ADVANCED TESTS OF NONLOCALITY
WITH ENTANGLED PHOTONS

BY

BRADLEY G. CHRISTENSEN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

Professor James Eckstein, Chair
Professor Paul Kwiat, Director of Research
Professor Anthony Leggett
Professor Nadya Mason

Abstract

In 1935, Einstein, Podolsky, and Rosen questioned whether quantum mechanics can be complete, as it seemingly does not adhere to a natural view of reality: local realism, which is the notion that an event can only be influenced by events in the past lightcone, and can only influence events in the future lightcone. This question sparked a philosophical debate that lasted for three decades, until John Bell demonstrated that not only are quantum mechanics and local realism philosophically incompatible, but they predict different statistical results for an appropriate set of measurements on entangled particles, which changed the debate to a scientific discussion. Since then, Bell inequality violations have occurred in a plethora of systems, hinting that local realism is indeed wrong. However, every experiment had imperfections that complicated the interpretation — the experiments had so-called "loopholes" which allowed local realism to persist.

In this manuscript, we present our work in using optimized sources of entangled photons to perform the long-sought loophole-free Bell test. This landmark experiment invalidates local realism to the best that science will allow. Beyond answering questions on reality, these Bell tests have a important application in generating provably-secure private random numbers, which then can be used as a seed for cryptographic applications.

Not only do we demonstrate that nonlocality must exist, but we begin an experimental exploration in an attempt to understand and quantify this nonlocality. We do so by considering all theories that obey no-signaling (or relativistic causality). In our experiments, we observe the counter-intuitive feature of measuring more nonlocality with less entangled states. We also place a bound on the predictive power of any theory that obeys relativistic causality. And finally, we are able to measure quantum correlations only attainable through complex qubits. This work merely begins to probe the quantum boundary, beginning a journey that may someday find evidence of a beyond-quantum theory.

Acknowledgments

Table of Contents

List of Tables

List of Figures

List of Abbreviations

APD   Avalanche photodiode

AR   Anti-reflection

CH   Clauser-Horne

CHSH   Clauser-Horne-Shimony-Holt

DIRE   Device-independent randomness expansion

GPS   Global positioning system

HWP   Half-wave plate

PBS   Polarizing beam splitter

PPKTP   Periodically poled potassium titanyl phosphate

PR   Popescu-Rohrlich

QRNG   Quantum random number generator

RNG   Random number generator

SMF   Single-mode fiber

SNSPD   Superconducting nanowire single-photon detector

SPDC   Spontaneous parametric downconversion

SQUID   Superconducting quantum interference device

TDC   Time-to-digital converter

TES   Transition-edge sensor

Chapter 1

Introduction

Non-classical states, such as n-particle states or entangled states, have a wide range of applications. The field of quantum information endeavors to produce these quantum states with high quality, and precise control. By doing so, one can enhance classical information processing tasks. Quantum computing, for example, has been shown to vastly increase computing speed for certain classical algorithms (assuming quantum computers can be realized) [1]. Quantum key distribution offers an alternative method of secure communications, where the security is dependent on assumptions of devices rather than assumptions on computational power (for optimal security, one would combine both classical and quantum cryptography methods) [2]. Quantum metrology uses these quantum states for increased measurement sensitivity [3]. And there are many more applications of precise quantum state engineering.

Beyond the immediate practical applications, quantum information also seeks to answer questions on the foundations of quantum mechanics. One could even argue that the original goal of precise entanglement creation was to perform a loophole-free Bell test, which would definitively rule out a classical theory of reality [4]. The Bell test experiment, first proposed over 51 years ago, can experimentally answer the question Einstein, Podolsky, and Rosen asked 30 years prior: *can* quantum mechanics be complete [5]? Only recently has the technology allowed this experiment to finally be performed. In this manuscript, we discuss our work in performing this landmark experiment, showing that at least the concerns Einstein, Podolsky, and Rosen had with quantum mechanics are unfounded. By showing that a more intuitive local realistic description is not correct, and that therefore quantum mechanics could be correct, we now begin a journey in answering the next question: *is* quantum mechanics correct?

1.1   Local realism: A natural view of reality

Suppose there are four events that we can measure, events A, B, C, and D, with outcomes either 0 or 1 (for example, an event could be whether or not you have a runny nose, with an outcome of 1 if you do, an outcome of 0 if you do not; or an event could be a measurement of a photon's polarization, where a

1 outcome corresponds to the photon transmitting through a polarizer and being detected, and 0 if there is no detected photon). Further assume that if events A and B are measured, their outcomes are always correlated (if the outcome of event A is 1, then the outcome of event B is also 1). Suppose further that if events B and C are measured, then their outcomes will also be correlated. Finally, if events C and D are measured, their outcomes will also be correlated (see Fig. 1.1). Given these correlations, can we infer the correlation between events A and D? As an example, let A be if you have a runny nose, event B is if you have a sore throat, event C is if you have a cough, and event D is if you have a headache, with outcome 1 if you have the symptom, and 0 otherwise. If you have a runny nose, you always have a sore throat, and if you have a sore throat, then you always have a cough; furthermore, if you have a cough, you always have a headache. Now, given that you have a runny nose, do you also have a headache? The seemingly obvious answer to this problem would be "yes", that the outcomes of events A (having a runny nose) and D (having a headache) must be correlated. However, to reach this conclusion that A and D are correlated, a "local realism" assumption has to be made. That is, we need to assume any system has preexisting values for all possible measurements of the system (realism), and that these preexisting values depend only on events in the past light cone of the system (locality). If we assume local realism, then it is the case that measuring event A to be 1 means that B has a predefined outcome of 1, which then implies C also must have a predefined value of 1, which also implies D must have a predefined value of 1 (in order to preserve the known correlations). The correlation between A and D (under local realism) therefore does not depend on whether or not the events in the correlation chain are actually measured.

Intuitively, there are only two ways that events can be correlated, either the events have a common cause, or one event causally influences the other event. The notion of these two ways of correlations is also known as Reichenbach's principle [6], and is an equivalent notion of local realism. In the example above, the correlation between the symptoms could be the result of a common cause: having a cold. In this case, we know if you have a cold, you will have all of those symptoms, and therefore the events are all correlated. The other reason for the correlations of the symptoms is a causal influence; that is, having a runny nose results in a sore throat (perhaps from sniffling too often), and having a sore throat is what causes you to start coughing (the throat becomes irritated, causing a cough); additionally, coughing results in having a headache, resulting in a correlation between a runny nose and a headache through causal influence.

However, without assuming local realism, no statement can be made about the correlations between A and D If, for example, B does not have a predefined outcome, then measuring A to be 1 does not imply B has an outcome of 1, B only has an outcome of 1 when we measure the outcome of B (regardless that we *know* the outcome of B will be 1, we must actually make the measurement to be able to assign the
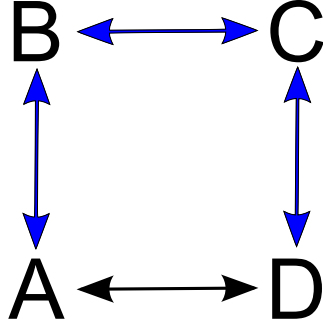
Figure 1.1: A diagram depicting a correlation chain. The outcomes of events A and B are correlated (blue arrow), the outcomes of events B and C are correlated, and the outcomes of events C and D are also correlated. Under a local realistic assumption, then the outcomes of events A and D must be correlated. However, entangled particles do not necessarily have a correlation between the outcomes of events A and D. In the general case, the correlation chain creates a twice-iterated triangle inequality, which results in a Bell inequality.

outcome). The idea that we cannot assign outcomes to events that we have not measured (even when we do know what the outcome will be) runs counter to our natural view of reality, as it (seemingly) should not matter whether we actually perform a measurement on B and C if we know the result that will be measured. In 1935, Einstein, Podolsky, and Rosen introduced what they believed should correspond to physical reality: "If, without in any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity" [5] (i.e., reality should obey local realism). Furthermore, they noted that quantum mechanics does not obey this definition of reality. As an example, consider two spatially-separated particles with position-momentum entanglement, where a measurement is performed on one of the particles. If the measurement is made on the momentum of the particle, then to preserve the uncertainty relations imposed by quantum mechanics, this measurement must project the other particle into a well-defined momentum state. If instead a position measurement were made, then the other particle must be projected into a well-defined position state. Thus whether the position or the momentum is the element of reality of one particle (i.e., whether the wavefunction of the particle is in the eigenstate of the position or in an eigenstate of the momentum) depends on the measurement made on the other particle (without in any way disturbing the first particle). Einstein, Podolsky, and Rosen concluded from this example, that "the quantum-mechanical description of physical reality given by wave functions is not complete."

1.2 A test of local realism

Whether quantum mechanics is complete remained merely a philosophical question for 30 years. In 1965[1], however, John Bell demonstrated that tests of local realism were in fact possible by analyzing the limit of allowed correlations between measurements made on an ensemble of any classical system [4], the limit of the allowed correlations results in a so-called Bell test or Bell inequality. If these measurements are performed under *sufficiently ideal conditions*, a violation of a Bell inequality would conclusively rule out all possible local realistic theories.

In a typical two-party Bell test, a source generates particles and sends them to two distant parties, Alice and Bob. Alice and Bob independently and randomly choose properties of their individual particles to measure, e.g., polarization in a particular basis. Later, they compare the results of their measurements. Local realism constrains the joint probability distribution of their choices and measurements. For example, in Fig. 1.1, the three perfectly correlated events (blue arrows) requires the fourth set of events (black arrow) to also be perfectly correlated. The basis of a Bell test is an inequality that is obeyed by probability distributions for any local realistic model, but can be violated by the probability distributions of certain entangled quantum particles [4].

While the only known (physically realizable) way to violate a Bell inequality is with entangled particles, to derive a Bell test only requires the assumption of locality and realism. The violation of the Bell inequality then forces us to abandon locality or realism (or both)[2]. As such, a Bell inequality is only a test of local realism, and does not make any statement about quantum mechanics, this in turn allows the use of the framework of a Bell inequality to go beyond the realm of quantum mechanics, allowing us to probe a general class of physical theories limited only by no-signaling constraints (that is, any theory that obeys causality).

1.3 Spontaneous parametric downconversion

To generate the correlations necessary to violate a Bell inequality, we use polarization-entangled photons. To do so, we employ spontaneous parametric downconversion (SPDC), which is a second-order ($\chi^{(2)}$) nonlinear process where a single pump photon splits into two daughter photons (for a full discussion on this process,

---

[1]The paper was published in the December issue of 1964, which was most likely printed in 1965. Even Bell himself (usually) cites his own paper as 1965[7].

[2]There are many ways a Bell inequality can be derived through different (but equivalent) sets of assumptions. A very common alternate assumption to make in deriving a Bell inequality is "local causality" [7] (often simply called "locality"), which is the notion that an event can only be influenced by events in the past lightcone, and can only influence events in the future lightcone. In this case, the violation of a Bell inequality implies that local causality is invalid (compared to local realism, where there are two distinct assumptions). This interpretation on Bell inequalities is the viewpoint of the author; however, for historical purposes, we will continue the discussion in this section using local realism. It should be noted then, later when we use the term "nolocality," we implicitly take the viewpoint of local causality being wrong, rather than stating the local assumption is the incorrect assumption of local realism.

see, e.g., Ref. [8]). Here, we will discuss a few critical points for our purpose. An approximation of the 2-photon state can be written as:

$$\psi_{2-photon} \propto L \int d\mathbf{k}_p \phi_p(\mathbf{k}_p) \int d\mathbf{k}_s d\mathbf{k}_i \chi^{(2)} \delta(\omega_p - \omega_s - \omega_i) \, \text{sinc} \, (\Delta\kappa L/2) \, |\mathbf{k}_s, \mathbf{k}_i\rangle, \qquad (1.1)$$

where $\mathbf{k}_p$, $\mathbf{k}_s$, and $\mathbf{k}_i$ represent the wave-vectors of the pump, signal, and idler photons outside the crystal, $\Delta\kappa$ is the phase mismatch, defined as $\Delta\kappa \equiv \kappa_p - \kappa_s - \kappa_i$, where κ_p, κ_s, and κ_i represent the wave-vectors of the pump, signal, and idler photons inside the crystal (where $\kappa_j = (\kappa_j^{(x)}, \kappa_j^{(y)}, \kappa_j^{(z)})$, and $\kappa_j^{(m)} = n^{(m)}(\omega_j) k_j$), $n^{(m)}(\omega_p)$, $n^{(m)}(\omega_s)$, and $n^{(m)}(\omega_i)$ are the refractive indices along the three axes of the crystal, $\omega_p(= |\mathbf{k}_p| c)$, ω_s, and ω_i represent the angular frequencies, $\phi_p(\mathbf{k}_p)$ represents the input momentum distribution of the pump, L is the length of the crystal along the z-axis (which we take to be aligned along the center of the distribution $\phi_p(\mathbf{k}_p)$ and we assume the crystal is infinite in length in the other two directions), and $\chi^{(2)}$ is the second-order nonlinear coefficient, which governs the strength of the SPDC interaction[3]. For a more thorough description of the 2-photon state, see Ref. [10]. In this form, we have neglected higher photon numbers per mode, which is important for understanding noise from multiple photon events in Bell tests. A description of the higher photon number state (in particular for Bell tests) is given in Ref. [9].

Now, as this is a spontaneous process (stimulated via vacuum fluctuations), and since $\chi^{(2)}$ is small, the probability of an event produced in a desired mode is very small ($\sim 10^{-13}$), and as such we use bright lasers as a source of pump photons to ensure a significant number of downconversion events into the desired mode per second. The SPDC process is constrained by conversation of energy ($\delta(\omega_p - \omega_s - \omega_i)$) and conservation of momentum ($\text{sinc} \, (\Delta\kappa L/2)$) (which is called the phase-matching condition), as shown in Fig. 1.2. Most materials exhibit normal dispersion over the range of frequencies we use (i.e., they have higher indices of refraction at high frequencies), and by conservation of energy the pump photon must have higher frequency than the downconverted photons; therefore, to satisfy phase-matching one must use birefringent materials (where the two polarization states see different indices of refraction). By using a birefringent (and nonlinear) material, the phase-matching conditions can be met by aligning the pump along the fast axis (the direction with the smaller of the two indices of refraction)[4]. Achieving proper phase matching can be done by adjusting the cut angle of the crystal, which allows for the necessary tunability (changing the crystal axis on order of $0.2°$ can result in $1°$ change in the emission angle).

As long as one of the two daughter photons has a polarization along the slow axis (the higher of the

---

[3]In reality, $\chi^{(2)}$ is a tensor that depends on the polarization of the 3 fields (pump, signal, and idler) and the crystal symmetries.

[4]This is just one of the ways to accomplish phase-match.
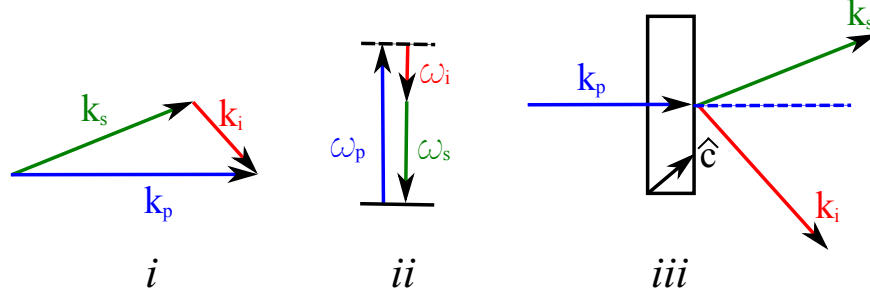
Figure 1.2: The phase-matching conditions. *i*) Momentum conservation, where the momentum of the pump (k_p) is equal to the momentum of the two daughter photons, called the signal (k_s) and idler (k_i) photons. If the momentum conservation is not met exactly, then there is a phase mismatch, and the field oscillates as shown in Fig. 1.3. In the case presented in this figure, we also see that the higher the frequency, the smaller the emission angle must be to conserve momentum (in general, the dependence between emission angle and frequency depends on the dispersion of the fast and slow indices). This fact results in undesirable additional correlations between the photon pairs, which reduce coupling efficiencies. *ii*) energy conservation is depicted: a pump photon is split into two daughter photons, facilitated by the nonlinear material (which remains unaltered throughout the process, i.e., it is a parametric process). Finally, Fig. 1.2*iii* demonstrates the process happening inside a crystal. A single pump photon passes through a crystal cut with the crystal axis along $\hat{c}$ and downconverts to the two daughter photons. Due to the symmetry of phase matching, both the signal and idler photons are emitted into a cone (centered on the pump). However, by detecting one of the photons in the pair, the other photon is heralded to be on opposite side of the conjugate cone, allowing for high heralding efficiency.

two indices in a birefringent material), then phase matching can be met (i.e., there can exist a crystal cut which satisfies phase matching). This allows for two different SPDC processes to occur: a fast pump photon downconverting to two slow daughter photons (called Type-I), or a fast pump photon to a slow and a fast daughter photon (called Type-II). Note that this discussion is for normal dispersion materials; in anomalous dispersion materials the situation is reversed, and in some cases the polarization of the pump and daughter photons can even be the same (Type-0 phase matching).

In some instances, depending on the birefringence of the material and the desired wavelength, satisfying the phase-matching constraints is inefficient or might even be impossible. To solve this issue, we can periodically-pole the crystal to achieve (quasi-)phase matching. Here, the direction of the crystal axis is rotated every half-wavelength of the downconverted photon amplitude oscillations, as depicted in Fig. 1.3, which allows for coherent buildup of the downconverted field. For periodically poled crystals, tuning the phase-matching conditions is accomplished through adjusting the temperature of the crystal (instead of adjusting the crystal cut).

One can use downconversion to generate pairs of photons using any of the above techniques. For most of the cases discussed here, we desire very high heralding efficiency systems (the detection of one of the downconverted photons heralds the presence of a second photon, and the probability of the detection of this second photon, given the first, is called the heralding efficiency), while maintaining high entanglement quality.
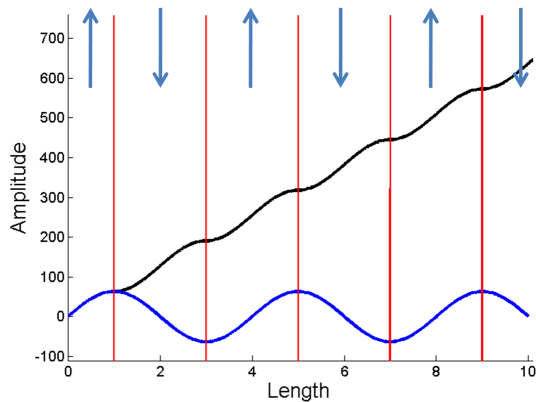
Figure 1.3: An illustration of the effects of periodic poling. Ordinarily, due to phase mismatch in the SPDC crystal, the downconversion amplitude oscillates along the length of the crystal (blue line). However, if the crystal axis (dark blue arrows) is rotated 180° every half wavelength of the oscillation, coherent buildup in the crystal is recovered (black curve). Both axes in this plot are in arbitrary units.

Here, we desire heralding efficiencies exceeding 75% probability[5]. Due to the phase-matching conditions, the photons have momentum and frequency correlations which can be used to efficiently collect the photon pairs. However, not all correlations are desirable, as there are additional correlations between the k-vector and frequency (if one of the downconverted photons has higher energy, then to conserve momentum, that photon's k-vector must exit the crystal at an angle closer to the pump), and these additional correlations actually degrade our capability to efficiently collect the photon pairs.

To preserve the entanglement quality we couple both photons into a single-mode fiber (which approximately supports a Gaussian spatial mode). If the pump has a small k-vector uncertainty (relative to the collected daughter photon's k-vector uncertainty), then the detection of one photon projects the other photon into a well-defined, mostly Gaussian spatial mode. For the sources considered here, the detection of one photon projects the other photon into a state that has 93-96% Gaussian mode content. In this case, the heralded photon suffers a 4-7% loss. Furthermore, in some instances where a large bandwidth is collected (in one of our systems, we collect photons within the 700-nm to 720-nm band), then the chromatic dispersion in the lenses cause the different wavelengths to have different focal lengths, this results in another 2% loss by collecting into single-mode fibers.

The k-vector and frequency correlations mean that the single-mode fiber also acts as a spectral filter with a Gaussian spectral profile. This effective Gaussian filter introduces a significant amount of loss (around 30%). As an example of how this causes loss, consider the case where there are infinitely many spectral modes (in our lab, we can easily see 100 spectral modes or more). If we detect a photon at frequency $\omega_o + \omega$,

---

[5]Heralding efficiency is defined many different ways. As our interest is in the detection of a photon given the detection of its conjugate photon, this is what we use as our definition. Some definitions divide out the detector efficiency, but we do not do so.

where ω_o is the central frequency, at the tail end of the Gaussian distribution, then the conjugate photon will have a frequency of $\omega_o - \omega$ (by conservation of energy), and will thus be at the opposite tail end of the distribution, and therefore unlikely to be transmitted through our spectral filter[6]. There are two way to remove this unwanted spectral-momentum entanglement; the first is to apply additional spectral filtering such that we only collect the photons which correspond to spectral modes with a high transmission (i.e., only collecting from the *peak* of the Gaussian distribution). In this case, energy conservation still ensures that if one photon transmits through the narrow spectral filter, then the other photon will also transmit through the conjugate filter. However, this necessarily means discarding most of the pairs, reducing source brightness.

The second method relies on using the finite bandwidth of the pump. In the above example, we neglected the energy uncertainty of the pump (we assumed nearly infinite many spectral modes, which is only attainable with a single frequency pump). If the energy uncertainty of the pump is approximately the same energy uncertainty of the Gaussian spectral filter we create with the single-mode fiber, then this results in collecting single spectral mode. In this case, by detecting one of the photons, the other photon is projected into the conjugate single spectral mode and single spatial mode, meaning it will be collected perfectly into the single-mode fiber. In the experiments presented here, we use both of these techniques (in different systems) to efficiently collect the entangled photon pairs.

So far we have only discussed efficiently collected photon pairs from downconversion. Downconversion itself, however, does not typically result in polarization entanglement[7]. In the two cases discussed previous, we generate the state $|ss\rangle$ or $|sf\rangle$ (where s corresponds to slow polarization, and f is fast polarization), neither of which are entangled. Instead, to generate entanglement, we need the pump photons to be in a superposition between downconverting in two different crystals or locations, which we then can convert to polarization entanglement (see Fig. 1.4). For example, if we embed a downconversion crystal in a polarization interferometer, then we can convert the superposition of the pump photon's path into a superposition between two different polarization states. Alternatively, we can put two orthogonally oriented crystals next to each other (such that, in the lab frame, one crystal downconverts $|H\rangle \rightarrow |VV\rangle$ and the other crystal produces $|V\rangle \rightarrow |HH\rangle$, where H is horizontal polarization and V is vertical). Since we are coupling into single-mode fibers, any information on the spatial distinction between the two crystals is lost (the outputs of both crystals

---

[6]The loss from the Gaussian spectral filter and Gaussian spatial mode cannot be discussed independently. We have taken liberties, in the interest of describing the problems, by discussing them as different forms of loss. To be perfectly clear, there is simply a 70% coupling efficiency into the fiber due to the spectral-momentum correlations in the current example. In the previous paragraph where we stated a 93-96% coupling efficiency, this number is taken after removing these unwanted correlations.

[7]A Type-II downconversion source can produce the entanglement through a superposition of the $f \rightarrow sf$ and $f \rightarrow fs$ processes.
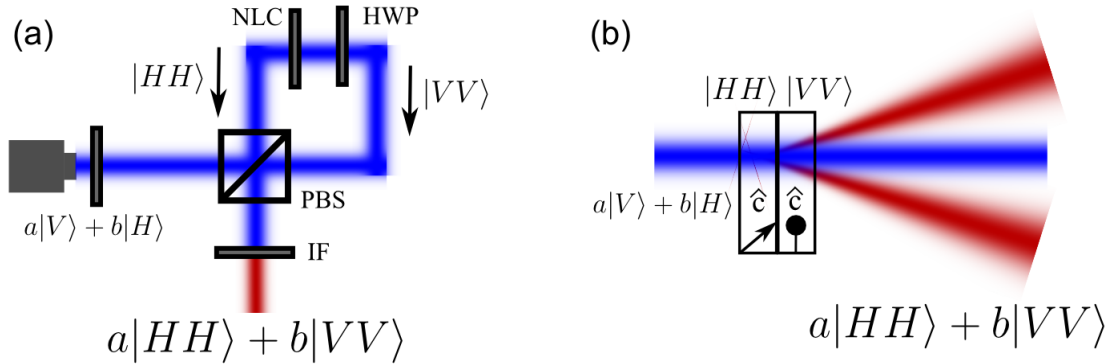
Figure 1.4: A diagram of two ways of producing entanglement. a) The pump is placed in a superposition state: $a|V\rangle+b|H\rangle$, which is transferred to a superposition of two paths at the polarizing beam splitter (PBS). The $|H\rangle$ (transmission) path of the pump is rotated to $|V\rangle$ via a half-wave plate (HWP). The nonlinear crystal (NLC) downconverts $|V\rangle \to |HH\rangle$, and the HWP rotates the clockwise path from $|HH\rangle \to |VV\rangle$. At the PBS, both states take the bottom path, resulting in an entangled state $a|HH\rangle + b|VV\rangle$ after filtering out the blue pump with an interference filter (IF). b) Entanglement is generated with a paired crystal. Again the pump is placed in a superposition state: $a|V\rangle + b|H\rangle$. Then in the first crystal, the $|V\rangle \to |HH\rangle$ downconversion process occurs. The second crystal is rotated such that the crystal axis is out of the page, which now has the $|H\rangle \to |VV\rangle$ process. If the two crystals are indistinguishable sources (by coupling into a single-mode fiber to remove spatial information), then the resultant entangled state is $a|HH\rangle+b|VV\rangle$. In both diagrams, the full entangled state (the relative amplitude and phase between the two terms) is easily controlled by altering the pump polarization with standard optical components.

are projected onto a single spatial mode, which means the two crystals are indistinguishable sources). In this case, we have two different polarization states generated from the two different crystals, resulting in polarization entanglement. This paired-crystal technique can be realized while introducing minimal loss, provided the crystals are sufficiently thin (in our case, we use 200 µm-thick crystals). We again will employ these two techniques in different systems, achieving high heralding efficiency in both cases.

1.4   Overview

In this manuscript we will use polarization-entangled photons generated via SPDC to perform measurements of different Bell inequalities, for a wide range of purposes. First, a formal description of a generalized Bell inequality will be presented. Here we will discuss the issues arising from experimentally testing the Bell inequalities, along with a discussion of typical assumptions that must be made and the validity of those assumptions (Chapter 2). Next, in Chapter 3, we will describe our experiment which violates a Bell inequality with a classical source by exploiting hidden assumptions in standard data analysis, and discuss methods to ensure these issues are not present in an actual Bell test. In Chapter 4, we will describe our system of entangled photons, which we use to legitimately violate a Bell inequality without making any fair-sampling assumptions, closing the so-called "detection loophole". We then use this result to perform

device-independent randomness expansion. Then we discuss possible improvements to the system. Next, in Chapter 5, we will discuss our experiment which performs a loophole-free Bell test to definitively invalidate local realism with entangled photons. Then in Chapter 6, we will go beyond the simple case of using a Bell test to rule out local realism, using Bell tests to perform experimental tests on a general class of physical theories consistent with the no-signaling principle (that is, any further extension of quantum mechanics). In this section, we will discuss the notion of a "PR-box" to be a fundamental unit of nonlocality, and use this concept to make a plethora of tests, all with rather interesting results. Finally, we will summarize our work in Chapter 7, as well as discuss the future of Bell tests.

Chapter 2

Bell Tests

Bell tests have become ubiquitous, they have been performed in many different physical systems and on many different degrees of freedom. Even more so, Bell tests are now performed in undergraduate lab environments [11]. These Bell tests almost always use the most famous Bell inequality, the CHSH Bell inequality. In this manuscript, however, this simple version of the Bell inequality is insufficient to perform our experiments. As Bell tests form the basis of every experiment presented here, in this chapter we introduce a generalized description of a Bell test, which we will return to again in Chapter 6. Furthermore, while Bell tests are performed in a myriad of experiments, nearly every experiment is plagued by experimental imperfections which invalidates the experiment from truly arguing against local realism (there have only been three so-called loophole-free Bell tests which do not have these issues, one of which is presented in Chapter 5). Finally, in this chapter, we also discuss the common imperfections in detail and the methods to fix those issues.

2.1 Generalized Bell tests
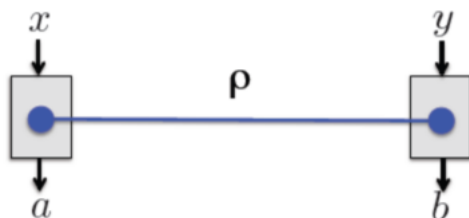
Figure 2.1: Bell test scenario. Alice and Bob perform "black box" measurements on a shared (quantum) state ρ. The experiment is characterized by the data $\{p(a,b|x,y)\}$, i.e., a set of conditional probabilities for each pair of measurement outputs (a and b) given measurement settings x and y. Based on the data $p(a,b|x,y)$, Bell inequalities (see Eq. (2.3)) can be tested.

Consider two separated observers, Alice and Bob, performing local measurements on a shared quantum state ρ. Alice's choice of measurement settings is denoted by x and the measurement outcome by a. Similarly, Bob's choice of measurement is denoted by y and its outcome by b (see Fig. 2.1). The experiment is thus

characterized by the joint distribution

$$p(a, b|x, y) = \text{Tr}(\rho M_{a|x} \otimes M_{b|y}), \tag{2.1}$$

where $M_{a|x}$ ($M_{b|y}$) represents the measurement operators of Alice (Bob); see Fig. 1(a). In his seminal work, Bell introduced a natural concept of locality, which assumes that the local measurement outcomes only depend on a pre-established strategy and the choice of local measurements [4]. Specifically, a distribution is said to be local if it admits a decomposition of the form

$$p(a, b|x, y) = \int d\lambda \, q(\lambda) p(a|x, \lambda) p(b|y, \lambda), \tag{2.2}$$

where λ denotes a shared local (hidden) variable, distributed according to the density $q(\lambda)^1$, and Alice's probability distribution—once λ is given—is notably independent of Bob's input and output (and vice versa). A Bell inequality is then an inequality that is obeyed by local realistic probability distributions. These inequalities can be written as

$$\mathcal{S} = \sum_{a,b,x,y} \beta_{a,b,x,y} p(a, b|x, y) \overset{\mathcal{L}}{\leq} L, \tag{2.3}$$

where $\beta_{a,b,x,y}$ are integer coefficients, and L denotes the local bound of the inequality—the maximum of the quantity $\mathcal{S}$ over distributions from $\mathcal{L}$, i.e., of the form 2.2.

