the strength and time dependence of the parametric gain and nonlinear loss. We propose using a second PPLN crystal phase-matched at a wavelength 9.5 nm shifted from the current pump, as shown in Figure 2.34. We can use a broadband pulsed 1550 nm source and split it into two parallel amplification paths - one at the pump frequency and one at the new, shifted frequency. If we rotate the SHG output of one crystal and launch them along the same beam path with a controllable delay, we can create two simultaneous, orthogonally polarized pulses of different strengths that we can use in a true pump-probe experiment. Because they have a controllable delay between them, it is feasible to use this source to probe the exact mechanism of the thermal decoupling as well.

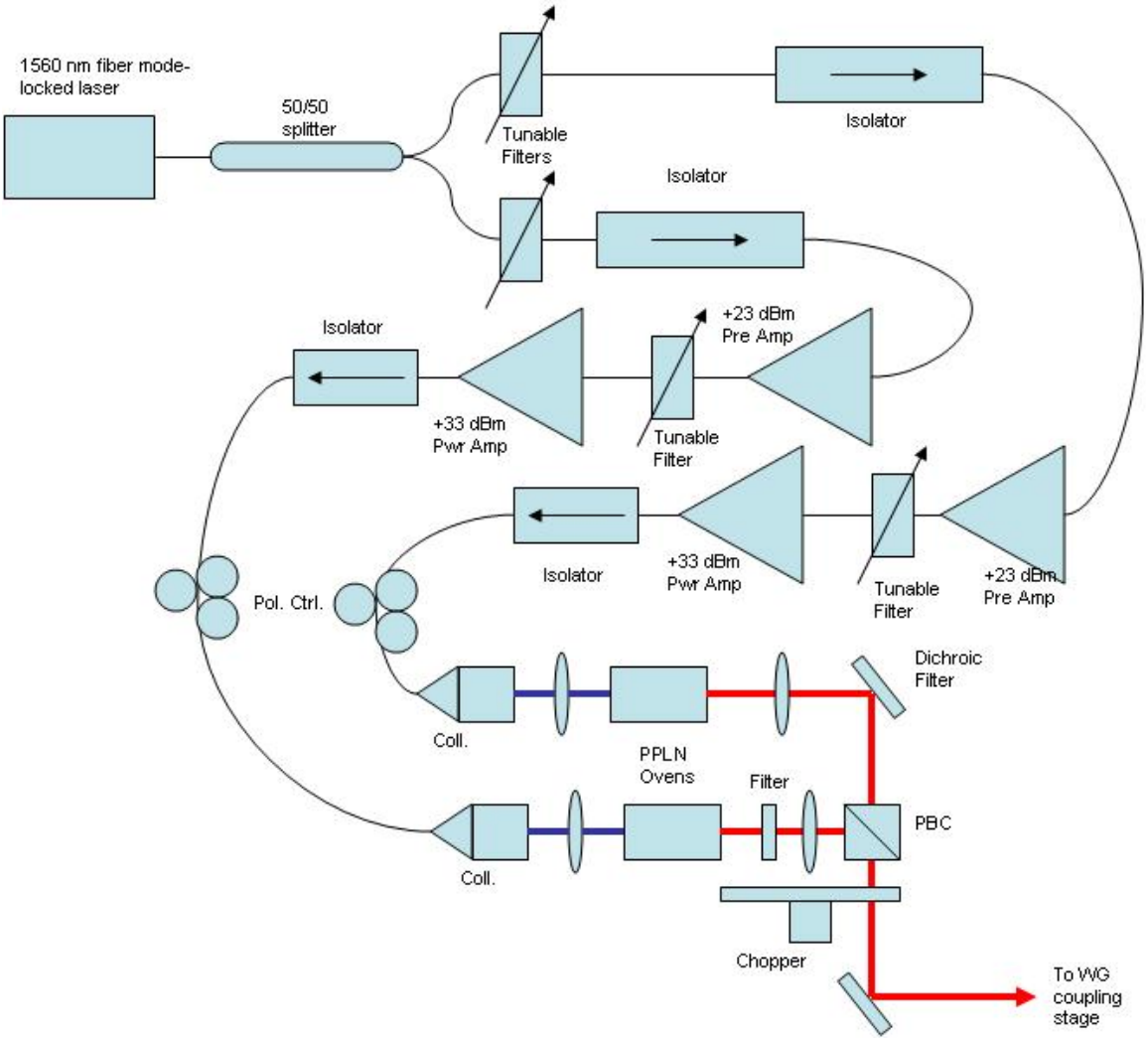

Figure 2.34: A schematic of the proposed source for a more detailed pump probe setup. Both the pump and probe beams are generated from the same broadband source, creating synchronized pulses that can be adjusted in time, wavelength, and power, allowing for maximum flexibility in measuring the nonlinearity.

Once we have thoroughly quantified the parametric gain, we can perform correlation measurements using a simple coincidence counting technique. The simplest approach would be to separate the signal and idler photons with only a long pass filter at 45^o incidence, reflecting one half of the pair to one SPAD and transmitting the other half to a second SPAD. If better filtering and precision are required, a grating can be used to separate the wavelengths. This presents a greater experimental challenge.

2.5.3 From correlation to entanglement

The final question that remains in this project is the exact method to create entangled photon pairs from the copolarized, correlated pairs that result from the four wave mixing process in the device. A number of possibilities exist, including using a two-photon interferometer or doing a double pass in opposite directions, using two counterpropagating pump beams to create two pairs in different directions. These possibilities are shown in Figure 2.35 and certainly warrant further investigation.

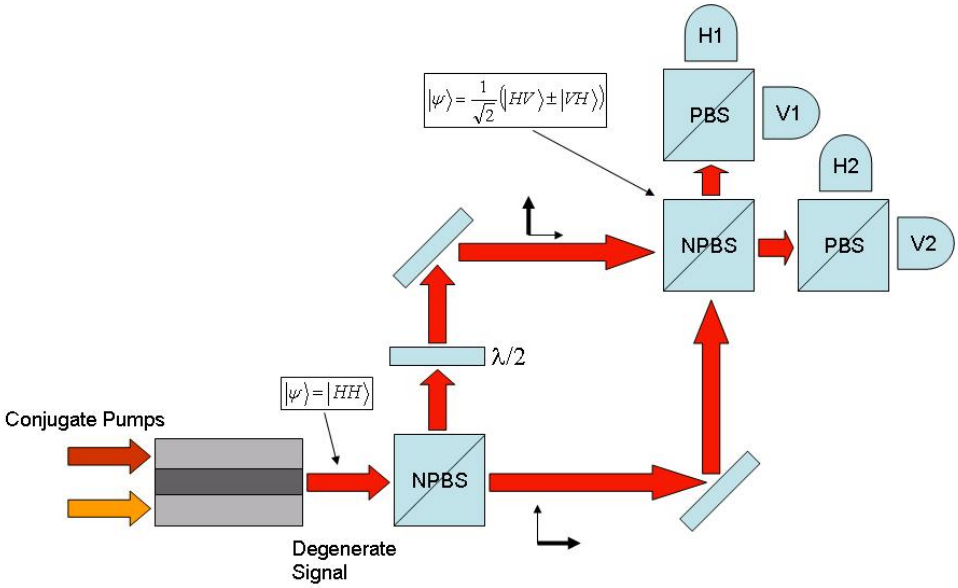

Figure 2.35: One concept for possibly generating polarization entanglement from this device. Note that a source of conjugate pump beams is required.

Another more straightforward way to glean entanglement from the 4WM process is simply to use the generated photons as a frequency entangled pair for a single sideband encoded QKD system. In this QKD implementation, described in [41], the qubit is encoded in the frequency of photon rather than the polarization. The basis choices are then made using Mach-Zender interferometers and acousto-optic modulators rather than polarization optics. This implementation is much more compatible with optical fiber channels than polarization, as the frequency stability of the fiber is orders of magnitude better than the polarization stability. In addition, unlike phase encoding approaches, this encoding scheme does not require that one balance two interferometers on either side of the link, an experimentally challenging task. Hence this technique holds a lot of promise for fiber-based QKD, and the 4WM source directly provides a source of frequency entanglement for such an encoding scheme.

2.6 Conclusions

This project is only a first step toward a more thorough effort to develop an entanglement source based on birefringent four wave mixing in a waveguide. However, it has laid the foundations for further efforts by providing a starting design and fundamental set of data about the device operation. The current set of preliminary results suggest the strength of the nonlinear interaction is significantly smaller than expected, but there are still a number of fundamental questions that remain. However, we have created a framework and test setup for testing other devices and materials. Should such further experiments succeed, they have the potential to

create a compact, convenient, fast source of entanglement for QKD and beyond.

Chapter 3

Detector Dead-Time Effects and Paralyzability in Broadband QKD

"I don't want to achieve immortality through my work; I want to achieve immortality through not dying." -Woody Allen

3.1 Problems encountered in the high-speed regime

We previously discussed (and will expound upon in Chapter 4) the speed limitation resulting from the detector's timing jitter. However, there are other properties of real single-photon detectors that become significant when transmitting quantum channel bits at high speeds. One of the most important parameters is recovery time. Silicon SPADs have a finite recovery time, τ, that is typically of the order of 100 ns. This interval, known as the *dead time*, is initiated when a detection event triggers an avalanche in the SPAD, after which the detector is unresponsive. The amplified avalanche current must be quenched and free charge carriers must be removed from the SPAD before it can be reset to its active state. This process limits the maximum count rate of such devices to less than τ^{-1}. It is worthwhile to note other types of detectors, such as superconducting single-photon detectors can support significantly higher count rates, but they still exhibit finite reset times

due to kinetic inductance. For most QKD systems, dead-time effects are reasonably assumed to have a negligible impact on overall performance; typical transmission rates, ρ_{TX}, and link losses, L, are such that most systems operate in a regime where the detection rate is low with respect to the maximum count rate, i.e. $\rho_{RX} \ll \tau^{-1}$. However, in broadband QKD links that transmit with GHz clock rates, effects due to transmitting significantly faster than the dead time become critical to both efficient operation and fundamental security [51].

The most common detector configuration for QKD in the BB84 protocol [6] is one in which the receiver, Bob, has a separate single-photon detector for each bit value in each basis. We restrict our discussion to this configuration and further assume that the detectors are free-running SPADs whose low noise allows them to be used without active gating. This is often the case in free-space QKD systems and fiber QKD systems with up-conversion detectors [75].

In this configuration, when the quantum-channel transmission rate satisfies $\rho_{TX} > \tau^{-1}$ photons can arrive and be detected at the receiver at a time when one or more of the SPADs is recovering from a prior detection event. If two such detection events occur in the same basis they necessarily correspond to opposite bit values in the key and are completely correlated [74]. This is an obvious security violation; Eve, having full access to the measurement basis, only needs to guess which detector started the alternating sequence in each basis. Thus any key sequence that is sifted while one detector in a basis is dead only contains, at most, *one bit* of secure information.

To better illustrate this effect, let us examine correlations that occur in the

sifted key as the transmission rate is increased. From a simple Monte-Carlo simulation of standard, ideal BB84 QKD, we can measure a parameter of the sifted key called the *transition probability*, or the probability that, given a bit valued 1, the next bit will be 0, and vice versa. Mathematically, this is defined as

$$P_{trans} = \frac{1}{N-1} \sum_{i=1}^{N-1} \left[(\text{bit}\,[i] + \text{bit}\,[i]) \bmod 2 \right] \tag{3.1}$$

where bit $[i]$ is the i^{th} bit in the sifted key sequence, N is the number of bits in the sifted key and the addition is performed modulo 2. Figure 3.1 shows the value of P_{trans} as the the transmission rate is increased beyond τ^{-1}.

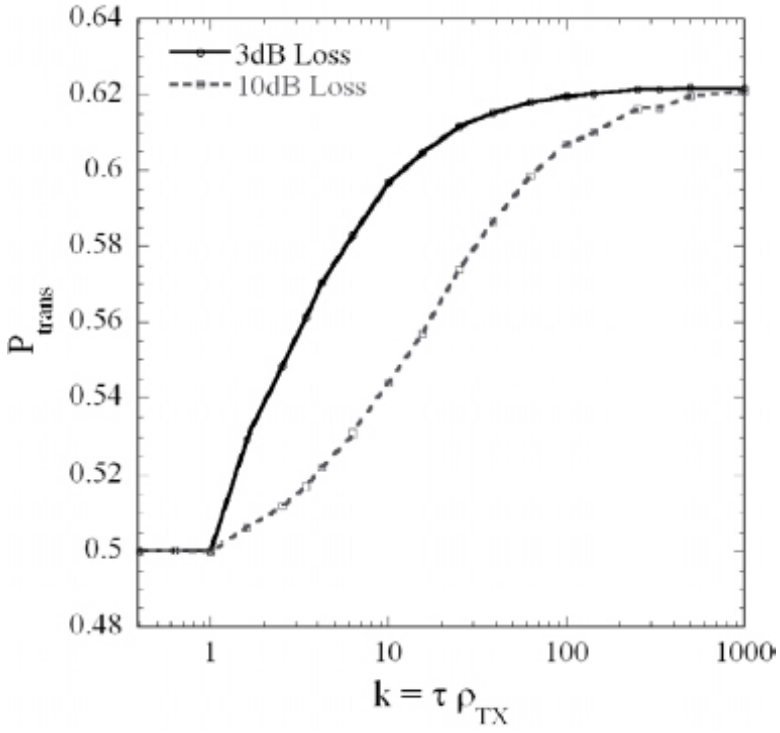

Figure 3.1: Monte-Carlo results showing the transition probability in a 1MB sifted key versus the normalized transmission rate for various link losses

Note that we introduce the normalized transmission rate, $k = \tau\rho_{TX}$, which is unitless and more convenient for our calculations. Also, note that, for the regime where $\rho_{TX} < \tau^{-1}$, P_{trans} remains constant at the expected value of 50%. As the clock rate is increased past the dead time ($k > 1$), P_{trans} quickly deviates from the deesired value of 50% for a random key, asymptotically approaching 62.2%. This value is unique to BB84 QKD with four detectors and can be understood from the following calculation [74]. At high photon-arrival rates, detection events tend to occur in fixed sequences; the detectors recover and then fire again in order. Without loss of generality, we arbitrarily choose one detector to produce the first sifted bit. After this event there are six possible detection sequences, which are listed in Table 3.1. Consider as an example the detection sequence 1-3-4-2, with detectors 1 and 3 representing bit value '1' in their respective bases and detectors 2 and 4 representing '0' in their respective bases. For this ordering, the probability, P_3, that the next detection event on detector 3 will produce the next sifted bit is $P_3 = (1/2)^1$. Similarly, $P4 = (1/2)^2$, $P_2 = (1/2)^3$, $P_1 = (1/2)^4$ and so on, are the probabilities that detectors 4, 2 and 1, respectively will produce the next sifted bit after detector 1 produces the first sifted bit (i.e. all detection events in between are not included in the sifted key). Given the sifted-bit value of '1' from the first detection event and the subsequent infinite sequence of the 1-3-4-2 firing order, we calculate the probability that next sifted bit is a '0' to be:

$$P_{1342} = \sum_{n=0}^{\infty} \left[\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3 \right] \left(\frac{1}{2}\right)^{4n} = \frac{2}{5} \tag{3.2}$$

Repeating this calculation for all six possible detection sequences gives us the in-

dividual transition probabilities for each of the six detection sequences, shown in Table 3.1. Since each sequence is equally likely to occur within a long sifted key, we average the resulting transition probabilities to obtain

$$P(0) = \frac{1}{6}\left(\frac{2}{3} + \frac{4}{5} + \frac{2}{5} + \frac{2}{5} + \frac{4}{5} + \frac{2}{3}\right) \approx 62.2\% \tag{3.3}$$

Table 3.1: Individual transition probabilities for each detection sequence. Boldface indicates a detection event that corresponds to a '0' bit value.