2.2   Experimental Bell test

By performing judiciously chosen local measurements on an entangled quantum state, one can obtain distributions 2.1 which violate one or more Bell inequalities, and hence do not admit a decomposition of the form 2.2. The simplest Bell tests are bipartite, binary input, binary output inequalities (Bell tests with more inputs will be discussed in Chapter 6, but for now we will focus on the simplest scenario.). In this case, there are two standard forms of the Bell inequality, the first introduced by Clauser, Horne, Shimony and Holt (CHSH) [12]:

$$\mathcal{S}_{\text{CHSH}} = E_{0,0} + E_{0,1} + E_{1,0} - E_{1,1} \overset{\mathcal{L}}{\leq} 2, \tag{2.4}$$

where $E_{x,y} \equiv p(a = b|x, y) - p(a \neq b|x, y)$ denotes the correlation function. Fig. 2.2 extends Fig. 1.1 to the

---

[1]$q(\lambda)$ is normalized, such that $\int q(\lambda) d\lambda = 1$
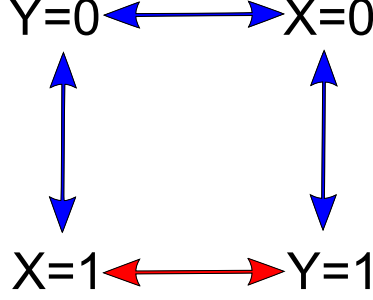
Figure 2.2: A diagram depicting a correlation chain for the CHSH Bell test scenario. A Bell test measures each link along the correlation chain to determine if the correlations could have been produced by a local realistic model. The CHSH Bell test sums the three blue correlations, and subtracts the red correlation, producing a local realistic bound of 2. This bound is due to the allowed correlations of the final link in a correlation chain, given the first three links.

case in Eq. 2.4. Here, we can see that this inequality is measuring the three blue correlations and subtracting the red correlation. A local realistic model can only measure values between $-2 \leq \mathcal{S}_{\mathrm{CHSH}} \leq 2$ for Eq. 2.4. In our experiments, we focus on violating the bound at 2, and therefore write Eq. 2.4 with only one of the bounds. Quantum correlations, however, can violate the above inequality up to $\mathcal{S}_{\mathrm{CHSH}} = 2\sqrt{2}$, the so-called Tsirelson bound [13]. The second (equivalent) set of forms of the simplest Bell inequality scenario were introduced by Clauser and Horne (CH) [14]:

$$\mathcal{S}_{\mathrm{CH}} = p(1,1|0,0) + p(1,1|0,1) + p(1,1|1,0) - p(1,1|1,1) - p(a = 1|x = 0) - p(b = 1|y = 0) \overset{\mathcal{L}}{\leq} 0, \qquad (2.5)$$

$$\mathcal{S}_{\mathrm{CH}} = p(1,1|0,0) - p(1,0|0,1) - p(0,1|1,0) - p(1,1|1,1) \overset{\mathcal{L}}{\leq} 0, \qquad (2.6)$$

where $p(i,j|k,l)$ is the probability of measuring $a = i$ and $b = j$, given settings $x = k$ and $y = l$, and $p(a = 1|x = 0)$ is the probability of measuring $x = 1$ for setting $a = 0$ (and similarly for $p(b = 1|y = 0)$). For Eq. 2.5 and Eq. 2.6, local models are limited to $-1 \leq \mathcal{S}_{\mathrm{CH}} \leq 0$, where we focus on the bound at 0. The Tsirelson bound for these inequalities is $\mathcal{S}_{\mathrm{CH}} = \frac{2\sqrt{2}-2}{4} \approx 0.207$.

The local bounds to these inequalities can be found by inserting the distribution from Eq. 2.2 into the Bell inequality. Another method to find the local bound is to simply maximize over all possible local strategies. Alice, for example, locally must choose to output either $a = 0$ or $a = 1$ for each of her chosen inputs of $x = 0$ and $x = 1$, i.e., she has 4 possible strategies. We only need to consider cases where Alice always makes the same decision (e.g., Alice always choses $a = 0$ if $x = 1$), as randomly changing her strategy cannot produce a Bell value larger than the value of any particular strategy. We can likewise consider Bob's four possible local strategies, for a total of $2^2 \mathrm{x} 2^2 = 16$ possible local strategies. In almost all cases, the symmetry of the

13

problem will significantly reduce the total number of unique local strategies, making this an efficient method for calculating the local bound.

It is worth noting that the three Bell inequalities above (Eqs. 2.4,2.5,2.6) are all identical Bell inequalities, as each inequality can generate the rest through only transformations that obey no-signaling. No-signaling is the statement that all local probability distributions do not depend on a nonlocal setting choice. An example of a no-signaling transformation is $p(x = 0|a = i, b = j) = p(x = 0|a = i, b = k)$, as Alice's local distribution is unaltered by Bob's setting choice. No-signaling is essentially a statement of communication: if a setting choice is able to alter the probability distribution, then it can be used to transfer information. All Bell tests thus require this assumption, as without it, a realistic theory could easily mimic a violation of a Bell inequality by simply communicating to reproduce any possible result.

Trial	x	a
1	0	1
2	0	0
3	1	0
4	0	1

Trial	y	b
1	1	1
2	1	0
3	1	1
4	0	1
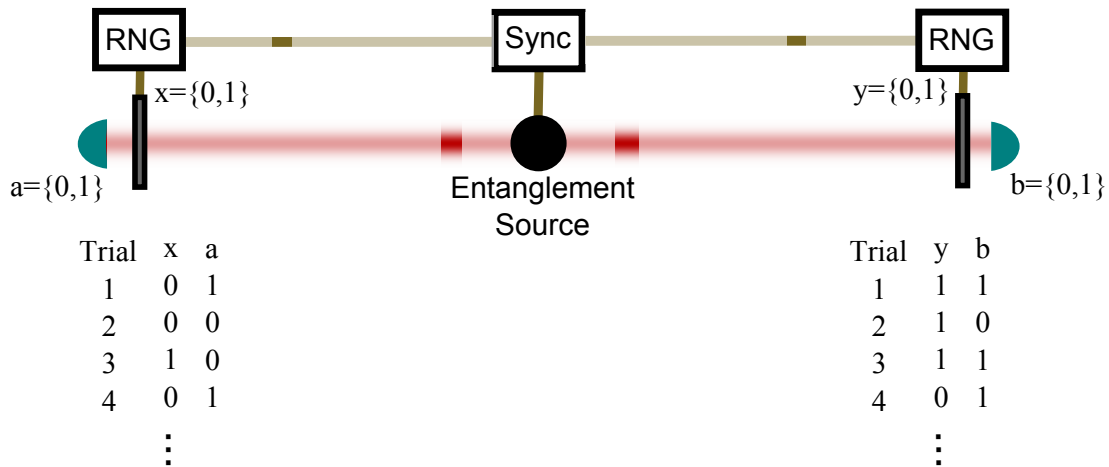
Figure 2.3: A diagram of an idealized experimental Bell test. A synchronization signal (gold line) is sent to Alice and Bob to inform them that they will soon receive entangled photons, and thus to trigger their setting choice (RNG). Based on the setting choice, a polarizer is rotated to a specific setting to perform the projective measurement. Later, the entanglement source will emit entangled photons to Alice and Bob. If they see a detected photon, they assign the outcome of the trial to be 1; if no photon is detected the trial outcome is set to 0. Alice and Bob each record the results of the trial (which consists of both their selected input and measured output). Later, they will compare the results of their trials to determine if local realism can explain the correlations.

While these inequalities appear simple, probabilities are not observable quantities, and any experiment will have to estimate probabilities after measuring an ensemble of entangled particles. As an example of an experimental version of the idealized Bell test described above, consider two parties, Alice and Bob, who each have spatially separated labs. Their labs consist of a polarizer, a single-photon detector (with a time-to-digital converter to register when events are recorded), and a method for choosing a setting. A third lab (Charlie) sends a message to Alice and Bob, stating that he is about to send them a photon to measure[2].

---

[2]Note that in reality such "event-ready" operation is difficult to achieve

Charlie then prepares polarization-entangled photons and sends one to Alice and one to Bob. Alice, upon receiving the message from Charlie, will then decide whether she will choose measurement setting 0 or measurement setting 1. If she chooses measurement setting 0, she will adjust her polarizer to transmit, say, light polarized along 20° above the horizontal. If she chooses measurement setting 1, she will orient her polarizer to be 5° above the horizontal. Bob will likewise do the same, though choosing his own polarization settings. Then upon receiving their respective photons from Charlie, Alice and Bob will each send their own photon through the polarizer and onto the single-photon detector. If the detector sees a photon they assign the output of 1, and if no detection event occurs they assign the output of 0; they will record the results of each trial (that is, they record both their chosen input and measured output). This type of experiment is displayed in Fig. 2.3.

After a predetermined number of trials[3] (say, after Alice and Bob have run their experiment 10^6 times), Alice and Bob will compare their results to determine the correlations between their measurement outcomes, given their inputs, by estimating each of the probability terms from their data. For example, consider the Bell test in Eq. 2.6; to estimate the first term, $p(1, 1|0, 0)$, Alice and Bob compare the number of times they jointly measured output of 1 while both choosing input 0 (defined as $C(1, 1|0, 0)$), to the total number of times they jointly choose input 0 (defined as $N(0, 0)$), regardless of their measurement results. Thus, they estimate

$$p(1, 1|0, 0) = \frac{C(1, 1|0, 0)}{N(0, 0)},\qquad(2.7)$$

and likewise for the other terms.

After analyzing the data, Alice and Bob can then conclude whether or not they believe a local realistic model could have yielded the correlations that occurred. Since the goal of a Bell test is to definitively rule out all local realistic models, we want to disprove any model, even conspiratorial models where the particles can communicate or force our detectors to behave in strange ways. That is, for the Bell test experiment, we need to carefully document any assumption that is explicitly (or implicitly) made in both the system design and in the data analysis. A significant violation of a Bell inequality implies that either local realism is false, or that one or more of the assumptions made about the experiment are not true; thus, every assumption in an experiment opens a so-called "loophole."

In the above example, there are a few subtle but key points that should be addressed further. First, each experiment must always assign an outcome, given an input. If, for example, the single-photon detector has a 65% detection efficiency (true for a typical silicon avalanche photo-diode (APD)), then every time we failed

---

[3]For a discussion on the need to predefine the number of trials, see Section 3.3.

to detect a photon, we still must assign an output of 0. Second, *how* and *when* we choose the measurement settings is critical, and the experiment can have unverified assumptions if this is not done properly. Next, the data analysis needs to be done properly, as estimating $C(a, b|x, y)$ can be nontrivial for a real system with timing uncertainties. In fact, standard data analysis techniques performed in most optical quantum information experiments to determine $C(a, b|x, y)$ have hidden assumptions. While these assumptions are very minor (e.g., the detector's time response is the same regardless of the measurement setting), they still allow local realistic theories to explain the results. As such, care must be taken in the data analysis to avoid unnecessary assumptions. Finally, while this example is based on how we would use quantum mechanics to violate a Bell inequality, the idea of a Bell test should still be considered as a black-box experiment, so how the measurements are made and how the correlations are generated does not matter (beyond the specific assumptions that the system introduces).

2.3   Fair-sampling assumption

The first experimental violations of the Bell inequality was demonstrated in 1972 by Freedman and Clauser [15], and Aspect, Grangier, Dalibard, and Roger in 1981 and 1982 [16, 17, 18], using pairs of polarization-entangled photons generated by an atomic cascade. However, due to technological constraints, these Bell tests and those that followed (see [19] for a review), were forced to make additional assumptions to show local realism was incompatible with their experimental results. That is, these experiments were required to assume that their measurement devices were fairly sampling the source, i.e., that the detection system behaves equally regardless of the measurement setting choice and outcome. This fair-sampling assumption (which opens the so-called "detection loophole") allows the experimenters to disregard all events where no detection event occurred, contrary to what was mentioned earlier where the experimenter must always determine an output for each given input. To do so, Alice and Bob replace their polarizer and single-photon detector in the earlier example with a waveplate (to rotate between the different bases) and a polarizing beam splitter (PBS) (which transmits one polarization, and reflects the orthogonal polarization) and two single-photon detectors (one in each port of the PBS). Now, a 1 outcome corresponds to a detection in the transmitted single-photon detector, a 0 outcome corresponds to a detection in the reflected single-photon detector, and a -1 outcome if no photon was detected[4]. If either party receives a -1 outcome, then the trial is discarded, allowing Alice and Bob to post-select on trials where they have a joint detection event (see Fig. 2.4). We can then rewrite our estimate of a probability as

---

[4]Standard notation when applying a fair-sampling assumption is ± 1 for detection events and 0 for no detection event. However, for the cases we consider in this manuscript, the notation defined in the text is more convenient.

$$p(1,1|0,0) = \frac{C(1,1|0,0)}{C(1,1|0,0) + C(1,0|0,0) + C(0,1|0,0) + C(0,0|0,0)}, \qquad (2.8)$$

where we now normalize by the number of detected events at the given setting, instead of normalizing by the number of times the settings were chosen. If the system has less than unit efficiency, then $C(1,1|0,0) + C(1,0|0,0) + C(0,1|0,0) + C(0,0|0,0) < N(0,0)$.

Trial	x	a
1	0	-1
2	0	0
3	1	0
4	0	1
⋮		

Trial	y	b
1	1	1
2	1	0
3	1	1
4	0	-1
⋮		
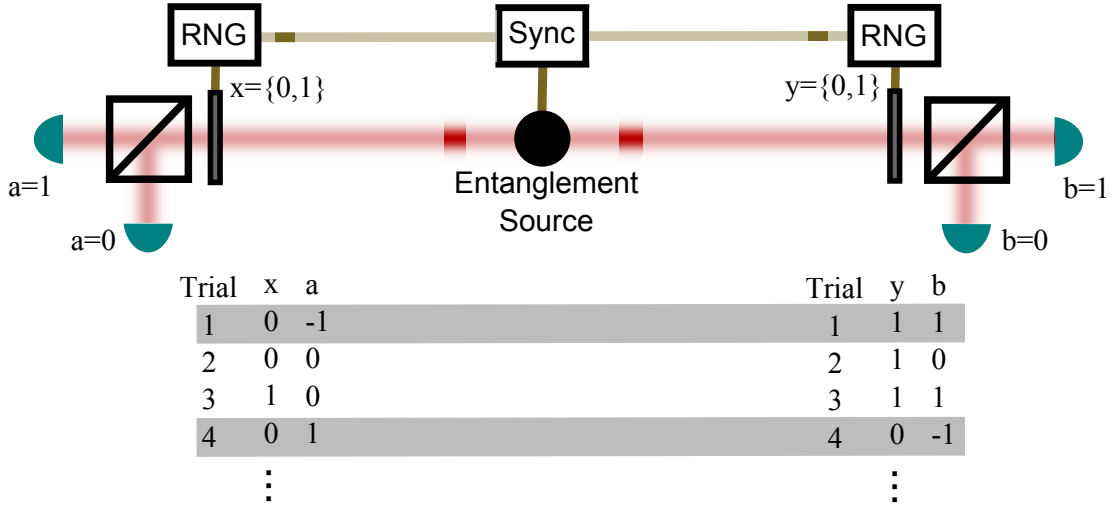
Figure 2.4: A diagram of an experimental Bell test that assumes fair-sampling. Similar to Fig. 2.3, a synchronization signal is sent to Alice and Bob (gold line) to trigger their setting choice (RNG), which is now performed by rotating waveplate. When Alice and Bob receive the entangled photons, they send them onto a polarizing beamsplitter after the waveplate. Now, if they see a detected photon in the transmitted port, they assign the outcome of the trial to be 1, if the photon is detected in the reflected port, then the trial outcome is set to 0. If no photon is detected, then the outcome is -1. Alice and Bob each record the results of the trial. Later, when Alice and Bob compare the results of their trials, they discard any trial where either party recorded a -1 event (gray bar). While this technique removes stringent system requirements, it can be exploited by a local realistic model to produce false Bell violations.

However, the fair-sampling assumption can exploited to mimic (false) nonlocal correlations using only local distributions. As an example, consider the Bell inequality in Eq. 2.6,

$$\mathcal{S}_{\text{CH}} = p(1,1|0,0) - p(1,0|0,1) - p(0,1|1,0) - p(1,1|1,1) \overset{\mathcal{L}}{\leq} 0,$$

and let Charlie prepare a state which obeys the following probability distribution,

17

$$p(a,b|x,y) = \frac{1}{4}(\delta_{x,0}\delta_{a,1} + \delta_{x,1}\delta_{a,-1})(\delta_{y,0}\delta_{b,1} + \delta_{y,1}\delta_{b,-1})$$
$$+\frac{1}{4}(\delta_{x,0}\delta_{a,0} + \delta_{x,1}\delta_{a,-1})(\delta_{y,0}\delta_{b,-1} + \delta_{y,1}\delta_{b,0})$$
$$+\frac{1}{4}(\delta_{x,0}\delta_{a,-1} + \delta_{x,1}\delta_{a,0})(\delta_{y,0}\delta_{b,0} + \delta_{y,1}\delta_{b,-1})$$
$$+\frac{1}{4}(\delta_{x,0}\delta_{a,-1} + \delta_{x,1}\delta_{a,0})(\delta_{y,0}\delta_{b,-1} + \delta_{y,1}\delta_{b,0}), \tag{2.9}$$

where $\delta_{i,j}$ is the Kronecker delta function. After Alice and Bob measure an ensemble of particles prepared with this distribution, they will estimate the CH Bell parameter ($\mathcal{S}_{\mathrm{CH}}$ in Eq. 2.6) as

$$\mathcal{S}_{\mathrm{CH}} = p(1,1|0,0) - p(1,0|0,1) - p(0,1|1,0) - p(1,1|1,1) \overset{\mathcal{L}}{\leq} 0,$$
$$\mathcal{S}_{\mathrm{CH}} = \frac{1}{4} - 0 - 0 - 0 = \frac{1}{4}, \tag{2.10}$$

which is a seemingly violation of the Bell inequality.

However, the distribution in Eq. 2.9 has a decomposition of the form in Eq. 2.2, and is therefore a local distribution! Essentially, this local hidden variable model can be thought of as a model where Charlie attempts to guess Alice and Bob's setting choice. If the particle sees the setting Charlie guessed, it will be detected as Charlie desires. If instead the particle sees the opposite setting, then it will not be detected at all, and thus the entire trial is discarded. In this way, Charlie can force Alice and Bob to observe a Bell inequality violation with a local realistic probability distribution since Alice and Bob only consider trials where Charlie correctly guessed their settings. This strategy critically relies on the fair-sampling assumption, and will not work without the assumption. In this example, Charlie exploits the fair-sampling assumption by forcing both Alice and Bob to only detect 50% of their particles. The relationship between loss (or missed detection events) and the fair-sampling assumption is important. If any experiment detects less than 2/3 of the prepared entangled particles (in a perfectly noiseless system), then a Bell inequality violation cannot be observed without a fair-sampling assumption [20]. Any detected noise will increase this minimum system efficiency. The efficiency requirement places very hard technical constraints on a system. For photonic systems, commonly used single-photon detectors (avalanche photo-diodes), typically have efficiencies of 65% , which already is too low to perform a Bell test. Furthermore, even extremely high quality systems will only have efficiencies around 80% of collecting and sending the photons to the single-photon detectors. As such, given a low-noise system, to perform a Bell test with photons, one must use detectors with > 90%

efficiency. Detectors of this quality (superconducting nanowire single-photon detector — SNSPD [21], and transition edge sensors — TES [22]) only recently became available, and require extensive cryogenic cooling (i.e., they are expensive).

The fair-sampling assumption is notably invalid if there is measurement-setting-dependent and outcome-dependent loss. Physically, as just one possible example, this could happen if the waveplate used to change basis has a wedge (the front and back face of the element are not parallel, which deviates the beam when the waveplate is rotated), and one of the output single-photon detectors is more sensitive to alignment. More maliciously, in terms of applicability to generate secure randomness, this assumption is critical since it is possible to experimentally realize the hidden-variable model in Eq. 2.9 by "blinding" the detectors. In the experiment in Ref. [23], bright laser pulses are used to alter the mode of operation of the single-photon detectors, enabling a (false) violation a Bell inequality that assumes fair-sampling. In fact, in the early Bell inequality experiments using atomic cascades (by Freedman, Clauser, Aspect, Grangier, Dalibard, and Roger), the fair-sampling assumptions made were demonstratively false [24] — the experiments only collected photons emitted back-to-back, which display unrepresentatively high polarization correlations.

Bell test experiments without fair-sampling assumptions have been performed in several matter systems: ions [25], superconductors [26], and atoms [27], whereas high-efficiency tests with photons have been lacking. Performing a Bell test without the fair-sampling assumptions has even more interest by noting its applicability to cryptography. Intuitively, if we have shown that the correlations exhibited between Alice and Bob's outputs do not obey local realism, or alternatively that the local systems do not have preexisting measurement results, then the measurement results must necessarily be unknown prior to measurement. Thus, we can use the results to generate provably secure random numbers (i.e., no one else could possible know what the random numbers are, otherwise the results would have been known in advance, and no Bell inequality violation would have been possible). In this case, conspiratorial models are necessary perfectly valid, as a adversary that is attacking the system would naturally attempt to implement such a model. In this case, it is critical to not make any fair-sampling assumption, as we cannot verify that an adversary is not controlling our source (in which case, the fair-sampling assumption would be invalid). In Chapter 4, we will discuss our experiment that performs a Bell test without any fair-sampling assumptions with photons, which we subsequently use for private randomness expansion.

It is worth noting that while we can perform a Bell inequality without assuming fair-sampling in the simplest (CH or CHSH Bell inequality) case, for Bell inequalities with additional inputs, discussed in Chapter 6, the technological constraints to remove the fair-sampling assumption are too challenging to achieve currently. In some of the measurements we make, we would require 99% total system efficiency (assuming

no noise), where the current best measured total system efficiency (including detection) is 85% [28]. For the more challenging Bell tests, we assume fair-sampling, as we are not attempting to generate random numbers, and instead we only wish to characterize nonlocality (and therefore there should be no malicious attacker who intentionally exploits the assumption). In this case, so long as we validate the fair-sampling assumption by demonstrating equal system efficiency regardless of setting and outcome, then in the absence of malicious adversaries, we can still make claims on nonlocality despite the technological constraints.

2.4   Measurement setting independence assumption
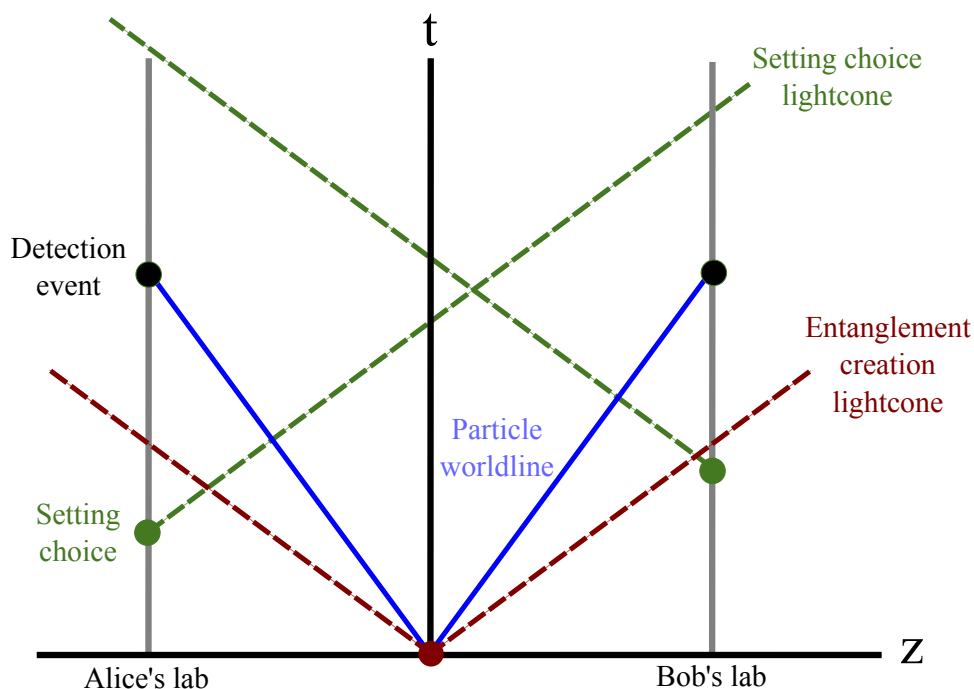
Figure 2.5: The spacetime diagram for a valid Bell test, which ensures the locality loophole and freedom-of-choice loopholes are properly addressed. The entanglement creation and the setting choice events should be spacelike separated (i.e., the setting choice should be below the red lightcone) to ensure the setting choices are not influenced by a signal sent by the entangled particles. Furthermore, the detection event on Alice's side and Bob's setting choice must also be spacelike separated events (i.e., the detection event must occur outside the green lightcone), to ensure that no entangled photons has information about the other particle's measurement setting.

As mentioned above, *how* and *when* we choose a measurement setting is important. In Alice's lab, she will need to have a random number generating device which chooses a measurement setting. In order for the experiment to rule out local realism, Alice must be locally determining her measurement setting (e.g., without any influence from Bob's measurement setting choice or outcome). If Bob's measurement outcome can influence Alice's measurement setting choice, then the explanation for the correlations in the

Bell inequality could be explained by a collaboration between Alice and Bob's measurement choices and detectors, where they communicate with each other to form statistics that violate a Bell inequality. It is, unfortunately, strictly impossible to prove the independence of Alice's setting choice from Bob's measurement setting or outcome, as the particles in Alice and Bob's measurement device, as well as the device that chooses the measurement setting, could have interacted at some point in their past lightcone. As such, there could always be residual correlations between their measurement settings and outcomes. Therefore, we are required to make an additional assumption when performing a Bell test. That is, our assumptions must always include locality, realism, and independent setting choice, with any violation implying at least one of the assumptions is incorrect. If, for example, one were to believe the independence assumption were invalid for a given experiment, then the experiment would not make any claims as to the validity of local realism, and the independence assumption is then referred to as a "locality loophole".

While the independence assumption cannot be avoided, and therefore the loophole will always exist, we can make arguments towards the validity of the assumption. For example, we can force Alice's measurement choice event and Bob's measurement detection event to be spacelike-separated events. From Reichenbach's Principle, correlations between events (under local realism) must either be from a causal influence or a common cause, so if the events are spacelike separated, then the only locally causal correlation between the outcomes must come from a common cause. Since, in our case, the entangled photons are generated at the same time, a natural local realistic theory could have the creation of the entangled photons as the common cause for the correlations between the measurement settings and outcomes, that is, when the entangled photons are created, the photons immediately send signals to the measurement setting choice devices to influence what setting is to be chosen. Thus, to exclude this local realistic model, we should also spacelike-separate the entanglement creation event from the measurement setting choices events in Alice and Bob's lab. Note that, in this model, the source of entangled photons can still be the source of a common cause so long as it sends out the signal to the two labs prior to creating the entangled photons, and so we still have to assume there are no common cause events (this specific component of the independence assumption is called the "free-choice" loophole). By forcing the separation of all of these relevant events, as depicted in Fig. 2.5, we can argue towards the validity of the independence assumptions.

Most local realistic models that benefit from the independence assumption require rather conspiratorial models, e.g., entangled particles communicating after they have been measured, and we are still able to scientifically argue towards the validity of the assumption. By doing so, we can provide the best scientific test of local realism possible, even against paranoid models. That is, if one wishes to still perceive reality in a local realistic manner, they must adopt a framework that cannot be experimentally tested (e.g., a super-

deterministic model, where the results of Bell tests were determined in advance such that we still *perceive* violations of local realism, despite existing in a local realistic universe).

In addition to assuming fair sampling, the early experiments also did not spacelike-separate all relevant events. Later experiments were able to address both of the locality assumptions, by using entangled photons to validate the independent-setting choice [29], and to validate the assumption that the setting choices are independent of the particles being measured [30]. However, to do so, these experiments still needed to assume fair-sampling. Simultaneously closing the locality and detection loopholes imposes competing requirements — the experiment requires vast separation between events, but this separation forces additional losses upon an already constrained system.

In terms of cryptography and private random number expansion, the locality loopholes are less critical. If the detectors cannot signal to each other to falsely produce a Bell violation, then an adversary can instead arrange to have the detectors send signals about the results of a legitimate experiment, which would still compromise the security of the randomness. However, depending on the model of eavesdroppers one wants to protect again, it is possible the experimenters still want to implement a system that does not have these locality loopholes.

2.5 Coincidence-time loophole

The fair-sampling and independence assumptions are typically made in Bell tests, as they ease the constraints on system design (not needing challenging synchronization, remote laboratories, and precise position and time measurements to validate the independence of settings) and technological requirements (not requiring $> 2/3$ system efficiency to remove the fair-sampling assumption). And while these assumptions allow one to make arguments against local realism, they present loopholes that could be exploited by a local realistic model to violate a Bell inequality.

However, beyond these two assumptions, there can also be implicit assumptions within the data analysis itself if it directly or tacitly assumes no-signaling or fair-sampling. Even worse, the data analysis may directly violate an assumption, thereby invalidating the analysis technique. The issue can be subtle; for example, in the case of the "coincidence-time loophole"[31], the implicit assumptions can come from an otherwise standard coincidence counting method, where the coincidence windows are centered on one party's detection events (the implicit assumption is that the local hidden-variable model has no time-dependence) instead of using a predefined coincidence window (see Fig. 2.6). In Chapter 3, we present our classical source which exploits this assumption to produce false Bell violations. Finally, additional loopholes can arise from the

assumed source statistics. Two analysis assumptions are noteworthy. The first is that most analyses assume that the source emits particles with independent and identical states. The second assumes that the average violation has a Gaussian distribution; in particular, nearly all reported Bell violations are cited in terms of numbers of standard deviations of violation, whose interpretation requires that the relevant distributions are Gaussian for many standard deviations from the mean, which fails to hold no matter how many particles are detected (for a discussion, see Section 3.3 and Ref. [32]).
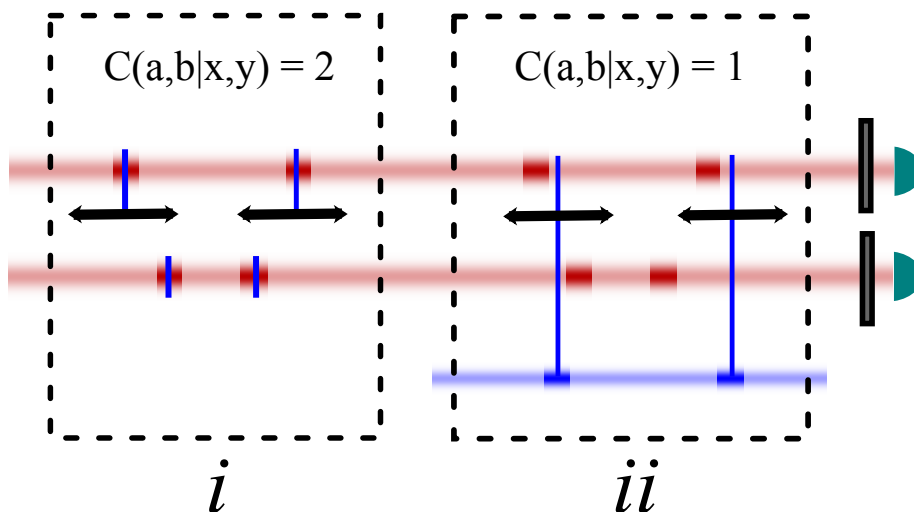
Figure 2.6: A display of two different coincidence counting methods. Due to timing jitter in the detectors (timing uncertainty), the output of Alice and Bob's detector can lead to ambiguities of the trial outcome. To solve this issue, Alice and Bob need to use a coincidence window (black arrow) to determine the outcome. Here, we depict two methods for assigning an outcome to a trial. In Fig. 2.6i, determining $C(a,b|x,y)$ is done through a traditional coincidence counting method, where a detection event on Alice's side is the center for a coincidence window. If Bob has an event falling within that window, his outcome is considered coincident, and the outcome of his event is assigned to the same trial as Alice's. However, this is a nonlocal method for determining trial outcomes, and therefore is not allowed. The hidden assumption in this case is that detectors have the same time response for all measurement settings and outcomes. In Chapter 3, we exploit this coincidence counting technique to falsely violate a Bell inequality. Fig 2.6ii is a valid form of coincidence counting, where the trials are decided based on an external clock (e.g., a reference laser pulse sent jointly to Alice and Bob, shown in blue). Here, the outcome of the trial is determined locally based on the external clock, rather than the other party's outcome.

In summary, as Bell tests can be a resource for cryptographic protocols, such as device-independent random number generation [33] and device-independent quantum key distribution [34], these issues raised here are critical to the security of the device, as each loophole allows for an avenue of attack. If the device satisfies a loophole-free Bell test, i.e., violates a Bell inequality with no extra assumptions, then the device can be trusted regardless of the manufacturer of the device or possible hacking technique. Thus, it is important to minimize any extra assumptions required by the analysis or its interpretation.

Chapter 3

Faking Bell Violations

According to quantum theory, Alice and Bob must share entanglement to obtain results that are not compatible with local realism. Traditionally, experimenters have endeavored to engineer their systems to deliver exactly one pair of entangled photons to Alice and Bob during each trial. This goal is technically challenging (due to the probabilistic nature of SPDC), so some experiments have allowed violations of the definition of a "trial." For example, in some experiments, several measurements happened before Alice and Bob chose new measurement settings, and those measurements were then analyzed as if they were from separate trials each with a randomly chosen setting. In some, Alice and Bob non-locally determined when trials have occurred and what their outcomes were. These allowances complicate the interpretation of the experiment so that local realism can be rejected only if other (often implicit) assumptions hold. In this chapter we demonstrate a system without entanglement that exploits these allowances and thus can appear to violate local realism when Alice and Bob determine trials and measurements non-locally. We then discuss proper analysis methods to ensure an experiment can definitely rule out local realism. The results in this chapter are published in Ref. [35][1].

3.1   Trials and assumptions

To explain possible reasoning for using relaxed definitions of a trial, consider again the CH Bell parameter (from Eq. 2.5),

$$\mathcal{S}_{\text{CH}} = p(1,1|0,0) + p(1,1|0,1) + p(1,1|1,0) - p(1,1|1,1) - p(a=1|x=0) - p(b=1|y=0), \qquad (3.1)$$

where $-1 \leq \mathcal{S}_{\text{CH}} \leq 0$ for any local realistic model (in this chapter, we will be concerned with both the -1 and 0 bound). The above Bell inequality is in a form that only need consider events where at least 1 party received a 1 outcome (i.e., there are no $p(0,0|x,y)$ terms). We can thus rewrite it in terms of coincidence counts ($J(x,y) = C(1,1|x,y)$), where $J(x,y)$ are coincidence (or joint) detection events at settings x and y),

---

[1]Portions of this chapter ©2015, American Physical Society. Used with permission.

and singles counts ($S_A(x) = C(1,0|x,0)+C(1,1|x,0)+C(1,0|x,1)+C(1,1|x,1)$, where $S_A(x)$ are the singles counts for Alice at setting x, and similarly for Bob). Singles counts and coincidence counts correspond to seemingly simple measurements: the total number of counts on a detector is the singles counts, and the total number of occurrences where Alice and Bob jointly received a detection event is the coincidence counts. Then the Bell inequality becomes

$$\mathcal{S}_{\text{CH}} = \frac{J(0,0)}{N(0,0)} + \frac{J(0,1)}{N(0,1)} + \frac{J(1,0)}{N(1,0)} - \frac{J(1,1)}{N(1,1)} - \frac{S_A(0)}{N_A(0)} - \frac{S_B(0)}{N_B(0)}, \tag{3.2}$$

where $N_A(0)$ ($N_B(0)$) is the number of times setting x (y) was chosen. This form is convenient to work with, since the resultant data from a Bell experiment is a string of detection times. That is, Alice and Bob do not record if a trial event is a 0 or 1; instead their record the *time* of each detection event, from which they must determine if a trial outcome was a 0 (no detected photon[2]) or 1 (a detected photon), or in the Bell inequality above, decide if a singles count was also a coincidence count (based on the time of the detection event). They must determine if the event was coincident despite the temporal uncertainty of their measurements.

Not only do all single-photon detectors have an intrinsic uncertainty in the precise arrival time of a detected photon (timing jitter), but downconversion is also a probabilistic process, where the emission can occur at any time when the pump laser has a non-zero amplitude. This is most notable for continuous-wave lasers, where downconversion events happen randomly, uniformly in time. These uncertainties cause individual photon-pair events to be difficult to reconstruct from the longer measurement record, so typical quantum optics experiments determine coincidence events by allowing for a coincidence window around one party's - say Alice's - detection events: if Bob has a detection event within the coincidence window determined by Alice's detection event, then it is called a coincident detection.

In a Bell test, however, this seemingly reasonable (but non-local) method for determining coincidences cannot exclude all local realistic models, as it opens up a loophole that can be exploited by a hacker to produce an apparent Bell inequality violation without any actual quantum correlations. The loophole, called the coincidence-time loophole, allows for a time-dependent local hidden-variable model [31]. Consider an experiment where the times of photon-pair arrivals at the two parties are unknown. To exploit the coincidence loophole, a hacker who has full control of the photon source can send a group of four pulses (two to Alice and two to Bob, as shown in Fig. 3.1) with each pulse offset by a little less than the Alice-detection-centered coincidence window used by Alice and Bob. In doing so, the pulses that result in detections for

---

[2]We will only be applying a fair-sampling assumption in Chapter 6, in all other chapters a 0 outcome will correspond to no detection event.

settings $x = 1$ and $y = 1$ are separated by nearly three coincidence window "radii" and therefore do not result in any coincidence counts, whereas at every other setting combination the detected pulses fall within the coincidence window. Consequently, a hacker can achieve an apparent Bell violation $\mathcal{S}_{\mathrm{CH}} > 0$, since the $p(1, 1|1, 1)$ term in Eq. 3.1 can be made to vanish.

It is interesting to consider the *size* of the apparent violation. The problem is somewhat subtle, as there is no longer a clear method to estimate a probability, since this loophole exists because of an invalid definition of a trial. Previously,

$$p(a, b|x, y) = \frac{C(a, b|x, y)}{N(x, y)} \tag{3.3}$$

was used to estimate the probability. However, in this case, there is no well-defined value $N(x, y)$, and therefore Alice and Bob must now first estimate $N(x, y)$ in order to be able to estimate $p(a, b|x, y)$. One natural way to do so would be to assume that the pairs arrive at a constant rate r_P. Given the constant rate assumption, we can express $p(1, 1|x, y) = C(1, 1|x, y)/(r_P \times t)$, where t is the duration of the measurement of $C(1, 1|x, y)$, and similarly for the other probabilities. The value of $\mathcal{S}_{\mathrm{CH}}$ can be inferred accordingly, and whether or not it violates the inequality $\mathcal{S}_{\mathrm{CH}} \leq 0$ does not depend on the rate r_P. Thus, it is not necessary to know the rate to observe such a violation. By exploiting the coincidence-time loophole, sending pulses at a rate r_H, a hacker can force an apparent violation of up to $\mathcal{S}_{\mathrm{CH}} = r_H/r_P > 0$, given the experimenter's assumed photon-pair rate r_P. If the experimenter attempts to measure r_P independently, this measurement may also be subject to the hacker's manipulations, and so the experimenter must estimate r_P from the data. However, the data set itself only yields a lower bound on r_P. That is, assuming (wrongly) that the detections arise from constant-rate photon pairs, we have that r_P should be at least the sum of the rate of detections by Alice and Bob, minus the rate of coincident detections at any given setting combination (i.e., if Alice detects N_A photons, Bob detects N_B, and they jointed detected N_{AB}, then they would conclude the source emitted at least $N_A + N_B - N_{AB}$ pairs). This rate is maximized for settings $x = 1$ and $y = 1$, where it is $2r_H$ (since $N_A = r_p$, $N_B = r_p$, but $N_{AB} = 0$). Accordingly, $r_P \geq 2r_H$. Setting $r_P = 2r_H$ gives a maximum inferred violation of $\mathcal{S}_{\mathrm{CH}} = 1/2$, which actually exceeds not only the local realistic limit, but even the maximal quantum mechanically allowed value of $\mathcal{S}_{\mathrm{CH}} \approx 0.207$! In this, this level of (false) extreme violation exactly matches the maximum allowed by no-signaling (no-signaling theories will be discussed in Chapter 6) [36]. When quantifying the violation (for example, in Fig. 3.1, Fig. 3.3, and Fig. 3.4) we use this normalization for the values of $\mathcal{S}_{\mathrm{CH}}$, i.e., we set $r_P = 2r_H$.

One method to close the coincidence loophole is to produce photon pairs only during well-defined time

windows by using a pulsed laser to pump the downconversion crystal. The arrival of the photon pairs at Alice and Bob can be synchronized, e.g., with a separate laser pulse sent to each party. The trial begins when Alice and Bob choose measurement settings just before the possible arrival of the entangled photon. Each party's measurement outcome is equal to 0 if no photons were detected and 1 if one or more photons were detected during the trial's duration. This strategy is used in the experiments discussed in Chapters 4 and 5. An alternative approach is to use a distance-based analysis method [37], which we applied to the data sets from this chapter as well as the data sets from Chapter 4 in Ref. [35]; we found no violation for the data sets from this chapter.

3.2 Experimental realization of the coincidence-time loophole

We realized the coincidence-time loophole experimentally by combining two attenuated lasers on a beam splitter for both Alice and Bob (Fig. 3.2). For Alice, one laser is polarized orthogonal to the polarizer setting for measurement setting $x = 0$, while the other laser is polarized orthogonal to the polarizer setting for $x = 1$, and similarly for Bob. This allows the source to address the measurement settings independently (i.e., when we send a laser pulse polarized along $(x = 1)^\perp$, we should only receive detection events for measurement setting $x = 0$). We then attenuate the sources to a mean photon number per pulse of around 10. The relatively high mean photon number offsets the loss in the measurement and detection process, but is still small enough to minimize the effect of crosstalk in the polarizer (there is a small chance that the polarization state to be blocked is still transmitted through the polarizer). We then pulse the lasers as shown in Fig. 3.1, with adjacent pulses separated by $T = 1\,\mu\text{s}$. If we determine the number of coincidence events by a non-local method of checking if Bob had a detection event within a window (e.g., $2\,\mu\text{s}$) around Alice's detection events, then we observe Bell inequality violations up to $\mathcal{S}_{\text{CH}} = 0.49$ (with the normalization discussed earlier), where Alice and Bob use the optimal settings for an ideal maximally entangled state as specified in the caption of Fig. 3.2. A plot of the data analyzed in this way is displayed in Fig. 3.3. We see a "violation" of over 2700-σ (assuming Gaussian statistics). By altering the two laser polarizations and increasing the mean photon number to offset any additional losses, we have been able to exploit this loophole for a wide range of measurement settings, as we now discuss.

3.2.1 Controlling Violation Size

In an actual attempt of a Bell test, Alice and Bob would likely suspect the presence of a hacker if their estimated CH-Bell parameter exceeded the quantum mechanical limit of $(\sqrt{2} - 1)/2$. Even more so, if Alice
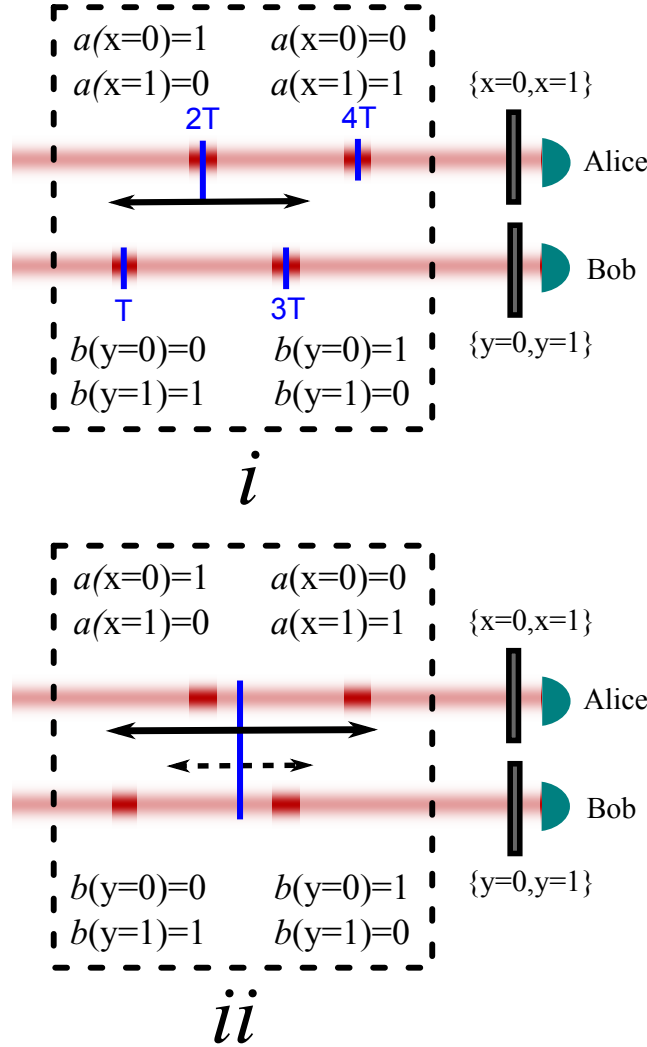
Figure 3.1: A diagram illustrating the coincidence-time loophole. Here $a(\text{x}=i)=1$ ($a(\text{x}=i)=0$) corresponds to a detection event (no detection event) when Alice chooses measurement setting i, and similarly for Bob. i) A coincidence window (black arrow) is selected based on Alice's detection event. A hacker can exploit this loophole by staggering pulses in time as shown. In this case, if the radius (half width) of the coincidence window is between T and $3T$, there are no $C(1,1|1,1)$ events (and thus $p(1,1|1,1)=0$), implying that $\mathcal{S}_{\text{CH}}$ in Eq. 3.1 is greater than 0, even for a classical source. ii) A well-defined trial is used, where the window is centered on a synchronization pulse (blue line); the loophole vanishes as there is no longer a way to address only the $C(1,1|1,1)$ term. If a large coincidence window (solid arrow) is used, then every measurement setting has a coincident event, resulting in $\mathcal{S}_{\text{CH}}=0$. If a short trial window (dashed arrow) is used, then there are only coincident and single events at settings $\{x=0,y=0\}$, giving a Bell value of $\mathcal{S}_{\text{CH}}=-1$; in neither case is $-1 \leq \mathcal{S}_{\text{CH}} \leq 0$ violated.
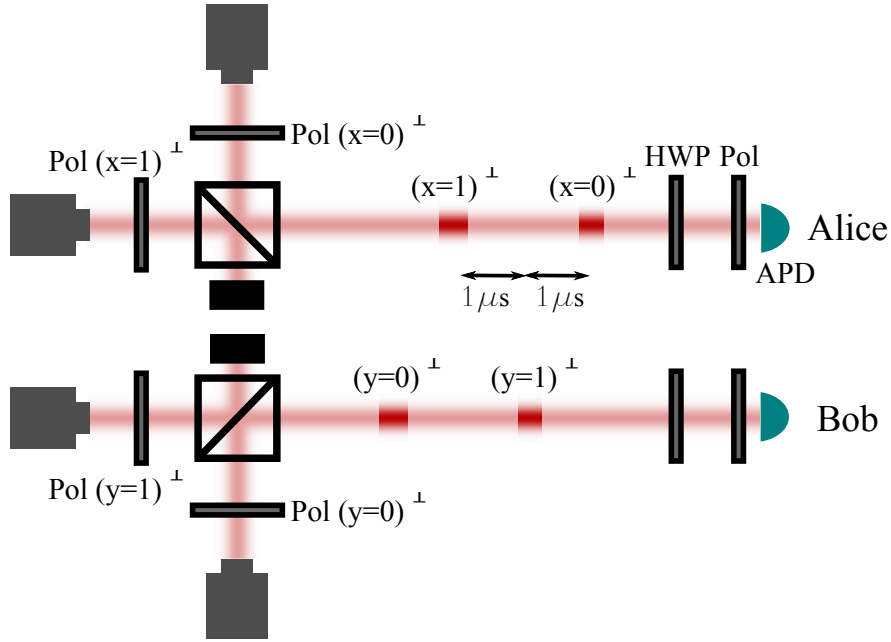
Figure 3.2: A diagram of our experimental setup to produce the local hidden-variable model described in Fig. 3.1. On Alice's side, we electrically pulse two 670-nm laser diodes with a pulse width of 100 ns; these pulses then pass through polarizers aligned orthogonally to her two measurement settings (Pol $(x = 0)^\perp$ and Pol $(x = 1)^\perp$). That is, we emit pulses that will *not* pass through one of the two measurement settings, ensuring only one of the measurement settings will detect our optical pulse. The laser pulse that passes through the $(x = 0)^\perp$ polarizer is emitted 2 μs before the $(x = 1)^\perp$ laser pulse. We attenuate the lasers enough so that after they are combined on a beam splitter, each pulse has a mean photon number of approximately 10, to offset any system loss while minimizing the noise due to crosstalk in the polarizers. Similarly on Bob's side, we combine two attenuated lasers on a beam splitter. Here, the $(y = 1)^\perp$ pulse is emitted 2 μs before the $(y = 0)^\perp$ pulse, and both are offset from Alice's photon pulses by 1 μs. The basis choice for the polarization analysis is implemented with a half-wave plate (HWP) and polarizer (Pol), where the settings are -11.25^o for $x = 0$, 33.75^o for $x = 1$, 11.25^o for $y = 0$, and -33.75^o for $y = 1$ (corresponding to the optimal CH-Bell-inequality-violating settings for a perfect maximally entangled state), thus, e.g., $\text{Pol}(x = 0)^\perp$ would correspond to setting the polarizer to $-11.25^\circ + 90^\circ = 78.75^\circ$. The photons are then detected by avalanche photo-diodes (APDs), with efficiencies lower than 66%, the outputs of which are recorded using time-to-digital converters. The results of analyzing the data both with a coincidence window determined by Alice's detection event, as well as a predefined coincidence window, are displayed in Fig. 3.3.
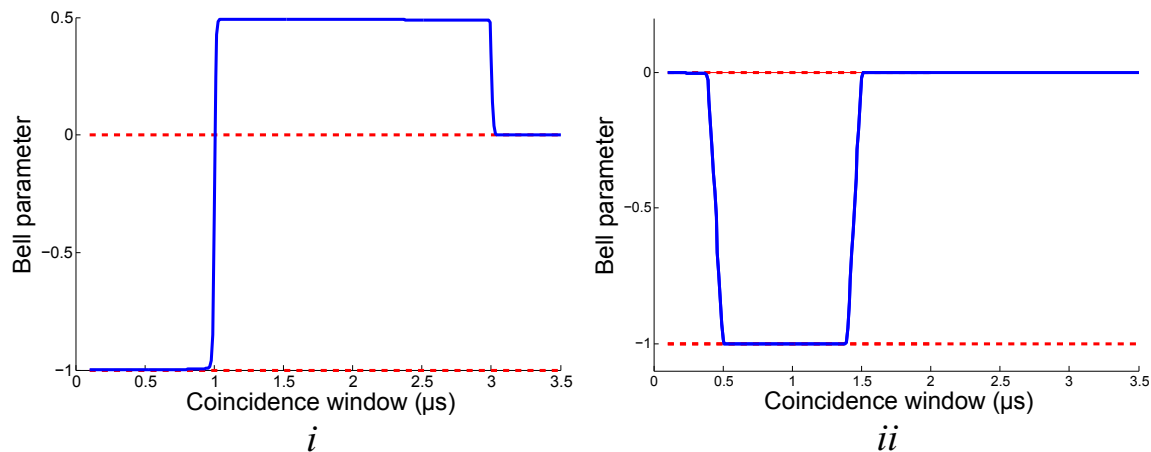
Figure 3.3: Two plots of the measured Bell parameter, $\mathcal{S}_{\mathrm{CH}}$ (Eq. 3.1) (solid blue line), as a function of the coincidence window radius for our experiment using a classical source to produce the local hidden-variable model shown in Fig. 3.1. When the data set is (improperly) analyzed with a coincidence window determined by a detection event (diagram i), the coincidence-time loophole can be exploited to produce a false Bell violation (values greater than 0). We separated each pulse by 1 µs, so with this model we see $\mathcal{S}_{\mathrm{CH}} > 0$ for any coincidence window radius between 1 µs and 3 µs. For coincidence windows less than 1 µs, we do not have any coincidence counts, but we still have single counts, resulting in a negative Bell parameter of $\mathcal{S}_{\mathrm{CH}} \geq -1$ (while this value depends on the chosen normalization, the minimal inferred value of r_p is at $r_p = 2r_h$, resulting in the most negative Bell parameter of -1). With window sizes larger than 3 µs, we assign coincident and single events to nearly every detection event, resulting in a Bell parameter of 0. In contrast, when the data is analyzed with a fixed predefined coincidence window (diagram ii), the Bell parameter remains between -1 and 0, and therefore never shows a violation of local realism. The results match well the predictions, given the structure of the classical source, as explained in Section 3.2.2. The positions of the transitions are due to the location of the predefined coincidence window relative to the pulse set. The transitions between 0 and -1 are not sharp because of a slow desynchronization between the fixed windows and the actual source pulse rate. That is, the window slowly drifted such that it was not always centered on the pulse set. For more details, see Section 3.2.2.
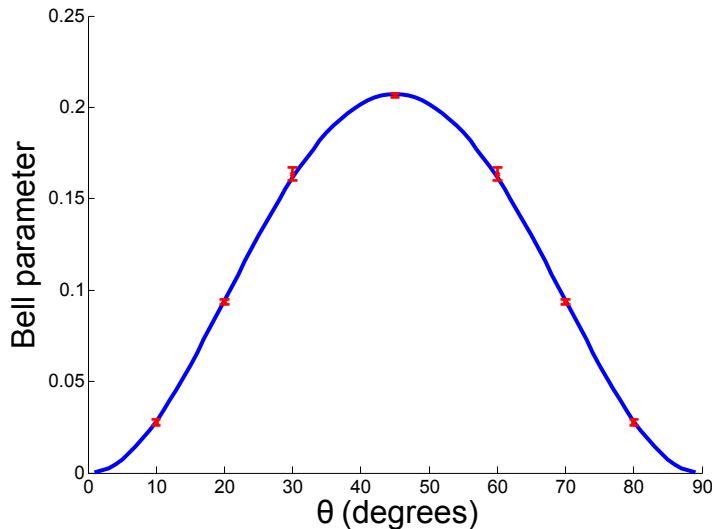
Figure 3.4: A plot of several measured data points from our classical source when analyzed incorrectly (susceptible to the coincidence-time loophole). The blue curve is the predicted quantum mechanical maximum given the state $\cos\theta|HH\rangle + \sin\theta|VV\rangle$; these states are critical to violate Bell inequalities without fair-sampling assumptions, as shown in Chapters 4 and 5. Here, we assumed Alice and Bob have a target θ and use the optimal measurement settings for their presumed input state. We then adjusted the angles of the source polarizers to match the quantum mechanically allowed maximum Bell parameter, given Alice and Bob's measurement settings. The resulting measurements (red data points) are indistinguishable from the quantum mechanical expectation, despite that the source is completely classical (and local).

and Bob know that they have low system efficiencies, then the value they expect is well below $(\sqrt{2}-1)/2$. In particular, with low efficiency, Alice and Bob design their system to use states of the form $\cos\theta|HH\rangle + \sin\theta|VV\rangle$ (see Chapters 4 and 5) to maximize the measured violation. Consequently, a hacker would want Alice and Bob to believe that they prepared a less entangled state (states with θ farther from $\pi/4$). If Alice and Bob estimate θ for their state, there is a maximum θ-dependent Bell parameter they expect. Ideally, the hacker would control the measured Bell parameter to match Alice and Bob's expectation and avoid suspicion. In our case, with the source depicted in Fig. 3.2, we can tune the source polarizers (and adjust the laser diode brightness to compensate for any increased loss) to create nearly any value of the Bell parameter. The results of several measurements using this technique are displayed in Fig. 3.4.

As a final note, while the plot in Fig. 3.3 has a well-defined structure (which would obviously make the hack easy to detect), it is possible to broaden the observed "violation range" by probabilistically switching between local hidden-variable models with different pulse spacings; therefore, one cannot simply look at a plot of the Bell violation versus coincidence window size to determine if the coincidence-time loophole is being exploited.

3.2.2 Predefined Window Analysis

In contrast to the above analysis, where the coincidence windows were defined *relative to Alice's detection*, if Alice and Bob determine each trial's measurement outcomes locally using a coincidence window of fixed duration centered on a predefined time, rather than one centered on a detection, we do not see a statistically significant Bell violation, as shown in Fig. 3.3.

To use a predefined window to analyze the data exploiting the coincidence-time loophole, we first add in a synchronization signal at the rate equal to the rate that the source emits a set of pulses, 100 kHz in our case (as there was no actual synchronization signal when the data set was taken, we implement this signal in post-processing). For comparison with Fig. 3.1*ii*, where the predefined coincidence window is in the center of the pulse set, we placed the first synchronization signal in the center as determined by the first two detection events in the data set. We then create a periodic signal by spacing each synchronization signal by 10 µs (= 1/100 kHz). To compensate for the relative temporal drift between the function generator and the timetagging electronics, we reset the synchronization signal every 500 detection events to be re-centered in the pulse set. If the separation between adjacent pulses is 1 µs (see Fig. 3.2), then we would expect a Bell parameter close to 0 for windows less than 0.5 µs, since there will be neither single nor coincident events (other than occasional dark counts, no event will fall within the predefined window)[3]. For windows between 0.5 µs and 1.5 µs we would expect a Bell parameter close to -1, since we see events primarily from $\{x = 0, y = 0\}$. That is, $p(1, 1|0, 0) = 1$, $p(a = 1|x = 0) = 1$, and $p(b = 1|y = 0) = 1$ in Eq. 3.1, while all other terms are 0. Finally, for predefined window sizes larger than 1.5 µs, all terms in Eq. 3.1 are equal to 1, leading to a Bell parameter of 0. The results of analyzing the classical data with a predefined coincidence window of variable width are displayed in Fig. 3.3*ii*.

While the data set in this case is contrived to be clearly determined by a local hidden-variable model, in real experiments the issues are far more subtle. For example, avalanche photo-diodes can have a count-rate-dependent latency, and since each measurement setting can have different detection rates (for example, in Ref. [38], the count rates differed by a factor of 3), it is critical that the analysis is not susceptible to these minor latency shifts. To show that these issues are relevant, Ref. [39] proposes a coincidence-loophole-exploiting scheme whose statistics closely match those of a standard photon-pair source.

---

[3]A single event, when analyzed using fixed windows, is an event that is in coincidence with the external clock. A coincident event now requires a three-fold coincidence between the external clock, Alice, and Bob.

3.3 Hypothesis testing

Exploiting the coincidence-time loophole demonstrates the need to be careful, not only when one designs a system, but also when the data of that system is analyzed. If the security of a private random number expansion system is dependent on a Bell violation, one has to be well aware of all possible attacks on the system and the analysis of the data. Furthermore, because the traditional coincidence counting method (that we exploit to produce the false Bell violations) is so common, this classical system further shows the need to truly understand the assumptions made in systems being used in practice.

For example, another common assumption is the Gaussian-statistics assumption. That is, typically an assumption is made to quantify the certainty of a Bell violation (e.g., "we observed a M-σ violation"), but this is another assumption that should be addressed. This assumption is most critical for large "σ" violations, where upwards of 10 standard deviation violations are often claimed (e.g., 244-σ [26]). Here, the normal distribution assumption has surely failed at the tail ends of the distribution[4], and therefore this bound is a clear overstatement of the certainty of a Bell violation. Furthermore, the standard deviation of a violation provides only information on the certainty of the measured quantity, but no information on the probability that a competing hypothesis (in this case, local realism) could have produced the results of the experiment. Instead, to determine the desired information, one should use a hypothesis test.

A hypothesis test is a statistical test where two data sets are compared to determine the probability the null hypothesis could have produced the data set or one more extreme (this probability, or p-value, is the statistical confidence of ruling out the null hypothesis) [40]. As an example, consider a coin that is believed to be biased towards heads. To demonstrate the coin is not fair, we want to disprove (through statistical testing) that the coin is fair (i.e., $p(\text{heads}) = p(\text{tails})$). So we start with the null hypothesis that the coin is fair. Next, to perform hypothesis testing we require a test statistic, that is, a single measure of the data set; for the biased coin, a test statistic could be the number of times a heads was the outcome of a coin toss. Before performing the coin-tossing experiment, we first must determine the number of times we will flip a coin, which is called the "stopping criteria"[5]. The stopping criteria is based on our expectation of the bias and our desired level of confidence in our result (what failure rate will be allowed for the experiment). If we believe the coin is only weakly biased, then more coin flips will be required; additionally, requiring a higher confidence in our measurement will result in choosing a larger number of coin flips as the stopping criteria.

Note that, before performing the experiment, we should also decide on how we will analyze the data, so

---

[4]To make claims at the level of 10-σ, the Gaussian approximation would have to be valid at the $\sim 10^{-23}$ level. By this point, there is almost certainly some other dominant source of noise that does not follow the normal distribution.

[5]It is possible to have no stopping criteria, but the analysis must take that fact into account, which results in a severely weakened p-value.

that our analysis method is not influenced by the data. While this point is unnecessary for coin flips (surely we will not have a hard time from the data determining if a coin flip was heads or tails), for Bell tests it is critical that the analysis method be decided before viewing the data, the reason being that the analysis itself could produce stronger results by post-selecting on statistical fluctuations, if one optimizes over system parameters such as relative delays between detectors or the size of the trial window. As it can often be challenging to construct a full analysis prior to having any data, an alternative approach is to sacrifice, say, 10% of the data as "training data", which is used to determine optimal analysis parameters (and then discarded and not used as part of the final data set). The remaining 90% of the data then can be analyzed using the parameters determined by the training data to produce a valid (not overestimated) test statistic.

Returning to the coin example, based on the expected bias and desired confidence, a number of coin flips is chosen (we decide to flip 300 coins), and then the experiment is performed. After flipping 300 coins and seeing 181 heads, our next step is to calculate the p-value, the probability that the null hypothesis (the coin is fair, i.e., $p(\text{heads}) = p(\text{tails})$) could have produced our test statistic (181 heads) or anything more extreme. Thus, the p-value for this example is calculated as

$$\sum_{k=181}^{300} \binom{300}{k} (1/2)^k (1/2)^{300-k} = 2.1 \times 10^{-4}. \tag{3.4}$$

Smaller p-values can be interpreted as stronger evidence against the null hypothesis. A small enough p-value allows one to effectively rule out null hypothesis. In particular, based on our desired confidence (or allowed failure rate), we can decide the meaning of the p-value. Perhaps in the above example we had prior belief that the coin was likely biased, and therefore, based on prior knowledge, we will say the p-value rules out the coin was fair. Or perhaps we had a stash of $10{,}000$ coins where we believed some may be biased, and so we had performed a statistical hypothesis test on each coin; then the measured p-value may not be strong enough to demonstrate this particular coin is biased as we would expect a p-value of this magnitude to be measured more than once even for fair coins. Or perhaps whether this coin is biased would demonstrate the existence of a new particle, then surely such a important discover would require a significantly lower p-value, for example, the high-energy community requires p-values smaller than $\approx 3 \times 10^{-7}$, which is equivalent to a 5-σ violation assuming Gaussian statistics. We believe tests of local realism require similarly small p-values. P-values can be tricky to interpret, which further reinforces the need to be very strict when performing a hypothesis test.

In our case, beyond allowing one to rule out local realism, p-values from Bell inequality violations can

also be used as certificates for cryptographic applications, such as random number generation, that rely on a Bell test [41, 38]. And therefore any invalidity in the testing technique again results in loopholes an adversary can exploit. Furthermore, it should be noted that all of the methods discussed here in performing Bell tests (e.g., understanding assumptions, determining system parameters from training data which is then discarded, choosing the duration of the experiment and desired confidence prior to performing an experiment) are not only relevant for Bell tests. Every experiment should follow these guidelines. That is, every experiment should thoroughly consider all assumptions, acknowledge where the assumptions fail (e.g., at the tail ends of a Gaussian distribution, the normally-distributed approximation most likely has failed), and perform proper hypothesis testing where applicable. Finally, every experiment should be careful not to use final data sets to determine analysis parameters.

Chapter 4

Bell Violation With no Fair-Sampling Assumption

In this chapter, we report on a valid violation of a Bell inequality, which is published in Ref. [38][1]. Specifically, our experiment uses polarization entangled photons to fully close the detection loophole. Here, we discuss the system in detail, and describe the relevant aspects of collecting our entangled photons with high efficiency. Moreover, we apply the stronger-than-classical correlations to verify the creation of true random numbers, achieving rates over 4 orders of magnitude beyond all past experiments. Finally, we conclude with a discussion on changes that can be made to improve different aspects of the system.