Detection Sequence	Transition Probability
1-**2**-3-**4**	$\sum_{n=0}^{\infty}\left(\frac{1}{2}+\frac{1}{8}\right)\left(\frac{1}{2}\right)^{4n} = \frac{2}{3}$
1-**2**-**4**-3	$\sum_{n=0}^{\infty}\left(\frac{1}{2}+\frac{1}{4}\right)\left(\frac{1}{2}\right)^{4n} = \frac{4}{5}$
1-3-**2**-**4**	$\sum_{n=0}^{\infty}\left(\frac{1}{4}+\frac{1}{8}\right)\left(\frac{1}{2}\right)^{4n} = \frac{2}{5}$
1-3-**4**-**2**	$\sum_{n=0}^{\infty}\left(\frac{1}{4}+\frac{1}{8}\right)\left(\frac{1}{2}\right)^{4n} = \frac{2}{5}$
1-**4**-**2**-3	$\sum_{n=0}^{\infty}\left(\frac{1}{2}+\frac{1}{4}\right)\left(\frac{1}{2}\right)^{4n} = \frac{4}{5}$
1-**4**-3-**2**	$\sum_{n=0}^{\infty}\left(\frac{1}{2}+\frac{1}{8}\right)\left(\frac{1}{2}\right)^{4n} = \frac{2}{3}$

3.2 Secure high-speed QKD

The security implications of transmitting faster than the inverse of the detector dead time go well beyond issues with key correlations. Since Eve has access to the

classical channel, she knows when bits are detected and in which basis they are sifted. As discussed above, when sequences of two or more detection events occur in a single basis with spacing less than the dead time, the detectors within a single basis fire alternately. This phenomenon provides Eve with nearly all of the information about the sifted bit string except for one bit representing which detector fired first within a given basis. Therefore, such detection sequences, regardless of their length, can produce at most a single sifted bit. In other words, production of a sifted bit from a detection sequence of any length requires that the detection sequence begins when both detectors in a given basis are active. This requirement is necessary for the secure operation of a QKD system at transmission rates $\rho_{TX} > \tau^{-1}$ and must be imposed on the receiver either by some means of gating the detectors or by the sifting algorithm.

It is apparent that this security requirement has an effect on the sifted key generation rate. In the low count rate regime, the sifted bit rate increases with increasing transmission rate. However, as the count rate increases beyond the inverse detector dead time, longer and longer detection sequences occur during which only one detector is live. Each of these sequences can only produce one bit of sifted key, resulting in a diminishing return on subsequent increases in the transmission rate. This effect eventually outweighs any benefit of increased transmission rate, resulting in a *decrease* of the sifted bit rate. Thus there should exist an optimum transmission rate for obtaining the most sifted key from the system.

To compute where this optimum occurs, we start by calculating the probability that both detectors in a given basis are active when a photon is detected using the

state space model shown in Figure 3.2. In this two-dimensional model, the state of one of the receiver's bases, in this case the HV basis, is quantified by how many clock periods need to pass before each detector in the basis is active, e.g. the state (3, 7) would denote that the H detector is three clock cycles away from being alive while the V detector is seven. Assuming that the two detectors have the same dead time, the state space ranges from 0 to k, as shown. On each transmission period, or clock cycle, a given detector either moves one period closer to recovery, or, if already active, the detector remains so or undergoes a detection event and moves k periods away from recovery. The probability that both detectors are active is given by the probability $P_{0,0}$ that the basis is in the state (0, 0).

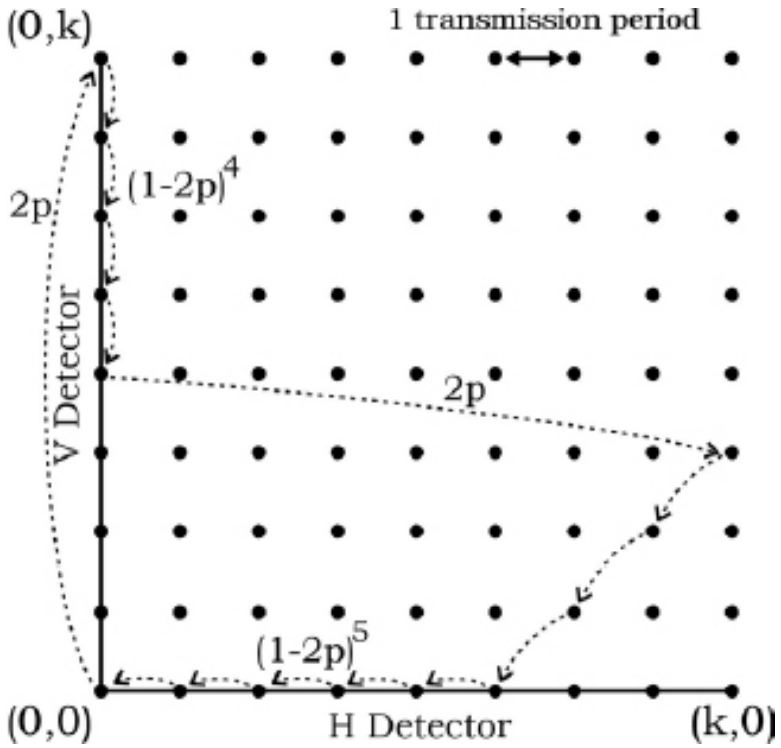

Figure 3.2: The state of Bob's HV detection basis depicting a hypothetical detection sequence and associated probabilities, as described in the text. The size of the space is determined by the value of k, in this case chosen to be 8. Note that, although they are depicted, the diagonal states are not accessible in the absence of noise, since they can only result from the simultaneous detection of the same photon by both detectors.

To quantify the probability of having a detection event during a given clock cycle, we find it useful to define a link loss parameter $p = L/8$, where L is the probability that a transmission event at Alice is detected at Bob (the detectors are assumed to be identical), commonly called the link loss. The factor of 8 accounts for Bob's basis choice (1/2), and Alice's state choice (1/4). Therefore, ignoring noise, p represents the probability that a particular detector produces a sifted bit on a given clock cycle. It should be noted that this particular definition incorporates all losses, including attenuation along the optical path, detector inefficiencies and even empty pulses due to a mean photon number less than unity. This overarching loss parameter then directly relates the transmission rate at Alice with the detection rate at Bob. Note that this analysis is thus parameterized in the transmission rate rather than the mean photon number. While the mean photon number is an important design parameter, the transmission rate is in fact more relevant to the analysis of dead-time effects. In fact, one can imagine a QKD system that transmits at $\rho_{TX} < \tau^1$ with link losses such that the photon arrival rate at Bob is equal to that of another system operating at a transmission rate $\rho_{TX} > \tau^1$, but with higher link losses. Only the latter of these two systems would exhibit the dead-time effects addressed by this analysis.

Given this definition of p, the probability that a particular detector fires on a given clock cycle is $2p$. Using this value, we can calculate as an example the likelihood of a hypothetical detection sequence, as depicted in Figure 3.2. The sequence starts with a detection event on the 'V' detector with probability $2p$, moving the basis from the origin to the state (0, k). For the next four clock cycles

the 'H' detector does not fire with probability $(1-2p)^4$, followed by a detection event on the 'H' detector with probability $2p$. The basis is now in the state (k, 3) and both detectors are inactive. The state evolves with unity probability for three clock cycles until the 'V' detector recovers, and then returns to the origin as the 'V' detector does not fire for the next five clock cycles with likelihood $(1-2p)^5$. The probability of this particular hypothetical detection sequence is therefore $(2p)^2(1-2p)^9$.

In general, the goal of using the state space model is the determination of the steady-state value of $P_{(0,0)}$, or the probability that both detectors will be alive at any given time. Knowing this value will ultimately allow use to calculate the expected sifted bit rate using this secure, high-speed BB84 scheme. To compute $P_{(0,0)}$, we begin by writing down the recursive expression describing the probability that both detectors are alive at the $(n+1)$ clock cycle:

$$P_{(0,0)}^{(n+1)} = (1-4p)P_{(0,0)}^{(n)} + (1-2p)P_{(0,1)}^{(n)} + (1-2p)P_{(1,0)}^{(n)} \tag{3.4}$$

where the first term represents the probability of no detection events occurring and the next two terms represent recovery from the (0, 1) and (1, 0) states, respectively. We ignore recovery from the state (1, 1) because such diagonal states require simultaneous detection events that will not occur in the absence of noise. In steady state, we drop the superscript and note that with random signals and identical detectors the steady-state behaviors of the H and V detectors are the same, allowing us to write $P_{(0,1)} = P_{(1,0)} = P_1$. Thus, we find

$$P_1 = \left(\frac{2p}{1-2p}\right) P_{(0,0)} \tag{3.5}$$

By the same argument one can write the probabilities $P_{(0,k)} = P_{(k,0)} = P_k$ as

$$P_k = 2pP_{(0,0)} + (2p)P_1 \tag{3.6}$$

which, with the substitution for $P_{(0,0)}$ from 3.5, reveals that $P_k = P_1$. In fact, similar calculations for $P_{(1,0)}$, $P_{(2,0)}$, etc., show that all 2k states lying upon the axes have the same steady state probability P_1. The states not lying on one of the axes represent instances when both detectors are dead. Omitting the states along the diagonal, the internal states are only accessible from on of the on-axis states. Since the on-axis states are all of equal probability one can show that the internal states, of which there are $(k^2 - k)$, are also of equal probability, in this case $(2p)P_1$.

The expressions above represent the steady-state probabilities of the basis being in each of the states in the entire state space. We normalize the sum of these probabilities, giving

$$P_{(0,0)} + (2k)P_1 + (k^2 - k)(2p)P_1 = 1 \tag{3.7}$$

Substituting for P_1 from 3.5, we can solve for $P_{(0,0)}$, the steady state probability that both detectors are alive for a given transmission event, as a function of the link loss parameter, p, and the dead time parameter, k:

$$P_{(0,0)}(p,k) = \left[1 + (2k)\left(\frac{2p}{1-2p}\right) + (k^2 - k)\left(\frac{(2p)^2}{1-2p}\right)\right]^{-1} \tag{3.8}$$

As stated above, a detection sequence can only produce a sifted bit from events that occur when both detectors are alive. Therefore, $P_{(0,0)}(p,k)$ should be used as an additional factor in the calculation of a system's sifted-bit rate. $P_{(0,0)}(p,k)$ is shown in Figure 3.3 as a function of the normalized transmission rate k, for three values

of the link loss $L = 8p$. It can be seen that as the transmission rate is increased $P_{(0,0)}(p,k)$ begins to roll off, approaching zero as k^2 at high count rates. The roll-off of $P_{(0,0)}(p,k)$ marks the onset of dead-time effects and the departure from the low-count-rate regime.

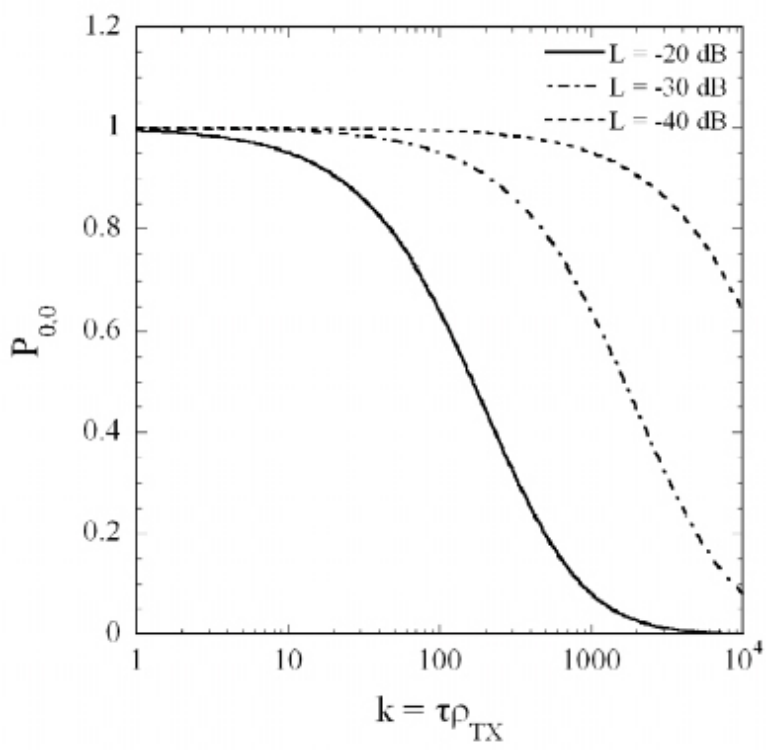

Figure 3.3: The likelihood $P_{(0,0)}$ that both detectors in a given basis are active when a photon arrives versus the number of transmission periods per dead time, k, for three values of the link loss L. The fact that $P_{(0,0)}$ tends to zero at high transmission rates demonstrates the paralyzability of the QKD receiver.

The behavior of $P_{(0,0)}(p,k)$ in the high-count-rate regime illuminates a characteristic unique to operation of QKD systems at transmission rates $\rho_{TX} > \tau^1$. From the standpoint of producing sifted bits, when the pair of SPADs in a given basis is considered as a whole, the QKD receiver becomes what is known as a paralyzable counter [52] [2] [30]. Signals that arrive at a paralyzable counter during recovery, though not counted, extend the necessary recovery time [30]. In contrast, non-paralyzable counting systems recover from each counting event regardless of signals that arrive during the dead time. Taken individually, SPADs are non-paralyzable detectors; with the exception of counts that occur just as the detector comes alive (referred to as 'twilight counts' in [71]), when the bias voltage is below the breakdown voltage, photons that arrive during the dead time have no significant effect on the detector. It is worthwhile to note that the response of paralyzable and non-paralyzable systems exhibit significant differences only in the regime of high count rates [30], and paralyzability has become relevant to QKD systems as they continue to increase in key-production rates.

Although each closely-spaced detection sequence can produce at most a single sifted bit, it is also true that as the length of the detection sequence grows the likelihood that a bit will be sifted from the sequence also grows. In the low-count-rate regime the average length of a detection sequence is 1 and the likelihood of sifting a bit from a sequence is 0.5. For a detection sequence of length 3, however, the likelihood that at least one of the detection events occurred in the correct basis is 7/8. This fact offsets some of the deleterious dead-time effects and must be included in the calculation of the sifted-bit rate.

At any count rate, we can write the probability of sifting a bit from a detection sequence that begins with both detectors active as

$$S(p,k) = \sum_{N=1}^{\infty} \left(1 - \left(\frac{1}{2}\right)^N\right) T_N(p,k) \tag{3.9}$$

where $T_N(p,k)$ is the probability that the detection sequence consists of N detection events. To calculate $T_N(p,k)$, we can use the state space model from Figure 3.2. For a detection sequence of length 1 (i.e., a single detection event) there is only one path through the state space. For longer sequences, we must sum the possible paths for a given number of detection events. For example, there are a total of $(k-1)$ ways to arrange two detection events before the basis returns to the (0,0) state; only one of these ways is depicted in Figure 3.2. The probabilities of a detection sequence having lengths up to $N = 4$ are

$$T_1(p,k) = (1-2p)^k \tag{3.10}$$

$$T_2(p,k) = \sum_{j=0}^{k-1} 2p(1-2p)^{2j+1} \tag{3.11}$$

$$T_3(p,k) = \sum_{j_1=0}^{k-1} \sum_{j_2=0}^{j_1} (2p)^2 (1-2p)^{2j_2+k} \tag{3.12}$$

$$T_4(p,k) = \sum_{j_1=0}^{k-1} \sum_{j_2=0}^{j1} \sum_{j_3=0}^{k-(j_1-j_2)-1} (2p)^3 (1-2p)^{2(j_2+j_3)+1} \tag{3.13}$$

The truncated geometric series in $T_N(p,k)$ can be evaluated with the standard techniques to yield analytic expressions for all N. While the sum over N in 3.9 is theoretically infinite, it is worthwhile to note that in practice, one needs to compute $T_N(p,k)$ only up to $N = 6$, as the probability of sifting a bit from a detection sequence longer than six events approaches unity. In addition, one interesting feature of $T_N(p,k)$ is the difference between even and odd values of N, as illustrated

in Figure 3.4. While all $N > 1$ sequences have low probability in the low count-rate regime, at high count rates the odd N sequences fall asymptotically to zero, but the even N sequences have constant finite probabilities. This behavior can be understood from the fact that an even N sequence minimizes the number of clock cycles during which the detector is active but does not fire. For an odd N sequence, the unlikely situation occurs where a live detector must not fire for at least k clock cycles before the basis returns to the (0,0) state.

Noise sources such as background counts and detector dark counts can also be included in the model in a straightforward manner. We define ϵ as the probability that a detector experiences a noise event during one clock cycle. The probability that a detector fires during a clock cycle, therefore changes from $(2p)$ to $(2p + \epsilon)$, which can be substituted into $P_{(0,0)}(p, k)$ and $T_N(p, k)$ accordingly. As mentioned above, a noise event on one detector can occur on the same clock cycle as a signal (or noise) event on the other detector. These simultaneous events put the basis in the state (k, k), after which the basis recovers with unity probability along the diagonal back to (0, 0). Thus, we find that when noise counts are included, the probability that both detectors are alive on the (n + 1) clock cycle becomes

$$P_{(0,0)}^{(n+1)} = (1 - 2(2p + \epsilon))P_{(0,0)}^{(n)} + 2(1 - (2p + \epsilon))P_1^{(n)} + (2p\epsilon + \epsilon^2)P_{(0,0)}^{(n-k)} \qquad (3.14)$$

where the third term accounts for the simultaneous detection events. The steady state calculation of $P_{(0,0)}(p, k)$ then proceeds in the same manner as described above. It should be noted that while noise sources can cause simultaneous detection events, no secure bits can be sifted from such occurrences.