4.1 Detection-loophole-free Bell inequality

Consider again the CH Bell inequality, which places the following constraint on any local realistic theory:

$$\mathcal{S}_{\text{CH}} = p(1,1|0,0) + p(1,1|0,1) + p(1,1|1,0) - p(1,1|1,1) - p(a=1|x=0) - p(b=1|y=0) \overset{\mathcal{L}}{\leq} 0. \quad (4.1)$$

The inequality can be violated using maximally entangled states (e.g.,$(|HH\rangle + |VV\rangle)/\sqrt{2}$, where H and V represent the polarization of the photons), assuming a detection efficiency $\eta > 2(\sqrt{2}-1) \approx 0.828$; this is the lower efficiency limit for any maximally entangled two-particle, two-measurement-setting system measured with a pair of detectors [42]. However, further analysis by Eberhard [20] showed that with non-maximally entangled states, e.g.,

$$|\psi_r\rangle = (|HH\rangle + r|VV\rangle)/\sqrt{1+r^2}, \quad (4.2)$$

the detector efficiency requirement could be reduced to 2/3, although no background counts in the detector can be tolerated in this limit. Essentially, using a small value of r, one can choose $x = 0$ and $y = 0$ to nearly block the vertically polarized single counts (thereby decreasing the $p(a=1|x=0)$ and $p(b=1|y=0)$

---

[1]Portions of this chapter ©2013, American Physical Society. Used with permission.

terms which contribute negatively to the Bell inequality in Eq. 4.1), while choosing $x = 1$ and $y = 1$ to maximize the positive terms. In this limit, the chosen measurements are essentially crossed polarizers (i.e., we produce mostly V-polarized photons, while our measurements settings are projecting onto H-polarization), and therefore any unpolarized noise or background noise is significantly amplified relative to the signal we are trying to detect.

As described in Chapter 3, we determine the probabilities in Eq. 4.1 by normalizing the measured singles and coincidence rates to the number of trials with the specific analyzer setting for each term. We can then write

$$\mathcal{S}_{\text{CH}} = \frac{J(0,0)}{N(0,0)} + \frac{J(0,1)}{N(0,1)} + \frac{J(1,0)}{N(1,0)} - \frac{J(1,1)}{N(1,1)} - \frac{S_A(0)}{N_A(0)} - \frac{S_B(0)}{N_B(0)} \overset{\mathcal{L}}{\leq} 0, \tag{4.3}$$

which is the Bell inequality whose violation we describe in this chapter.

In order to avoid the coincidence-time loophole [31] (discussed in Chapter 3), we use a Pockels cell between crossed polarizers to periodically transmit short "bursts" of the pump laser. Each burst corresponds to a single well-defined "event", easily distinguished with the detectors. Care must still be taken, however, to guarantee that there is no temporally-correlated effect that unduly affects the measured counts. For example, laser power drift can lead to a violation even with non-entangled photons, if the order of the measurements is not made randomly. We address this issue by measuring each of the terms in Eq. 4.3 multiple times while randomly choosing the detector settings, and then determine the counts and relative errors (due to both finite counting statistics as well as multiple measurements of each term).

4.2 Entangled photon source

For our entanglement source (see Fig 4.1), we use a mode-locked frequency-tripled Nd:YAG laser (120-MHz repetition rate, 5-ps pulse width, $\lambda = 355$ nm). Our type-I phase-matched spontaneous parametric downconversion source [43, 44] consists of a pair of BiBO crystals, cut at $\theta = 141.8^o, \phi = 90^o$, and $\gamma = 0^o(90^o)$ for the first (second) crystal. Each crystal is only 200-μm thick to reduce the birefringent walkoff of the two downconversion beams ($\sim 20\mu$m). We precompensate the temporal walkoff [45] with a 550-μm-thick piece of BBO (cut at $\theta = 90^o$), resulting in very high entanglement quality (with an interference visibility $99.70 \pm 0.05\%$, i.e., how well the $|HH\rangle$ and $|VV\rangle$ terms interfere in an arbitrary basis). The degree of entanglement of the downconverted photons can be controlled using waveplates to manipulate the pump polarization, i.e., a pump polarization of $\left(|V\rangle + re^{i\phi}|H\rangle\right)/\sqrt{2}$ will produce the entangled state $\left(|HH\rangle + re^{i(\phi+\phi_c)}|VV\rangle\right)/\sqrt{2}$, where ϕ_c is the relative phase picked up in the nonlinear crystals [44].
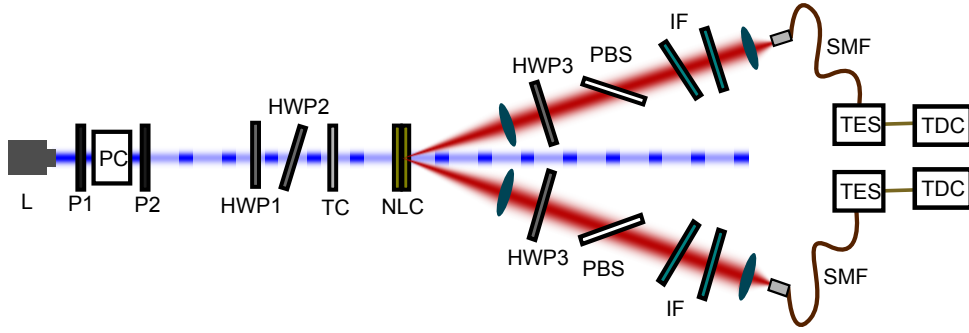
Figure 4.1: A diagram of the system used to violate the CH Bell inequality. We "pulse" our laser (L) by putting a Pockels cell (PC) between crossed polarizers (P1 and P2). The Pockels cell is periodically turned on for a short time to allow 240 laser pulses to transmit through P2, thus creating "event intervals" that the detectors can distinguish, which we can then use as well-defined trials. Downconversion is produced in paired nonlinear BiBO crystals (NLC). The produced state is controlled through the half-wave plates HWP1 and HWP2, which control the relative amplitude and phase of the $|HH\rangle$ and $|VV\rangle$ downconversion terms. We attain very high entanglement quality by compensating temporal decoherence caused by group-velocity dispersion in the downconversion crystals [45], using a BBO crystal (TC). The downconversion creation spot is imaged onto a single-mode fiber (SMF). HWP3 sets the basis for the polarization analysis (based on input from QRNG data) by the Brewster's angle polarizing beam splitter (PBS). Custom spectral interference filters (IF) are used to only detect spectrally conjugate photons. Finally, the photons are detected by transition-edge-sensor (TES) detectors; the output signals are sent to a time-to-digital converter (TDC) to record all detection event times.

The pump is focused to a 350-μm radius waist on the crystal, while the downconversion is collected from a 60-μm waist, which is located near the crystals. To image the downconversion crystal spot onto the collection fiber, we use a 250-mm focal length plano-convex lens with an 11-mm focal length asphere on each arm of the downconversion. All lenses are anti-reflection coated. Here, to a very high approximation, the collection of one photon into a SMF heralds a photon in a well-defined Gaussian mode on the opposite side of the downconversion cone. We are able to collect the conjugate photon in its own SMF with an estimated 90% efficiency (assuming no other losses). The SMF is then fusion-spliced to a fiber connected to a TES detector [22].

These detectors, supplied by NIST in Boulder, are made using a thin tungsten film embedded in an optical stack of materials to enhance the absorption [46]. Photons are delivered to the detector stack using a SMF for 1550 nm, which is anti-reflection (AR)-coated for 710 nm and fusion-spliced (with less than 5% loss[2]) to the 710-nm SMF used for downconversion collection. The detectors, cooled to $\sim$ 100 mK using an adiabatic demagnetization refrigerator, are voltage-biased at their superconducting transition so that absorbed photons cause a measurable change in the current flowing through the tungsten film. The change in current is measured with a superconducting quantum interference device (SQUID) amplifier, the output

---

[2]Splice loss is hard to measure, so we give an overestimation of the loss.

of which is connected to room-temperature electronics for processing before being sent to a time-to-digital converter (TDC). The stream of timetags from the TDC are sent to a computer and saved for later analysis.

At the 710-nm wavelength of our entangled photons, the TES detectors have a net efficiency (including loss in the input fiber, fusion splice, and from the AR coating being imperfect over our large bandwidth) of 0.95 ± 0.01. In fact, the development of such high efficiency detectors was the very advance that enabled both our experiment, and a competing one [47]. Unfortunately, the readouts of the TES detectors also have a rather large timing jitter, ~ 500 ns. For our experiment, a high-power PBS and HWP attenuate the laser power (from 4 W to ~ 20 mW) to reduce the number of pairs produced per second (multi-photon downconversion events harm the Bell violation, and the problem is greatly exacerbated by the 1-μs coincidence window to accommodate the large TES jitter). As the TES detector jitter is also much larger than the interpulse spacing of our pulsed laser (1 μs compared to 8.3 ns), we use a BBO Pockels cell (Extinction Ratio (ER) $> 300 : 1$) followed by a standard birefringent polarizer to create 2-μs-long pulse trains at a 25 kHz rate (approximately 240 laser pulses).

The polarization correlations for the Bell test are measured with a polarization analyzer consisting of a fixed Brewster's angle polarizing beam splitter (we detect the transmitted photons with an extinction ratio $> 8700{:}1$ and transmission $> 99\%$), preceded by a motion-controlled AR-coated half-wave plate (at 710 nm) to choose the basis of the projective measurement. In the experiment we randomly choose the measurement settings ($x = 0$ or $x = 1$, $y = 0$ or $y = 1$, see Table 4.1) using the output from a separate photon-arrival-time-based quantum random number generator (QRNG) [48].

Finally, in each arm we use a combination spectral filter, comprised of two interference filters (see Fig. 4.2), centered on twice the pump wavelength (710 nm). One of the filters has a lower bandedge at ~ 700 nm (this was custom set at the company, Semrock, by heat treating the filter), while the other filter has an upper bandedge that can be tuned (to lower wavelengths) by tilting the filter. We then tilt this filter to maximize the heralding efficiency, which will set the upper bandedge to ~ 720 nm (to satisfy energy conservation, the wavelength of the upper edge for the signal photon must be conjugate to the lower edge of the idler photon and vice versa, i.e. $\frac{1}{\lambda_{signal,upper}} = (\frac{1}{\lambda_{pump}} - \frac{1}{\lambda_{idler,lower}})^{-1}$). We were thus able to achieve a spectral heralding efficiency (the heralding efficiency if there is no other loss in the system) of 95%. After accounting for all losses, the total system efficiency is $\sim 76\%$ from the $|HH\rangle$ crystal, and $\sim 69\%$ efficiency from the $|VV\rangle$ crystal (due to birefringent transverse walkoff, which reduces the collection of this mode). However, because the measurements are made at angles near H, we estimate the net effective system efficiency to be $75\% \pm 2\%$, sufficient to violate a Bell inequality without needing any extra fair-sampling assumption
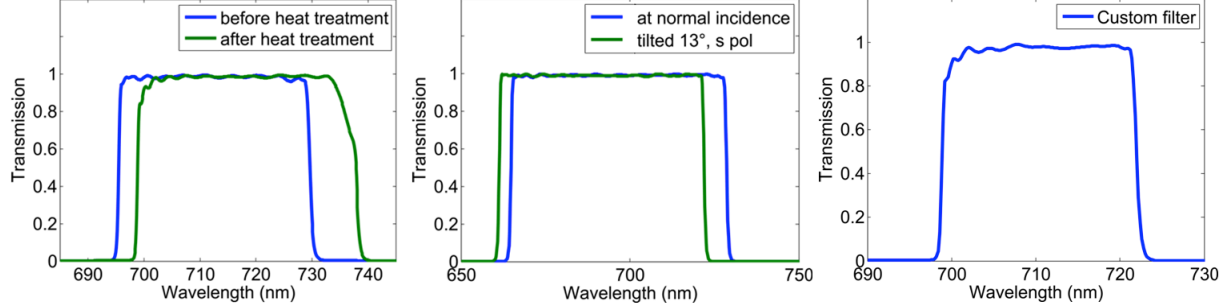
Figure 4.2: The combined filter spectrum. The lower bandedge (at 700 nm) is set by a fixed filter, while the upper bandedge (at 720 nm) is set by a tunable filter. The upper bandedge is controlled by tilting the tunable filter, and adjusted to maximize the heralding efficiency. The resultant spectral heralding is 95%.

4.3 Background counts and desired state

To determine the optimal r value (in Eq. 4.2), we first need to characterize the system noise. The state that maximally violates a CH Bell inequality is a compromise between non-unit system efficiency (which pushes to smaller r values) and non-zero (unpolarized) background counts (contributing to the singles rate, which limits the minimum usable r value) [20]. The noise in our experiment is dominated by stray photons (e.g., counts from LEDs in the room) and of broadband fluorescence from the crystal. The fluorescence rate (broadband emission believed to be mostly from the anti-reflection coating on the BiBO crystal surfaces) was measured by pumping the crystal with only an H(V)-polarized pump, and orienting the downconversion photon polarizers to H(V) as well. By accounting for the extinction ratio of the polarizer, we can estimate the number of unpolarized broadband (i.e., not downconversion) photons produced from the crystal. We observed rates ranging from 0.15% to 0.25% of the singles counts (i.e., if we had 10,000 singles counts per second, measured with V-polarizers when the pump was H-polarized, we would see between 15-25 fluorescence counts per second, after H-polarizers). The rate would vary depending on the spot being pumped on the crystal, and also increased over time; we suspect the coating on the crystal surface started to degrade under exposure to the UV laser.

In addition, we also detected approximately 8 cps of stray photons coupled into the SMF, i.e., even with the pump laser off. While these noise counts may seem minor, they increase the required detection efficiency from 2/3 (with no noise) to $\sim 73\%$. For these background levels in our experiment, and estimated system detection efficiency, a value of $r = 0.26$ allows us to maximally violate the CH inequality[3].

---

[3]Note that while using a non-maximally entangled state allows us to reduce the required system efficiency to achieve *any* violation, the actual *value* of the Bell parameter (according to quantum mechanics) achievable with such a state approaches the local realistic bound as $r \to 0$, which is the reasoning for the form of the plot in Fig. 3.4.

Table 4.1: The accumulated measurements. We used the settings $(x = 0) = 3.8^o$, $(x = 1) = -25.2^o$, $(y = 0) = -3.8^o$, $(y = 1) = 25.2^o$, and $r = 0.26$, see Appendix A for a discussion on the calculation of the settings. We cycled through these measurement settings randomly (using QRNG data [48] to pick the basis for a given run), changing the measurement settings in intervals of 1 second, with approximately 25,000 trials in each 1 second interval. The data corresponds to $\mathcal{S}_{CH} = 5.4 \times 10^{-5}$, larger than the local realistic bound of $\mathcal{S}_{CH} \leq 0$), and a p-value of 1.16×10^{-10}.

Settings	Singles(A)	Coincidences	Singles(B)	Trials
$x = 0, y = 0$	46,068	29,173	46,039	27,153,020
$x = 0, y = 1$	48,076	34,145	146,205	28,352,350
$x = 1, y = 0$	150,840	34,473	47,447	27,827,318
$x = 1, y = 1$	150,505	1,862	144,070	27,926,994

4.4 Data for detection-loophole-free Bell test

We collected timetags for the Bell test in "blocks" of 1 second (25,000 trials per second) at each measurement setting, for a total of 4450 blocks. Note that this is a slight violation of the definition of a trial we have been using, in that we do not have a new unique setting for each output. However, this fact is actually taken into account in the p-value analysis [35], where the trial output now consists of the entire 1-s measurement block (instead of a binary 0 or 1 output). Therefore, to progress to a full loophole-free Bell test, we "only" need to spacelike separate all relevant events. The data, summarized in Table 4.1, shows a p-value of $2^{-33} = 1.16 \text{x} 10^{-10}$ (for comparison, this p-value is a 6.3-σ standard deviation equivalent) [35], which we conclude is sufficient to reject our null hypothesis that local realism explains the results of our experiment. An interesting way to "interpret" this p-value, would be to ask the following: assuming local realism is correct and given that we have 1.2 hrs of data, how long, on average, would it take a local realistic model to reproduce our test statistic? In this case, one would need to perform the experiment for $1.2 \text{x} 10^6$ years (we would have to have started the experiment around the same time the first humans appeared!). The $\mathcal{S}_{CH}$ parameter, for use in private randomness expansion, was measured to be $\mathcal{S}_{CH} = 5.4 \text{x} 10^{-5}$. Our results are in good agreement with those predicted using our measured entangled state, after accounting for the measured background and fluorescence noise.

As we have made a no-signaling assumption, that is we assumed, e.g, $p(a|x, y) = p(a|x, y')$, we should make a check on the data to verify that we have not clearly violated the assumption. To do so, we can look at the all of the local probability terms, and check their dependence on the non-local setting choice. As the counts are rare events, they should (assuming no additional noise) follow Poissonian statistics. Any additional noise in the system will only increase the uncertainty beyond Poissonian statistics; therefore, as long as the local count rates between two different non-local setting choices are within 1 to 2 standard deviations of each other (assuming Poissonian statistics), then we can say that the no-signaling assumption

Table 4.2: The no-signaling checks. Here, we apply Eq. 4.4 to the four possible combinations and report the standard deviation distance between the two local distributions. For visual clarity, we define $S_A^{(N)}(x|y) \equiv S_A(x|y)/N(x,y)$ as the normalized singles counts. The values here are well within the expected range, implying that there is no clear violation of no-signaling.

No-signaling term	Standard deviation distance
$\dfrac{S_A^{(N)}(0|y=0) - S_A^{(N)}(0|y=1)}{\sqrt{S_A(0|y=0)}}$	0.12
$\dfrac{S_A^{(N)}(1|y=0) - S_A^{(N)}(1|y=1)}{\sqrt{S_A(1|y=0)}}$	2.25
$\dfrac{S_B^{(N)}(0|x=0) - S_B^{(N)}(0|x=1)}{\sqrt{S_B(0|x=0)}}$	-1.20
$\dfrac{S_B^{(N)}(1|x=0) - S_B^{(N)}(1|x=1)}{\sqrt{S_B(1|x=0)}}$	-0.16

has not clearly been violated. Thus we expect to see

$$\frac{S_A(x|y=0)/N(x,0) - S_A(x|y=1)/N(x,1)}{\sigma_{x|y=0}/N(x,0)} \approx \frac{S_A(x|y=0)/N(x,0) - S_A(x|y=1)/N(x,1)}{\sqrt{S_A(x|y=0)}/N(x,0)} \approx 1. \quad (4.4)$$

where $S_A(x|y=0)$ is the singles counts on Alice side at setting x, given Bob choosing setting y, and $\sigma_{x|y=0}$ is the uncertainty of $S_A(x|y=0)$ (under Poissonian statistics, $\sigma_{x|y=0} = \sqrt{S_A(x|y=0)}$). The denominator here can be either of the two singles term's uncertainty, as it does not change the calculation significantly. The no-signaling check is performed similarly for Bob. We expect to see the absolute values to be mostly less than 1 ($\sim 68\%$ of the terms should have absolute values < 1), and occasionally values as high as 2 ($\sim 95\%$ of the absolute values should be < 2). When we start seeing multiple values around 3 or more, then we should begin to check the data for systematic errors (as we have only 4 no-signaling checks, see Table 4.2 for the 4 possible terms, the probability of a single 3-σ event is $4 \times 0.25\% \approx 1\%$, the probability of multiple values near 3 is on order of 10^{-4}, a very unlikely random occurrence). It is important to emphasize again, that passing this check does not guarantee the assumption was correct — we can only check if the assumption is invalid. The four different no-signaling checks are shown in Table 4.2, indicating there is no clear violation of the no-signaling assumption.

4.5 Device-independent randomness expansion

The high entanglement quality, along with the detection-loophole-free capability, offers interesting possibilities for applications, notably for "device-independent" quantum information processing. Here the goal is to implement a certain protocol, and to guarantee its security, without relying on assumptions about the internal functioning of the devices used in the protocol. Being device-independent, this approach is more robust to device imperfections compared to standard protocols, and is in principle immune to "side-channel attacks[4]" (which were shown to jeopardize the security of some experimental quantum cryptography systems).

One prominent example is device-independent randomness expansion (DIRE) [33, 49, 50, 51], which uses the random outcomes in a Bell test to produce a string of random numbers. Strictly speaking, we do not generate randomness, as we always require a "seed" of some sort; in our case, we require a random seed to set the inputs into the Bell test (if done correctly, future inputs can be determined by previous outputs). In that sense, we only securely expand the random seed into a longer secure random string. By performing local measurements on entangled particles, and observing nonlocal correlations between the outcomes of these measurements, it is possible to certify the presence of genuine randomness in the data in a device-independent way. DIRE was previously demonstrated in a proof-of-principle experiment using entangled atoms located in two traps separated by one meter [33]; however, the resulting 42 random bits required a month of data collection! Here we show that our setup can be used to implement DIRE much more efficiently. The intrinsic randomness of the quantum statistics can be quantified as follows. The probability for any observer (hence, for any potential adversary) to guess the measurement outcome (of a given measurement setting) is bounded by the amount of violation of the CH inequality: $p_{guess} \leq (1 + \sqrt{2 - (1 + 2\mathcal{S}_{\mathrm{CH}})^2})/2$ [33] (neglecting finite size effects); i.e., $p_{guess} = 1$ when $\mathcal{S}_{\mathrm{CH}} = 0$, and $p_{guess} = 1/2$ when $\mathcal{S}_{\mathrm{CH}} \approx 0.207$ (the maximum allowed by quantum mechanics). In turn, this leads to a bound on the min-entropy per bit of the output data, $H_{min} = -log_2(p_{guess})$. Finally, secure private random bits can be extracted from the data (which may in general not be uniformly random) using a randomness extractor [52]. At the end of the protocol, a final random bit string of length $L \approx N * H_{min} - S$ is produced, where N is the length of the raw output data, and S includes the inefficiency and security overhead of the extractor.

Over the 4450 measurement blocks (each block features 25,000 events), we acquire 111,259,682 data points for 3 hours of data acquisition. The average CH violation of $\mathcal{S}_{\mathrm{CH}} = 5.4\mathrm{x}10^{-5}$ gives a min-entropy of $H_{min} = 7.2\mathrm{x}10^{-5}$. Thus, we expect ~ 8700 bits of private randomness, of which one could securely extract at least 4350 bits. The resultant rate (0.4 bits/s) improves by more than 4 orders of magnitude over the bit

---

[4]A side-channel attack is an attack based on the actual implementation, rather than a theoretical weakness, of the protocol. For example, exploiting the coincidence-time loophole to fake a Bell test in Chapter 3 was a side-channel attack, because the system implemented the Bell test improperly.

rate achieved in [33] ($1.5\text{x}10^{-5}$ bits/s). This shows that efficient and practical DIRE can be implemented with photonic systems.

4.6   Prospects and shortcomings of the system

While the system described above had incredibly high visibility and impressive heralding efficiency at 710 nm, for future Bell tests (e.g., a completely loophole-free Bell test that also incorporates spacelike separations, and Bell tests with larger inputs), the system does have limitations. In this section, we will look at the two important metrics, heralding efficiency and interference visibility, for our system.

4.6.1   Heralding efficiency limitations

To transition from a detection-loophole-free Bell test to a full loophole-free Bell test, one must separate Alice and Bob's detection systems such that the relevant events are spacelike separated. First, the large increase in distance the photons must travel introduces a non-negligible loss, either due to transmission losses in optical fibers, or significantly increased sensitivity to vibrations, which even with active feedback will introduce loss when coupling the photons to the detectors. And second, the spacelike separation requires active switching elements, which introduces more loss since they typically have more interfaces and the photons must travel through more material (than just a AR-coated halfwave plate); in our lab the best Pockel cells still have $\sim 1\%$ loss. To view how much additional loss the current system can handle, it is informative to normalize the Bell inequality in a slightly different manner:

$$\mathcal{S}'_{\text{CH}} = \frac{\frac{J(0,0)}{N(0,0)} + \frac{J(0,1)}{N(0,1)} + \frac{J(1,0)}{N(1,0)} - \frac{J(1,1)}{N(1,1)}}{\frac{S_A(0)}{N_A(0)} + \frac{S_B(0)}{N_B(0)}} \overset{\mathcal{L}}{\lessgtr} 1. \qquad (4.5)$$

Here, the numerator consists of the coincidence counts, which are proportional to ϵ^2, where ϵ is the system efficiency, and the denominator is porportional to ϵ. Thus, the Bell parameter $\mathcal{S}'_{\text{CH}}$ is proportional to the system efficiency. A plot of our measured values of $\mathcal{S}'_{\text{CH}}$ versus the state-entanglement parameter (r) is shown in Fig. 4.3. We achieve a maximum of $\mathcal{S}'_{\text{CH}} = 1.015 \pm 0.002$, implying a 1.5% efficiency overhead (i.e., we could tolerate 1.5% more loss in each channel, with the current level of background). If we consider a low-loss active switch, e.g., Pockels cells, of 1%, then the photon's trip to Alice and Bob's remote labs must be done with $< 1\%$ loss. Ideally, one would use optical fiber as the channel to the remote labs (using mirrors introduces significant beam-pointing control problems due to the long lever arms of each mirror). Unfortunately, 710 nm is a poor wavelength for low-loss fiber; the lowest loss available are pure silica core
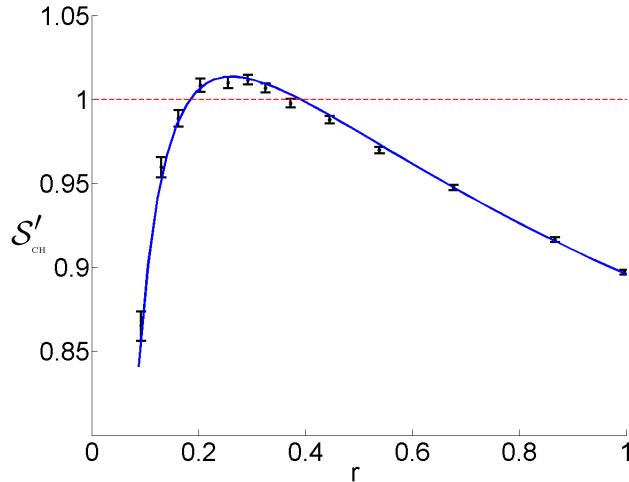
Figure 4.3: A plot of the Bell parameter $\mathcal{S}'_{\text{CH}}$ (Eq. 4.5) as a function of the produced entangled state. $\mathcal{S}'_{\text{CH}} > 1$ (red dashed line) is not possible for any local realistic theory. Data points in black are the measured $\mathcal{S}'_{\text{CH}}$ as r is varied in the state $(r|HH\rangle + |VV\rangle)/\sqrt{1+r^2}$; $r = 1$ (0) corresponds to a maximally entangled (separable) state. For this plot, every data point was measured for 30 seconds at each measurement setting; the particular settings were optimally chosen based on the model of our source for each value of θ. The blue line represents the Bell parameter we expect from the model of our source. Here, to improve the statistics, we did not pulse our source with the Pockels cell. We see violations for $0.20 < r < 0.33$.

fibers from Nufern, which we measured to have a core attenuation of ~ 3.6 dB/km (Nufern specs the fiber at < 10 dB/km). With this fiber, we expect approximately 2.5% loss in the transmission to the remote labs (assuming 30 m to each lab). Thus, to achieve a loophole-free Bell test, we would need to further increase our heralding efficiency.

With our system, we have a known loss due to chromatic dispersion in the 11-mm lenses (across the 20-nm bandwidth, the different colors see a different focal length, and thus some wavelengths couple worse into the fiber than others, leading to an average 2% loss). By replacing the lenses, we would have a 3.5%-efficiency overhead, which is approximately equal to the amount of additional loss we expect to enter the system. Another possible method to recover additional efficiency overhead would be to use adaptive optics (e.g., a deformable mirror) to improve coupling into the single-mode fiber. However, the improvement is likely to be negligible, based on theoretical estimates of the optimal coupling our source could achieve — although we appear to still be able to improve the fiber coupling by approximately 1%, the deformable mirror itself introduces approximately the same amount of loss (the reflectivity is only 99%). Thus we are within the uncertainty of our measurements as to whether our system can perform the loophole-free Bell test; without additional overhead, it would be unlikely to succeed. Instead, as we will see in Chapter 5, we can use a 1550-nm source of entangled photons, where fiber losses are < 0.5 db/km.

4.6.2 Visibility limitations
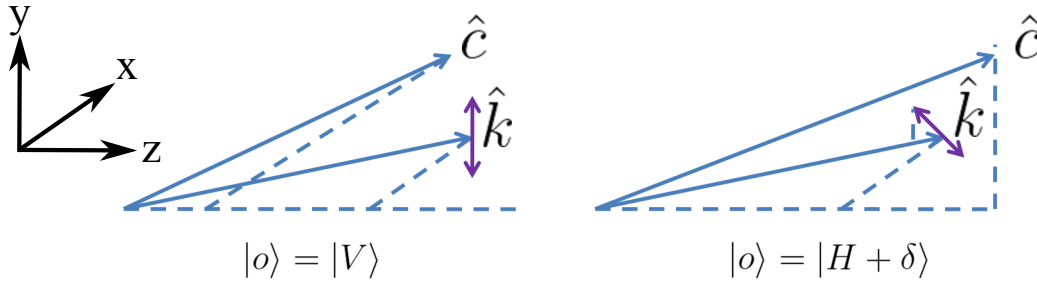
Figure 4.4: A depiction of the ordinary polarization ($|o\rangle$) for the two downconversion crystals. In a negative uniaxial crystal, the Type-I downconversion process is $e \to oo$, where the downconverted photon's ordinary eigen-polarization obeys $\hat{o} \cdot \hat{c} = 0$ and $\hat{o} \cdot \hat{k} = 0$, where $\hat{o}$ is the ordinary polarization unit vector, $\hat{c}$ is the crystal axis unit vector, and $\hat{k}$ is the propagation direction of the downconverted photon (and the extraordinary eigen-polarization of the pump is orthogonal to the pump's ordinary polarization). Here, if we collect downconversion with a non-zero opening angle ($\hat{k}$ not parallel to the pump), and with a crystal axis ($\hat{c}$) that is not vertical, then the eigen-polarizations in the crystal for the downconverted photons are not necessarily the H/V basis (where $|H\rangle$ is defined to be a polarization in the $x - z$ plane, and $|V\rangle$ is defined to be a polarization entirely in the y direction. In the left crystal, $\hat{k}$ and $\hat{c}$ are co-planar (both are in the $x - z$ plane), thus the ordinary polarization is simply $|V\rangle$ (it must be orthogonal to the $x - z$ plane, and thus it must be entirely in the y direction). However, if $\hat{k}$ and $\hat{c}$ do not form the $y - z$ plane, as in the right crystal, then the ordinary polarization is slightly rotated out of the H/V basis (i.e., its polarization is not in the $x - z$ plane). This effect causes us to no longer make a perfect maximally entangled state, which is the dominant source of entanglement degradation in our system, as we have corrected the other major sources of decoherence.

The incredible visibility (99.7%) of our entanglement source allows for the challenging Bell tests discussed in Chapter 6. However, it is important to note why there is a "missing" 0.3%. In our system, by coupling into a single-mode fiber, we have removed any spatial-mode decoherence. By temporally compensating for the group velocity dispersion in the crystal, we have also removed any spectral decoherence, thus leaving us with the dominant docoherence being from not creating the correct polarization states. That is, one of the crystals does not actually produce the desired state $|HH\rangle$, whereas the other crystal accurately produces $|VV\rangle$. To understand this effect, consider a negative uniaxial crystal (while BiBO, our crystal of choice due to its high nonlinearity, is biaxial, a similar but more complicated analysis can still be done). In this case, the downconversion process is $e \to oo$ (i.e., an extraordinary polarized pump downconverts to two ordinary polarized photons). Now, the extraordinary polarization is along the crystal optic axis, whereas the ordinary ray is orthogonal to the optic axis, as shown in Fig. 4.4. The issue is this: since we collect downconversion at a 3° angle (relative to the pump), and since our crystal axis ($\hat{c}$) is also cut at a somewhat large angle (141.8°), this results in the $|H + \delta\rangle = |H + 1.5°\rangle$ polarization state. If we consider the other photon in the pair, its k-vector is at $-3°$, which results in the polarization state of $|H - \delta\rangle = |H - 1.5°\rangle$, for a final

entangled state of $|V, V\rangle + |H + 1.5°, H - 1.5°\rangle$. This state has a 99.7% visibility, and cannot be unitarily transformed into a maximally entangled state (or our desired nonmaximally entangled state), as the two terms have different eigen-bases.

This effect can be eliminated (or significantly reduced) in a few ways. First, the effect depends on having a sizable opening angle, so decreasing the opening angle from 3° reduces the effect, and collinear downconversion (both photons propagate in the same direction as the pump) completely eliminates it. However, we typically work at 3° so that it is easy to separate the two photons; moving towards collinear downconversion requires other techniques to split the two photons, e.g., using non-degenerate downconversion and a dichroic mirror to separate the colors. Alternatively, periodically-poled crystals do not observe this effect, as the crystal axis is not at an angle (however, periodically-poled downconversion sources are almost always required to be collinear). One could also use the crystal temperature to adjust the indices of refraction to achieve "non-critical" phase matching, from which again crystal axis is then vertical (though, this method may be impractical due to the potentially high temperatures required). Finally, one could use Type-II downconversion (i.e., $f \rightarrow sf$), as it can require smaller crystal cuts (in this case, the temporal compensation becomes more challenging due to the 3 different group velocities one has to consider). Of these methods, the most practical is most likely producing collinear downconversion with periodically-poled crystals to achieve the next level of visibilities ($> 99.99\%$). Nevertheless, our existing system still has very high visibility, which allows us to make the intriguing advanced quantum nonlocality measurements presented in Chapter 6.

Chapter 5

A Loophole-Free Bell Test

With the detection-loophole closed with a photonic source, the next experimental step is to address all loopholes simultaneously. To this end, in collaboration with Krister Shalm, and many others, we performed a loophole-free Bell test at NIST-Boulder, which is published in Ref. [28][1]. In this chapter, we present our loophole-free violation of local realism using entangled photon pairs. We ensure that all relevant events in our Bell test are spacelike separated by placing the parties far enough apart and by using fast random number generators and high-speed polarization measurements. A high-quality polarization-entangled source of photons, combined with high-efficiency, low-noise, single-photon detectors, allows us to make measurements without requiring any fair-sampling assumptions. We compute p-values as small as 5.9×10^{-9} for our Bell violation while maintaining the spacelike separation of our events. Based on our experimental results, along with the two other loophole-free Bell test experiments, one using nitrogen vacancy centers in diamonds [53], and another with entangled photons [54], we reject the hypothesis of local realism.

5.1   Loophole-free Bell inequality

Our Bell test uses the version of the Clauser-Horne inequality presented in Eq. 2.6 [14, 20, 55] where, according to local realism,

$$p(11 \mid 00) \quad \overset{\mathcal{L}}{\leq} \quad p(10 \mid 01) + p(01 \mid 10) + p(11 \mid 11). \tag{5.1}$$

The terms $p(11 \mid 00)$ and $p(11 \mid 11)$ correspond to the probability that both Alice and Bob record detection events when they choose the measurement settings $\{x = 0, y = 0\}$ or $\{x = 1, y = 1\}$, respectively. Similarly, the terms $P(10 \mid 01)$ and $P(01 \mid 10)$ are the probabilities that only Alice or Bob record an event for settings $\{x = 0, y = 1\}$ and $\{x = 1, y = 0\}$, respectively. A local realistic model can at best saturate this inequality, while the probability distributions of entangled quantum particles can violate it.

To quantify our Bell violation we construct a hypothesis test based on the inequality in Eq. (5.1). The null hypothesis we test is that the measured probability distributions in our experiment are constrained by local realism. Our evidence against this null hypothesis of local realism is quantified in a p-value that we compute from our measured data using a test statistic. Our test statistic is similar to the example presented in Section 3.3. Here, the null hypothesis achieves (at best) $p(11 \mid 00) = p(10 \mid 01) + p(01 \mid 10) + p(11 \mid 11)$, similar to the fair-coin hypothesis having $p(\text{heads}) = p(\text{tails})$. Our "heads" is then $C(11|00)$, and a coin flip occurs if there is an event in the set $\{C(11|00), C(10|01), C(01|10), C(11|11)\}$. The p-value is calculated somewhat similar as before. In our case, to ensure proper analysis for computing the p-value, we use a martingale binomial technique from [55] that makes no assumptions about the distribution of events and does not require that the data be independent and identically distributed [56] as long as appropriate stopping criteria are determined in advance. This technique is, in essence, the same as used in Section 3.3, where $p(\text{heads}) = p(11|00)$, and $p(\text{tails}) = p(10|01) + p(01|10) + p(11|11)$. In this case, the "number of heads" is $C(11|00)$ and total coin flips is $C(11|00) + C(10|01) + C(01|10) + C(11|11)$, the p-value is then calculated accordingly[2].

5.2   The entanglement source

In our experiment, the source creates polarization-entangled pairs of photons and distributes them to Alice and Bob, located in distant labs. At the source location a mode-locked Ti:Sapphire laser running at repetition rate of approximately 79.3 MHz produces picosecond pulses centered at a wavelength of 775 nm (Fig. 5.1). These laser pulses pump an apodized[3] periodically poled potassium titanyl phosphate (PPKTP) crystal to produce photon pairs at a wavelength of 1550 nm via the process of spontaneous parametric downconversion [57]. The downconversion system was designed using the tools available in [58]. The PPKTP crystal is embedded in the middle of a polarization-based Mach-Zehnder interferometer that enables high-quality polarization-entangled states to be generated [59]. While here we use beam-displacers, the idea is similar to the interferometer shown Fig. 1.4a. In this case, we form an interferometer with polarizing beam displacers (instead of polarizing beam splitters), since they are typically higher quality.

Varying the polarization analyzer angles at Alice and Bob, we measure the visibility of coincidence detections for a maximally entangled state to be 0.999 ± 0.001 in the horizontal/vertical polarization basis

---

[2]One notable difference between the coin flips and the Bell test is that the Bell test has a distribution of setting choices, which one can take into account by increasing the likelihood a local realistic model would be able to produce $C(11|00)$ events.

[3]In a normal periodically poled crystal, the nonlinearity is a rectangular function in time, which causes the spectrum of downconversion to be a sinc function (Fourier transform of the rectangular function). This is undesirable as the secondary lobes of the sinc function cause additional spectral modes, thereby reducing the net collection efficiency. To solve this issue, the crystal is apodized, which means the crystal's nonlinearity is made to look like a Gaussian (typically by altering the poling period across the crystal), which then has a Fourier transform of a Gaussian, and thus does not have the secondary lobes.
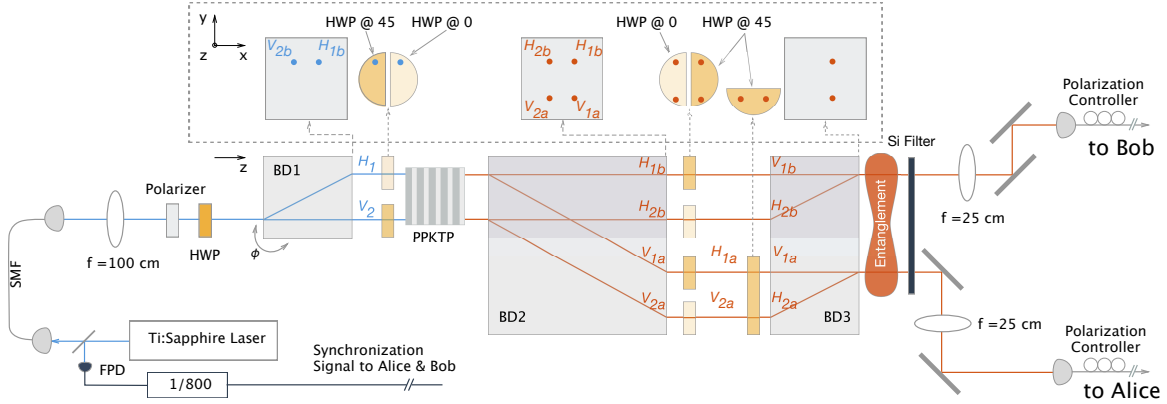
Figure 5.1: Schematic of the entangled photon source. A pulsed 775 nm-wavelength Ti:Sapphire picosecond mode-locked laser running at 79.3 MHz repetition rate is used as both a clock and a pump in our setup. A fast photodiode (FPD) and divider circuit are used to generate the synchronization signal that is distributed to Alice and Bob. A polarization-maintaining single-mode fiber (SMF) then acts as a spatial filter for the pump. After exiting the SMF, a polarizer and half-wave plate (HWP) set the pump polarization. To generate entanglement, a periodically poled potassium titanyl phosphate (PPKTP) crystal designed for collinear Type-II phasematching is placed in a polarization-based Mach-Zehnder interferometer formed using a series of HWPs and three AR-coated calcite beam displacers (BD). At BD1 the pump beam is split in two paths (1 and 2): the horizontal (H) component of polarization of the pump translates laterally in the x direction while the vertical (V) component of polarization passes straight through. Tilting BD1 slightly sets the phase, ϕ, of the interferometer to 0. After BD1 the pump state is $(\cos(16°)\,|H_1\rangle + \sin(16°)\,|V_2\rangle)$. To address the polarization of the paths individually, semi-circular waveplates are used. A HWP in path 2 rotates the polarization of the pump from vertical (V) to horizontal (H). A second HWP at 0° is inserted into path 1 to keep the path lengths of the interferometer balanced. The pump is focused at two spots in the crystal, and photon pairs at a wavelength of 1550 nm are generated in either path 1 or 2 through the process of spontaneous parametric downconversion. After the crystal, BD2 walks the V-polarized signal photons down in the y direction (V_{1a} and V_{2a}) while the H-polarized idler photons pass straight through (H_{1b} and H_{2b}); the x–y view shows the resulting locations of the four beam paths. HWPs at 45° correct the polarization while HWPs at 0° provide temporal compensation. BD3 then completes the interferometer by recombining paths 1 and 2 for the signal and idler photons. The two downconversion processes interfere with one another, creating the entangled state in Eq. (5.2). A high-purity silicon wafer with an anti-reflection coating is used to filter out the remaining pump light. The idler (signal) photons are coupled into a SMF and sent to Alice (Bob). When the SMFs are instead fusion-spliced to a superconducting nanowire single-photon detector (SNSPD) (instead of being spliced to the fiber that sends the photons to the remote labs), we achieve system efficiencies around 85%, close to the theoretical limit of the system. There is approximately 1% loss due to the clipping (loss due to the beam overfilling an aperture) on the beam displacers.
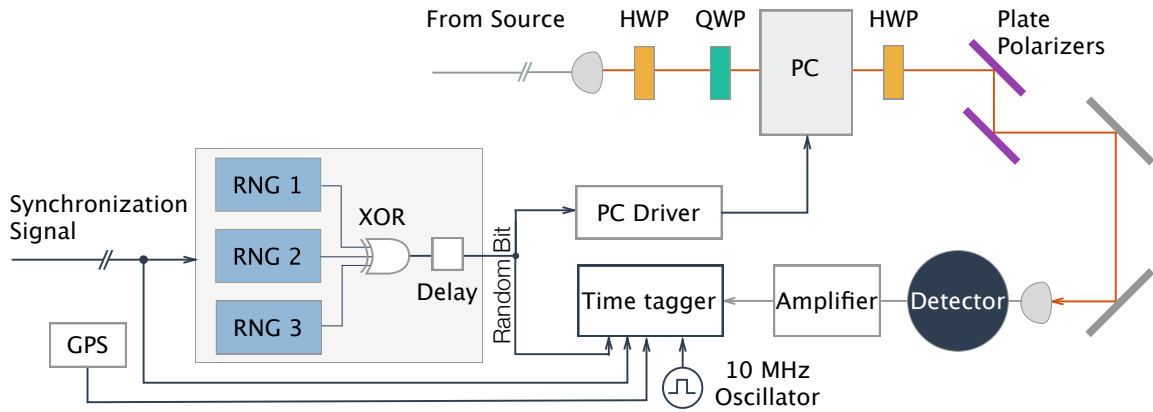
Figure 5.2: Receiver station setup for Alice and Bob. A photon arrives from the source. Two half-wave plates (HWP), a quarter-wave plate (QWP), a Pockels cell (PC), and two plate polarizers together act to measure the polarization state of the incoming photon. The polarization projection is determined by a random bit from XORing the outputs of two random number generators (RNG1 and RNG2) with pre-determined pseudorandom bits (RNG3). If the random bit is "0", corresponding to measurement setting a (b) for Alice (Bob), the Pockels cell remains off. If the final random bit is "1", corresponding to measurement setting a' (b') for Alice (Bob), then a voltage is applied to the Pockels cell that rotates the polarization of the photons using a fast electro-optic effect. The two plate polarizers have a combined contrast ratio $> 7000 : 1$. The photons are coupled back into a single-mode fiber (SMF) and detected using a superconducting nanowire single-photon detector (SNSPD). The detector signal is amplified and sent to a time-tagging unit where the arrival time of the event is recorded. The time tagger also records the measurement setting, the synchronization signal, and a one pulse-per-second signal from a global positioning system (GPS). The pulse-per-second signal provides an external time reference that helps align the time tags Alice and Bob record. To synchronize the internal clocks on Alice's and Bob's time taggers, we drive a laser diode at $10\,\text{MHz}$, which is split in a fiber beam splitter and sent to Alice and Bob, where they detect the signal on a standard photodiodes. The synchronization pulse from the source is used to trigger the measurement basis choice.

and 0.996 ± 0.001 in the diagonal/antidiagonal polarization basis[4]. The entangled photons are then coupled into separate single-mode optical fibers with one photon sent to Alice and the other to Bob. Alice, Bob, and the source are positioned at the vertices of a nearly right-angle triangle (see Fig. 5.4a). Due to constraints in the building layout, the photons travel to Alice and Bob in fiber optic cables that are not positioned along their direct lines of sight. While the photons are in flight toward Alice and Bob, their random number generators each choose a measurement setting. Each choice is completed before information about the entangled state, generated at the PPKTP crystal, could possibly reach the random number generators. As shown in Fig. 5.2, when the photons arrive at Alice and Bob, they are launched into free space, and each photon passes through a Pockels cell and polarizer that perform the polarization measurement chosen by the random number generators. After the polarizer, the photons are coupled back into a single-mode fiber and sent to superconducting nanowire single-photon detectors, each with a detection efficiency of $91 \pm 2 \%$ [21]. The detector signal is then amplified and sent to a time tagger where the arrival time is recorded. We assume the measurement outcome is fixed when it is recorded by the time tagger[5], which happens before information about the other party's setting choice could possibly arrive, as shown in Fig. 5.3(b).

Alice and Bob have system detection efficiencies of $74.7 \pm 0.3\%$ and $75.6 \pm 0.3\%$, respectively. Background counts from blackbody radiation and room lights reduce our observed violation of the Bell inequality. Every time a background count is observed it counts as a detection event for only one party. As discussed in Chapter 4, these background counts increase the heralding efficiency required to close the detector loophole above 2/3 [20]. To reduce the number of background counts, the only detection events considered are those that occur within a window of approximately $625 \, \mathrm{ps}$ at Alice and $781 \, \mathrm{ps}$ at Bob, centered around the expected arrival times of photons from the source[6]. The probability of observing a background count during a single window is 8.9×10^{-7} for Alice and 3.2×10^{-7} for Bob, while the probability that a single pump pulse downconverts into a photon pair is $\approx 5 \times 10^{-4}$. These background counts in our system raise the efficiency needed to violate a Bell inequality from 2/3 to 72.5 %. Given our system detection efficiencies, our entangled photon production rates, entanglement visibility, and the number of background counts, we estimate the optimal non-maximally entangled state and measurement settings for Alice and Bob. The optimal state is given by:

---

[4]Although the final experiment was a non-maximally entangled state, high visibility (particularly in the diagonal basis) proves that the beam-displacer-based system has sufficient coherence and stability.

[5]This assumption is untestable. In quantum mechanics, a measurement never occurs; to measure a state we merely entangle the state to pointer, but to measure the pointer we must entangle it to another pointer, and so on. But to argue that our detection events are spacelike separated, we must assume a measurement occurred at some point. As one motivation for this experiment is randomness expansion, we allow the adversary to access the signal until it reaches our timetagger. If she has access to our timing electronics, then she can hack the system in a plethora of other ways, so we then assume she only has access up to this point.

[6]The two nanowires were not from the same fabrication run, and therefore have different characteristics. To get the high detection efficiency, we measured many SNSPD detectors and chose the highest efficiencies.
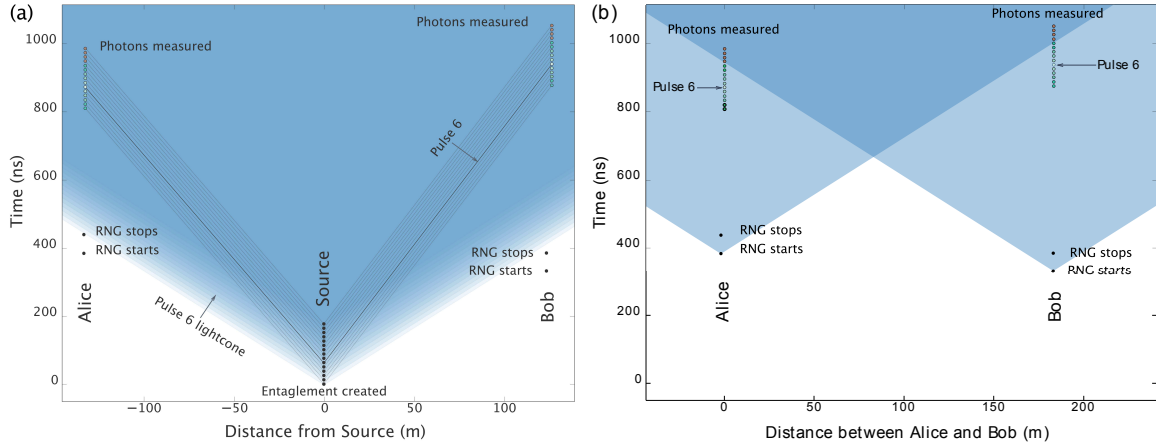
Figure 5.3: Minkowski diagrams for the spacetime events related to Alice (A) and the source (S) and Bob (B) and the source (a), and Alice and Bob (b). All lightcones are shaded blue. Due to the geometry of Alice, Bob, and the source, more than one spacetime diagram is required. In a) the random number generators (RNGs) at Alice and Bob must finish picking a setting outside the lightcone of the birth of an entangled photon pair. A total of 15 pump pulses have a chance of downconverting into an entangled pair of photons each time the Pockels cells are on. The events related to the first pulse are not spacelike separated, because Alice's RNG does not finish picking a setting before information about the properties of the photon pair can arrive; pulses 2 through 11 are spacelike separated. As shown in (b), pulses 12 through 15 are not spacelike separated as the measurement is finished by Alice and Bob after information about the other party's measurement setting could have arrived. In our experiment the events related to pulse 6 are the furthest outside of all relevant lightcones.
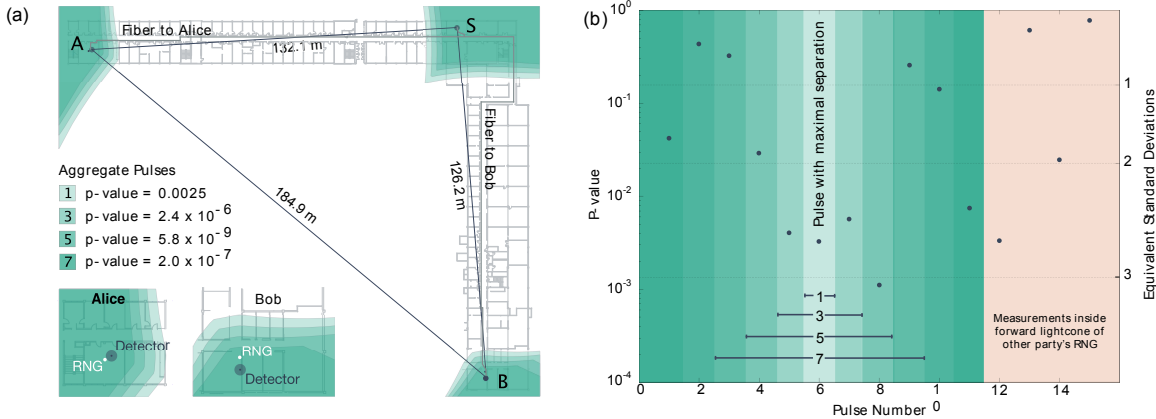
Figure 5.4: (a) The positions of Alice (A), Bob (B), and the source (S) in the building where the experiment was carried out. The insets show a magnified (×2) view of Alice's and Bob's locations. The white dots are the location of the random number generators (RNGs). The larger circle at each location has a radius of 1 m and corresponds to our uncertainty in the spatial position measurements. Alice, Bob, and the source can be located anywhere within the green shaded regions and still have their events be spacelike separated. Boundaries are plotted for aggregates of one, three, five, and seven pulses. Each boundary is computed by keeping the chronology of events fixed, but allowing the distance between the three parties to vary independently. In (b) the p-value of each of the individual 15 pulses is shown. Overlayed on the plot are the aggregate pulse combinations used in the contours in (a). The statistical significance of our Bell violation does not appear to depend on the spacelike separation of events. For reference and comparison purposes only, the corresponding number of standard deviations for a given p-value (for a one-sided normal distribution) are shown.

$$|\psi\rangle = 0.961\,|H_A H_B\rangle + 0.276\,|V_A V_B\rangle, \tag{5.2}$$

where H (V) denotes horizontal (vertical) polarization, and A and B correspond to Alice's and Bob's photons, respectively. From the simulation we also determine that Alice's optimal polarization measurement angles, relative to a vertical polarizer, are $\{a = 4.2^o, a' = -25.9^o\}$ while Bob's are $\{b = -4.2^o, b' = 25.9^o\}$.

Synchronization signals enable Alice and Bob to define trials based only on local information. The synchronization signal runs at a frequency of $99.1\,\mathrm{kHz}$, allowing Alice and Bob to perform 99,100 trials/s ($79.3\,\mathrm{MHz}/800$). This trial frequency is limited by the rate the Pockels cells can be stably driven. When the Pockels cells are triggered they stay on for $\approx 200\,\mathrm{ns}$. This is more than 15 times longer than the $12.6\,\mathrm{ns}$ pulse-to-pulse separation of the pump laser. Therefore, photons generated by the source can arrive in one of 15 slots while both Alice's and Bob's Pockels cells are on. Since the majority of the photon pulses arriving in these 15 slots satisfy the spacelike separation constraints, it is possible to aggregate multiple adjacent pulses to increase the event rate and statistical significance of the Bell violation. However, including too many pulses will cause one or more of the spacelike separation constraints to be violated. Because the probability per pulse of generating an entangled photon pair is so low, given that one pair has already arrived, the chance of getting a second event in the same Pockels cell window is negligible ($< 1\,\%$).

Alice and Bob each have three different sources of random bits that they XOR together to produce their random measurement decisions. The first source, produced by the research group of Prof. Morgan Mitchell, is based on measuring optical phase diffusion in a gain- switched laser that is alternatively driven above then below the lasing threshold. A new bit is produced every $5\,\mathrm{ns}$ by comparing adjacent laser pulses [60]. Each bit is then XORed with all past bits that have been produced. The second source is based on the detection of light attenuated to low mean-photon numbers. By sampling the output of a detector over an interval of time in which the optical-state and vacuum-state contributions are balanced, the detection probability can be set to 0.5 and used for random bit generation. This source, developed by a member of the Kwiat research group, Michael Wayne, produces a bit on demand and is triggered by the synchronization signal. Finally, Alice and Bob each have a different predetermined pseudorandom source, created by Kwiat group members Joseph Chapman and Malhar Jere that is composed of various popular culture movies (such as the movie *Back to the Future*) and TV shows, as well as the digits of π, XORed together[7]. It is important to remember

---

[7]To generate the file for Alice, we XORed the binary data string of "Back to the Future 1", "Back to the Future 3" (which was used in reverse order), a concatenation of episodes of "Saved by the Bell", and a concatenation of 1×10^9 digits of π after applying a modulo 2 operation with "Monty Python and the Holy Grail". For Bob, we XORed episodes of "Doctor Who" (which were all concatenated together), "Back to the Future 2", and the concatenation of *Leonard Nimoy: Star Trek Memories* with episodes of "Star Trek" (this complete file was used in reverse order). More details, e.g., tests of randomness, are in [28].

that the randomness of the measurement decision is to validate the independence assumption between the photons and the setting choices. One could image a "quantum field" which influences both quantum random number generators, as well as the entanglement source; however, it is hard to image this quantum field also influencing movies produced 20 years prior to our measurement [61]! Now, suppose that a local-realistic system with the goal of producing a violation of the Bell inequality was able to manipulate the properties of the photons emitted by the entanglement source before each trial. Provided that the randomness sources correctly extract their bits from the underlying processes of phase diffusion, optical amplitude sampling, and the production of cultural artifacts, this powerful local realistic system would be required to predict the outcomes of all of these processes well in advance of the beginning of each trial to achieve its goal. Such a model would have elements of superdeterminism—the fundamentally untestable idea that all events in the universe are preordained.

5.3 Loophole-free Bell test data

Over the course of two days we took a total of 6 data runs with differing configurations of the experimental setup [62]. Here we report the results from the final dataset that recorded data for 30 minutes (see Ref. [28] for descriptions and results from all datasets). This is the dataset where the experiment was most stable and best aligned; small changes in coupling efficiency and the stability of the Pockels cells can lead to large changes in the observed violation. The events corresponding to the sixth pulse out of the 15 possible pulses per trial are the farthest outside all the relevant lightcones. Thus, we say these events are the most spacelike separated. To increase our data rate we aggregate multiple pulses centered around pulse number 6. We consider different Bell tests using a single pulse (number 6), three pulses (pulses 5, 6, and 7), five pulses (pulses 4 through 8), and seven pulses (pulses 3 through 9). The joint measurement outcomes and corresponding p-values for these combinations are shown in Table 5.1. For a single pulse we measure a p-value $= 2.5 \times 10^{-3}$, for three pulses a p-value $= 2.4 \times 10^{-6}$, for five pulses a p-value $= 5.8 \times 10^{-9}$, and for seven pulses a p-value $= 2.0 \times 10^{-7}$, corresponding to a strong violation of local realism. Surprisingly, aggregating 7 pulses produced worse data than aggregating 5 pulses, this is most likely due to instability in the applied voltage of the Pockels cell.

Satisfying the spacetime separations constraints in Fig. 5.3 requires precise measurements of the locations of Alice, Bob, and the source, as well as the timing of all events. Using a combination of position measurements from a global positioning system (GPS) receiver and site surveying, we determine the locations of Alice, Bob, and the source with an uncertainty of $< 1\,\mathrm{m}$. This uncertainty is set by the physical

Table 5.1: P-value results for different numbers of aggregate pulses. Here $C(11 \mid 00)$ refers to the number of times Alice and Bob both detect a photon with settings $x = 0$ and $y = 0$, respectively. Before analyzing the data a stopping criteria, N_{stop}, was chosen. This stopping criteria refers to the total number of events considered that have the settings and outcomes specified by the terms in Eq. (5.1), $N_{\text{stop}} = C(11 \mid 00) + C(10 \mid 01) + C(01 \mid 10) + C(11 \mid 11)$. As data set presented here was a 30 minute long measurement, we chose N_{stop} to be the number of expected events in a 27 minute long experiment, to ensure that the data set would reach the stopping criteria (if we do not reach N_{stop}, we can not use the data set). After this number of trials the p-value is computed and the remaining trials discarded. Such pre-determined stopping criteria are necessary for the hypothesis test we use. The total trials include all trials up to the stopping criteria regardless of whether a photon is detected, and is used to normalize the Bell parameter, $\mathcal{S}_{\text{CH}} = p(11|00) - p(10|01) - p(01|10) - p(11|11)$. The time difference between Bob finishing his measurement and the earliest time at which information about Alice's measurement choice could arrive at Bob sets the margin of timing error that can be tolerated and still have all events guaranteed to be spacelike separated.

Aggregate Pulses	$C(11 \mid 00)$	N_{stop}	Total trials	P-value	$\mathcal{S}_{\text{CH}}$	Timing Margin (ns)
1	1257	2376	175,654,992	2.5×10^{-3}	3.14×10^{-6}	63.5 ± 3.7
3	3800	7211	175,744,824	2.4×10^{-6}	8.85×10^{-6}	50.9 ± 3.7
5	6378	12127	177,358,351	5.9×10^{-9}	1.42×10^{-5}	38.3 ± 3.7
7	8820	16979	177,797,650	2.0×10^{-7}	1.49×10^{-5}	25.7 ± 3.7

size of the cryostat used to house our detectors and the uncertainty in the GPS coordinates. There are four events that must be spacelike separated: Alice's and Bob's measurement choice must be fixed before any signal emanating from the photon creation event could arrive at their locations, and Alice and Bob must finish their measurements before information from the other party's measurement choice could reach them. Due to the slight asymmetry in the locations of Alice, Bob, and the source, the time difference between Bob finishing his measurement and information possibly arriving about Alice's measurement choice is always shorter than the time differences of the other three events, as shown in Fig. 5.3(b). This time difference serves as a kind of margin; our system can tolerate timing errors as large as this margin and still have all events remain spacelike separated. For one, three, five, and seven aggregate pulses this corresponds to a margin of 63.5 ± 3.7 ns, 50.9 ± 3.7 ns, 38.3 ± 3.7 ns, and 25.7 ± 3.7 ns, respectively as shown in Table 5.1. The uncertainty in these timing measurements is dominated by the $1\,\text{m}$ positional uncertainty (see [28] for further details on the timing measurements).

A way to visualize and further quantify the the spacelike separation of events is to compute how far Alice, Bob, and the source could move from their measured positions and still be guaranteed to satisfy the locality constraints, assuming that the chronology of all events remains fixed. In Fig. 5.4(a) Alice, Bob, and the source locations are surrounded by shaded green regions. As long as each party remains anywhere inside the boundaries of these regions their events are guaranteed to be spacelike separated. There are specific configurations where all three parties can be outside the boundaries and still be spacelike separated, but here we consider the most conspiratorial case where all parties can collude with one another. The boundaries are

overlayed on architectural drawings of the building in which the experiment was performed. Four different boundaries are plotted, corresponding to the Bell test performed with one, three, five, and seven aggregate pulses. Minimizing over the path of each boundary line, the minimum distance that Alice, Bob, and the source are located from their respective boundaries is $9.2\,\mathrm{m}$, $7.3\,\mathrm{m}$, $5.4\,\mathrm{m}$, and $3.5\,\mathrm{m}$ for aggregates of one pulse, three pulses, five pulses, and seven pulses, respectively. For these pulse configurations we would have had to place our source and detection systems in different rooms than the rooms that they actually occupied (or even move outside of the building) to compromise our spacelike separation. Aggregating more than seven pulses leads to boundaries that are less than three meters away from the measured positions. In these cases we are not able to make strong claims about the spacelike separation of our events.

Finally, as shown in Fig. 5.4(b), we can compute the 15 p-values for each of the time slots we consider that photons from the source can arrive in every trial. Photons arriving in slots 2 through 11 are spacelike separated while photons in slots 12 through 15 are not. The photons arriving in these later slots are measured after information from the other party's random number generator could have arrived, as shown in Fig. 5.3(b). It appears that spacelike separation has no discernible effect on the statistical significance of the violation. However, we do see large slot-to-slot fluctuation in the calculated p-values. We suspect that this is due to instability in the applied voltage when the Pockels cell is turned on; consequently, photons receive slightly different polarization rotations depending on which slot they arrive in, leading to non-ideal measurement settings at Alice and Bob. It is because of this slot-to-slot variation that the aggregate of seven pulses has a computed p-value larger than the five-pulse case. Fixing this instability and using more sophisticated hypothesis test techniques [63, 32, 37] will enable us to robustly increase the statistical significance of our violation for the seven pulse case.

5.4   Concluding remarks

The experiment reported here is a commissioning run of a Bell test machine to certify randomness. The ability to include multiple pulses in our Bell test highlights the flexibility of our system. Our machine is capable of high event rates, making it well suited for generating random numbers required by cryptographic applications [41]. Currently, the amount of data collected is insufficient for randomness extraction. Future work will focus on improving the system to increase the rate of violation, with an end goal of incorporating the Bell test machine as an additional source of real-time randomness into the National Institute of Standards and Technology's public random number beacon (https://beacon.nist.gov).

It has been 51 years since John Bell formulated his test of local realism. In that time his inequality

has shaped our understanding of entanglement and quantum correlations, led to the quantum information revolution, and transformed the study of quantum foundations. Until very recently it has been impossible to carry out a complete and statistically significant loophole-free Bell test. Using advances in random number generation, photon source development, and high-efficiency single-photon detectors, we are able to observe a strong violation of a Bell inequality that is loophole free, meaning that we only need to make a minimal set of assumptions. These assumptions are that our measurements of locations and times of events are reliable, that Alice's and Bob's measurement outcomes are fixed at the time taggers, and that during any given trial the random number generators at Alice and Bob are physically independent of each other and the properties of the photons being measured. It is impossible, even in principle, to eliminate a form of these assumptions in any Bell test. Under these assumptions, as stated by John Bell, "if a hidden variable theory is local it does not agree with our results, and if it agrees with our results then it is not local" [64].