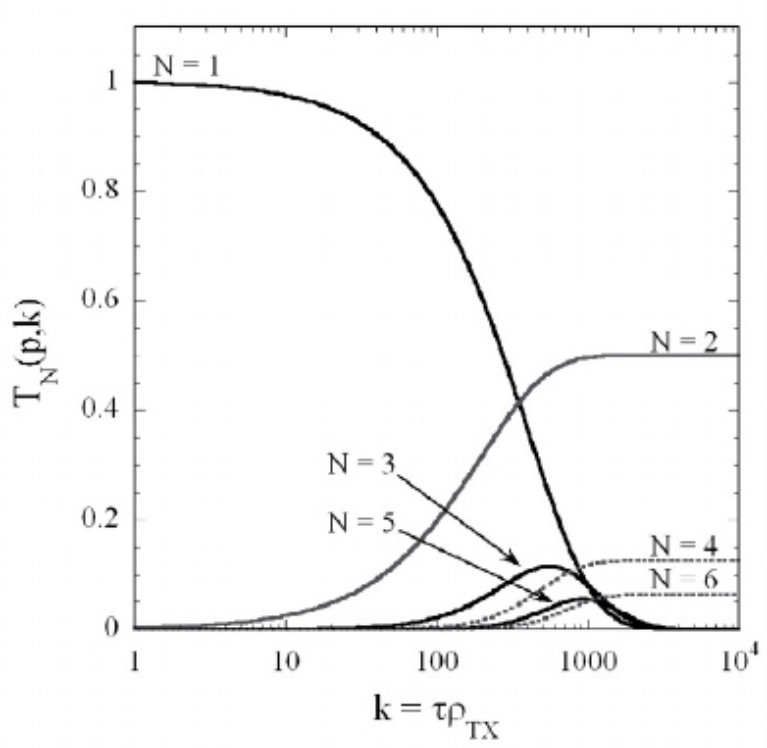

Figure 3.4: The probabilities $T_N(p, k)$ of a detection sequence having N detection events versus the normalized transmission rate k, for link losses L = - 20 dB. There is a characteristic difference in even and odd numbers of detection events in the high-count rate regime.

3.3 A Monte-Carlo simulation of QKD

In addition to the analytical model outlined above, we have also implemented a Monte-Carlo simulation of BB84 QKD, both using the standard protocol (the code used to examine the transition probability) and also a version incorporating the modifications described above to maintain security in the high-speed regime. The code for both implementations is listed in Appendix B.

The first version of the code, qkd_trans_prob.c, implements the traditional BB84 protocol without considering the security implications of the detector dead time. It operates in time units that are normalized to the detector's dead time. It is a Monte-Carlo simulation that uses the built-in ANSI C random number generator seeded by the machine time. The code generates seven random variables it then uses to choose the transmission basis, bit value, and detection basis. It also uses these random variables to implement simulated link loss, detector inefficiency, and even QBER (which was kept at 0 for these simulations). The code maintains flag variables to determine which detectors are dead at what points throughout the key generation process and uses an array of counters to keep track of the four detectors' respective dead times. Since it runs in 'unitless' time, it is designed to continuously generate sifted key bits until it has accumulated *NBITS* of key (set to *NBITS=1048576* in the version of the code in Appendix B). Once it accumulates this much sifted key, it then analyzes the transition probability of the key, providing the data plotted in Figure 3.1.

The second version of the code incorporates the modification to BB84 that is

required to maintain security in the presence of long dead times. It does this by defining a second flag variable that corresponds to the state of the opposite detector in the given detection basis during a detection event. If one detector goes dead, all events originating from its pair in the basis are not incorporated into the sifted key. One major difference between this version of the code and the previous version is that the new code, fc.c, operates in simulated 'real time.' This means that there is a fundamental clock unit of 100 ps around which everything else in the simulation is designed. Thus the clock rate, dead times, etc., can all be defined in real time units rather than in units of dead time. This is critical to properly simulating dead time effects and accurately measuring the sifted bit rates that result.

Both versions of the code were run on the NIST RARITAN Linux-based computing cluster. The cluster consists of over 400 nodes connected via ethernet. The cluster is fully managed and includes all of the high-performance compilers and MPI interfaces required to do parallel computing. However, instead of parallelizing the code directly, it was simply run in a 'task farming' way, where various parameters were run on individual machines. As the code is not prohibitively intensive in computation, this method was sufficient to obtain a full set of results within a few days of computing time on a small number of machines.

3.4 The sifted bit rate in high-speed QKD

We have now described all of the necessary factors to incorporate dead time effects into the sifted bit production rate. Returning to the noiseless picture, we

write the sifted bit rate as

$$SBR = \rho_{TX} 8p P_{(0,0)}(p,k) S(p,k) \tag{3.15}$$

The sifted-bit rate is shown in Figure 3.5 as a function of the transmission rate for various detector dead times (a) and link losses (b). The lines indicate the results from the analytic state-space model presented above. The symbols indicate results from the BB84 Monte-Carlo simulation that incorporates the modified sifting algorithm described above, sifting at most a single bit from sequences of closely-spaced detection events. As illustrated in Figure 3.5, dead-time effects induce a maximum value on the sifted-bit rate, above which further increases in transmission rate actually reduce the sifted-bit rate. The maximum value of the sifted-bit rate is a complicated function of the link parameters. However, we find this maximum is not strongly dependent on the link losses. As demonstrated in Figure 3.5(b), it may be accurately approximated as a function of dead time alone by

$$SBR_{Max} \approx \frac{1.433}{2\tau} \tag{3.16}$$

where the constant of proportionality was found by a least-squares fit. The factor of $(2\tau)^{-1}$ represents the maximum sifted bit rate for the case of an actively gated receiver in which all of the detectors are disabled when any of one of them fires [74]. Most significantly, the numerator is greater than 1, indicating that one can achieve secure key production rates over 40% faster than previously thought.

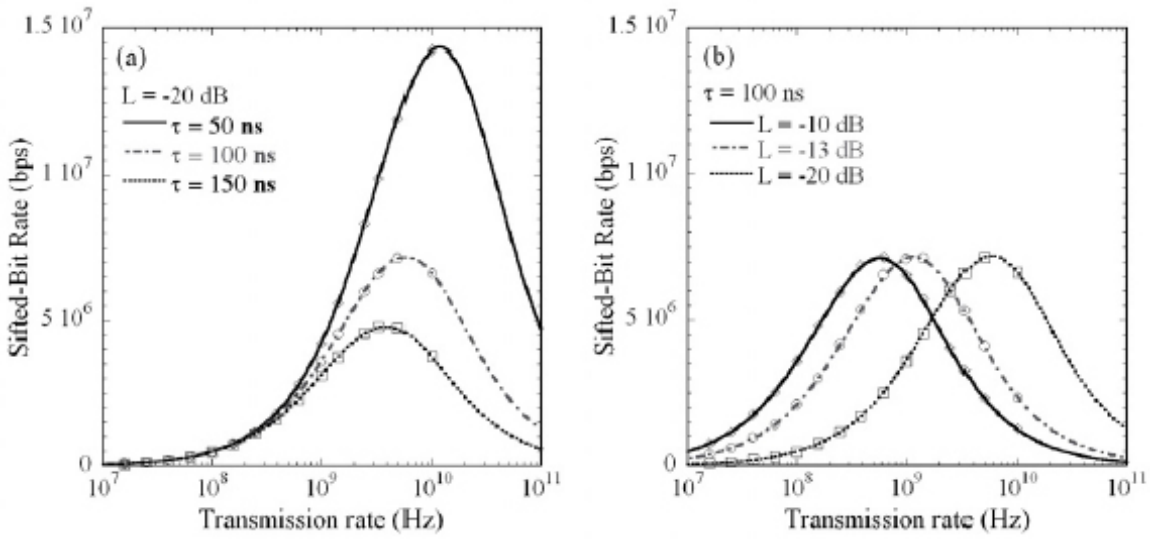

Figure 3.5: The sifted-bit rate including dead-time effects, showing excellent agreement between the model (lines) and the simulation (symbols). The effect of varying the dead time with fixed link loss is shown in (a). The effect of varying the link loss with fixed dead time is shown in (b).

The transmission rate at which the sifted-bit rate is maximized is also a complicated function of the dead time and link losses. However, for typical link losses and detector dead times we find that it may be approximated to an accuracy better than 1% by

$$\rho_{TX}^{Max} \approx \frac{5.92}{8p\tau} \tag{3.17}$$

For most QKD links the loss and dead time are such that detector timing resolution plays a dominant role in determining the optimum transmission rate [9] [25][63][64]. However, as the disparity between the detector timing resolution and recovery time grows with improved timing resolution, transmission-rate limitations imposed by dead-time effects will become more significant.

3.5 Hardware approaches to addressing dead time effects

There are a variety of methods that may be employed to address the security issue that arises in the $\rho_{TX} > \tau^1$ regime. The algorithmic solution modeled above is, to our knowledge, the most efficient with respect to the production of sifted bits. The communications overhead associated with implementing are comparable to traditional implementations of QKD, which can be anywhere from 17 to 100 times the quantum channel data load. However, in the event that some implementation would preclude the software scheme implementation proposed above, a hardware solution would be required. An active hold-off scheme has been previously proposed, in which all the detectors are disabled when any one of them fires [74]. Actively disabling the detectors by some electronic means can be technically challenging,

particularly as transmission rates exceed 1 GHz. As an alternative, we propose the self-disabling receiver shown in Figure 3.6. In this system, the states in each of Bob's bases are sent to the same detector, though with different propagation delays depending on the state. The states of the photons incident on each detector are distinguished by their arrival times, much in the same manner that time-division-multiplexed communications links distinguish various channels. Detection schemes similar to this have also been implemented for basis discrimination in QKD links [47]. The receiver proposed in Figure 3.6, however, would go further than basis discrimination and shut down a basis entirely after a detection event. With only one detector in each basis, the entire basis is disabled for the duration of the dead time and sequences of closely-spaced detection events are eliminated. This QKD receiver is a non-paralyzable counter capable of producing sifted bits at rates up to τ^{-1}s in the high-count-rate regime.

As a consequence of operating the receiver bases in the self-disabling format shown in Figure 3.6, each transmission event from Alice is analyzed in two time bins at Bob's receiver. If these time bins are limited by the SPAD's ability to distinguish photon-arrival times, i.e., by the SPAD timing resolution, then Alice's transmission period must be at least twice as long. Thus the maximum transmission rate as determined by the detector timing resolution is reduced by one half. The reduction in transmission rate makes this type of receiver useful only in cases in which the algorithmic implementation described above is somehow impractical.

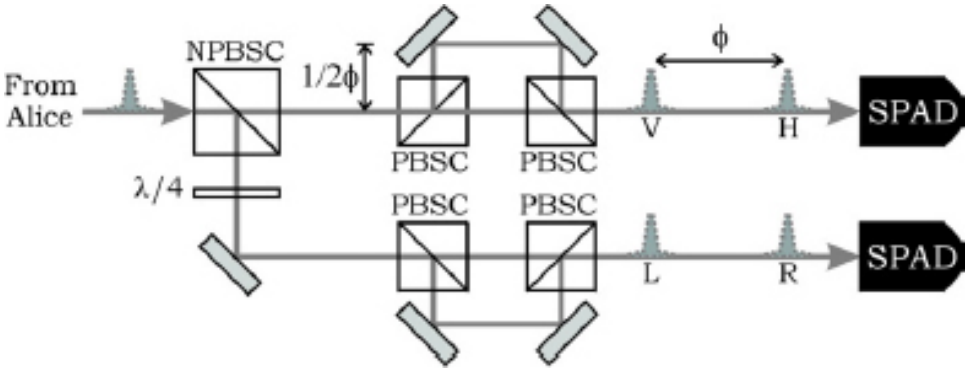

Figure 3.6: BB84 receiver with self-disabling bases. In this configuration the individual states in each measurement basis are distinguished by their arrival times at the SPADs. (N)PBSC is a (non) polarizing beam-splitting cube.

3.6 Conclusions

We have presented a model for the sifted-bit production rate of BB84-type QKD systems operating at transmission rates that exceed the maximum count rate of the component single-photon detectors. This model addresses critical security concerns that must be considered when operating in this regime and quantifies the onset of dead-time effects. We have established that, with free-running SPADs, high-speed QKD systems are paralyzable counting systems. This phenomenon emerges from the collective behavior of the pair of detectors in a given basis, as SPADs are non-paralyzable counting systems when considered individually. We have shown with both analytic modeling and Monte Carlo simulation that dead-time effects cause there to be an optimum transmission rate that maximizes the sifted-bit production rate. The functional dependence of the maximum sifted-bit rate on the link parameters has been presented, and these relations will be useful in the design of QKD systems and single-photon detection systems.

This chapter has focused on polarization-encoded BB84 QKD. A useful extension of the analysis presented here would be the application of the state-space model to other protocols and encoding schemes. In particular, the differential-phase-shift encoding scheme used in [63] readily lends itself to extremely high transmission rates. Detector dead times are likely to have significant influence on the performance of such systems and the analysis of such influence in the context of the current understanding would be useful.

Chapter 4

High-speed QKD in the Hα Fraunhofer Window

"Far out in the uncharted backwaters of the unfashionable end of the Western Spiral Arm of the Galaxy lies a small critically unregarded yellow sun. Orbiting this, at a distance of roughly ninety-eight million miles is an utterly insignificant little blue-green planet whose ape descended life forms are so amazingly primitive that they still think digital watches are a pretty neat idea." Douglas Adams

4.1 Speed limits in broadband QKD

In 2004, NIST demonstrated a free-space QKD system that achieved a secret key rate in excess of 1 MB/s [9]. This and similar systems, such as those described in [75] and [63], represent the fastest QKD links currently in existence and are even fast enough to stream one-time-pad encrypted video. The NIST systems achieve their high key rates by incorporating a number of techniques, many of which have been adapted from high-speed telecommunications engineering. Examples of this include performing clock synchronization between the transmitter and receiver over the classical communications channel, as well as implementing all of the control and sifting code on a dedicated FPGA, utilizing tailored forward error correcting

algorithms to minimize the bidirectionality of the classical conversation, and multithreading the error correction and privacy amplification steps in order to speed up key distillation. However, in all of these systems there remains a fundamental technology hurdle to performing faster QKD: The timing resolution of the single photon detectors.

Figure 4.1 shows a histogram of detection events from a clocked 1.25 GHz pulse. As evidenced from the data, the ubiquitous Perkin-Elmer single photon counting modules used in the system have a -3 dB timing jitter of 250 ps. Nominally this would imply that the detector could differentiate between 4 GHz pulses. However, because the QKD protocol is inherently sensitive to errors, we must transmit at a period greater than the -20 dB timing jitter in order to avoid data being detected in overlapping time bins. Because the Perkin-Elmer detectors have a -20 dB timing jitter tail in excess of 1.6 ns, the NIST 2004 QKD system described in [9] must be limited to a 625 MHz clock rate.

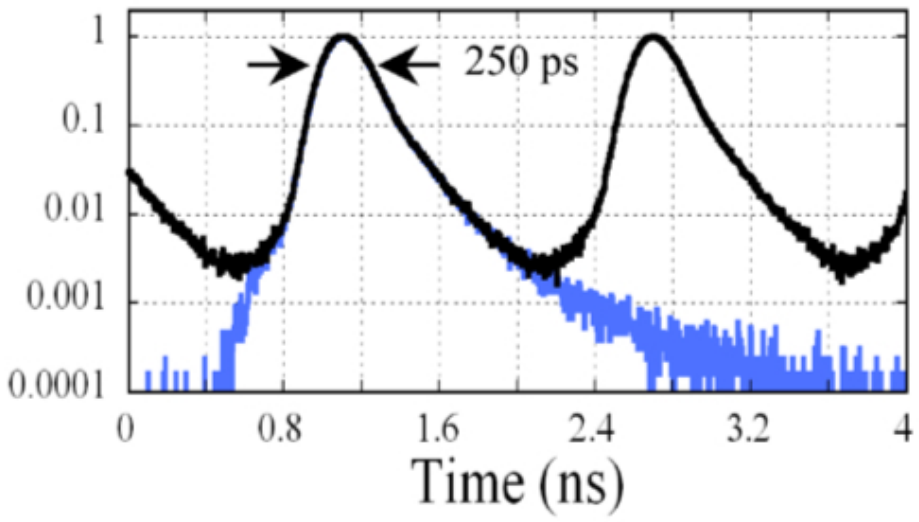

Figure 4.1: A histogram of detection events from the Perkin-Elmer SPCM, showing a -3 dB timing jitter of 250 ps and a -20 dB timing jitter of over 1.6 ns. This intrinsic uncertainty in the detection event timing information is the current speed limitation in the NIST QKD system.