Chapter 6

Exploring Nonlocality

Though the experiments described in Chapter 5 effectively spell the end for local realism, we are just beginning to explore the nonlocality that quantum mechanics offers us. Nonlocality is arguably among the most counter-intuitive phenomena predicted by quantum theory. In recent years, the development of an abstract theory of nonlocality has brought a much deeper understanding of the subject, revealing a rich and complex phenomenon.

In this chapter, we present a systematic experimental exploration of the limits of quantum nonlocality. Using the polarization-entanglement source from Chapter 4, we explore the boundary of quantum correlations, demonstrate the counter-intuitive effect of more nonlocality with less entanglement, present the most nonlocal correlations ever reported, and achieve quantum correlations requiring the use of complex qubits. All our results are in remarkable agreement with quantum predictions, and thus represent a thorough test of quantum theory. Pursuing these experiments is nevertheless highly desirable, as any deviation may provide evidence of new physics beyond the quantum model. The work presented here is published in Ref. [65][1].

6.1 Introduction

Distant observers sharing a well-prepared entangled state can establish correlations which cannot be explained by any theory compatible with a natural notion of locality, as witnessed via a suitable Bell inequality violation [4]. Once viewed as marginal, nonlocality is today considered as one of the most fundamental aspects of quantum theory [66, 67], and represents a powerful resource in quantum information science, in particular in the context of the device-independent[2] approach [68, 33, 69].

From a more abstract perspective, the recent years have been marked by several developments providing a much deeper understanding of the phenomenon of quantum nonlocality. A generalized theory of nonlocality

---

[1]This article is available under the terms of the Creative Commons Attribution 3.0 License. Published by the American Physical Society. Used with permission.

[2]Previously we discussed DIRE as a device-independent approach to randomness expansion. The notion of device-independence is a protocol for the circumstance in which you do not trust the manufacturers of your equipment. A loophole-free Bell test, where we supply the measurement settings, can certify that the devices are truly random, regardless of who manufactured them. These device-independent protocols are not limited to only randomness, but also extend to other protocols, such as quantum key distribution. They do, however, always require a loophole-free Bell test.

was developed [36, 70, 66], aimed at characterizing correlations satisfying the no-signaling principle (hence not in direct conflict with relativity). Importantly, there exist no-signaling correlations—the most notable example being the highly nonlocal box of Popescu-Rohrlich [36]—which are stronger than any correlations realizable in quantum theory. Characterizing the boundary of quantum correlations (i.e., the separation from more general no-signaling correlations) is an important area of research today [13, 71, 72]. Intense theoretical research effort is also devoted to explain why super-quantum correlations are unlikely to exist in nature [73, 74, 75, 76, 77, 78, 79]. These ideas have clear potential to deepen our understanding of the foundations of quantum theory (see e.g., [80] for a recent review), and may give a first glimpse of physics beyond the quantum model [81].

Another fundamental issue is the relationship between entanglement and nonlocality. While early work showed that the two concepts are genuinely different, it was shown recently that entanglement and nonlocality are in fact not even monotonically related. Hence, it is possible to obtain 'more nonlocality with less entanglement' not only at the qubit level [20], but also when there is no restriction on the size of the entanglement [82, 83, 84], a strikingly counter-intuitive effect[3]. In particular, there exist portions of the boundary of quantum correlations which can only be accessed using weakly entangled states [82, 83], i.e., these correlations are provably impossible to reach using maximally entangled states of whatever Hilbert space dimension, even with perfect detectors. Perhaps even more surprisingly, very weakly entangled states can lead to strongly nonlocal correlations, stronger still than the highly nonlocal PR box [85].

Although the above ideas triggered considerable attention from the theoretical community, they remain essentially unexplored at the experimental level. Indeed, most Bell experiments performed so far [18, 86, 29, 25, 26, 27] focus on the simplest (and most famous) CHSH Bell inequality [12]; only few exploratory works considered Bell tests in the multipartite setting [87, 88, 89], or for high-dimensional systems [90, 91, 92].

The goal of the work presented in this chapter is to start a systematic experimental exploration of the limits of quantum nonlocality. Using a high-quality entangled-photon source, we perform a wide range of Bell tests. In particular, we probe the boundary of quantum correlations in the simplest Bell scenario. We demonstrate the phenomenon of more nonlocality with less entanglement; specifically, using weakly entangled states, we observe (i) nonlocal correlations which could provably not have been obtained from *any finite-dimensional* maximally entangled state, and (ii) nonlocal correlations which could not have been obtained using a single PR box. Moreover, we observe the most nonlocal correlations ever reported, i.e., featuring

---

[3]We saw a hint of this feature when we used non-maximally entangled states to reduce the required efficiency to avoid the detection loophole. However, one could argue that using non-maximally entangled states might be a relic of our inability to produce and detect the entangled state, and non-maximally entangled states are simply more forgiving when measuring nonlocality; perhaps we only observe more nonlocality with less entanglement because we struggle to observe any nonlocality at all. For the Bell tests discussed in this chapter, even with perfect detectors we would observe more nonlocality with a single non-maximally entangled state than with a maximally entangled state of any finite dimension — a truly bizarre result.

the smallest local content [93], and provide the strongest bounds to date on the outcome predictability in a general class of physical theories consistent with the no-signaling principle [94]. Finally, we observe nonlocal correlations which certify the use of complex qubit measurements [95].

However, to demonstrate the desired tests of nonlocality, one requires exceptionally high system qualities (some of the measurements considered here require system efficiencies > 99%), so it is useful to once again consider the fair-sampling and no-signaling assumptions. For example, fair-sampling is invalid if there is measurement-setting dependent loss, which we can verify. Additionally, to not violate a fair-sampling assumption, we must make repeated measurements in a random order to avoid periodic system noise (such as laser power drift). Thus, by verifying there is no measurement-dependent loss, by making repeated measurements in random orders, and by demonstrating there are no conspiratorial hidden-variable models that exploit the fair sampling assumption (which has been shown by the loophole-free Bell test), we can make a fair-sampling assumption to reduce the technological constraints[4]; furthermore, to reduce constraints on system design, we can also make a no-signaling assumption, that is, we no longer spacelike separate the relevant events and instead assume that the photons are not signaling each other or controlling the measurement settings. Again, this assumption should be valid, given the disproof of conspiratorial hidden-variable models by the loophole-free Bell test.

The results described below provide the most comprehensive experimental study of quantum nonlocality performed so far. All our results are in remarkable agreement with quantum predictions. Nevertheless, we believe that pursuing such a large scale exploration is of prime importance. These tests provide stringent verifications of quantum predictions, as any deviation could indicate new physics beyond quantum theory.

6.2 Concepts and notations

Recall from Chapter 2 that a distribution is said to be local if it admits a decomposition of the form

$$p(a,b|x,y) = \int d\lambda \, q(\lambda) p(a|x,\lambda) p(b|y,\lambda). \tag{6.1}$$

For a given number of settings and outcomes the set of local distributions forms a polytope $\mathcal{L}$, the facets of which correspond to Bell inequalities [66]. These inequalities can be written as

$$\mathcal{S} = \sum_{a,b,x,y} \beta_{a,b,x,y} p(a,b|x,y) \overset{\mathcal{L}}{\leq} L, \tag{6.2}$$

---

[4]There are still ways that the fair-sampling assumption can be invalid, such as detector saturation. However, in our case, we believe that the most likely issue with the fair-sampling assumption is a wedge on the waveplates
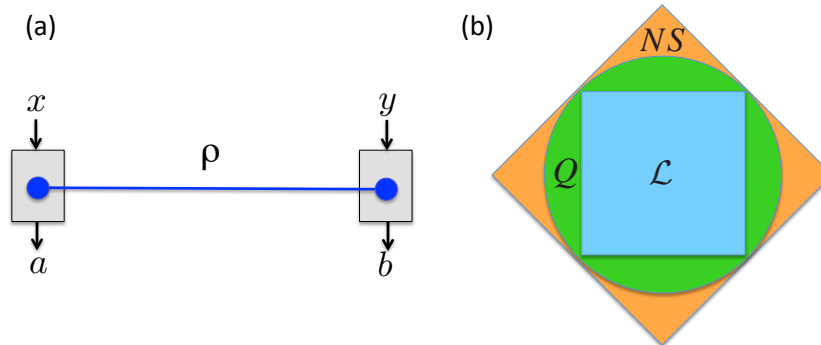
Figure 6.1: (a) Bell test scenario. Alice and Bob perform "black box" measurements on a shared (quantum) state ρ. The experiment is characterized by the data $\{p(a,b|x,y)\}$, i.e., a set of conditional probabilities for each pair of measurement outputs (a and b) given measurement settings x and y. Based on the data $p(a,b|x,y)$, Bell inequalities (see Eq. (2.3)) can be tested. (b) Geometrical representation of non-signaling correlations. The set of local ($\mathcal{L}$), quantum ($\mathcal{Q}$), and non-signaling ($\mathcal{NS}$) distributions are projected onto a plane, where the following inclusion relations are clear: $\mathcal{L} \subset \mathcal{Q} \subset \mathcal{NS}$.

where $\beta_{a,b,x,y}$ are integer coefficients, and L denotes the local bound of the inequality—the maximum of the quantity S over distributions from $\mathcal{L}$, i.e., of the form (6.1).

By performing judiciously chosen local measurements on an entangled quantum state, one can obtain distributions of the form

$$p(a,b|x,y) = \mathrm{Tr}(\rho M_{a|x} \otimes M_{b|y}), \tag{6.3}$$

which violate one (or more) Bell inequalities, and hence do not admit a decomposition of the form (6.1). Therefore, the set of quantum correlations $\mathcal{Q}$, i.e., those admitting a decomposition of the form (6.3), is strictly larger than the local set $\mathcal{L}$. Characterizing the quantum set $\mathcal{Q}$, or equivalently the limits of quantum nonlocality, turns out to be a hard problem [71, 72]. In their seminal work, Popescu and Rohrlich [36] asked whether the principle of no-signaling (or relativistic causality) could be used to derive the limits of $\mathcal{Q}$ and surprisingly found this not to be the case. Specifically, they proved the possible existence of no-signaling correlations — the so-called "PR box" correlations — which are not achievable in quantum theory. Therefore, the set of no-signaling correlations, denoted by $\mathcal{NS}$, is strictly larger than $\mathcal{Q}$, and we get the relation $\mathcal{L} \subset \mathcal{Q} \subset \mathcal{NS}$ (see Fig. 1(b)). The study and characterization of the boundary between $\mathcal{Q}$ and $\mathcal{NS}$ is today a hot topic of research [80]. A central question is whether the limits of quantum nonlocality could be recovered from a simple physical principle (i.e., is it possible to derive quantum mechanics from just causality and another axiom).

6.3 PR boxes

An important concept for this chapter is the notion of no-signaling theories and PR boxes. The idea is to use the framework of Bell inequalities to analyze the results of measurements without being constrained by assumptions of quantum mechanics. To begin, it is useful to consider the CHSH Bell inequality,

$$\mathcal{S}_{\mathrm{CHSH}} = E_{0,0} + E_{0,1} + E_{1,0} - E_{1,1} \overset{\mathcal{L}}{\leq} 2. \tag{6.4}$$

Now, the algebraic bound of this inequality is 4, which leads one to question what sorts of theories could achieve $\mathcal{S}_{\mathrm{CHSH}} = 4$. The answer, surprisingly, is a theory only limited by no-signaling (i.e., relativistic causality). Instead of using entangled states which are partially correlated (or anti-correlated) depending on the basis chosen, a no-signaling theory can use PR boxes to reach this bound. A PR box is an object whose correlations obey

$$xy = a \oplus b. \tag{6.5}$$

In essence, the outputs will be correlated so long as either party chose measurement input 0, and anti-correlated if both parties choose measurement setting 1. Putting these correlations into Eq. 6.4 results in $\mathcal{S}_{\mathrm{CHSH}} = 4$. As a final note, as these PR boxes obey no-signaling, it must be the case that the outputs (a, b) must be completely random; if they are not, then one could use PR boxes to send signals, in direct violation of the axiom of the theory. In the context of considering "super-quantum" correlations, we consider PR boxes as our fundamental unit of nonlocality [96].

6.4 Experimental setup

Here, we experimentally explore the limits of quantum nonlocality using the high quality source of entangled photons from Chapter 4. To briefly recap, the entanglement source consists of a 355-nm pulsed laser focused onto two orthogonal nonlinear BiBO crystals to produce polarization-entangled photon pairs at 710 nm, via spontaneous parametric down-conversion: the first (second) crystal has an amplitude to create horizontal (vertical) polarized photon pairs, which interfere to produce the entangled state [43] (see Fig. 6.2). Using wave plates to control the polarization of the pump beam, we create polarization entangled states with arbitrary degree of entanglement

$$|\psi_\theta\rangle = \cos\theta \, |H, H\rangle + \sin\theta \, |V, V\rangle. \tag{6.6}$$

In addition to the ability to precisely tune the entanglement of the source, which is crucial for many of the Bell tests we perform, we also achieve extremely high state quality. To do so, we pre-compensate the temporal decoherence from group-velocity dispersion in the down-conversion crystals with a BBO crystal [45], resulting in an interference visibility of 0.9970 ± 0.0005 in all bases. The high state quality (along with the capability of creating a state with nearly any degree of entanglement) allows us to make measurements very close to the quantum mechanical bound in a large array of different Bell tests.
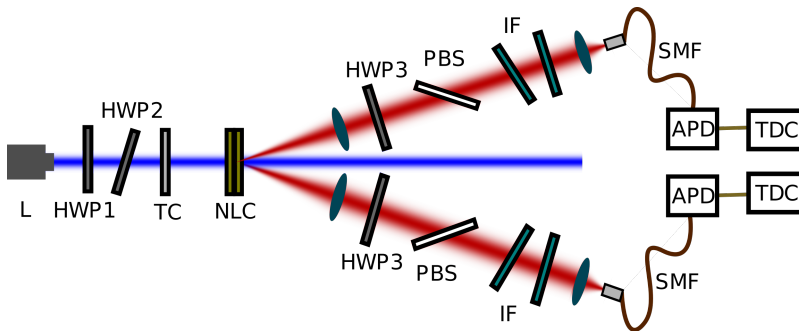
Figure 6.2: A diagram of the entanglement source. The high-power laser (L) is prepared in a specific polarization state (depending on the Bell test) by two half-wave plates (HWP1 and HWP2). We pre-compensate for the temporal decoherence (arising from the group velocity dispersion in the downconversion crystals) by passing the laser through a crystal (TC) designed to have the opposite group velocity dispersion. Passing the pump through a pair of nonlinear crystals (NLC), orthogonal to each other, produces the entangled photons. The measurements are performed using a motorized half-wave plate (HWP3) and a polarizing beam splitter (PBS). We then spectrally filter (IF) the photons to limit the collected bandwidth to 20 nm, as well as spatially filter the photons using a single-mode fiber (SMF) to remove any spatial decoherence. Finally, the photons are detected using avalanche photodiodes (APD), the events of which are recorded on a time-to-digital converter (TDC) and saved on a computer for analysis.

For the Bell tests, the local polarization measurements are implemented using a fixed Brewster-angle polarizing beam splitter, preceded by an adjustable half-wave plate, and followed by single-photon detectors to detect the transmitted photons. This allows for the implementation of arbitrary projective measurements of the polarization, represented by operators $A = \vec{\alpha} \cdot \vec{\sigma}$ and $B = \vec{\beta} \cdot \vec{\sigma}$, where $\vec{\alpha}$ and $\vec{\beta}$ are the Bloch vectors and $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ denotes the vector of Pauli matrices. Measurement outcomes are denoted by $a = \{0, 1\}$ and $b = \{0, 1\}$, where in our experiments the 0 outcome is measured by projecting onto the orthogonal polarization (since we are assuming fair-sampling, we require a detection event to occur to assign an outcome). To remove any potential systematic loopholes (e.g., seemingly better results due to laser power fluctuations), we measure each Bell inequality multiple times, where the measurements settings are applied in a different randomized order each time. Finally, to ensure the validity of the results, we do not perform any post-processing of the data (e.g., accidental subtraction).

6.5 Experiments and results

We start our investigation by considering the simplest Bell scenario, featuring two binary measurements each for Alice and Bob. The set of local correlations, i.e., of the form (6.1), is fully captured by the CHSH inequality [12]:

$$\mathcal{S}_{\text{CHSH}} = E_{00} + E_{01} + E_{10} - E_{11} \overset{\mathcal{L}}{\leq} 2, \tag{6.7}$$

where $E_{xy} \equiv p(a = b|x, y) - p(a \neq b|x, y)$ denotes the correlation function. Quantum correlations can violate the above inequality up to $\mathcal{S}_{\text{CHSH}} = 2\sqrt{2}$, the so-called Tsirelson bound [13]. More generally, quantum correlations must also satisfy the following family of inequalities

$$\mathcal{S}_{\text{CHSH}} \cos\theta + \mathcal{S}'_{\text{CHSH}} \sin\theta \overset{\mathcal{Q}}{\leq} 2\sqrt{2}, \tag{6.8}$$

parametrized by $\theta \in [0, 2\pi]$, and where $\mathcal{S}'_{\text{CHSH}} = -E_{00} + E_{01} + E_{10} + E_{11}$ is a different representation (or symmetry) of the CHSH expression (the difference is in which term is negated). The above quantum Bell inequalities are tight, in the sense that quantum correlations can achieve $2\sqrt{2}$ for any $\theta \in [0, 2\pi]$; specifically, inequality (6.8) can be saturated by performing appropriate local measurements on a maximally entangled state $|\psi_{\pi/4}\rangle$ (see Ref. [65] for more details). Therefore, the set of quantum correlations $\mathcal{Q}$ forms a circle (of radius $2\sqrt{2}$) in the plane defined by $\mathcal{S}_{\text{CHSH}}$ and $\mathcal{S}'_{\text{CHSH}}$. Fig. 6.3 presents the experimental results which confirm these theoretical predictions with high accuracy. To make the measurements, we kept the entangled state fixed and varied the settings for 180 different values of θ. The average radius of our measurements was 2.817, with 8 data points falling beyond the limit of quantum mechanics.

To calculate the likelihood, given our measurements, of having the observed 8 (or more) events beyond the quantum limit, first we considered our expected radius to be the mean of all measured radii, calculated to be 2.817. Next, we assume the calculated error bars (based on Poissonian statistics) to correspond to one σ from the mean (that is, each data point comes from a Gaussian distribution with a mean of the mean radius, and a standard deviation given by the error bar). From here, we can compute the probability of each data point being measured greater than $2\sqrt{2}$ (the probability of the ith data point being measured beyond the quantum bound is p_i). We then assume that the probability distribution of seeing k events beyond the quantum limit follows a Gaussian distribution with mean $\mu = \sum_i p_i = 8.677$ and variance of $\sigma^2 = \sum_i p_i(1 - p_i) = 7.928$. This distribution then determines the probability of seeing 8 or more events to be 0.66, i.e., there is no statistically significant conflict with quantum predictions.
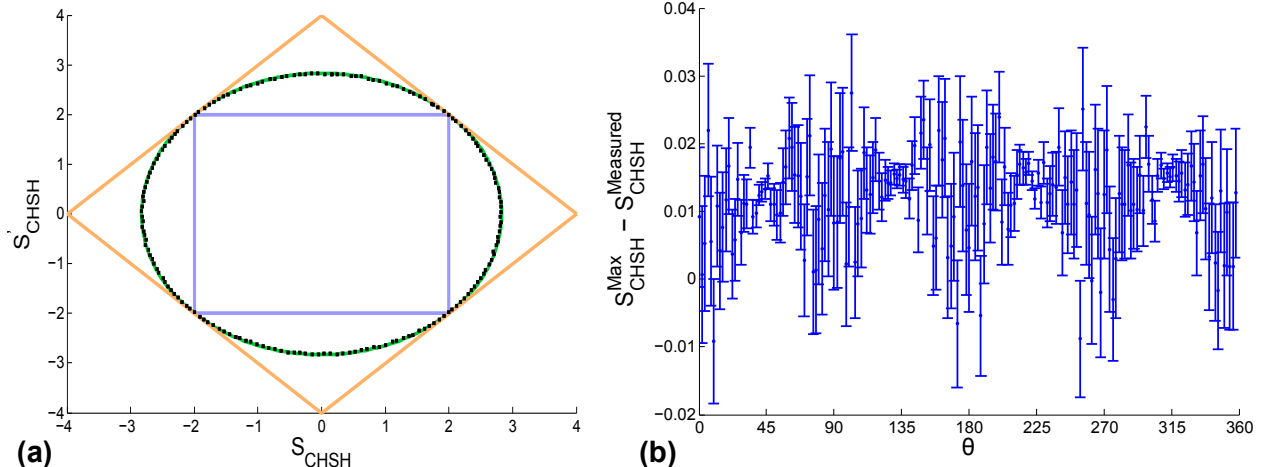
Figure 6.3: Testing the boundary of quantum correlations in the simplest Bell test scenario. (a) Plot of the experimental measurements of the curve Eq. (6.8). Here, local correlations ($\mathcal{L}$) form the inner blue square, no-signaling correlations form the outer orange square (the vertices represent the PR-box and its symmetries), and quantum correlations form the green circle; the black dots are the 180 measured data points, all of whose error bars lie within the thickness of the dot. (b) A comparison of the analyzed data with the quantum mechanical maximum ($2\sqrt{2}$). Here, θ is defined in Eq. (6.8), and corresponds to rotating around the circle in Fig. 6.3(a). The vertical axis is the distance from $2\sqrt{2}$ of the radius of curve formed by $\mathcal{S}'_{\text{CHSH}}$ and $\mathcal{S}_{\text{CHSH}}$ (i.e., the radius of the data point at a given θ). Plotted values greater than zero correspond to measured values less than $2\sqrt{2}$; negative values could correspond to results beyond the quantum limit. Here, the uncertainty is smaller for measurements around the vertices of the local correlations (at $45°$, $135°$, $225°$, and $315°$) due to large correlations between $\mathcal{S}'_{\text{CHSH}}$ and $\mathcal{S}_{\text{CHSH}}$. We observe 8 data points beyond the quantum limit, which we expect to happen with a probability of 0.66.

6.5.1 Tilted Bell inequalities

It turns out, however, that the complete boundary of $\mathcal{Q}$ cannot be fully recovered by considering only maximally entangled states. That is, there exist sections of the no-signaling polytope where the quantum boundary can only be reached using partially entangled states. While studies [20, 97] on minimal detection efficiency required to close the detection loophole have already hinted at this for qubit entanglement, a more general observation with no restriction on the size of the entanglement has only been theoretically established recently in [83, 84, 82]. To demonstrate this bizarre phenomenon experimentally, we shall first strengthen the theoretical result obtain in [82].

Specifically, consider the projection plane defined by the parameters $\mathcal{S}_{\text{CHSH}}$ and $-E_0^A - E_0^B$, where $E_0^A \equiv \sum_{a=\{0,1\}} a\, p(a|x=0)$ denotes Alice's marginal (similarly for Bob's marginal E_0^B). In order to find the quantum boundary in this plane, we consider the family of Bell inequalities (called the tilted Bell inequalities)

$$\mathcal{S}_\tau = \mathcal{S}_{\text{CHSH}} + 2(1-\tau)[E_0^A + E_0^B] \overset{\mathcal{L}}{\leq} 2(2\tau - 1), \tag{6.9}$$

with $1 \leq \tau \leq 3/2$. For $\tau = 1$, we recover CHSH, while for $1 < \tau < 3/2$ the inequality has the peculiar feature

that the maximal quantum violation can *only* be obtained using partially entangled states [82]. Moreover, for $1/\sqrt{2} + 1/2 \leq \tau \leq \frac{3}{2}$, the inequality *cannot* even be violated using *any* finite-size maximally entangled state (see the red and blue curves in Fig. 6.4b) [65]. This illustrates the fact that weak entanglement can give rise to nonlocal correlations which cannot be reproduced using strong entanglement.

We achieved violations of the above inequalities (for several values of the parameter τ) extremely close to the theoretically predicted maximum, by adjusting the degree of entanglement and using the corresponding settings (see App. B; see Fig. 6.4. For instance, tuning our source to produce weakly entangled states, we obtain clear violation of the inequality $\mathcal{S}_{\tau=1.300} \overset{\mathcal{L}}{\leq} 3.2$, where we measure $\mathcal{S}_\tau = 3.258 \pm 0.002$, which is impossible using maximally entangled states. Our results thus clearly illustrate the phenomenon of 'more nonlocality with less entanglement' [83, 84, 82]. It is worth noting that, due to the small difference between the quantum maximum and the local bound $2(2\tau - 1)$, an experimental demonstration of the above phenomenon using inequality (6.9) is essentially impossible without extremely precise control of high quality weakly entangled states.

6.5.2 Chained Bell inequalities

In the remainder of the paper, we go beyond the CHSH scenario and consider Bell inequalities featuring $n > 2$ binary-outcome measurements per observer. This will allow us to investigate other aspects of the phenomenon of quantum nonlocality. We start by considering the family of chained Bell inequalities (see App. C for a discussion of this family of inequalities[5]) [98, 99]

$$I_n = \sum_{a,b=\{0,1\}} \left[p(a = b|n-1, 0) + p(a \neq b|n-1, n-1) + \sum_{x=0}^{n-2} \sum_{y=x}^{x+1} p(a \neq b|x, y) \right] \overset{\mathcal{L}}{\geq} 1, \qquad (6.10)$$

where $a = 1$ is the a detection for Alice at setting x, and $a = 0$ is a detection for Alice at setting $x^\perp$ (similarly for Bob), and $x(y)$ takes on values from 0 to $n - 1$. Using a maximally entangled state $|\psi_{\pi/4}\rangle$, quantum theory allows one to obtain values up to $I_n = n(1 - \cos \frac{\pi}{2n})$. Note that, as n increases, the quantum violation approaches the bound imposed by the no-signaling principle, namely $I_{n \to \infty} = 0$ (here given by the algebraic minimum of I_n). Using our setup we obtain violations of the chained inequality up to $n = 45$. Because I_n becomes increasingly sensitive to any noise in the system as n increases, we found the strongest violation at $n = 18$, with a value of $I_{18} = 0.126 \pm 0.001$, see Fig. 6.5. For comparison, the previous best measurement of

---

[5]The chained Bell inequality is a generalization of $\mathcal{S}_{\text{CHSH}}^{0011} = E_{00} + E_{01} + E_{10} - E_{11}$. The CHSH Bell inequality is I_2, and the next chained Bell inequality, I_{n+1}, is formed by adding $\mathcal{S}_{\text{CHSH}}^{n-1,n-1,n,n}$ to I_n. As an example, for $n = 3$, we add $\mathcal{S}_{\text{CHSH}}^{0011} + \mathcal{S}_{\text{CHSH}}^{1122} = E_{00} + E_{01} + E_{10} + E_{21} + E_{12} - E_{22}$, and $I_4 = I_3 + \mathcal{S}_{\text{CHSH}}^{2233}$. As each I_n will have a single negative term, and $2n - 1$ positive terms, we then want to make $2n - 1$ measurements with high correlations, and a single measurement with a large anti-correlation; we discuss this in more detail in App. C. The I_n in Eq. 6.10 is written in a form where the local bound is constant (at 1), and the nonlocal bound is also constant (at 0).
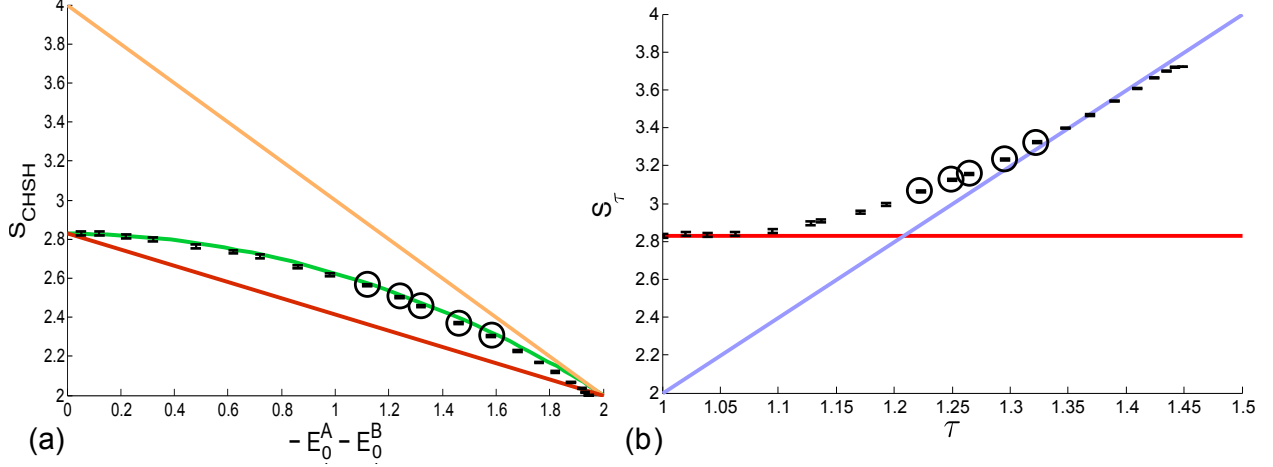
Figure 6.4: Testing the boundary of quantum correlations with the tilted Bell inequality, Eq. 6.9. (a) A plot of the measured values for a projection where the quantum boundary can only be attained using partially entangled states. The orange line is the boundary for no-signaling correlations (the PR box sitting at the top), and the red line is the boundary of the set of correlations achievable by a maximally entangled state [65], whereas the horizontal axis at $\mathcal{S}_{\text{CHSH}} = 2$ coincides with the boundary of the local set $\mathcal{L}$. The green curve represents the quantum boundary, with the black points corresponding to measured data points. For large values of $-E_0^A - E_0^B$ (corresponding to less entangled states), the system becomes increasingly sensitive to system noise (i.e., slight state-creation and measurement imperfections), resulting in the measured values deviating slightly from the quantum curve. (b) A plot of the measured values for the tilted Bell inequalities. The red line is the bound of maximally entangled sates, the blue line is the local bound, and the black points are the analyzed data. The red and blue lines cross at $1/\sqrt{2} + 1/2$, where maximally entangled states can no longer violate a tilted Bell inequality. Here, for the measured points up to $\tau = 1.323$, we see a value of $\mathcal{S}_\tau$ at least three standard deviations above the local bound; notably, we have violations for $\tau = 1.223, 1.250, 1.265, 1.296,$ and 1.323 (circled data points in both plots), as well as $\tau = 1.300$ (this data point is described in the text, but since it was taken separately from the other measurements, we do not place it in the plots), none of which are possible for maximally entangled states in any dimension, implying that with less entanglement, we have more nonlocality.

I_n was $I_7 = 0.324 \pm 0.0027$ [100].