4.2 The origin of timing jitter in Geiger mode avalanche photodiodes

Avalanche photodiodes biased just below the breakdown voltage (the so-called *Geiger mode* of operation) have been enormously successful photon counting devices for a number of applications, ranging from the obvious quantum optical experiments to biophysical applications such as single-molecule spectroscopy. However, all of these applications demand precise timing information, and the intrinsic structural properties of the devices themselves currently limit how well they can meet those demands. A schematic of the device is shown in Figure 4.2. As the figure illustrates, there is a finite volume, called the *depletion region*, in which the incident photon is absorbed to create the electron-hole pair. This volume must be relatively large in order to increase the probability of photoabsorption and improve the detector's overall efficiency. However, a side effect of such a large absorption volume is an intrinsic, statistical uncertainty in the time between when the photon is incident on the front of the detector and when it is absorbed, resulting in an avalanche and final detection pulse. This uncertainty is the origin of the detector's timing jitter.

Also shown in Figure 4.2 is a schematic of a new design from Micro Photon Devices, a company that has formed out of the group of Sergio Cova at the Politechnico di Milan [26]. In this design, the depletion region of the device has been reduced in thickness from 30-40 μm down to between 1 and 4 μm. Thinning this region has the desired effect of reducing the timing jitter to as low as 35 ps, as shown in figure 4.3.

Better timing resolution immediately allows for faster single photon trans-

mission rates in our QKD system. However, because silicon has a higher extinction coefficient at shorter wavelengths [44], a thinner depletion region implies that device will also have higher quantum efficiency at shorter wavelengths. In fact, the efficiency curves shown in Figure 4.4 verify that, for the thinner devices, the quantum efficiency is not only lower overall, but the wavelength of peak detection efficiency has shifted from the near-infrared into the visible.

The implications of lower overall quantum efficiency are obvious - lower detection efficiency means higher link loss. But more importantly, the shift from peak efficiency in the NIR to peak efficiency in the visible implies that free-space QKD links using these detectors will be more susceptible to solar background noise. The solar spectrum has significantly higher light levels at visible wavelengths than at NIR ones. This is likely the factor that drove humans, a very visually dominated species, to evolve a strong sensitivity in this spectral region. But what makes us able to see also introduces significant errors into free-space QKD links operating at visible wavelengths. In fact, it is likely that, without mitigating this effect, any gain in performance from faster transmission rates will be negated by the increased noise from the solar background. Thus any free-space QKD system whose design is centered around these new SPADs must incorporate some additional techniques to mitigate the background noise. In the new design presented here, two additional features are added. First, we design our system around a Fraunhofer line in the solar spectrum where the solar background is reduced and second, we incorporate a sub-clock gating circuit that forces the detector to ignore a significant number of background counts within each clock bin.

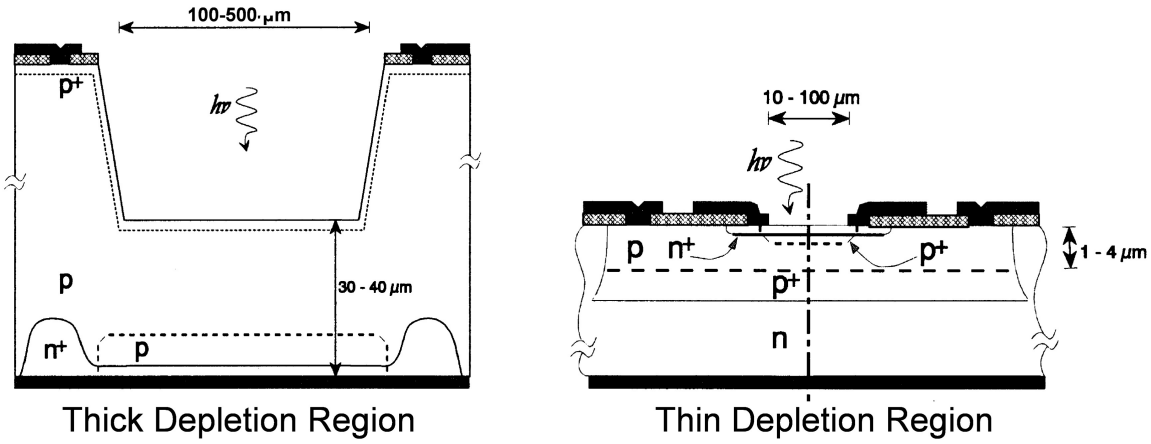

Figure 4.2: A schematic of the structure of two different APDs [26]. On the left is the APD in the Perkin-Elmer SPCM with a thicker depletion region and thus higher timing jitter. On the right is a newer design from the Cova group, incorporating a thinner depletion region and thus improved timing resolution.

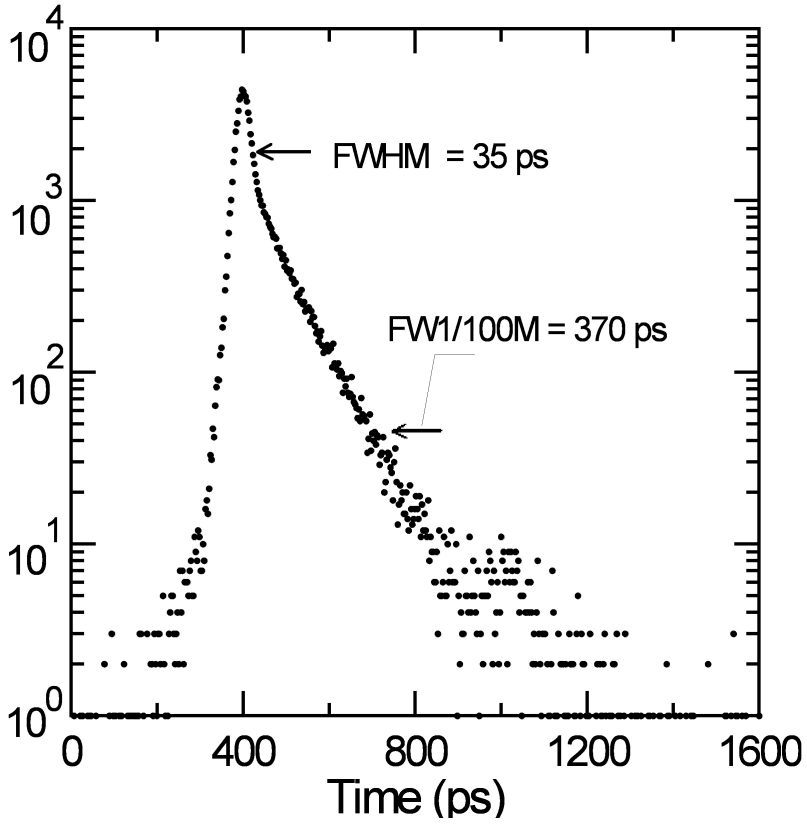

Figure 4.3: A histogram of counts from the thin MPD SPAD, showing a 35 ps timing jitter [26].

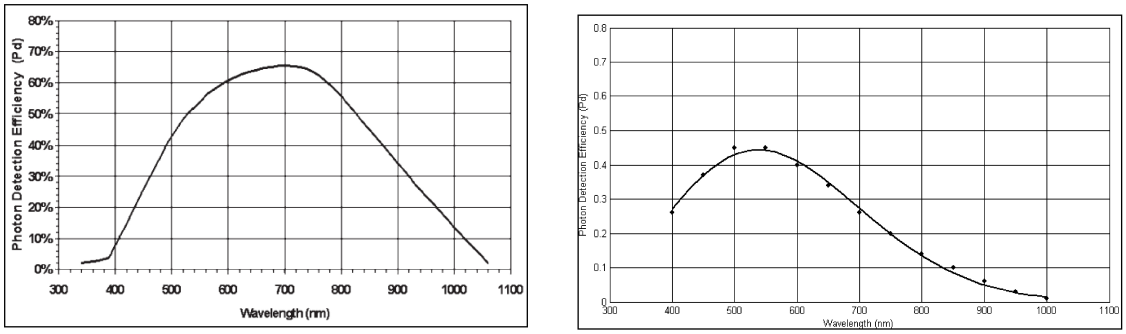

Figure 4.4: A comparison of the quantum efficiency curves of the Perkin-Elmer SPCM versus the MPD device. By thinning the depletion region, the overall quantum efficiency is lowered and the peak efficiency wavelength is shifted into the visible region of the spectrum.

4.3 Josef von Fraunhofer and the Hα line

Josef von Fraunhofer was born on March 6th, 1787, in the Bavarian town of Straubing[1]. After becoming an orphan at the age of 11, he began work as an apprentice in a Munich company making scientific instruments, where he learned to make lenses and mirrors. Following a major accident that resulted in the collapse of the building where he was working, Fraunhofer was rescued by Maximillian I Josef, then the Prince Elector of Bavaria. This future ruler of Bavaria took Fraunhofer under his wing, providing young Josef with books and encouraging his studies. After eight months, Fraunhofer went to work at the Benediktbeuern Abbey's Optical Institute, which specialized in glassmaking. There he developed scientific instruments of such quality that even Michael Faraday could not compete, shifting the center of optical instrument manufacturing from England to Bavaria. By 1818, Fraunhofer became the Institute's director and eventually earned an honorary doctorate from the University of Erlangen.

As part of his work, Fraunhofer often used sunlight and a homemade slit spectrometer to look for impurities in his glass. He noticed that certain lines in the transmission spectrum of his glass were always present, no matter the quality of the material he was testing. As a result, Fraunhofer examined his source directly, realizing that the dark lines in the spectrum were actually part of the sunlight rather than the impurities in his glass. These lines had been missed by Newton

[1]This section is based on the High Altitude Observatory's online biography of Josef von Fraunhofer

in his prism work over a hundred years earlier, but William Wollaston had noticed them previously. However, it was Fraunhofer who closely studied and cataloged the hundreds of lines which would eventually bear his name.

Figure 4.5: Josef von Frauhofer, 1787-1826

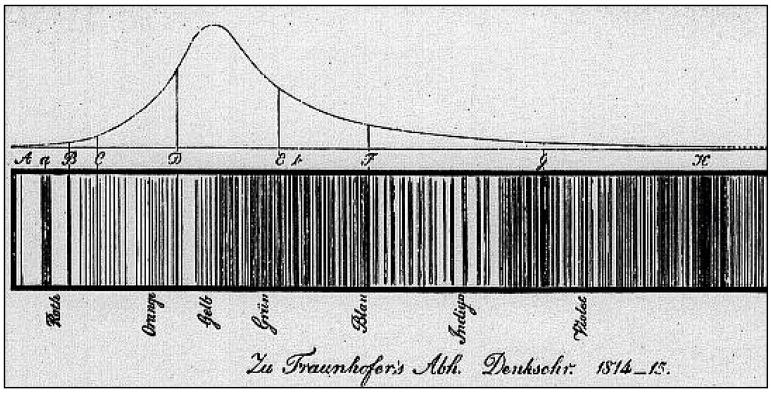

Figure 4.6: Fraunhofer's original drawing of the lines in the solar specturm. He noticed that certain lines were always present when he used sunlight to test the glass that he made, no matter the quality of the specimen under test. He labeled the most prominent of these lines with letters A, B, C, etc., and some of them, such as the Sodium D line, still bear these labels.

4.4 Free-space optical communication in the Fraunhofer window

Today these lines are known to be the result of atomic absorption in the sun's atmosphere and have been studied in much more depth. The visible spectrum of the sun, as taken in false color and 'unwrapped' by the National Optical Astronomy Observatories in Figure 4.7, shows hundreds of such lines. One of these lines, though, stands out as a particularly dark hole in the solar spectrum - the H-α, or Ballmer α line, in the red region at 656.28 nm. This line is a result of the $n = 2$ to $n = 3$ absorption transition in the atomic hydrogen contained in the solar atmosphere [49]. It is 0.12 nm wide and 7-8 dB deep. As there is some reemission by the hydrogen atoms, the line is not completely dark. However, from a free-space optical communication point of view, it is the equivalent of operating at perpetual twilight.

Exploiting the Fraunhofer lines to do free-space optical communication during daylight hours is not a particularly new idea. For example, it has been considered for links to submarines and in deep space [29]. However, in the context of QKD this technique can be applied specifically to compensate for the extra background counts caused by operating in the visible region. In fact, narrowband optical filters are rather easy to come by, as they are vital for solar photography and are often stocked as surplus items by manufacturers. Calibrations performed at the NIST SIRCUS facility, the results of which are presented in Figure 4.8 show some of these filters to be as narrow as 0.15 nm with a peak transmission of 70%. This performance is achieved by using a *rugate* filter design. Standard dielectric stack filters are fabricated by depositing alternating layers of materials with different

indices of refraction. The resulting index structure has a transmission edge at the fundamental period, but since the index alternates between two discreet values, the edge is not ideally sharp and there are other transmission peaks at higher harmonics of the index period. Rugate filters, on the other hand, consist of a continuous sinusoidally varying index layer in lieu of the dielectric stack. By modulating the index of refraction sinusoidally, the filter achieves a single transition edge that is transform limited and is usually only about 0.1 nm wide, as shown in Figure 4.8. Despite the difficulties in fabricating such filters, their utility in solar photography results in mass production and their relative ubiquity.

In addition to calibrating the filter characteristics, we have directly measured the reduction in solar background counts due to the filter's presence. We constructed a simple test consisting of a collimator collecting sunlight, going through the H-α filter, then into a SPAD connected to a TTL counter. The output of the counter was normalized to the output of another SPAD that was collecting sunlight without a filter in order to mitigate any temporal variations in sunlight. The filter was mounted on a computer-controlled, motorized rotation stage. The data shown in Figure 4.9 shows that the background counts are reduced by the expected amount as the filter is tuned across normal incidence. In fact, as the data show, the minimum background count rate occurs just off of normal incidence, indicating that the filter is not precisely centered at the H-α wavelength and will require some slight tuning during operation.

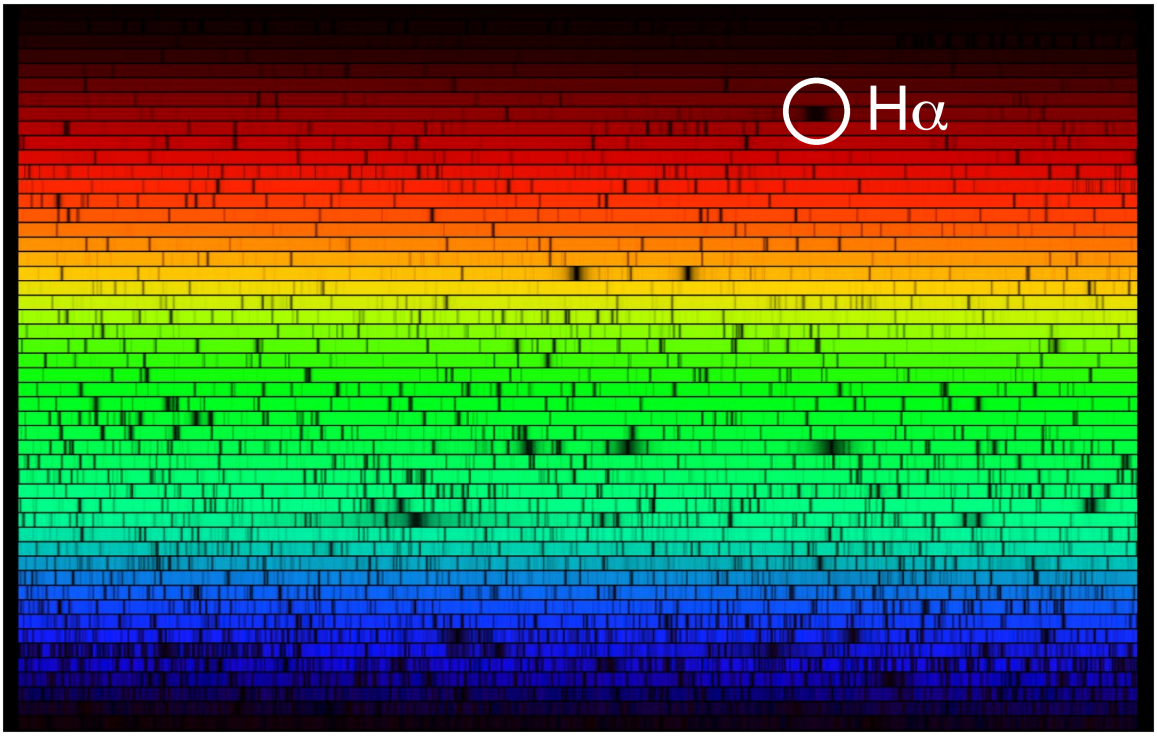

Figure 4.7: A false color, unwound depiction of the visible solar spectrum created by the National Optical Astronomy Observatories. Fraunhofer noticed hundreds of lines occurring in the spectrum, with a particularly dark region in the red. This line, at 656.28 nm, is the result of the n=2 to 3 transition in atomic hydrogen in the solar atmosphere and is particularly useful to performing free-space QKD in the daytime with reduced background counts.