These violations have interesting consequences. First, they allow us to put strong lower bounds on the nonlocal content of the observed statistics $\mathbf{p}_{obs} = \{p_{obs}(ab|xy)\}$. Following the approach of Ref. [93], we can write the decomposition

$$\mathbf{p}_{obs} = (1-q)\mathbf{p}_L + q\,\mathbf{p}_{NS}, \tag{6.11}$$

where $\mathbf{p}_L$ is a local distribution (inside $\mathcal{L}$) and $\mathbf{p}_{NS}$ is a no-signaling distribution (achieved, e.g., via PR boxes), and then minimize $q \in [0, 1]$ over any such decomposition. The minimal value q_{min} is then the nonlocal content of $\mathbf{p}_{obs}$, and can be viewed as a measure of nonlocality. That is, we can think of q_{min} as being the likelihood that some nonlocal resource (e.g., a PR box) would need to be used in order to replicate the results. For example, consider the maximum quantum mechanical violation of the CHSH Bell
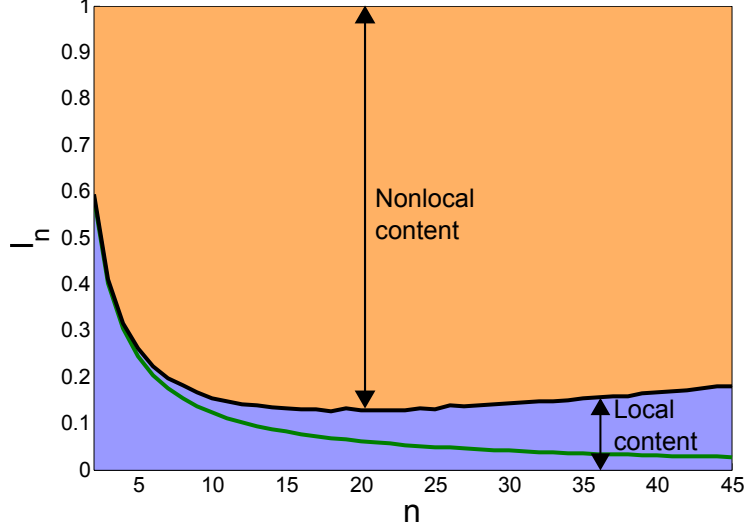
Figure 6.5: A plot of the measured chained Bell inequality values for $n = 2$ to $n = 45$. Here, the local limit is $I_n = 1$ and the no-signaling limit is $I_n = 0$. The quantum boundary in this case is the green line; our measured Bell inequality values are connected by the black line, with the error bars lying within the thickness of the line. The "local content" for a given n is represented by distance from 0 to the measured I_n value (black line), which is colored blue, and the nonlocal content is the distance from the measured value to 1, colored orange. As the value of I_n approaches 0, the correlations present in the system match those of a PR box—if $I_n = 0$ were measured, the system would require the use of a PR box for every measurement. Our measured points deviate from the quantum boundary due to the 0.3% noise from imperfect state preparation, which becomes more noticeable with larger number of measurements (e.g., I_{45} requires 360 specific measurements along the Bloch sphere, for the specific measurements see App. B).

inequality, $\mathcal{S}_{\text{CHSH}} = 2\sqrt{2}$. If instead of using entangled states, we were to use a local strategy ($\mathcal{S}_{\text{CHSH}} = 2$) and PR boxes ($\mathcal{S}_{\text{CHSH}} = 4$), we would only need to use PR boxes on (at most) 41% of the trials to recover the quantum mechanical bound ($0.41 \times 4 + 0.59 \times 2 \approx 2\sqrt{2}$), thus $q_{min} = 0.41$. For an observed violation of the chained inequality, we can place a lower bound on the nonlocal content: $q_{min} \geq 1 - I_n$ [101]. Notably, for the case $n = 18$, we obtain $q_{min} = 0.874 \pm 0.001$ which represents the most nonlocal correlations ever produced experimentally. For comparison, the previous best bound was $q_{min} = 0.782 \pm 0.014$ [102, 103].

Moreover, following the work of [94], we can place bounds on the outcome predictability in a general class of physical theories consistent with the no-signaling principle. While quantum theory predicts that the measurement results are fully random (e.g., one cannot predict locally which output port of the polarizing beam splitter each photon will be detected), there could be a super-quantum theory that could predict better than quantum theory (that is, with a probability of success strictly greater than $1/2$) in which port each photon will be detected. This predictive power, represented by the probability δ of correctly guessing the output port, can be upper bounded from the observed violation of the chained Bell inequality. In our experiments, the best bound is obtained for the case $n = 18$, for which we obtain $\delta = 0.5702 \pm 0.0005$ (that is, given any possible extension of quantum theory satisfying the free-choice assumption [100], the measurement

Table 6.1: A table of the measured values from two different Bell inequalities, M_{3322} and M_{4322}, as defined in Eq. (6.12). For these inequalities, correlations from $\mathcal{L}$ and those augmented with the use of a single PR-box (represented as $\mathcal{L} + 1PR$) give rise to the same bound. Any measured values above the corresponding bound imply that the data is not only incompatible with Bell-locality, but also with a single use of a PR box. Instead, two PR boxes must be used to replicate the data. The approximate quantum mechanical maximums (obtained using the tools of [71, 72] and [104, 105]) and the quantum mechanical maximums for two qubits are given as a reference.

Bell inequality	Measured value	2-qubit maximum	Quantum maximum
$M_{3322} \overset{\mathcal{L}+1PR}{\leq} 6$	6.016 ± 0.0003	6.024	6.130
$M_{4322} \overset{\mathcal{L}+1PR}{\leq} 7$	7.004 ± 0.0004	7.041	7.127

result could be guessed with a probability at most 57%)[6], which is the strongest experimental bound (closest to 50%) to date; the previous bound was $\delta = 0.6644 \pm 0.0014$ [100].

6.5.3 M-class Bell inequalities

The above results on the chained Bell inequality show that in order to reproduce the measured correlations, nonlocal resources (such as the PR box) must be used in more than 87% of the experimental rounds (i.e., $q_{min} 0.874 \pm 0.001$). While the chained Bell inequality provides an interesting metric of nonlocal content, there are, however, even more nonlocal correlations achievable using two-qubit entangled states, which can provably not be reproduced using a single PR box [85]. Interestingly, such correlations can arise only from partially entangled states, since maximally entangled states can always be perfectly simulated using a single PR box [106]. The accuracy of our experimental setup allows for the study of Bell inequalities which require the use of more than a single PR box. Specifically, consider the inequalities from Ref. [85] (for $n = 3$ and $n = 4$):

$$\mathcal{M}_{3322} \equiv E_{00} + E_{01} + E_{02} + E_{10} + E_{11} - E_{12} \tag{6.12a}$$

$$+ E_{20} - E_{21} - E_0^A - E_1^A - E_0^B + E_1^B \overset{\mathcal{L}+1PR}{\leq} 6,$$

$$\mathcal{M}_{4322} \equiv E_{00} + E_{01} + E_{02} + E_{10} - E_{12} + E_{13} \tag{6.12b}$$

$$+ E_{20} - E_{21} - E_{23} - E_0^A - E_1^A - E_2^A - E_0^B \overset{\mathcal{L}+1PR}{\leq} 7,$$

---

[6]This may seem contradictory to the early example of DIRE, where $p_{guess} = 1/2$ for a maximal quantum mechanical Bell violation. However, for DIRE we assumed quantum mechanics is correct. Here we only assume a no-signaling theory, so this is a much stronger bound since it does not depend on the validity of quantum mechanics. These results still apply to quantum mechanics, as quantum is a no-signaling theory, thus we can certify our randomness even if quantum mechanics is incorrect!

which cannot be violated by any local correlations supplemented by a single maximally nonlocal PR box ($\mathcal{L} + 1PR$), which is viewed as a unit of nonlocality. Nevertheless, by performing well-chosen measurements on a very weakly entangled state ($|\psi_{\approx 3\pi/7}\rangle$), we observed violations of the above inequalities (see Table 6.1). Note that since the observed statistics (leading to $\mathcal{M}_{3322} > 6$ and $\mathcal{M}_{4322} > 7$) could not have been obtained using a single PR box, they also cannot be obtained using a maximally entangled state $|\psi_{\pi/4}\rangle$, and required the use of a weakly entangled state (or two PR boxes). Hence, we provide a second experimental verification of the phenomenon of more nonlocality with less entanglement.

As these results are all nonintuitive, it is worthwhile to recap the nonlocal "hierarchy" measured from the previous Bell inequalities. To do so, we will define $\mathcal{PR}$ to correspond to a single PR box ($2\mathcal{PR}$ to correspond to two PR boxes where it is different from a single PR box), $\mathcal{M}$ to be the nonlocality of maximally entangled particles, $\mathcal{NM}$ for non-maximally entangled particles, and $\mathcal{L}$ for local correlations. Now, for the $\mathcal{S}_{\text{CHSH}}$ Bell inequality, with a noiseless system at unit system efficiency, then the nonlocality ordering is:

$$\mathcal{PR} \supset \mathcal{M} \supset \mathcal{NM} \supset \mathcal{L}. \tag{6.13}$$

However, when the system has reduced system efficiency ($< 83\%$) with no noise, then the order is slightly different:

$$\mathcal{PR} \supset \mathcal{NM} \supset \mathcal{M} \supset \mathcal{L}. \tag{6.14}$$

The above results also hold for the chained Bell inequalities. The hierarchy, however, for the M-class Bell inequalities described in this section is drastically different:

$$2\mathcal{PR} \supset \mathcal{NM} \supset \mathcal{PR} = \mathcal{M} = \mathcal{L}, \tag{6.15}$$

which shows the complicated landscape of the nonlocal structure. Depending on the measured Bell inequality, different states exhibit significantly different amounts of nonlocality.

6.5.4 Elegant Bell inequality

Finally, we consider a Bell inequality which can certify the use of complex qubits (versus real qubits) [95], given by

$$\mathcal{S}_E \equiv E_{00} + E_{10} + E_{20} + E_{01} - E_{11} - E_{21} \tag{6.16}$$

$$-E_{02} + E_{12} - E_{22} - E_{03} - E_{13} + E_{23} \overset{\mathcal{L}}{\leq} 6.$$

The optimal quantum violation is $\mathcal{S}_E = 4\sqrt{3} \simeq 6.928$, which can be obtained by using a maximally entangled two-qubit state $|\psi_{\pi/4}\rangle$, and a set of highly symmetric qubit measurements. The measurements of Alice are given by three orthogonal vectors on the sphere ($\vec{a}_0 = (1,0,0)$, $\vec{a}_1 = (0,1,0)$, and $\vec{a}_2 = (0,0,1)$), and Bob's measurements are given by the four vectors of the tetrahedron: $\vec{b}_0 = \frac{1}{\sqrt{3}}(1,1,1)$, $\vec{b}_1 = \frac{1}{\sqrt{3}}(1,-1,-1)$, $\vec{b}_2 = \frac{1}{\sqrt{3}}(-1,1,-1)$, $\vec{b}_3 = \frac{1}{\sqrt{3}}(-1,-1,1)$. To perform these measurements, we added an additional quarter-wave plate after HWP3 in Fig. 6.2 (see App. B for the waveplate values). Implementing this strategy experimentally, we observe a violation of $\mathcal{S}_E = 6.890 \pm 0.002$, close to the theoretical value. Interestingly, such a violation could not have been obtained using a real qubit strategy. Indeed, the use of measurement settings restricted to a great circle of the Bloch sphere, i.e., *real* qubit measurements, can only provide violations up to $\mathcal{S}_E = 2 + 2\sqrt{5} \simeq 6.472$ [95]. Thus, the observed violation requires the use of complex qubit measurements, spanning the Bloch sphere. Note, however, that any strategy involving complex qubit measurements can be mapped to an equivalent strategy involving two real qubits[7] [107, 108].

6.6 Conclusion

To summarize, we have reported the exploration of various facets of the rich phenomenon of quantum nonlocality. The results of our systematic experimental investigation of quantum nonlocal correlations are in extremely good agreement with quantum predictions; nevertheless, we believe that pursuing such tests is of significant value, as Bell inequalities are not only fundamental to quantum theory, but also can be used to discuss physics outside of the framework of quantum theory. By doing so, one can continue to place bounds on the features of theories beyond quantum mechanics, as we have here. Such continued experiments investigating the bounds of quantum theory are important, as any valid deviation from quantum predictions, e.g., by observing stronger correlations than predicted by quantum theory, would provide evidence of new

---

[7]A complex qubit is described by three parameters (to assign a position on the Bloch sphere). Two real qubits are described by four parameters (two parameters for each qubit to assign a position on a circular disk). One of the many ways to do this mapping would be to assign the first real qubit the parameters (a, b) and the second real qubit the parameters (c, d), then we can map this onto a complex qubit described by the three parameters (a, b, c), and simply ignoring parameter d.

physics beyond the quantum model. Furthermore, nonlocality has important applications for quantum information protocols, though the optimal way to quantify the nonlocality present in a system is still an open question (see, e.g., [109]). Here, we experimentally verified, for the first time, that for certain correlations from non-maximally entangled states, two PR boxes (i.e., two units of the nonlocal resource) are required to recreate the correlations from these weakly entangled states. A natural question then is if these systems could be used advantageously for certain quantum information tasks, e.g., perhaps there is some novel quantum cryptographic protocol relying on non-maximally entangled states to provide enhanced security against an eavesdropper.

Chapter 7

Conclusions

We have presented our research using a high quality source of entangled photons to close the detection-loophole for a Bell inequality. We also presented our work in performing a landmark experiment, a completely loophole-free Bell test, free from the usual assumptions made by all experiments until now; this test definitively rules out all local realistic theories. By demonstrating that nonlocality must exist, we can now attempt to quantify this nonlocality. Here, we presented experiments exploring different avenues of nonlocality, showing counter-intuitive features and placing the tightest bounds to date on the predictability of all no-signaling theories.

There are clear extensions to all of the experiments presented here. First, the next step past the loophole-free Bell test would be perform a loophole-free version of the chained-Bell inequality experiment presented in Chapter 6. As the high n chained-Bell inequalities have more nonlocal content, and therefore will provide more secure randomness for the device-independent randomness expansion protocol, the desire is to perform the experiment with the highest n possible. In this case, the detection efficiency requirement of I_n is

$$\epsilon_n \geq \epsilon_n^{\min} = \frac{2n-2}{2n-1}, \tag{7.1}$$

where $\epsilon_n^{\min}$ is the minimum (no noise) required efficiency[1] (see Fig. 7.1). With the source presented in Chapter 5, we should be able to violate up to I_4, which requires $\epsilon = 6/7 = 0.857$, if we do not spacelike separate all relevant events (which is not as critical for device-independent randomness expansion).

Additional interesting experiments could also be performed while making valid assumptions, as was done in Chapter 6. Here, we explored a very few of the massive set of possible inequalities. There are surely many more tests with interesting interpretations for future theories than we were aware of when we completed our experiments. Ideally one would perform these tests with high visibilities ($> 99.99\%$), as it will place stronger bounds on all no-signaling theories. Additionally, from a theoretical point of view, people will continue to search for "almost-quantum" theories (self-consistent theories that are not quantum mechanics, but could

---

[1]This formula was found through numerical simulation.
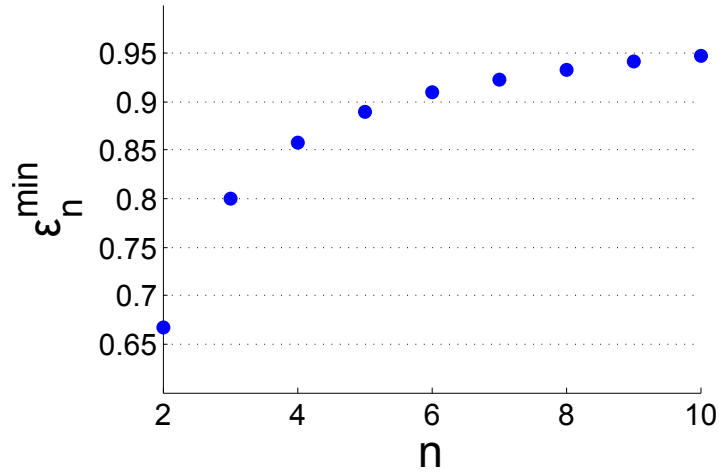
Figure 7.1: A plot of required efficiency to close the detection loophole for the chained Bell inequalities. As $n \to \infty$, the required efficiency also tends to ∞. With the system described in Chapter 5, we estimate a possible system efficiency (after some system changes) of around 0.86 to 0.87, which would be sufficient to violate a detection-loophole-free version of I_4.

still explain our observed results), and the difference between quantum mechanics and these new theories can be tested through a Bell-like inequality. Most likely these tests will require unparalleled quality sources (perhaps requiring visibilities on order of $1 - 10^{-10}$). Therefore, designing extremely high quality sources to pursue these tests will be absolutely critical for the possibility to find any deviation from quantum theory, such results could herald the next physical theory beyond quantum mechanics.

Appendix A

Optimal State and Settings Calculation

In this Appendix, we give a description of the method we use to calculate the optimal state and settings of the different Bell inequalities. To do so, we will describe the method for determining the optimal state and settings for the CH Bell inequality,

$$\mathcal{S}_{\text{CH}} = p(1,1|0,0) + p(1,1|0,1) + p(1,1|1,0) - p(1,1|1,1) - p(a=1|x=0) - p(b=1|y=0) \overset{\mathcal{L}}{\leq} 0, \quad \text{(A.1)}$$

and then describe the generalization to the more complex Bell inequalities.

For the CH Bell inequality, we model the system (given the estimated noise and efficiency) and perform a numerical maximization to determine the settings and state. That is, we optimize over the state parameter r (where $|\psi_r\rangle = (|HH\rangle + r|VV\rangle)/\sqrt{1+r^2}$), and over settings $(x=0) = \theta_0^A$, $(x=1) = \theta_1^A$, $(y=0) = \theta_0^B$, and $(y=1) = \theta_1^B$ setting choices. Note, however, that due to the symmetry of the CH Bell inequality and the state we produce, the settings have a symmetry about the $0°$ polarization setting. Thus, we assume that $\theta_0^A = -\theta_0^B = \theta_0$ and $\theta_1^A = -\theta_1^B = \theta_1$, which reduces the problem to an optimization over only 3 parameters (r, θ_0, θ_1).

As mentioned in Section 4.6.2, our state is actually best described as: $|\psi_r\rangle \propto |H+\delta, H-\delta\rangle + r|VV\rangle$, where $\delta \approx 3°$. In the H/V basis, this is then written as:

$$|\psi\rangle \propto (1 - r\sin^2\delta)|VV\rangle + r(\cos^2\delta|HH\rangle + \sin\delta\cos\delta|VH\rangle - \sin\delta\cos\delta|VH\rangle). \quad \text{(A.2)}$$

From the density matrix, $\rho = |\psi\rangle\langle\psi|$, we can calculate the probability of measuring a joint detection event (a joint outcome of 1) from such an entangled photon pair through polarizers at θ^A on Alice's side and θ^B on Bob's side via $p(1,1|\theta^A, \theta^B) = \text{tr}(\rho J)$, where J is the matrix representation of the joint polarization measurement.

We then determine Alice's (Bob's) marginal probability distribution, $p(1|\theta)$, which can be calculated with the reduced density matrix, $\rho_A = \text{tr}_B(\rho)$. Thus, $p(1|\theta) = tr(\rho_A S)$, where S is the matrix representation of the polarization measurement on Alice's side.

Now we need to include noise to the above terms, as well as the finite system efficiency. The three noise terms to include are: dark counts (here we define a dark count as any count that does not scale with the laser power, i.e., including room light background), fluorescence counts (counts that scale linearly with the pump power), and accidental counts (counts that scale quadratically with pump power). To consider dark counts, we modify the marginal distribution by adding in an additional term that is independent of pump power, $N \ p_m(1|\theta) = N \ \epsilon \ p(1|\theta) + dcr$, where N is a laser power parameter, $p_m(1|\theta)$ is the measured probability distribution with a polarizer at θ, ϵ is the system efficiency, and dcr is the measured number of counts with the laser blocked[1]. The parameter N can be interpreted as the number of pairs emitted from the source into our desired mode, where we can experimentally set N by noting that $N \ p_m(1|\theta) = S(1|\theta)$. To include fluorescence counts, we further modify the marginal distribution: $N \ p_m(1|\theta) = N \ \epsilon \ p(1|\theta) + dcr + N \ flr$. This term, while being the most critical, is also unfortunately the hardest to measure precisely (the measurement of this value is described in Section 4.3).

Finally, the accidental counts only affect the joint probability distribution. These counts are due to two photon events (e.g., Alice detects one photon from one pair, whereas Bob detects a photon from the other pair), which will degrade the polarization correlations. The probability distribution of photon pairs follows:

$$p_n = \frac{m^m \langle n \rangle^n \Gamma(m+n)}{(m + \langle n \rangle)^{m+n} \Gamma(m) \Gamma(1+n)}, \tag{A.3}$$

where n is the number of pairs in a trial, $\langle n \rangle$ is the mean number of pairs per trial, and m is the number of modes. If $m = 1$, this reduces to the usual thermal distribution, and if $m = \infty$, this follows Poissonian statistics. Here, due to the (typically) low probability of even a single pair being emitted from the source, we only consider accidentals from two-pair events. That is, we alter the joint detection distribution by $N \ p_m(1,1|\theta^A, \theta^B) = N \ \epsilon^2 \ p(1,1|\theta^A, \theta^B) + acc$, where acc includes all possible detections for the 2-pair emission. As the above calculation is involved and tedious, an alternate approximation can be used. TO do so, we can find an overestimate of the accidental counts by determining the number of coincidence counts one would expect with a source of completely random photons (that is, the source does not output pairs of photons, it instead randomly emits photons to Alice and Bob with no correlation on the emission times). Here, we would find $acc_u = N^2 \ p_m(1|\theta^A) \ p_m(1|\theta^B)$, where acc_u is the upper bound on the accidentals[2]. As we are interested in ensuring the violation of a Bell inequality, it is acceptable to overestimate the noise, and therefore we typically set $acc = acc_u$.

---

[1] As we experimentally measure counts, putting in the noise terms is easiest by writing all terms as counts, i.e., multiplying the probabilities terms by N.

[2] We can similarly find a lower bound on acc by accounting for the known emission correlations, which results in a quadratic equation.

We can then insert the expected measured probabilities (p_m) into $\mathcal{S}_{\text{CH}}$, and then maximize over $(r, \theta_0, \theta_1, N)$ to determine the optimal solution. It is worth noting, that while we can make any value of r, θ_0, and θ_1, the parameter N will need to limited based on the maximum laser power and pair emission from the photon source, e.g., the source in Chapter 4 can emit (into the desired mode) up to 15 MHz (when we run with a pulse rate of 120 MHz), thus N cannot be larger than $N = S(1|\theta)/p_m(1|\theta)$

For the Bell tests considered in Chapter 6, we do not need to consider any system noise to determine the optimal state and settings (since we assume fair-sampling, and therefore post-select on coincidence counts, these noise terms are negligible). However, in this case we can no longer assume symmetry to reduce the number of optimization parameters. Instead, we allow the state and all polarization settings to be parameters to maximize over. In all of the cases, both here and for the CH Bell inequality with noise, these maximizations can be done with the built-in tools available in both Matlab and Mathematica.

Appendix B

Measurements and Settings for Bell Tests

In this Appendix, we give the results of the analyzed data and quantum states used for each Bell test presented in this paper. First, for the data collected for the projection onto the $\mathcal{S}_{\text{CHSH}}\cos\theta$ and $\mathcal{S}'_{\text{CHSH}}\sin\theta$ axes, we used maximally entangled states, altering the measurement settings to rotate around the circle in Fig. 6.3. Here, each point requires 16 total measurement combinations; and we collected data for 1s at each setting.

For the plot of the tilted Bell inequality in Fig. 6.4 (Eq. (6.9)), we collected data for 15 s for each setting (again, 16 total measurement setting combinations). We used states of varying degree of entanglement, which we cite by listing the θ value in the state $\cos\theta|H,H\rangle + \sin\theta|V,V\rangle$. The analyzed data is displayed in Table B.1. As a note, the value in the text listed for $\mathcal{S}_{\tau=1.3}$ had separately optimized settings (instead of automatically generated settings), as well as was measured for 100 s (i.e., a total measurement time of 100×16 s).

For the chained Bell inequality [Eq. (6.10)], I_n is the chained Bell parameter, with ν_n being the measurement bias. The measurement bias is the deviation of Alice's (or Bob's) individual measurements from being completely random, that is, the difference in probability of measuring output 0 to measuring output 1 (calculated by $p(1|x) - p(-1|x)$). The bias given in Table B.2 (and the bias used in calculating δ_n) is the maximum bias over all possible measurement settings. Finally, δ_n is the bound on the predictive power, with $\Delta\delta_n$ being the uncertainty. The uncertainty of I_n is approximately twice as large as $\Delta\delta_n$ (since $\delta_n \propto I_n/2$). For these measurements, we used a maximally entangled state, with the settings described in Appendix C, and collected data for 5 s at each measurement setting, except from $n = 18$ to $n = 21$, where we collected for 20 s at each setting, as the first scan through all values of n showed the lowest value in that region. The analyzed data for the chained Bell inequality is shown in Table B.2.

Next, in Table B.3 we list the measurement settings and states for the M_{3322} and M_{4322} Bell inequalities. The settings are given as the angle in the projection onto the state $\cos a_i|H\rangle + \sin a_i|V\rangle$ (and similarly for b_j). Here, data was collected data for 1200 s at each measurement setting.

Finally, the waveplate settings for the "Elegant" Bell inequality are displayed in Table B.4. We give the

Table B.1: Analyzed data and estimated parameters for the tilted Bell inequality (Eq. (6.9)). Here, the estimate of the uncertainty of $\mathcal{S}_\tau$ is given by $\Delta\mathcal{S}_\tau$.

τ	$\mathcal{S}_{\text{CHSH}}$	$-E_1 - E_2$	$\mathcal{S}_\tau$	$\Delta\mathcal{S}_\tau$	Local bound	θ
1.001	2.828	0.052	2.828	0.011	2.004	45.0
1.020	2.827	0.120	2.837	0.010	2.080	46.5
1.039	2.816	0.220	2.833	0.010	2.156	48.1
1.063	2.800	0.320	2.840	0.010	2.252	49.8
1.095	2.764	0.480	2.855	0.009	2.380	52.2
1.128	2.736	0.620	2.895	0.009	2.512	54.9
1.137	2.712	0.720	2.909	0.008	2.548	55.5
1.171	2.660	0.860	2.954	0.008	2.684	58.5
1.193	2.616	0.980	2.994	0.007	2.772	60.2
1.223	2.564	1.120	3.064	0.007	2.892	62.7
1.250	2.504	1.240	3.124	0.006	3.000	65.2
1.265	2.456	1.320	3.156	0.006	3.060	66.4
1.296	2.368	1.460	3.232	0.005	3.184	69.2
1.323	2.304	1.580	3.325	0.005	3.292	71.7
1.348	2.228	1.680	3.397	0.004	3.392	74.2
1.369	2.168	1.760	3.467	0.003	3.476	76.3
1.390	2.120	1.820	3.540	0.003	3.560	78.4
1.410	2.064	1.880	3.606	0.002	3.640	80.5
1.424	2.036	1.920	3.664	0.002	3.696	81.9
1.435	2.016	1.932	3.697	0.002	3.740	82.9
1.442	2.000	1.946	3.721	0.002	3.768	83.7
1.449	1.964	1.957	3.722	0.002	3.796	84.6

wave plate settings for both the setting (for output of 1) and the orthogonal setting (for output of 0). As described in the text, these measurements include an additional quarter-wave plate in the system, where the ordering of the optical elements (as seen by the downconverted photons) is the half-wave plate, followed by the quarter-wave plate, then the polarizer.

Table B.2: Analyzed data for the chained Bell inequality (Eq. (6.10)).

n	I_n	ν_n	δ_n	$\Delta\delta_n$
2	0.5931	0.0062	0.8028	0.0016
3	0.4115	0.0058	0.7116	0.0014
4	0.3148	0.0055	0.6629	0.0013
5	0.2624	0.0068	0.6380	0.0012
6	0.2230	0.0058	0.6173	0.0012
7	0.1965	0.0065	0.6048	0.0011
8	0.1812	0.0059	0.5964	0.0011
9	0.1667	0.0073	0.5906	0.0011
10	0.1539	0.0066	0.5836	0.0011
11	0.1479	0.0069	0.5809	0.0011
12	0.1419	0.0069	0.5778	0.0011
13	0.1396	0.0065	0.5763	0.0011
14	0.1357	0.0064	0.5742	0.0011
15	0.1324	0.0077	0.5739	0.0010
16	0.1312	0.0061	0.5718	0.0010
17	0.1294	0.0064	0.5711	0.0010
18	0.1262	0.0065	0.5702	0.0005
19	0.1318	0.0070	0.5714	0.0005
20	0.1290	0.0075	0.5722	0.0005
21	0.1279	0.0074	0.5709	0.0005
22	0.1291	0.0071	0.5717	0.0010
23	0.1287	0.0065	0.5708	0.0010
24	0.1325	0.0072	0.5734	0.0010
25	0.1312	0.0074	0.5730	0.0010
26	0.1380	0.0067	0.5757	0.0011
27	0.1372	0.0070	0.5755	0.0010
28	0.1389	0.0073	0.5768	0.0011
29	0.1409	0.0073	0.5777	0.0011
30	0.1429	0.0069	0.5783	0.0011
31	0.1456	0.0075	0.5803	0.0011
32	0.1474	0.0066	0.5803	0.0011
33	0.1475	0.0070	0.5808	0.0011
34	0.1506	0.0083	0.5836	0.0011
35	0.1547	0.0073	0.5846	0.0011
36	0.1573	0.0066	0.5853	0.0011
37	0.1577	0.0081	0.5870	0.0011
38	0.1594	0.0072	0.5869	0.0011
39	0.1655	0.0072	0.5899	0.0011
40	0.1665	0.0070	0.5903	0.0011
41	0.1698	0.0073	0.5922	0.0011
42	0.1716	0.0065	0.5923	0.0011
43	0.1750	0.0067	0.5942	0.0011
44	0.1810	0.0069	0.5974	0.0011
45	0.1801	0.0079	0.5980	0.0011

Table B.3: Details for the M_{3322} and M_{4322} Bell inequalities (Eqs. (6.12)).

Inequality	θ	a_1	a_2	a_3	a_4	b_1	b_2	b_3
M_{3322}	77.2	-1.2	27.2	-35.2	N/A	-0.7	9.2	-20.3
M_{4322}	76.6	0	61	45	119	15.6	164.3	0

Table B.4: The waveplate settings for the "Elegant" Bell inequality (Eq. (6.16)). All angles refer to angle of the fast axis of the waveplate relative to H.

Setting	HWP angle	QWP angle
$\vec{a}_0$	$45°$	$0°$
$\vec{a}_0^\perp$	$0°$	$0°$
$\vec{a}_1$	$22.5°$	$0°$
$\vec{a}_1^\perp$	$67.5°$	$0°$
$\vec{a}_2$	$0°$	$-45°$
$\vec{a}_2^\perp$	$0°$	$45°$
$\vec{b}_0$	$2.4°$	$72.4°$
$\vec{b}_0^\perp$	$-24.9°$	$-72.4°$
$\vec{b}_1$	$-2.4°$	$-72.4°$
$\vec{b}_1^\perp$	$24.9°$	$72.4°$
$\vec{b}_2$	$42.6°$	$-72.4°$
$\vec{b}_2^\perp$	$-20.1°$	$72.4°$
$\vec{b}_3$	$-42.6°$	$72.4°$
$\vec{b}_3^\perp$	$20.1°$	$-72.4°$

Appendix C

Additional Discussion on the Chained Bell Inequality

In this Appendix, we discuss the intuitive reason that quantum mechanics approaches the PR-box limit for the chained Bell inequalities.