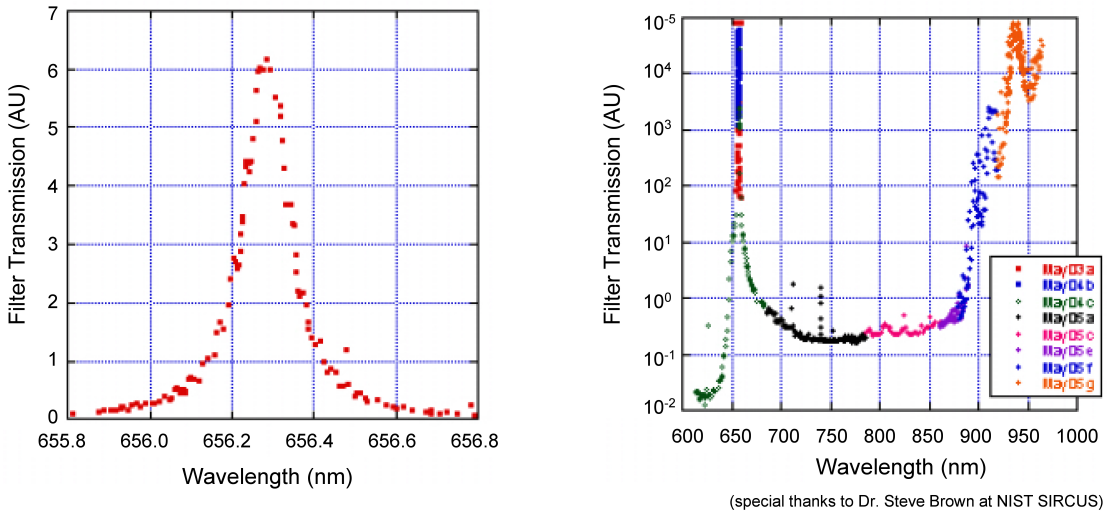

Figure 4.8: Calibration data taken at NIST SIRCUS (many thanks to Dr. Steve Brown) showing the performance of H-α Rugate filters, both specifically around 656.28 nm and across the broader optical spectrum.

4.5 Single-photon sources at 656 nm

Despite the ease of obtaining filters at the H-α wavelength, there are very few fast pulsed optical sources at that wavelength. The 2004 NIST system is based on gain-switched VCSEL diodes that are designed for high-speed modulation. However, for market reasons these devices are fabricated at 850 nm, and no comparable device exists in the visible spectrum. One approach that we have taken to generating red pulses has been to implement a pulse carving scheme using a CW tunable laser and an external intensity modulator. The laser is a Sacher external cavity tunable diode laser designed for Hydrogen spectroscopy and the modulator is a custom Mach-Zender type Electro-optic intensity modulator designed by EOSpace, Inc., to operate in the visible spectrum. While we have demonstrated the feasibility of this approach, as demonstrated by the histogram in Figure 4.10, we have observed several limitations. First, we have found that the external cavity diode laser is not stable enough in frequency for the long operation required by a QKD system. This effect is not unique to our laser and has been observed in other lasers and even in units built by other manufacturers. Additionally, the modulator is highly sensitive to wavelength, temperature and polarization effects, resulting in low on-off extinction ratios over long periods of operation.

To work around both of these effects, we have identified a number of possible solutions. The most obvious but most difficult approach would be to directly fabricate 656 nm VCSELs. However, this approach is rather high risk, as little is known about the issues associated with high-speed visible diodes. The second, more feasi-

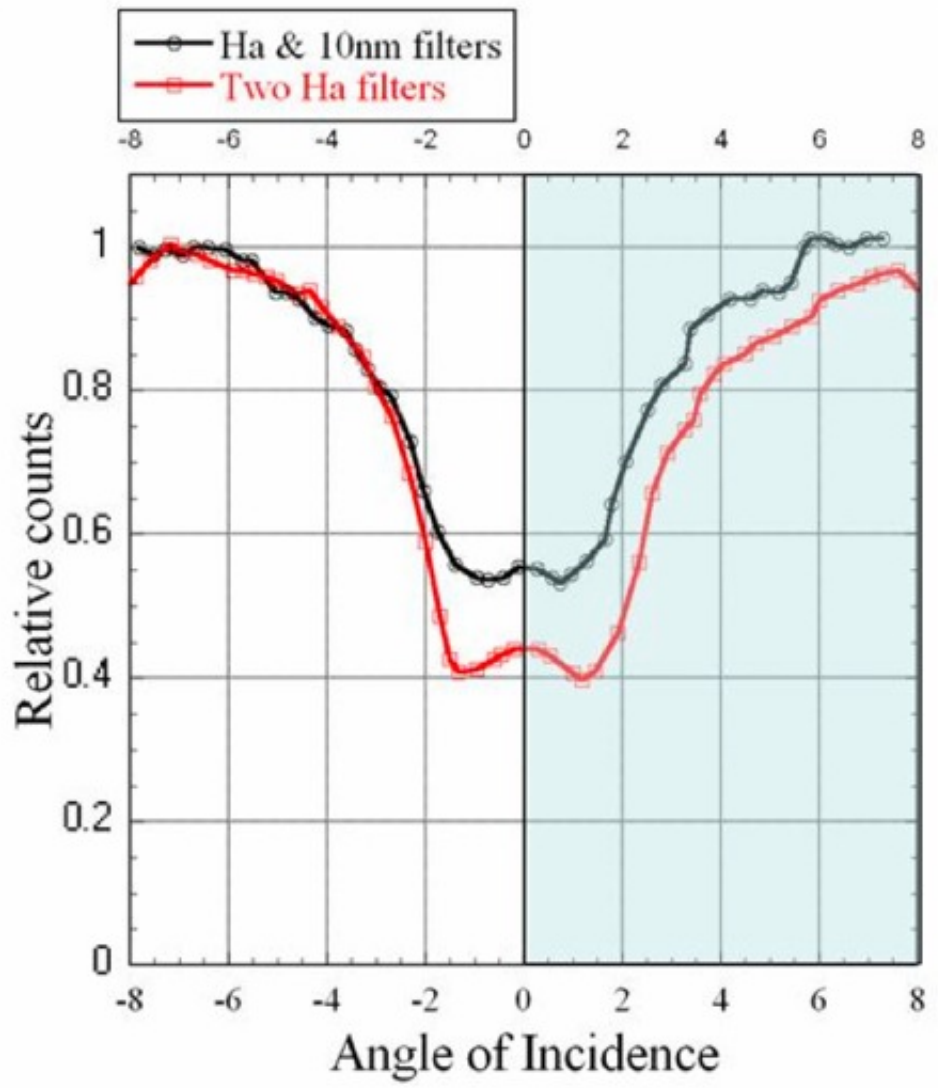

Figure 4.9: Data showing solar background counts versus tuning angle of the H-α filter. Note that the minimum does not occur precisely at normal incidence, indicating that the filter is not quite centered on the H-α line and will require some tuning during operation.

ble approach, is to use off-the-shelf visible diodes that are selected to match the H-α wavelength and then to fabricate custom electro-absorption modulators to do pulse carving. EA modulators can exploit intrinsic nonlinearities in the device to achieve extinction ratios approaching -60 dB. This approach seems to be a better alternative to both the scheme already implemented and the idea of fabricating visible VCSELs from scratch.

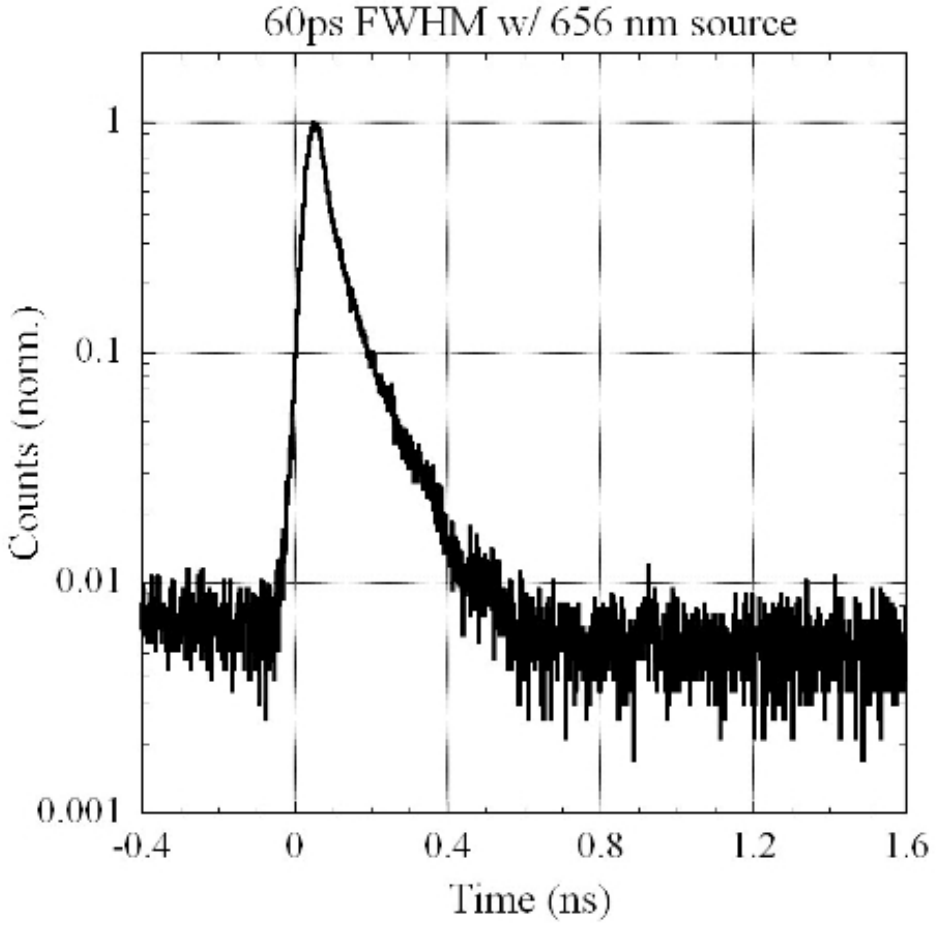

Figure 4.10: A count histogram showing a 60 ps pulse generated using an external cavity laser diode and a custom visible EO modulator. While the extinction ratio is nominally -20 dB, we have observed that the laser is not stable and the modulator suffers from oversensitivity to environmental fluctuations.

4.6 Sub-clock gating

In general, the engineering challenge in free-space quantum cryptography reduces to the problem of finding and measuring the single photon of interest among the slew of solar background photons incident on Bob's receiver aperture during any given pulse. Thus any information you can provide Bob's receiver about the single photon's wavelength, position, and time of arrival will aid in setting the signal apart from the noise. In addition to good spectral filters, narrow receiver apertures, and careful wavelength selection, another tool we can use is time gating. Even though we are limited in transmission speed by the detector's timing jitter and we synchronize Alice and Bob's clocks via the classical channel, we have much finer control over where the pulse occurs within the timing bin than simply the clock speed. In fact, as Figures 4.11 and 4.12 show, most of the detection counts fall within a very narrow window within the clock bin. Employing a technique we call *sub-clock gating*, we can essentially instruct the detector to ignore any detection events that occur outside of this small window within the clock cycle. To do this, we implement a rather simple AND gate circuit show in Figure 4.13, albeit one that must operate with very fast rise times. To achieve this, we incorporate fast InP logic that can operate at clock speeds up to 40 GHz. The histogram in Figure 4.14 shows the effects of such a circuit, eliminating much of the background noise between pulses.

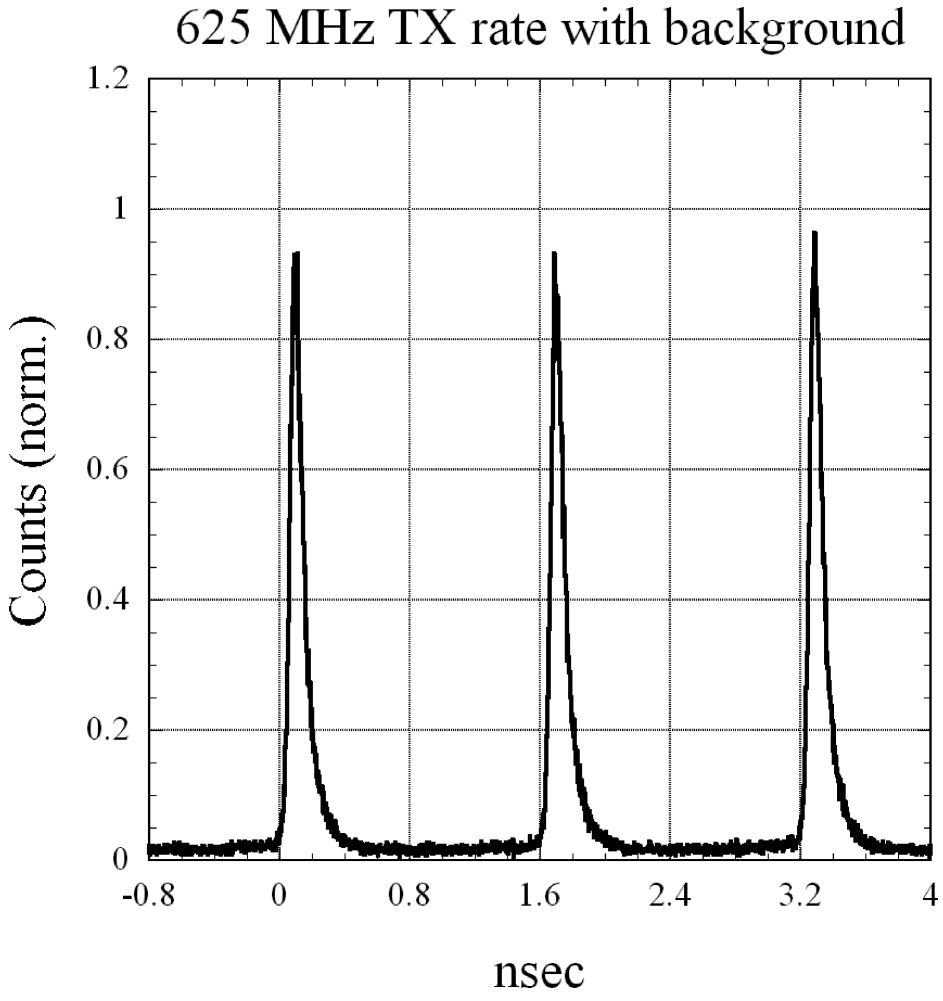

Figure 4.11: A *linear* scale histogram showing 625 MHz pulses. On this scale it is apparent that the pulses are very well localized in the transmission clock bin.