First, consider the CHSH Bell inequality: $\mathcal{S}_{\text{CHSH}}^{0011} = E_{00} + E_{01} + E_{10} - E_{11}$, with a noiseless, unit-efficiency system. Given the entangled state, $|HH\rangle + |VV\rangle$, then the correlations can be written as:

$$E_{i,j} = \cos^2 \theta_{i,j} - \sin^2 \theta_{i,j} = \cos 2\theta_{i,j}, \tag{C.1}$$

where $\theta_{ij} \equiv \theta_i^A - \theta_j^B$ is the angle separation between the two polarization settings. Essentially, with the entangled state of $|HH\rangle + |VV\rangle$, parallel measurements ($\theta_{i,j} \approx 0°$) result in high correlations, whereas orthogonal measurements ($\theta_{i,j} \approx 90°$) result in high anti-correlations. So the goal of the measurements is to evenly space the polarization angles between the two poles of the Bloch sphere (see Fig. C.1) to maximize the three positive terms in $\mathcal{S}_{\text{CHSH}}$, and minimize the negative term, resulting in the angles[1]:

$$\theta_1^A = 0 \times (90/4)° = 0°, \tag{C.2}$$

$$\theta_0^B = 1 \times (90/4)° = 22.5°,$$

$$\theta_0^A = 2 \times (90/4)° = 45°,$$

$$\theta_1^B = 3 \times (90/4)° = 67.5°.$$

With these angles, the three positive correlations (E_{00},E_{01},E_{10}) have polarization settings separated by 22.5°, resulting in $E_{i,j} = 1/\sqrt{2} \approx 0.707$, and the negative term in $\mathcal{S}_{\text{CHSH}}$ has $\theta_{11} = 67.5°$, resulting in $E_{11} = -1/\sqrt{2} = -0.707$, and thus $\mathcal{S}_{\text{CHSH}} = 2\sqrt{2} \approx 2.828$ (i.e., the quantum mechanical bound). Recall for this inequality that the local bound ($\mathcal{L}$) is $\mathcal{S}_{\text{CHSH}} = 2$ and the no-signaling bound ($\mathcal{NS}$) is 4; thus, the

---

[1]Since there is no preferred axis of these measurements, 0° is not a special angle. We can add any fixed angle to all four measurements to reach the same outcome.
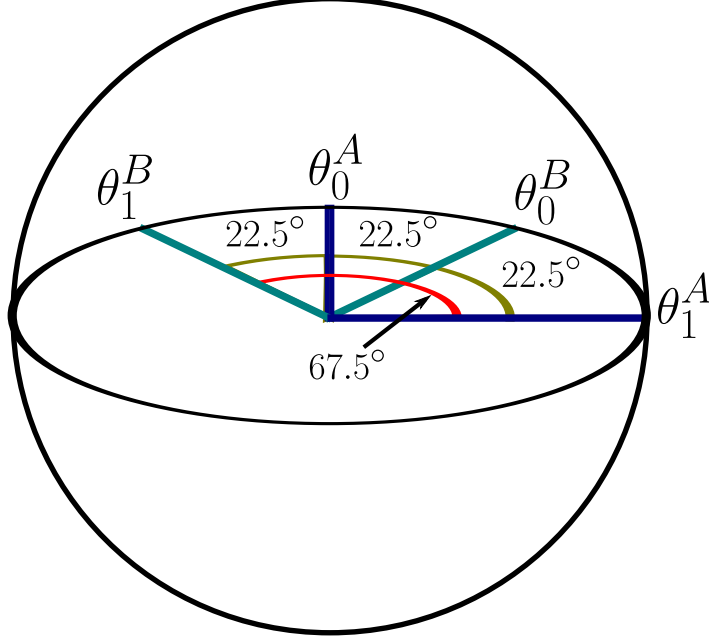
Figure C.1: The Bloch sphere representation of the measurements for $\mathcal{S}_{\text{CHSH}}$ with a perfect system. Given the state $|HH\rangle + |VV\rangle$, the optimal settings evenly divide the two poles. In real space, the angular separation is then $22.5°$ between settings (note the factor of 2 difference between Bloch sphere settings). The result of these settings is three high correlations ($> 1/2$) represented by the olive lines, and one large anti-correlation ($< -1/2$), represented by the red line. These correlations result in a violation of the $\mathcal{S}_{\text{CHSH}}$ Bell inequality. As we "chain" multiple $\mathcal{S}_{\text{CHSH}}$ Bell inequalities together, the distance between any two measurement settings decreases, resulting in higher correlations between the measurements, while the final (negative) correlation tends to -1. Thus, quantum mechanics tends to the PR-box limit as we increase the number of "links" in a chained Bell inequality.

"nonlocal content" is defined in terms of $\mathcal{S}_{\text{CHSH}}$ ($2\sqrt{2}$), $\mathcal{L}$ (2), and $\mathcal{NS}$ (4) as

$$\text{Nonlocal content} = \frac{\mathcal{S}_{\text{CHSH}} - \mathcal{L}}{\mathcal{NS} - \mathcal{L}} = \frac{2\sqrt{2} - 2}{4 - 2} \approx 0.414. \tag{C.3}$$

Now, if we "chain" two $\mathcal{S}_{\text{CHSH}}$ Bell inequalities together: $I_3 = \mathcal{S}_{\text{CHSH}}^{0011} + \mathcal{S}_{\text{CHSH}}^{1122} = E_{00} + E_{01} + E_{10} + E_{21} + E_{12} - E_{22}$, we have a similar situation as above. In this case, we have five positive terms (instead of three), but still only a single negative term. So again, similar to the case for $I_2 = \mathcal{S}_{\text{CHSH}}$, we evenly divide the Bloch

sphere between the two poles, giving angles:

$$\theta_2^A 0 \times (90/6)^\circ = 0^\circ, \tag{C.4}$$

$$\theta_1^B 1 \times (90/6)^\circ = 15^\circ,$$

$$\theta_0^A 2 \times (90/6)^\circ = 30^\circ,$$

$$\theta_0^B 0 \times (90/6)^\circ = 45^\circ,$$

$$\theta_1^A 0 \times (90/6)^\circ = 60^\circ,$$

$$\theta_2^B 3 \times (90/6)^\circ = 75^\circ.$$

These angles result in positive correlations of $E_{i,j} = \sqrt{3}/2 \approx 0.866$, and the negative correlation term of $E_{22} = -\sqrt{3}/2 \approx -0.866$, giving $I_3 \approx 5.196$. For I_3, the local bound is 4, and the no-signaling bound is 6, thus resulting in a nonlocal content (for the maximum quantum mechanical violation) of 0.598, significantly higher than $\mathcal{S}_{\text{CHSH}} = I_2$.

As we continue chaining $\mathcal{S}_{\text{CHSH}}$ inequalities together, the positive correlation measurements have decreasing angle separation (resulting in higher correlations), and the single negative term has an increasing separation (resulting in a higher anti-correlation). For this version of the chained Bell inequality, the local bound is $I_n \overset{\mathcal{L}}{=} 2(n-1)$, the no-signaling bound is $I_n \overset{\mathcal{NS}}{=} 2 + 2(n-1)$, and the quantum bound is $I_n \overset{\mathcal{Q}}{=} 2n \cos \frac{\pi}{2n}$, resulting in a nonlocal content of

$$\text{Nonlocal content} = \frac{I_n - \mathcal{L}}{\mathcal{NS} - \mathcal{L}} = 1 - n + n \cos \frac{\pi}{2n} \underset{n \to \infty}{=} 1. \tag{C.5}$$

Thus, in the infinite limit, all positive terms will have correlations of 1, and the single negative term will have an (anti-)correlation of -1, which is also the best that a PR box would be able to accomplish, resulting in complete nonlocal content.

Appendix D

Notes on Fusion Splicing

In this Appendix, we give a few comments on the fusion splices performed in the experiments presented in the main text[1]. We assume basic knowledge of fusion splicing, which can be learned, e.g., by watching an introductory youtube video to fusion splicing[2]. The primary goal of the fusion splices for our experiments is high transmission through the splice, but we do have a secondary goal of preserving the physical length of fiber. In many of the splices, we have a fixed length of fiber to use (e.g., there is a fixed amount of fiber going to the cryogenic detectors); once this length is used, a new feed-through will have to be installed.

There are a few key points on performing a high-quality fusion splice (though these notes have not definitive, following the guidelines here has resulted in a high yield of quality fusion splices). The initial step in the fusion splice process is stripping the fiber. It is important that the stripper is adjusted properly, with the set screw, to be the correct size for the fibers one desires to splice, and that the recommended $\sim 30°$ angle is maintained when stripping. If one of these conditions is not met, then the fiber very frequently breaks, and some people have suggested that the fiber might form micro-fractures (and not break) if stripped improperly, which would scatter light in the fiber, resulting in loss. After stripping the fiber, the next step is to clean the fiber with methanol (or some other alcohol); I personally believe this step to be the most crucial. Cleaving the fiber should only be performed on a thoroughly cleaned fiber (e.g., wiping the fiber > 5 times is typical). The belief here is that when the fiber is not cleaned properly, the cleaver can become dirty, and that dirt can get on the tip of the cleaved fiber. And worse, the cleaver may also dirty the next fiber that is cleaved, resulting in multiple poor splices. After cleaning the fiber, the cleaving process is mostly automated. One should, however, ensure there is sufficient extra fiber in the cleaver so that the cleaving does not fail. It is typically better to lose an extra centimeter of fiber in the cleaving process than to risk a failed cleave, which could lose significantly more than a centimeter of fiber (if the cleave fails, the fiber will still be damaged by the failed cleave, and therefore the cleaving must be retried at a lower spot on the fiber, thereby increasing the loss of fiber length). After the cleaving has been performed, it is important that the pristine tip does not touch any surface, as it will likely fracture (or get dirty), which will result in

---

[1]In our experiments, we used the Fitel S177 and S178 fusion splicers to perform the splice.
[2]One possible video is available at: https://www.youtube.com/watch?v=Wu0j9ql6XBE.

a low-quality splice.

Adhering to the above guidelines, we have been able to consistently maintain $> 99\%$ transmission through fusion splices of 1550-nm fiber (e.g., 4 out of every 5 splices see this transmission), and an estimated 98% transmission through 630-HP fiber (though the variance of these fibers is higher, we have achieved $> 99\%$ splices with reasonable likelihood ($\sim 1/3$)). We believe the reasoning for the lower quality splices for the 630-HP fiber is simply due to lack of demand for high-quality splices at the shorter wavelengths, and therefore the splicing "recipes" (i.e., the settings of the fusion splicer) are not as well optimized.

References

[1] P. W. Shor, "Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer," *SIAM Journal on Scientific and Statistical Computing*, vol. 26, no. 2, p. 1484, 1995.

[2] C. Bennett and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing," *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*, pp. 175–179, 1984.

[3] V. B. Braginsky and F. Y. Khalili, "Quantum nondemolition measurements: the route from toys to tools," *Rev. Mod. Phys.*, vol. 68, pp. 1–11, Jan 1996.

[4] J. S. Bell, "On the Einstein-Podolsky-Rosen paradox," *Physics*, vol. 1, pp. 195–200, 1964.

[5] A. Einstein, B. Podolsky, and N. Rosen, "Can quantum-mechanical description of physical reality be considered complete?," *Phys. Rev.*, vol. 47, pp. 777–780, May 1935.

[6] H. Reichenbach, "The direction of time," *Berkeley, University of Los Angeles Press*, 1956.

[7] H. M. Wiseman, "The two bells theorems of john bell," *Journal of Physics A: Mathematical and Theoretical*, vol. 47, no. 42, p. 424001, 2014.

[8] R. W. Boyd in *Nonlinear Optics (Third Edition)* (R. W. Boyd, ed.), pp. 1 – 67, Burlington: Academic Press, third edition ed., 2008.

[9] V. Caprara Vivoli, P. Sekatski, J.-D. Bancal, C. C. W. Lim, B. G. Christensen, A. Martin, R. T. Thew, H. Zbinden, N. Gisin, and N. Sangouard, "Challenging preconceptions about bell tests with photon pairs," *Phys. Rev. A*, vol. 91, p. 012107, Jan 2015.

[10] L. E. Vicent, A. B. U'Ren, R. Rangarajan, C. I. Osorio, J. P. Torres, L. Zhang, and I. A. Walmsley, "Design of bright, fiber-coupled and fully factorable photon pair sources," *New Journal of Physics*, vol. 12, no. 9, p. 093027, 2010.

[11] D. Dehlinger and M. W. Mitchell, "Entangled photons, nonlocality, and bell inequalities in the undergraduate laboratory," *American Journal of Physics*, vol. 70, no. 9, pp. 903–910, 2002.

[12] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed Experiment To Test Local Hidden-Variable Theories," *Phys. Rev. Lett.*, vol. 23, no. 15, pp. 880–884, 1969.

[13] B. S. Tsirelson, "Quantum generalizations of bell's inequality," *Lett. Math. Phys.*, vol. 4, pp. 93–100, 1980.

[14] J. F. Clauser and M. A. Horne, "Experimental consequences of objective local theories," *Phys. Rev. D*, vol. 10, pp. 526–535, Jul 1974.

[15] S. J. Freedman and J. F. Clauser, "Experimental test of local hidden-variable theories," *Phys. Rev. Lett.*, vol. 28, no. 14, pp. 938–941, 1972.

[16] A. Aspect, P. Grangier, and G. Roger, "Experimental tests of realistic local theories via Bell's theorem," *Phys. Rev. Lett.*, vol. 47, no. 7, pp. 460–463, 1981.

[17] A. Aspect, P. Grangier, and G. Roger, "Experimental realization of einstein-podolsky-rosen-bohm *Gedankenexperiment*: A new violation of bell's inequalities," *Phys. Rev. Lett.*, vol. 49, pp. 91–94, Jul 1982.

[18] A. Aspect, J. Dalibard, and G. Roger, "Experimental test of Bell's inequalities using time-varying analyzers," *Phys. Rev. Lett.*, vol. 49, no. 25, pp. 1804–1807, 1982.

[19] M. Genovese, "Research on hidden variable theories: A review of recent progresses," *Phys. Rep.*, vol. 413, no. 6, pp. 319 – 396, 2005.

[20] P. H. Eberhard, "Background level and counter efficiencies required for a loophole-free Einstein-Podolsky-Rosen experiment," *Phys. Rev. A*, vol. 47, pp. R747–R750, Feb 1993.

[21] F. Marsili, V. B. Verma, J. A. Stern, S. Harrington, A. E. Lita, T. Gerrits, I. Vayshenker, B. Baek, M. D. Shaw, R. P. Mirin, and S. W. Nam, "Detecting Single Infrared Photons with 93% System Efficiency," *Nature Photonics*, vol. 7, no. February, pp. 210–214, 2013.

[22] A. J. Miller, S. W. Nam, J. M. Martinis, and A. V. Sergienko, "Demonstration of a low-noise near-infrared photon counter with multiphoton discrimination," *Applied Physics Letters*, vol. 83, no. 4, pp. 791–793, 2003.

[23] I. Gerhardt, Q. Liu, A. Lamas-Linares, J. Skaar, V. Scarani, V. Makarov, and C. Kurtsiefer, "Experimentally faking the violation of bell's inequalities," *Phys. Rev. Lett.*, vol. 107, p. 170404, Oct 2011.

[24] T. Marshall, E. Santos, and F. Selleri, "Local realism has not been refuted by atomic cascade experiments," *Physics Letters A*, vol. 98, no. 1, pp. 5 – 9, 1983.

[25] M. A. Rowe, D. Kielpinski, V. Meyer, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland, "Experimental violation of a Bell's inequality with efficient detection.," *Nature*, vol. 409, no. 6822, pp. 791–794, 2001.

[26] M. Ansmann, H. Wang, R. C. Bialczak, M. Hofheinz, E. Lucero, M. Neeley, A. D. O'Connell, D. Sank, M. Weides, J. Wenner, A. N. Cleland, and J. M. Martinis, "Violation of bell's inequality in josephson phase qubits," *Nature*, vol. 461, pp. 504–506, Sept. 2009.

[27] J. Hofmann, M. Krug, N. Ortegel, L. Gérard, and M. Weber, "Heralded Entanglement of Widely Separated Atoms," *Nature*, vol. 891, no. July, pp. 72–75, 2012.

[28] L. K. Shalm, E. Meyer-Scott, B. G. Christensen, P. Bierhorst, M. A. Wayne, M. J. Stevens, T. Gerrits, S. Glancy, D. R. Hamel, M. S. Allman, K. J. Coakley, S. D. Dyer, C. Hodge, A. E. Lita, V. B. Verma, C. Lambrocco, E. Tortorici, A. L. Migdall, Y. Zhang, D. R. Kumor, W. H. Farr, F. Marsili, M. D. Shaw, J. A. Stern, C. Abellán, W. Amaya, V. Pruneri, T. Jennewein, M. W. Mitchell, P. G. Kwiat, J. C. Bienfang, R. P. Mirin, E. Knill, and S. W. Nam, "Strong loophole-free test of local realism*," *Phys. Rev. Lett.*, vol. 115, p. 250402, Dec 2015.

[29] G. Weihs, T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger, "Violation of bell's inequality under strict einstein locality conditions," *Phys. Rev. Lett.*, vol. 81, pp. 5039–5043, Dec 1998.

[30] T. Scheidl, R. Ursin, J. Kofler, S. Ramelow, X.-S. Ma, T. Herbst, L. Ratschbacher, A. Fedrizzi, N. K. Langford, T. Jennewein, and A. Zeilinger, "Violation of local realism with freedom of choice," *Proc. Nat. Acad. Sci. USA*, vol. 107, no. 46, pp. 19708–19713, 2010.

[31] J.-A. Larsson and R. D. Gill, "Bell's inequality and the coincidence-time loophole," *Europhysics Letters*, vol. 67, no. 5, p. 707, 2004.

[32] Y. Zhang, S. Glancy, and E. Knill, "Efficient quantification of experimental evidence against local realism," *Phys. Rev. A*, vol. 88, p. 052119, Nov 2013.

[33] S. Pironio, A. Acin, S. Massar, A. B. de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe, "Random numbers certified by Bell's theorem," *Nature*, vol. 464, pp. 1021–1024, Apr. 2010.

[34] A. Acin, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, "Device-independent security of quantum cryptography against collective attacks," *Phys. Rev. Lett.*, vol. 98, p. 230501, Jun 2007.

[35] B. G. Christensen, A. Hill, P. G. Kwiat, E. Knill, S. W. Nam, K. Coakley, S. Glancy, L. K. Shalm, and Y. Zhang, "Analysis of coincidence-time loopholes in experimental bell tests," *Phys. Rev. A*, vol. 92, p. 032130, Sep 2015.

[36] S. Popescu and D. Rohrlich, "Quantum nonlocality as an axiom," *Foundations of Physics*, vol. 24, no. 3, pp. 379–385, 1994.

[37] E. Knill, S. Glancy, S. W. Nam, K. Coakley, and Y. Zhang, "Bell inequalities for continuously emitting sources," *Phys. Rev. A*, vol. 91, p. 032105, Mar 2015.

[38] B. G. Christensen, K. T. McCusker, J. B. Altepeter, B. Calkins, T. Gerrits, A. E. Lita, A. Miller, L. K. Shalm, Y. Zhang, S. W. Nam, N. Brunner, C. C. W. Lim, N. Gisin, and P. G. Kwiat, "Detection-loophole-free test of quantum nonlocality, and applications," *Phys. Rev. Lett.*, vol. 111, p. 130406, Sep 2013.

[39] E. Knill, S. Glancy, S. W. Nam, K. Coakley, and Y. Zhang, "Bell inequalities for continuously emitting sources," *Phys. Rev. A*, vol. 91, pp. 032105/1–17, 2015.

[40] J. Shao, *Mathematical Statistics*. Springer Texts in Statistics, New York: Springer, 1998. See 2nd edition pages 126-127.

[41] S. Pironio, A. Acín, S. Massar, A. B. de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe, "Random numbers certified by Bell's theorem," *Nature*, vol. 464, pp. 1021–1024, 2010.

[42] N. D. Mermin, "The epr experiment - thoughts about the loophole," *Annals of the New York Academy of Sciences*, vol. 480, no. 1, pp. 422–427, 1986.

[43] P. G. Kwiat, E. Waks, A. G. White, I. Appelbaum, and P. H. Eberhard, "Ultrabright source of polarization-entangled photons," *Phys. Rev. A*, vol. 60, pp. R773–R776, Aug 1999.

[44] A. G. White, D. F. V. James, P. H. Eberhard, and P. G. Kwiat, "Nonmaximally entangled states: Production, characterization, and utilization," *Phys. Rev. Lett.*, vol. 83, pp. 3103–3107, Oct 1999.

[45] R. Rangarajan, M. Goggin, and P. Kwiat, "Optimizing type-i polarization-entangled photons," *Opt. Express*, vol. 17, pp. 18920–18933, Oct 2009.

[46] A. E. Lita, A. J. Miller, and S. W. Nam, "Counting near-infrared single-photons with 95% efficiency," *Opt. Express*, vol. 16, pp. 3032–3040, Mar 2008.

[47] M. Giustina, A. Mech, S. Ramelow, B. Wittmann, J. Kofler, J. Beyer, A. Lita, B. Calkins, T. Gerrits, S. W. Nam, R. Ursin, and A. Zeilinger, "Bell violation using entangled photons without the fair-sampling assumption," *Nature*, vol. 497, pp. 227–230, May 2013.

[48] M. A. Wayne, E. R. Jeffrey, G. M. Akselrod, and P. G. Kwiat, "Photon arrival time quantum random number generation," *Journal of Modern Optics*, vol. 56, no. 4, pp. 516–522, 2009.

[49] R. Colbeck and A. Kent, "Private randomness expansion with untrusted devices," *Journal of Physics A: Mathematical and Theoretical*, vol. 44, no. 9, p. 095305, 2011.

[50] S. Pironio and S. Massar, "Security of practical private randomness generation," *Phys. Rev. A*, vol. 87, p. 012336, Jan 2013.

[51] S. Fehr, R. Gelles, and C. Schaffner, "Security and composability of randomness expansion from bell inequalities," *Phys. Rev. A*, vol. 87, p. 012335, Jan 2013.

[52] R. Shaltiel, "Recent developments in explicit constructions of extractors," *Bulletin of the European Association for Theoretical Computer Science*, vol. 77, p. 6795, 2002.

[53] B. Hensen, H. Bernien, A. E. Dreau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. L. Vermeulen, R. N. Schouten, C. Abellan, W. Amaya, V. Pruneri, M. W. Mitchell, M. Markham, D. J. Twitchen, D. Elkouss, S. Wehner, T. H. Taminiau, and R. Hanson, "Loophole-free bell inequality violation using electron spins separated by 1.3 kilometres," *Nature*, vol. 526, pp. 682–686, 10 2015.

[54] M. Giustina, M. A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.-A. Larsson, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, J. Beyer, T. Gerrits, A. E. Lita, L. K. Shalm, S. W. Nam, T. Scheidl, R. Ursin, B. Wittmann, and A. Zeilinger, "Significant-loophole-free test of bell's theorem with entangled photons," *Phys. Rev. Lett.*, vol. 115, p. 250401, Dec 2015.

[55] P. Bierhorst, "A robust mathematical model for a loophole-free Clauser-Horne experiment," *J. Phys. A*, vol. 48, no. 19, p. 195302, 2015.

[56] R. D. Gill, "Accardi contra Bell (cum mundi): The impossible coupling," in *Mathematical Statistics and Applications: Festschrift for Constance van Eeden* (M. Moore, S. Froda, and C. Léger, eds.), vol. 42, pp. 133–154, Institute of Mathematical Statistics. Beachwood, Ohio, 2003.

[57] P. B. Dixon, D. Rosenberg, V. Stelmakh, M. E. Grein, R. S. Bennink, E. A. Dauler, A. J. Kerman, R. J. Molnar, and F. N. C. Wong, "Heralding efficiency and correlated-mode coupling of near-IR fiber-coupled photon pairs," *Phys. Rev. A*, vol. 90, no. 4, p. 043804, 2014.

[58] L. K. Shalm, K. Garay, J. Palfree, A. L. Migdall, A. U'Ren, and S. W. Nam, "Spontaneous parametric downcoversion calculator." http://www.spdcalc.org.

[59] P. G. Evans, R. S. Bennink, W. P. Grice, T. S. Humble, and J. Schaake, "Bright source of spectrally uncorrelated polarization-entangled photons with nearly single-mode emission," *Phys. Rev. Lett.*, vol. 105, p. 253601, Dec 2010.

[60] C. Abellan, W. Amaya, D. Mitrani, V. Pruneri, and M. W. Mitchell, "Generation of fresh and pure random numbers for loophole-free Bell tests," *ArXiv e-prints*, June 2015.

[61] S. Pironio, "Random 'choices' and the locality loophole," *ArXiv e-prints*, Oct. 2015.

[62] Raw data from each experimental run will be made available online.

[63] Y. Zhang, S. Glancy, and E. Knill, "Asymptotically optimal data analysis for rejecting local realism," *Phys. Rev. A*, vol. 84, p. 062118, Dec 2011.

[64] J. S. Bell, "Locality in quantum mechanics: reply to critics," *Epistemological Letters*, pp. 2–6, Nov 1975.

[65] B. G. Christensen, Y.-C. Liang, N. Brunner, N. Gisin, and P. G. Kwiat, "Exploring the limits of quantum nonlocality with entangled photons," *Phys. Rev. X*, vol. 5, p. 041052, Dec 2015.

[66] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, "Bell nonlocality," *Rev. Mod. Phys.*, vol. 86, p. 419, 2014.

[67] P. Shadbolt, J. Mathews, A. Laing, and J. O'Brien, "Testing foundations of quantum mechanics with photons," *Nat. Phys.*, vol. 10, p. 278, 2014.

[68] A. Acin, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, "Device-independent security of quantum cryptography against collective attacks," *Phys. Rev. Lett.*, vol. 98, p. 230501, 2007.

[69] R. Colbeck, "Quantum and relativistic protocols for secure multi-party computation," *PhD Thesis, Univ. of Cambridge*, 2007.

[70] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, "Nonlocal correlations as an information-theoretic resource," *Phys. Rev. A*, vol. 71, p. 022101, 2005.

[71] M. Navascues, S. Pironio, and A. Acin, "A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations," *New J. Phys.*, vol. 10, p. 073013, 2008.

[72] A. C. Doherty, Y.-C. Liang, B. Toner, and S. Wehner *Proceedings of the 23rd IEEE Conference on Computational Complexity (IEEE Computer Society, College Park, MD)*, pp. 199–210, 2008.

[73] W. van Dam, "Implausible consequences of superstrong nonlocality," *Preprint at http://arxiv.org/abs/quant-ph/0501159*, 2005.

[74] M. Pawlowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Zukowski, "Information causality as a physical principle," *Nature*, vol. 461, pp. 1101–1104, 2009.

[75] T. Fritz, A. B. Sainz, R. Augusiak, J. B. Brask, R. Chaves, A. Leverrier, and A. Acin, "Local orthogonality as a multipartite principle for quantum correlations," *Nature Commun.*, vol. 4, p. 2263, 2013.

[76] M. Navascues and H. Wunderlich, "A glance beyond the quantum model," *Proc. Roy. Soc. Lond. A*, vol. 466, p. 881, 2009.

[77] J. Allcock, N. Brunner, M. Pawlowski, and V. Scarani, "Recovering part of the quantum boundary from information causality," *Phys. Rev. A*, vol. 80, p. 040103, 2009.

[78] D. Rohrlich, "Pr-box correlations have no classical limit," *Quantum Theory: A two Time Success Story, (Springer, NY)*, 2013.

[79] N. Gisin, "Quantum measurement of spins and magnets, and the classical limit of pr-boxes," *arXiv:1407.8122*, 2014.

[80] S. Popescu, "Nonlocality beyond quantum mechanics," *Nature Physics*, vol. 10, pp. 264–270, 2014.

[81] M. Navascues, Y. Guryanova, M. J. Hoban, and A. Acin, "Almost quantum correlations," *Nat. Commun.*, vol. 6, p. 6288, 2015.

[82] Y.-C. Liang, T. Vertesi, and N. Brunner, "Semi-device-independent bounds on entanglement," *Phys. Rev. A*, vol. 83, p. 022108, 2011.

[83] T. Vidick and S. Wehner, "More nonlocality with less entanglement," *Phys. Rev. A*, vol. 83, p. 052310, 2011.

[84] M. Junge and C. Palazuelos, "Large violation of bell inequalities with low entanglement," *Commun. Math. Phys.*, vol. 306, p. 695, 2011.

[85] N. Brunner, N. Gisin, and V. Scarani, "Entanglement and non-locality are different resources," *New J. Phys.*, vol. 7, p. 88, 2005.

[86] W. Tittel, J. Brendel, H. Zbinden, and N. Gisin, "Violation of Bell inequalities by photons more than 10 km apart," *Phys. Rev. Lett.*, vol. 81, no. 17, pp. 3563–3566, 1998.

[87] J.-P. Pan, D. Bouwmeester, M. Daniell, H. Weinfurter, and A. Zeilinger, "Experimental test of quantum nonlocality in three-photon greenberger-horne-zeilinger entanglement," *Nature*, vol. 403, p. 515, 2000.

[88] J. Lavoie, R. Kaltenbaek, and K. Resch, "Experimental violation of svetlichny's inequality," *New J. Phys.*, vol. 11, p. 073051, 2009.

[89] C. Erven, E. Meyer-Scott, K. Fisher, J. Lavoie, B. L. Higgins, Z. Yan, C. J. Pugh, J.-P. Bourgoin, R. Prevedel, L. K. Shalm, L. Richards, N. Gigov, R. Laflamme, G. Weihs, T. Jennewein, and K. J. Resch, "Experimental three-photon quantum nonlocality under strict locality conditions," *Nature Photonics*, vol. 8, pp. 292–296, 2014.

[90] R. T. Thew, A. Acin, H. Zbinden, and N. Gisin, "Bell-type test of energy-time entangled qutrits," *Phys. Rev. Lett.*, vol. 93, p. 010503, 2004.

[91] A. Mair, A. Vaziri, G. Weihs, and A. Zeilinger, "Entanglement of the orbital angular momentum states of photons," *Nature*, vol. 412, p. 313, 2001.

[92] A. C. Dada, J. Leach, G. S. Buller, M. J. Padgett, and E. Andersson, "Experimental high-dimensional two-photon entanglement and violations of generalized bell inequalities," *Nat. Phys.*, vol. 7, pp. 677–680, 2011.

[93] A. Elitzur, S. Popescu, and D. Rohrlich, "Quantum nonlocality for each pair in an ensemble," *Phys. Lett. A*, vol. 162, p. 25, 1992.

[94] R. Colbeck and R. Renner, "No extension of quantum theory can have improved predictive power," *Nature Commun.*, vol. 2, p. 411, 2011.

[95] N. Gisin, "Bell inequalities: many questions, a few answers," *quant-ph/0702021; Quantum reality, relativistic causality, and closing the epistemic circle*, pp. 125–140, 2009.

[96] J. Barrett and S. Pironio, "Popescu-rohrlich correlations as a unit of nonlocality," *Phys. Rev. Lett.*, vol. 95, p. 140401, Sep 2005.

[97] G. Vallone, G. Lima, E. S. Gomez, G. Canas, J.-A. Larsson, P. Mataloni, and A. Cabello, "Bell scenarios in which nonlocality and entanglement are inversely related," *Phys. Rev. A*, vol. 89, p. 012102, 2014.

[98] P. Pearle, "Hidden-variable example based upon data rejection," *Phys. Rev. D*, vol. 2, p. 1418, 1970.

[99] S. Braunstein and C. Caves, "Wringing out better bell inequalities," *Annals of Phys.*, vol. 202, p. 22, 1990.

[100] T. E. Stuart, J. A. Slater, R. Colbeck, R. Renner, and W. Tittel, "Experimental bound on the maximum predictive power of physical theories," *Phys. Rev. Lett.*, vol. 109, p. 020402, 2012.

[101] J. Barrett, A. Kent, and S. Pironio, "Maximally nonlocal and monogamous quantum correlations," *Phys. Rev. Lett.*, vol. 97, p. 170409, 2006.

[102] L. Aolita, R. Gallego, A. Acin, A. Chiuri, G. Vallone, P. Mataloni, and A. Cabello, "Fully nonlocal quantum correlations," *Phys. Rev. A*, vol. 85, p. 032107, 2012.

[103] T. Yang, Q. Zhang, J. Zhang, J. Yin, Z. Zhao, M. Zukowski, Z.-B. Chen, and J.-W. Pan, "All-versus-nothing violation of local realism by two-photon, four-dimensional entanglement," *Phys. Rev. Lett.*, vol. 95, p. 240406, 2005.

[104] Y.-C. Liang and A. C. Doherty, "Bounds on quantum correlations in bell-inequality experiments," *Phys. Rev. A*, vol. 75, p. 042103, 2007.

[105] Y.-C. Liang, C. W. Lim, and D. L. Deng, "Reexamination of a multisetting bell inequality for qudits," *Phys. Rev. A*, vol. 80, p. 052116, 2009.

[106] N. J. Cerf, N. Gisin, S. Massar, and S. Popescu, "Simulating maximal quantum entanglement without communication," *Phys. Rev. Lett.*, vol. 94, p. 220403, 2005.

[107] M. McKague, M. Mosca, and N. Gisin, "Simulating quantum systems using real hilbert spaces," *Phys. Rev. Lett.*, vol. 102, p. 020505, 2009.

[108] K. Pal and T. Vertesi, "Efficiency of higher-dimensional hilbert spaces for the violation of bell inequalities," *Phys. Rev. A*, vol. 77, p. 042105, 2008.

[109] C. Bernhard, B. Bessire, A. Montina, M. Pfaffhauser, A. Stefanov, and S. Wolf, "Non-locality of experimental qutrit pairs," *J. Phys. A: Math. Theo.*, vol. 47, p. 424013, 2014.