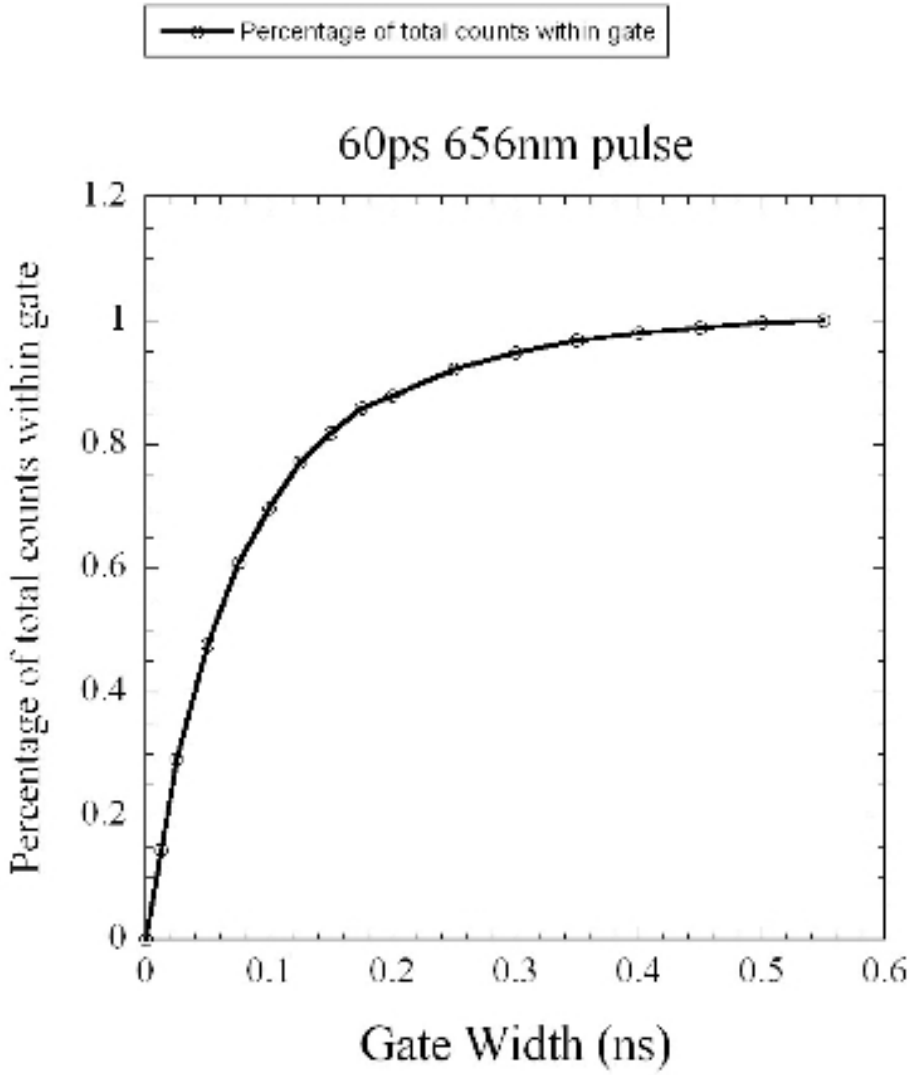

Figure 4.12: A plot showing the percentage of counts within a sub-clock gate as a function of the gate width. Note that for gate times greater than 150 ps, over 80% of the signal counts are retained while most of the background counts between transmission events are eliminated.

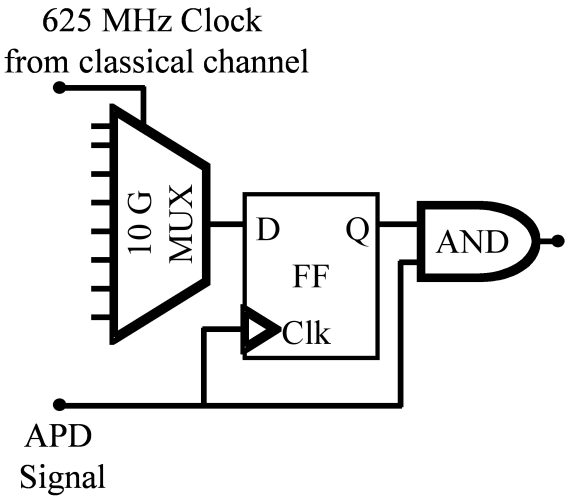

Figure 4.13: A schematic of the sub-clock gating circuit used to eliminate much of the background noise. The AND gate is a 40 GHz InP logic device.

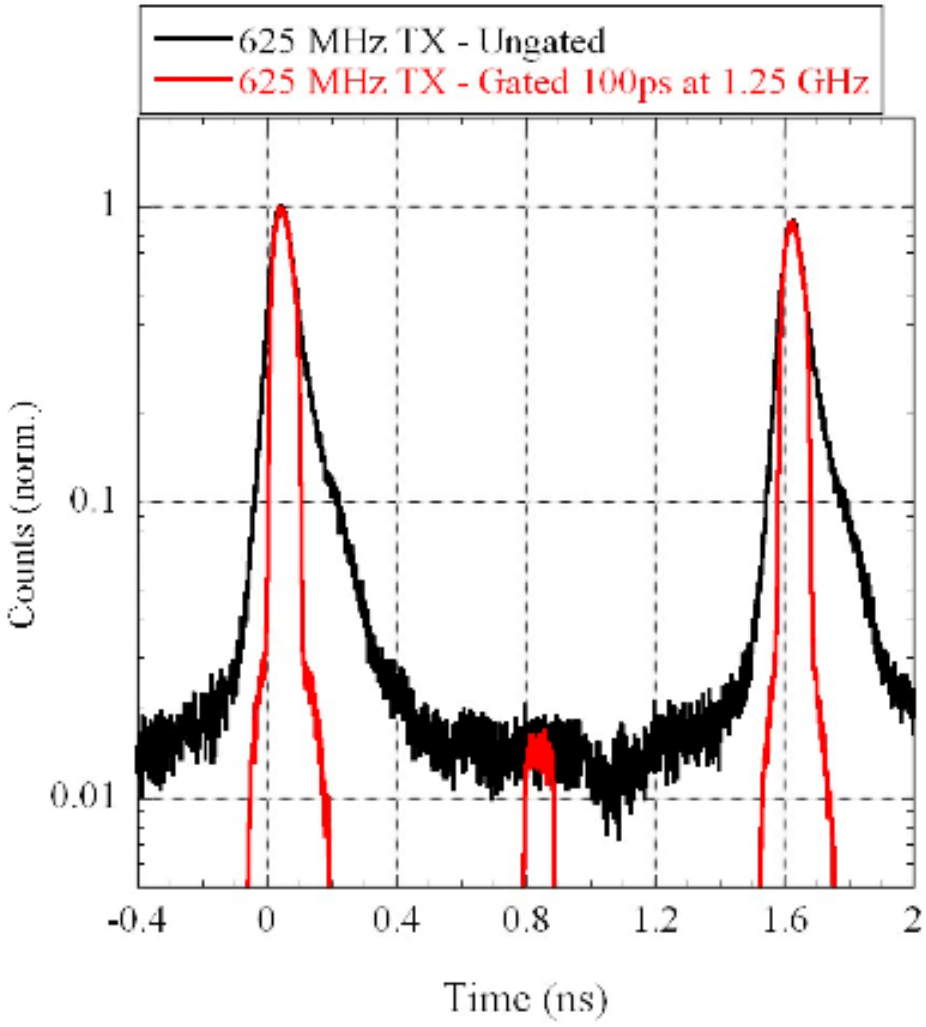

Figure 4.14: A histogram showing the receiver counts before and after the sub-clock gating circuit. Most of the background counts between pulses are eliminated.

4.7 Performance projections using MODTRAN

In considering a new free-space QKD system that incorporates all of these improvements, we can now use the various system parameters to predict its overall performance. However, one system parameter that we have not yet taken into account is the atmosphere. There are many subtleties in predicting the transmission through the atmosphere and the scattering of solar background radiation. Currently the most common model for such prediction is MODTRAN, an atmospheric modeling program created by the U.S. Air Force and used extensively to model electromagnetic wave propagation through the atmosphere. The details of the software can be found in [8]. It suffices that the model takes into account the various link parameters such as location, time of day, season, link geometry, and atmospheric conditions and provides predictions of both the atmospheric transmission and the scattered background spectrum over the wavelength range of interest. The wavelength resolution of the model is considered to be moderate, as compared with the LOWTRAN and HITRAN variants of the model. While the model does not contain the actual H-α line, we can combine the background level with the measured reduction in counts to determine the predicted value of scattered solar background photons. Figure 4.15 shows the MODTRAN outputs for our current link geometry at NIST. It assumes a mid-spring day with few clouds and good visibility.

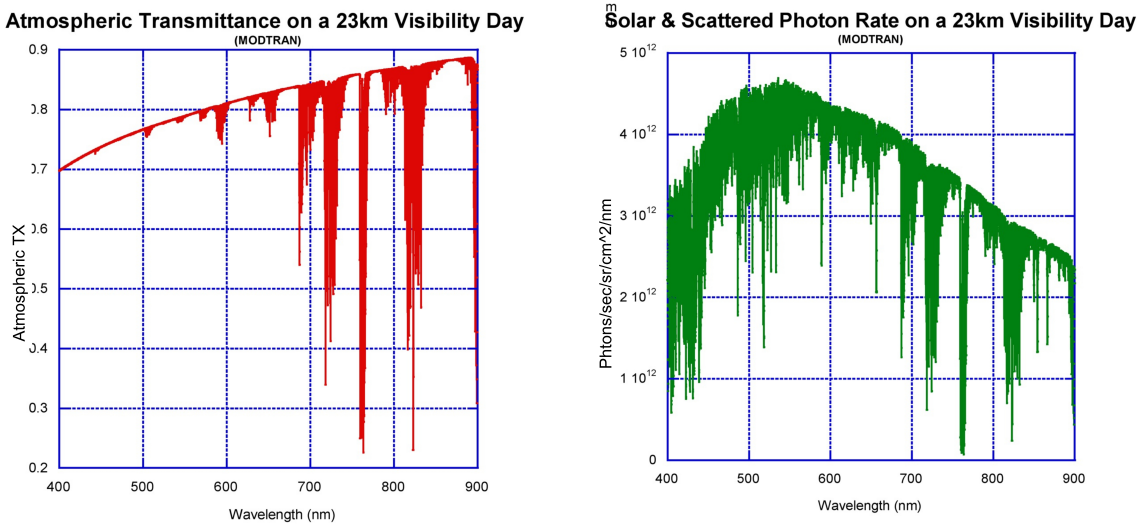

Figure 4.15: Atmospheric transmission and scattered solar background spectra from MODTRAN for the NIST free-space QKD link geometry. The model was run for the day April 15th, with few clouds and 23 km of visibility. Note the large number of water lines in the atmospheric transmission spectrum.

Using these data along with the rest of the system parameters, we now have all of the information we require to make a prediction of the link performance. The most fundamental performance metric of any QKD link is the sifted bit rate. We can determine the expected sifted bit rate for a BB84 system from the following expression:

$$\rho_{sift} = \frac{\mu T_{atmos} \eta}{2\Delta t} + 2D + \frac{R_{sky} A \Delta t' \lambda \Delta\lambda \Omega}{2hc\delta t} \tag{4.1}$$

In this expression, the first term represents the actual single photon transmissions from Alice, where μ is the mean photon number, assumed to be 0.1, T_{atmos} is the atmospheric transmission taken from MODTRAN, η is the detector efficiency, and Δt is the transmission period. The factor of two in the denominator accounts for the sifting. The second term accounts for the detector dark counts that occur when no photon is incident on the detector. D is the detector dark count rate as specified by the manufacturer, and the factor of 2 accounts for the events from four detectors, half of which are sifted. The final term accounts for solar background noise, where R_{sky} is the scattered solar background photon rate taken from combining the MODTRAN with the H-α filter data, A is the receiver aperture area, $\Delta t'$ is the gate width, λ is the wavelength, $\Delta\lambda$ is the spectral filter width, and Ω is the receiver aperture's solid angle.

From the sifted bit rate, we can then calculate the expected quantum bit error rate, or QBER, using

$$QBER = \frac{D + \frac{R_{sky} A \Delta t' \lambda \Delta\lambda \Omega}{4hc\delta t}}{\rho_{sift}} \tag{4.2}$$

As is evident from the numerator, errors come from both dark counts and scattered

sunlight.

Finally, using the QBER, we can then calculate the expected secret key rate. The secret key rate is determined directly from the QBER, as the EC/PA algorithm is limited only by the error rate. This speaks to the security of QKD, where any attempt to eavesdrop on the key will simply raise the error rate. Once the error rate goes above a few percent, whether due to dark counts, solar background, or even the presence of an eavesdropper, the EC/PA algorithm will no longer be able to distill secret bits from the sifted key and the communications will stall[2] . Below this cutoff error rate, however, the secret key rate is given by

$$\rho_{secret} = 2.8e^{(-28 \times QBER)} \tag{4.3}$$

This expression is determined purely from empirical tests of the current version of the EC/PA software running on the current Linux-based Xeon workstations and will undoubtedly change as the software is updated and ported to higher performance machines. However, using this expression, we can estimate the final secret key rate of a link built as described above. We assume the following parameters for the calculation: $\mu = 0.1$ photons/pulse, $\eta = 0.34$, $\Delta t = 400$ps, $D = 750$ counts per second, $A = 30$ cm^2, $\lambda = 656.28$ nm, $\Delta\lambda = 0.12$ nm, and the atmospheric visibility is 23 km. The results of the calculations are outlined in Table 4.1. The calculation is performed for various sub-clock gate times, showing the dramatic decrease in background noise resulting from the gating circuit.

[2]Note that, in the QKD security proof, it is assumed that all errors are due to an eavesdropper. Thus the system is not immune to a denial of service attack

Table 4.1: Performance predictions at various gate times for the proposed H-*alpha* QKD link

Gate Width (ps)	Sifted Key Rate (Mb/s)	QBER (%)	Secret Key Rate (Mb/s)
400 (ungated)	3.6635	4.10	0.8849
200	3.3040	2.28	1.4749
100	2.4586	1.55	1.8133

As the table shows, we expect to see up to a two-fold increase in the secret key rate using these new detectors at the H-α wavelength. As we increase the key rate, we will begin to approach the regime where a small number of multiplexed links can be used together to achieve the goal of a 10 Mb/S, OTP-encrypted QKD link that is compatible with first-generation ethernet.

Chapter 5

Conclusions - The future of broadband QKD

"You sort of start thinking anything's possible if you've got enough nerve." - J. K. Rowling

5.1 Advances in sources, detectors, and systems

In this thesis, I have outlined a number of advances relating to the various components of broadband QKD. I have demonstrated progress on a new source of entanglement for QKD. This source, based on birefringent phase matched four wave mixing, provides a number of advantages over existing sources of generating entanglement, including reduced Raman noise, more compact size, and stronger nonlinear interaction. However, as with any integrated photonic device, this approach poses unique challenges. The simple task of coupling light in and out of micro- and nano-scale photonics poses a particular challenge. In addition, the specific nonlinear process of interest is not well characterized compared to other nonlinear interactions. Overall, we have demonstrated that, at the very least, the ability to fabricate a device with the proper characteristics to perform birefringent phase-matched four wave mixing. We have also constructed a setup to investigate other materials that

may have stronger nonlinearity for their suitability as correlation sources.

We have also demonstrated a very important result relating to high-speed QKD using detectors with finite dead time. This result uncovers a key security vulnerability in fast QKD systems and provides a straightforward modification to the BB84 protocol that does not excessively limit the transmission rate. This modified protocol will become increasingly important as systems begin to operate at multi-gigahertz transmission rates. The counterintuitive existence of an optimal transmission rate is also an interesting result of this analysis.

Finally, I present a new concept of a QKD system based around faster detectors that will hopefully push the key rate toward the goal of 10 Mb/s. This system, when built, will represent the most advanced QKD link in the world.

5.2 Personal contributions to each project

I personally made significant contributions to each project. The nonlinear photonics project, in fact, was an original idea inspired by the research of Dr. Alan Migdall's group at NIST in microstructure fiber sources. I initiated the project and was the sole party responsible at each step, excluding the actual device fabrication, but including refining the concept of the approach, designing the device, and characterizing its properties and performance. I gained extensive experience in photonic device design using various simulation tools and did all of the work setting up the source and apparatus to perform the testing. In addition, I gained invaluable experience setting up a collaboration between NIST and LPS. Once the collaboration

was established, the actual MBE and fabrication work was left in the able hands of Dr. Chris Richardson.

On the project investigating detector dead-time issues, I was specifically responsible for the simulation work and the investigation of the key correlations. I made contributions to the analytical investigation, though the original concept of the state space was initiated by Dr. Anastase Nakassis and Dr. Joshua Bienfang. In addition, some of the calculations relating to transition probabilities were developed by Dr. Hai Xu.

My personal contributions to the Hα system design were extensive, though I was not the original inspiration for the idea. That concept was initiated by Dr. Bill Jeffrey, formerly the Director of NIST, upon hearing about the problems associated with the new SPADs. Once we began looking into operating in one of the Fraunhofer lines, I was responsible for all of the MODTRAN modeling and the calculations relating to the Hα line. The experiemental measurements of the solar background counts and the calculations predicting the system performance were an equal effort between Dr. Joshua Bienfang and myself. The calibration of the Hα filter was performed by Dr. Steve Brown of the NIST SIRCUS facility.

5.3 The future of free-space and fiber optic QKD

Recent demonstrations of free-space QKD have elicited strong interest in the technology for practical applications. The final chapter of this thesis proposed a new design for a free-space QKD system. While it may not end up that most fast

free-space QKD systems rely on a Fraunhofer line to reduce the solar background noise, the basic elements of the free-space system proposed will hopefully continue to push the speed limits of quantum cryptography for a number of years. Already there are plans to demonstrate this new system over a 16 km link provided by the Naval Research Laboratory. If successful, this will certainly be the most advanced demonstration of free-space QKD in terms of distance-bandwidth product. Such a system will be vital to implementing the concept of a satellite-based, global quantum encrypted network.

The other facet of QKD research focuses on short haul fiber based systems. While not addressed here, there are a number of research avenues that have the potential to make great advances in fiber-based QKD. One of the most promising approaches to fiber-based QKD is single-sideband encoding[41]. In this approach, the bits are encoded on the frequency of the single quanta, a property that is much more stable in a fiber than polarization or phase. While this approach has received little attention in the research community, it promises to be one of the most practicable implementations of QKD known to date. In fact, it even offers much more potential as an entangled system than other approaches. Generating frequency entanglement is significantly more straightforward than generating polarization entanglement and often involves only a single nonlinear process rather than two. Hence there is much potential to make advances in frequency-coding and develop frequency entanglement sources for significantly improved fiber-based quantum key distribution systems.

5.4 QKD in the marketplace

The final question that remains in the QKD community is whether there exists a real-world need for such a technology. In the modern IT security infrastructure, cryptography is certainly one of the strongest links in the chain. Often vulnerabilities occur from faulty implementations of security procedures or social engineering attacks on lapses in user vigilance. Thus it is almost always easier to attack another point of entry rather than the encrypted data set itself.

That said, there still exists a vulnerability. Not only is there no proof that one cannot break RSA encryption with a classical computer, but quantum computers that are on the long-term technology horizon have already been shown to readily break RSA. Even if a system implements perfect security with the most advanced classical cryptography, it is still vulnerable to attacks that record the ciphertext in order to break it in 50 years or so, when the technology required will very likely be available. Thus any customer who has information that needs to be secret for a long period of time, whether it be governments, financial institutions, or health care providers, has a need for communications that are immune to eavesdropping from *any* adversary, now or in the future.

Whether this market exists is still unknown. However, even it if did, there is still a significant amount of technical work that remains until QKD can keep pace with today's, and especially tomorrow's, IT infrastructure. Increases in speed and transmission distance, robustness to real-world perturbations, and validations of the security of the protocols all still need research attention if any market were

to become interested in quantum encryption.

The future of QKD is at a crossroads. If it becomes of interest in the IT marketplace, then we will look forward to many exciting advances as the technology gains its place in the everyday world. If it remains only an exotic sideshow in the IT security community, then it certainly has produced technology that will be of use in other arenas. Either way, science is science, and we continue to trudge on.

Appendix A

Index of refraction calculation for aluminum gallium arsenide

This C code implements the detailed calculation of the bulk index of refraction for the ternary zincblende semiconductor Al(x)Ga(1-x)As, for any temperature and aluminum fraction. The code is implemented as a MATLAB MEX file that can be compiled into a MATLAB-compatible function. For a detailed description of this code and its applications, see [23].

```
//file algaas_mex.c

/***********************************
MATLAB MEX file version of...

Calculator for index of refraction
of bulk Al(x)Ga(1-x)As including
temperature dependence

MATLAB format: n = algaas_mex(lambda, x, T)

Based on Gehrsitz, et al., J. Appl. Phys.,
```

```
Vol. 87, No. 11, June 2000


Created 29 Aug 2008

Modified 31 Aug 2008


by DR

***********************************/


#include <math.h>

#include <stdio.h>

#include <stdlib.h>


#include "mex.h"


#define kB 0.0861708 //in meV/K


double coth(double z)

{

        return ((exp(2.0*z)+1)/(exp(2.0*z)-1));

}


double R(double lambda, double x)  //lambda in microns

{
```

```
double E_sq, r;

const double E2_sq = 0.724e-3;
const double C2    = 1.55e-3;
const double E3_sq = 1.331e-3;
const double C3    = 2.61e-3;

E_sq = pow(lambda, -2.0);

r = (((1-x)*C2)/(E2_sq-E_sq)) + (x*C3/(E3_sq-E_sq));

return r;
}

double A0(double T) //temp in K
{
return 5.9613 + (7.178e-4*T) - (0.953e-6*T*T);
}

double E10_sq(double T) //temp in K
{
return 4.7171 - (3.237e-4*T) - (1.358e-6*T*T);
}
```

```
double E_Gamma(double T) //temp in K

{

double p,q,eg;


p = 1 - coth(15.9/(2.0*kB*T));

q = 1 - coth(33.6/(2.0*kB*T));


eg = 1.5192 + (1.8*15.9e-3*p) + (1.1*33.6e-3*q);


return eg/1.239865;

}


double A(double x, double T)

{

double c0;

const double c1 = -16.159;

const double c2 = 43.511;

const double c3 = -71.317;

const double c4 = 57.535;

const double c5 = -17.451;


c0 = A0(T);
```

```
return c0 + c1*x + c2*x*x + c3*x*x*x + c4*x*x*x*x + c5*x*x*x*x*x;
}

double C1(double x)
{
const double c0 = 21.5647;
const double c1 = 113.74;
const double c2 = -122.5;
const double c3 = 108.401;
const double c4 = -47.318;

return c0 + c1*x + c2*x*x + c3*x*x*x + c4*x*x*x*x;
}

double E1_sq(double x, double T)
{
double c0;
const double c1 = 11.006;
const double c2 = -3.08;

c0 = E10_sq(T);
```

```
return c0 + c1*x + c2*x*x;

}


double C0(double x)

{

double c;


const double c0 = 50.535;

const double c1 = -150.7;

const double c2 = -62.209;

const double c3 = 797.16;

const double c4 = -1125;

const double c5 = 503.79;


c = c0 + c1*x + c2*x*x + c3*x*x*x + c4*x*x*x*x + c5*x*x*x*x*x;


return (1.0/c);

}


double E0_sq(double x, double T)

{

double e0;
```

```
double c0;

const double c1 = 1.1308;

const double c2 = 0.1436;


c0 = E_Gamma(T);


e0 = c0 + c1*x + c2*x*x;


return (e0*e0);

}


double n(double lambda, double x, double T) //lambda in microns, x in %, T in K

{

double E_sq, n_sq;


E_sq = pow(lambda, -2.0);


n_sq = A(x,T) + (C0(x)/(E0_sq(x,T) - E_sq)) + (C1(x)/(E1_sq(x,T) - E_sq)) + R(lamb


return sqrt(n_sq);

}


/************************************
```

```
MATLAB equivalent of main()
used in creating a C-based MATLAB compatible function...

For fomatting and documentation, see MATLAB mex helpfile or http://cnx.org/
************************************/

void mexFunction (int nlhs, mxArray *plhs[], int nrhs, const mxArray *prhs[])
{
  double *xValues, *outArray;
  int, i, j;
  double lambda, x, T, index;

  if (nrhs == 3){
    xValues = mxGetPr(prhs[0]);
    lambda = xValues[0];
    xValues = mxGetPr(prhs[1]);
    x = xValues[0];
    xValues = mxGetPr(prhs[2]);
    T = xValues[0];

    index = n(lambda, x, T);

    plhs[0] = mxCreateDoubleMatrix(1, 1, mxREAL);
```

```
    outArray = mxGetPr(plhs[0]);
    outArray[0] = index;
  }

}
```

Appendix B

Monte-Carlo QKD Simulation Code

For a detailed description of this code and its applications, see Chapter 3.

B.1 Traditional BB84 protocol

The following code performs a simulation of traditional BB84 QKD using unitless time. It runs continuously until it accumulates *NBITS* of sifted key.

```
//file: qkd_trans_prob.c


#include <stdlib.h>

#include <time.h>

#include <math.h>


//===== OPERATING PARAMETERS =======


#define NBITS 1048576


#define DEAD_START 101

#define DEAD_STOP 500
```

```
#define DEAD_STEP 2


#define LINK_LOSS 0.99


#define QBER 0.0


#define H_EFF 0.6

#define V_EFF 0.6

#define L_EFF 0.6

#define R_EFF 0.6


//================================


//Model Constants


#define HV_BASIS 0

#define LR_BASIS 1


//H=L=0, V=R=1


#define H 0

#define V 1

#define L 2
```

```
#define R 3


//Global vars


int alice_key[NBITS];

int bob_key[NBITS];


double det_eff[4];


int dead_flags[4];

int dead_counts[4];


int a_basis, a_value, b_basis, b_value;


long int bit_count;


long double clk;


//=======Statistics to record============


int det_counts[4];


double transition_prob;
```

```
//=====================================


double NextBitTransitionProb(int key[])
{
  //calculates the probability that the next bit
  //is different from the previous bit...

  int i, count;

  count=0;
  for(i=0; i<NBITS-1; i++)
    if(key[i] != key[i+1])
      count++;

  return (double)count/(double)(NBITS-1);
}//=====================================


void doQKD(int dead_time)
{
  int i, d, b;
  double r1, r2, r3, r4, r5, r6, r7;

  //initialize everything
```

```
srand(time(NULL));


clk = 0.0;

bit_count = 0;


for(d=H; d<=R; d++){

  det_counts[d] = 0;

  dead_flags[d] = 0;

  dead_counts[d] = 0;

}


det_eff[H] = H_EFF;

det_eff[V] = V_EFF;

det_eff[L] = L_EFF;

det_eff[R] = R_EFF;


while (bit_count < NBITS){


  //roll seven dice...

  r1 = (double)rand()/((double)RAND_MAX + 1); //Alice's basis

  r2 = (double)rand()/((double)RAND_MAX + 1); //Alice's bit value

  r3 = (double)rand()/((double)RAND_MAX + 1); //link loss
```

```
    r4 = (double)rand()/((double)RAND_MAX + 1); //Bob's basis
    r5 = (double)rand()/((double)RAND_MAX + 1); //Bob's bit value (if req'd)
    r6 = (double)rand()/((double)RAND_MAX + 1); //Error introduction/Background Co
    r7 = (double)rand()/((double)RAND_MAX + 1); //Detector Efficiency

    //generate Alice's random bits...

    a_basis = (r1 < 0.5) ? HV_BASIS : LR_BASIS;
    a_value = (r2 < 0.5) ? 0 : 1;

    if (r3 > LINK_LOSS){
      //select Bob's basis
      b_basis = (r4 < 0.5) ? HV_BASIS : LR_BASIS;
      //sift
      if (a_basis == b_basis){  //Bit goes through and is recorded...
b_value = a_value;
b_value = (r6<QBER) ? (!b_value) : b_value;  //introduce errors...
d = (b_basis*2) + b_value;  //selects H,V,L or R... I know, I'm pretty clever...
if ((r7 < det_eff[d]) && (!dead_flags[d])){
  alice_key[bit_count] = a_value;
  bob_key[bit_count] = b_value;
  bit_count++;
  dead_flags[d] = 1;
```

```
  det_counts[d]++;
}
      }
      else{  //basis is wrong, and detector fires, but bit is not recorded...
b_value = (r5 < 0.5) ? 1 : 0;
d = (b_basis*2) + b_value;
        if ((r7 < det_eff[d]) && (!dead_flags[d])){
  dead_flags[d] = 1;
  det_counts[d]++;
}
      }
    }

    //increment dead time counters with the clock...

    for (d=H; d<=R; d++){
      if (dead_flags[d]){
if(dead_counts[d] < dead_time)
  dead_counts[d]++;
else{
  dead_flags[d] = 0;
  dead_counts[d] = 0;
        }
```

```
      }
    }
    clk = clk + 1.0;
  }
}

//====================================

int main(int argc, char* argv[])
{
  int i, j, result, dt;
  double avg, rate;

  int dstart, dstop, dstep;

  if (argc == 4){
    dstart = atoi(argv[1]);
    dstop = atoi(argv[2]);
    dstep = atoi(argv[3]);
  }
  else{
    dstart = DEAD_START;
    dstop = DEAD_STOP;
```

```
    dstep = DEAD_STEP;
  }

  for(dt=dstart; dt<=dstop; dt+=dstep){
    doQKD(dt);
    transition_prob = NextBitTransitionProb(bob_key);
    avg = (det_counts[H] + det_counts[V] + det_counts[L] + det_counts[R])/4.0;
    rate = avg/clk;
    printf("%d\t%f\t%f\n", dt, rate, transition_prob);
  }

  return 0;
}
```

B.2 Modified BB84 protocol

This version of the code incorporates the modification required to maintain security in the presence of long dead times (see Chapter 3). Unlike the previous version, this one runs the simulation in 'real time,' meaning it includes a fundamental time unit of 100 ps. All transmission events, dead times, etc., are scaled to this fundamental simulation time unit.

```
//file: fc.c
```

```
#include <stdlib.h>

#include <time.h>

#include <math.h>


//===== OPERATING PARAMETERS =======


#define RUN_TIME (100000000) //number of 0.1 nsec units to run simulation


#define TX_PER_START_EXP (0.0)

#define TX_PER_STOP_EXP (3.0)   //tx_per = 10 ^ TX_Exp...

#define TX_PER_STEP_EXP (0.2)


//we want to scan from 10 MHz to 10 GHz TX rate

//so we set the fundamental time unit to 100 ps

//and run for TX rates from 1 clock cycle to

// 1,000 clock cycles


#define DEAD_TIME 1500   //in 0.1 nsec units


#define LINK_LOSS (0.99)


#define QBER (0.0)
```

```
#define H_EFF (1.0)

#define V_EFF (1.0)

#define L_EFF (1.0)

#define R_EFF (1.0)


//================================


//Model Constants


#define HV_BASIS 0

#define LR_BASIS 1


//H=L=0, V=R=1


#define H 0

#define V 1

#define L 2

#define R 3


//Global vars (I know... don't say anything)


int dead_flags[4];
```

```
int dead_counts[4];

int sift_flags[2];

int a_basis, a_value, b_basis, b_value;
long int bit_count;
long int tossed_bits;

int bob_key[RUN_TIME];

double det_eff[4];

double transition_prob;

//============================================================================

void doQKD(long int tx_per, int dead_time) //both params in units of 100 ps
{
  int i, d, d2;
  double r1, r2, r3, r4, r5, r6, r7;
  long int clk;

  //initialize everything
```

```
srand(time(NULL));


bit_count = 0;

tossed_bits = 0;


for(d=H; d<=R; d++){

  dead_flags[d] = 0;

  dead_counts[d] = 0;

}


sift_flags[HV_BASIS] = 0;

sift_flags[LR_BASIS] = 0;


det_eff[H] = H_EFF;

det_eff[V] = V_EFF;

det_eff[L] = L_EFF;

det_eff[R] = R_EFF;


for (clk = 0; clk <= RUN_TIME; clk++){


  if((clk % tx_per) == 0){

   //roll seven dice...
```

```
      r1 = (double)rand()/((double)RAND_MAX + 1); //Alice's basis

      r2 = (double)rand()/((double)RAND_MAX + 1); //Alice's bit value

      r3 = (double)rand()/((double)RAND_MAX + 1); //link loss

      r4 = (double)rand()/((double)RAND_MAX + 1); //Bob's basis

      r5 = (double)rand()/((double)RAND_MAX + 1); //Bob's bit value (if req'd)

      r6 = (double)rand()/((double)RAND_MAX + 1); //Error introduction/Background C

      r7 = (double)rand()/((double)RAND_MAX + 1); //Detector Efficiency


      //generate Alice's random bits...


      a_basis = (r1 < 0.5) ? HV_BASIS : LR_BASIS;
a_value = (r2 < 0.5) ? 0 : 1;


      if (r3 > LINK_LOSS){
        //select Bob's basis
        b_basis = (r4 < 0.5) ? HV_BASIS : LR_BASIS;
        //sift
        if (a_basis == b_basis){  //Bit goes through and is recorded...
b_value = a_value;
b_value = (r6<QBER) ? (!b_value) : b_value;  //introduce errors...


d = (b_basis*2) + b_value;  //selects H,V,L or R...
```

```
//d2 is assigned other detector in basis - this is the modification to BB84...
d2 = ((d%2) == 0) ? d+1 : d-1;


if ((r7 < det_eff[d]) && (!dead_flags[d])){
     if (!dead_flags[d2]){
    bob_key[bit_count] = b_value;
    bit_count++;
 sift_flags[b_basis] = 1;   //a bit is recorded... set sift flag
     }
     else{
  if (sift_flags[b_basis] == 0){
    bob_key[bit_count] = b_value;
    bit_count++;
    sift_flags[b_basis] = 1;  //a bit is recorded... set sift flag
  }
}
dead_flags[d] = 1;
}
       }
       else{  //basis is wrong, and detector fires, but bit is not recorded...
b_value = (r5 < 0.5) ? 1 : 0;
d = (b_basis*2) + b_value;
         if ((r7 < det_eff[d]) && (!dead_flags[d])){
```

```
   dead_flags[d] = 1;
}
     } //Sifting
     }  //Link Loss
    } //TX block

    //increment dead time counters with the clock...

    for (d=H; d<=R; d++){
      if (dead_flags[d]){
if(dead_counts[d] < dead_time)
   dead_counts[d]++;
else{
   dead_flags[d] = 0;
   dead_counts[d] = 0;
}
      }
    }
    if ((!(dead_flags[H])) && (!(dead_flags[V])))
      sift_flags[HV_BASIS] = 0;
    if ((!(dead_flags[L])) && (!(dead_flags[R])))
      sift_flags[LR_BASIS] = 0;
  }
```

```
}


//====================================


int main(int argc, char* argv[])
{
  int tx_per;
  double t_ex, trans_prob;


  for(t_ex=TX_PER_START_EXP; t_ex <= TX_PER_STOP_EXP; t_ex += TX_PER_STEP_EXP){
   tx_per = (int)ceil(pow(10.0, t_ex));
    doQKD(tx_per, DEAD_TIME);
    //trans_prob = TransitionProb(bob_key, bit_count);
    printf("%d\t%d\n", tx_per, bit_count);
  }
  return 0;
}
```

Bibliography

[1] M. A. Afromowitz. Refractive index of Ga(1-x)Al(x)As. *Solid State Communications*, 15(59), 1974.

[2] G. E. Albert and Lewis Nelson. Contributions to the statistical theory of counter data. *Ann. Math. Statist.*, 24(1):9–22, 1953.

[3] Alain Aspect, Jean Dalibard, and Gérard Roger. Experimental test of Bell's inequalities using time- varying analyzers. *Phys. Rev. Lett.*, 49(25):1804–1807, Dec 1982.

[4] Alain Aspect, Philippe Grangier, and Gérard Roger. Experimental tests of realistic local theories via Bell's theorem. *Phys. Rev. Lett.*, 47(7):460–463, Aug 1981.

[5] J. S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1(3):195–200, November 1964.

[6] C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. In *IEEE International Conference on Computers, Systems, and Signal Processing, Bangalore, India*, page 175. IEEE, 1984.

[7] Charles H. Bennett. Quantum cryptography using any two nonorthogonal states. *Phys. Rev. Lett.*, 68(21):3121–3124, May 1992.

[8] Alexander Berk, Gail P. Anderson, Lawrence S. Bernstein, Prabhat K. Acharya, H. Dothe, Michael W. Matthew, Steven M. Adler-Golden, Jr. James H. Chetwynd, Steven C. Richtsmeier, Brian Pukall, Clark L. Allred, Laila S. Jeong, and Michael L. Hoke. Modtran4 radiative transfer modeling for atmospheric correction. *SPIE Optical Spectroscopic Techniques and Instrumentation for Atmospheric and Space Research III*, 3756(1):348–353, 1999.

[9] J. Bienfang, A. Gross, A. Mink, B. Hershman, A. Nakassis, X. Tang, R. Lu, D. Su, Charles Clark, Carl Williams, E. Hagley, and Jesse Wen. Quantum key distribution with 1.25 Gbps clock synchronization. *Opt. Express*, 12(9):2011–2016, 2004.

[10] J. C. Bienfang, Charles W. Clark, Carl J. Williams, E. Hagley, and Jesse Wen. Advantages of high-speed technique for quantum key distribution; reply to quant-ph/0407050, July 2004.

[11] L. Blum, M. Blum, and M. Shub. A simple unpredictable pseudo-random number generator. *SIAM Journal on Computing*, 15(2):364–383, 1986.

[12] M. Born and E. Wolf. *Principles of Optics*. Cambridge, 1999.

[13] R. W. Boyd. *Nonlinear Optics*. Academic Press, 2003.

[14] G. Brassard and L. Salvail. *Secret-key reconciliation by public discussion*, volume 765 of *Lecture Notes in Computer Science*, pages 410–423. Springer, 1994.

[15] Mazzoni D. Puleo M. Costa, B. and E. Vezzoni. Phase shift technique for the measurement of chromatic dispersion in optical fibers using leds. *IEEE Journal of Quantum Electronics*, 18(10):1509–1515, October 1982.

[16] M.D. Dvorak, W.A. Schroeder, D.R. Andersen, A.L. Smirl, and B.S. Wherrett. Measurement of the anisotropy of two-photon absorption coefficients in zincblende semiconductors. *Quantum Electronics, IEEE Journal of*, 30(2):256–268, Feb 1994.

[17] Artur K. Ekert. Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.*, 67(6):661–663, Aug 1991.

[18] A.B. Fallahkhair, K.S. Li, and T.E. Murphy. Vector finite difference modesolver for anisotropic dielectric waveguides. *Lightwave Technology, Journal of*, 26(11):1423–1431, June1, 2008.

[19] J. Fan, A. Migdall, and L. J. Wang. Efficient generation of correlated photon pairs in a microstructure fiber. *Optics Letters*, 30:3368–3370, December 2005.

[20] A. Fiore, S. Janz, L. Delobel, P. van der Meer, P. Bravetti, V. Berger, E. Rosencher, and J. Nagle. Second-harmonic generation at lambda = 1.6 microns in AlGaAs/Al(2)O(3) waveguides using birefringence phase matching. *Applied Physics Letters*, 72(23):2942–2944, 1998.

[21] C. Flytzanis. Third-order optical susceptibilities in IV-IV and III-V semiconductors. *Physics Letters A*, 31:273–274, March 1970.

[22] Mark A. Foster, Amy C. Turner, Reza Salem, Michal Lipson, and Alexander L. Gaeta. Broad-band continuous-wave parametric wavelength conversion in silicon nanowaveguides. *Opt. Express*, 15(20):12949–12958, 2007.

[23] S. Gehrsitz, F. K. Reinhart, C. Gourgon, N. Herres, A. Vonlanthen, and H. Sigg. The refractive index of Al(x)Ga(1-x)As below the band gap: Accurate determination and empirical modeling. *Journal of Applied Physics*, 87(11):7825–7837, 2000.

[24] J. Goldhar and W. Herman. publication in preparation.

[25] Karen Gordon, Veronica Fernandez, Gerald Buller, Ivan Rech, Sergio Cova, and Paul Townsend. Quantum key distribution system clocked at 2 GHz. *Opt. Express*, 13(8):3015–3020, 2005.

[26] A. Gulinatti, P. Maccagnani, I. Rech, M. Ghioni, and S. Cova. 35 ps time resolution at room temperature with large area single photon avalanche diodes. *Electronics Letters*, 41(5):272–274, March 2005.

[27] M. Holtz, R. Zallen, Art E. Geissberger, and R. A. Sadler. Raman-scattering studies of silicon-implanted gallium arsenide: The role of amorphicity. *Journal of Applied Physics*, 59(6):1946–1951, 1986.

[28] W. C. Huffman and V. Pless. *Fundamentals of Error-Correcting Codes*. Cambridge, 2003.

[29] E. L. Kerr. Fraunhofer filters to reduce solar background for optical communications. In *TDA Progress Report*. Jet Propulsion Laboratory, 1986.

[30] G. F. Knoll. *Radiation Detection and Measurement*. Wiley, 1979.

[31] Onur Kuzucu, Franco N. C. Wong, David E. Zelmon, Shrikrishna M. Hegde, Tony D. Roberts, and Philip Battle. Generation of 250 mw narrowband pulsed ultraviolet light by frequency quadrupling of an amplified erbium-doped fiber laser. *Opt. Lett.*, 32(10):1290–1292, 2007.

[32] Paul G. Kwiat, K. Mattle, H. Weinfurter, and A. Zeilinger. New high-intensity source of polarization-entangled photon pairs. *Physical Review Letters*, 75(24):4337–4341, December 1995.

[33] M. J. LaGasse, K. K. Anderson, C. A. Wang, H. A. Haus, and J. G. Fujimoto. Femtosecond measurements of the nonresonant nonlinear index in algaas. *Applied Physics Letters*, 56(5):417–419, 1990.

[34] Xiaoying Li, Jun Chen, Paul Voss, Jay Sharping, and Prem Kumar. All-fiber photon-pair source for quantum communications: Improved generation of correlated photons. *Opt. Express*, 12(16):3737–3744, 2004.

[35] Q. Lin, F. Yaman, and Govind P. Agrawal. Photon-pair generation by four-wave mixing in optical fibers. *Opt. Lett.*, 31(9):1286–1288, 2006.

[36] S. Lloyd. *Ancient Turkey: A Traveller's History*. Univ. of California Press, 1999.

[37] R. Loudon. *The Quantum Theory of Light*. Oxford, 2000.

[38] J. M. Manley and H. E. Rowe. General energy relations in nonlinear reactances. *Proc. I. R. E.*, 47:2115–2116, 1959.

[39] Ivan Marcikic, Antía Lamas-Linares, and Christian Kurtsiefer. Free-space quantum key distribution with entangled photons. *Applied Physics Letters*, 89(10):101122, 2006.

[40] Makoto Matsumoto and Takuji Nishimura. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Model. Comput. Simul.*, 8(1):3–30, 1998.

[41] Duraffourg L. Goedgebuer J. P. Soujaeff A. Patois F. Merolla, J. M. and W. T. Rhodes. Integrated quantum key distribution system using single sideband detection. *Eur. Phys. J. D*, 18:141–146, 2002.

[42] A. Migdall. Correlated-photon metrology without absolute standards. *Physics Today*, 52:41–46, January 1999.

[43] I. Newton. *Principia mathematica*. 1687.

[44] E. D. Palik. *Handbook of Optical Constants of Solids*. Academic Press, 1985.

[45] W. W. Peterson and E. J. Weldon. *Error-Correcting Codes*. MIT Press, 1961.

[46] J. Rarity, J. Fulconis, J. Duligall, W. Wadsworth, and P. Russell. Photonic crystal fiber source of correlated photon pairs. *Opt. Express*, 13(2):534–544, 2005.

[47] J. G. Rarity, P. C. M. Owens, and P. R. Tapster. Quantum random-number generation and key sharing. *J. Mod. Opt.*, 41(12):2435–2444, December 1994.

[48] J G Rarity, P R Tapster, P M Gorman, and P Knight. Ground to satellite secure key exchange using quantum cryptography. *New Journal of Physics*, 4:82, 2002.

[49] Joseph Reader. Reference wavelengths for strong lines of atomic hydrogen and deuterium. *Appl. Spectrosc.*, 58(12):1469–1474, 2004.

[50] D. Rogers, J. C. Bienfang, A. Mink, B. J. Hershman, A. Nakassis, L. Ma, X. Tang, D. H. Su, Charles W. Clark, and Carl J. Williams. High-speed photon counting techniques for broadband quantum key distribution. *SPIE Advanced Photon Counting Techniques*, 6372(1):637211, 2006.

[51] Daniel J. Rogers, Joshua C. Bienfang, Anastase Nakassis, Hai Xu, and Charles W. Clark. Detector dead-time effects and paralyzability in high-speed quantum key distribution. *New Journal of Physics*, 9(9):319, 2007.

[52] K. Schatzel, R. Kalstrom, B. Stampa, and J. Ahrens. Correction of detection-system dead-time effects on photon-correlation functions. *J. Opt. Soc. Am. B*, 6(5):937, 1989.

[53] Tobias Schmitt-Manderbach, Henning Weier, Martin Fürst, Rupert Ursin, Felix Tiefenbacher, Thomas Scheidl, Josep Perdigues, Zoran Sodnik, Christian Kurtsiefer, John G. Rarity, Anton Zeilinger, and Harald Weinfurter. Experimental demonstration of free-space decoy-state quantum key distribution over 144 km. *Physical Review Letters*, 98(1):010504, 2007.

[54] C. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28(4):656–715, 1949.

[55] Jay E. Sharping, Kim F. Lee, Mark A. Foster, Amy C. Turner, Bradley S. Schmidt, Michal Lipson, Alexander L. Gaeta, and Prem Kumar. Generation of correlated photons in nanoscale silicon waveguides. *Opt. Express*, 14(25):12388–12393, 2006.

[56] M. Sheik-Bahae, D. J. Hagan, and E. W. Van Stryland. Dispersion and band-gap scaling of the electronic kerr effect in solids associated with two-photon absorption. *Phys. Rev. Lett.*, 65(1):96–99, Jul 1990.

[57] Y. H. Shih, A. V. Sergienko, M. H. Rubin, T. E. Kiess, and C. O. Alley. Two-photon interference in a standard Mach-Zehnder interferometer. *Phys. Rev. A*, 49(5):4243–4246, May 1994.

[58] Peter W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, 1997.

[59] S. Singh. *The Code Book*. Anchor, 2000.

[60] A. V. Smith. How to select nonlinear crystals and model their performance using SNLO software. available at http://www.as-photonics.com/SNLO/.

[61] R. H. Stolen, M. A. Bosch, and Chinlon Lin. Phase matching in birefringent fibers. *Opt. Lett.*, 6(5):213, 1981.

[62] W. Stolz. Ph.d. thesis (unpublished). University of Stuttgart, 1986.

[63] Hiroki Takesue, Eleni Diamanti, Carsten Langrock, M. M. Fejer, and Yoshihisa Yamamoto. 10-GHz clock differential phase shift quantum key distribution experiment. *Opt. Express*, 14(20):9522–9530, 2006.

[64] R T Thew, S Tanzilli, L Krainer, S C Zeller, A Rochas, I Rech, S Cova, H Zbinden, and N Gisin. Low jitter up-conversion detectors for telecom wavelength GHz QKD. *New Journal of Physics*, 8(3):32, 2006.

[65] Amy C. Turner, Christina Manolatou, Bradley S. Schmidt, Michal Lipson, Mark A. Foster, Jay E. Sharping, and Alexander L. Gaeta. Tailored anomalous group-velocity dispersion in silicon channel waveguides. *Opt. Express*, 14(10):4357–4362, 2006.

[66] J. P. van der Ziel and A. C. Gossard. Optical birefringence of ultrathin Al(x)Ga(1-x)As-GaAs multilayer heterostructures. *Journal of Applied Physics*, 49(5):2919–2921, 1978.

[67] J. P. van der Ziel and A. C. Gossard. Two-photon absorption spectrum of AlAs-GaAs monolayer crystals. *Phys. Rev. B*, 17(2):765–768, Jan 1978.

[68] G. S. Vernam. Secret signaling system. U.S. Patent No. 1310719, Jul 1919.

[69] P Villoresi, T Jennewein, F Tamburini, M Aspelmeyer, C Bonato, R Ursin, C Pernechele, V Luceri, G Bianco, A Zeilinger, and C Barbieri. Experimental verification of the feasibility of a quantum channel between space and earth. *New Journal of Physics*, 10(3):033038 (12pp), 2008.

[70] L. J. Wang, C. K. Hong, and S. R. Friberg. Generation of correlated photons via four-wave mixing in optical fibres. *Journal of Optics B: Quantum and Semiclassical Optics*, 3(5):346–352, 2001.

[71] M. Ware, A. Migdall, J. C. Bienfang, and Sergey V. Polyakov. Calibrating photon-counting detectors to high accuracy: background and deadtime issues. *J. Mod. Opt.*, 54:361–372, January 2007.

[72] Stephen Wiesner. Conjugate coding. *SIGACT News*, 15(1):78–88, 1983.

[73] W. K. Wootters and W. H. Zurek. A single quantum cannot be cloned. *Nature*, 299(5886):802–803, 1982.

[74] H. Xu, L. Ma, J. C. Bienfang, and X. Tang. Influence of avalanche-photodiode dead time on the security of high-speed quantum-key distribution systems. In *CLEO-QELS*, 2006.

[75] Hai Xu, Lijun Ma, Alan Mink, Barry Hershman, and Xiao Tang. 1310-nm quantum key distribution system with up-conversion pump wavelength at 1550 nm. *Opt. Express*, 15(12):7247–7260, 2007.