On Position-Momentum Entanglement, Nonlocality, and Measurement

by

James Edward Schneeloch

Submitted in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Supervised by

Prof. John C. Howell

Department of Physics and Astronomy
Arts, Sciences and Engineering
School of Arts and Sciences

University of Rochester
Rochester, New York

2015

Dedicated to my family, my labmates, and my students.

Biographical Sketch

James Schneeloch is a graduate research assistant in quantum optics and a Ph.D. candidate at the University of Rochester. Originally from Westfield, Massachusetts, James began his undergraduate studies at the Commonwealth Honors College in the University of Massachusetts at Amherst, graduating summa cum laude in 2009. Before doing research full time, James worked as a Teaching Assistant, receiving accolades including the American Association of Physics Teachers Teaching prize in 2011. Since joining Prof. John Howell's quantum optics group in June of 2010, he has spearheaded theoretical research into quantum entanglement, quantum measurement, and continuous-variable quantum information.

James has written multiple articles on entropic EPR-steering inequalities, experimental demonstrations of position-momentum entanglement, and has contributed to articles on new approaches to weakly perturbing quantum measurement without weak values, as well as articles on directly measuring the complex density operator describing entangled photon pairs. He is currently working on a project to experimentally demonstrate that the position-momentum statistics of photon pairs violate a Bell inequality.

Publications

[1] James Schneeloch, and John C. Howell, "Introduction to the Transverse Spatial Correlations in Spontaneous Parametric Down-Conversion through the Biphoton Birth Zone", *arXiv* 1502.06996 (2015).

[2] James Schneeloch, Samuel H. Knarr, Gregory A. Howland, and John C. Howell, "Demonstrating Continuous-Variable EPR Steering in spite of Finite Experimental Capabilities using Fano Steering Bounds", *J. Opt. Soc. Am. B* **32** 4 (2015).

[3] James Schneeloch, Curtis J. Broadbent, and John C. Howell, "Uncertainty Relation for Mutual Information", *Phys. Rev. A* **90** 062119 (2014).

[4] Gregory A. Howland, James Schneeloch, Daniel J. Lum, and John C. Howell, "Simultaneous Measurement of Complementary Observables with Compressive Sensing", *Phys. Rev. Lett.* **112** 254602 (2014).

[5] James Schneeloch, Curtis J. Broadbent, and John C. Howell, "Improving Einstein-Podolsky-Rosen Steering with State Information", *Phys. Lett. A* **378** 766-769 (2014).

[6] James Schneeloch, "The Relationship Between Discrete and Continuous Entropy in EPR-steering Inequalities", *arXiv* 1312.2604 (2013).

[7] James Schneeloch, Curtis J. Broadbent, Stephen P. Walborn, Eric G. Cavalcanti, and John C. Howell, "Einstein-Podolsky-Rosen Steering Inequalities from Entropic Uncertainty Relations", *Phys. Rev. A* **87** 062103 (2013).

[8] James Schneeloch, P. Ben Dixon, Gregory A. Howland, Curtis J. Broadbent, and John C. Howell, "Violation of Continuous-Variable Einstein-Podolsky-Rosen

Steering with Discrete Measurements", *Phys. Rev. Lett.* **110** 130407 (2013).

[9] P. Ben Dixon, Gregory A. Howland, James Schneeloch, and John C. Howell, "Quantum Mutual Information Capacity of High-Dimensional Entangled States", *Phys. Rev. Lett* **108** 143603 (2012).

Acknowledgments

The research discussed in this thesis wouldn't've been possible without the help and guidance of many people.

First, I would like to thank my family for their support, fostering my interest in science, and inspiring me to continue on in the face of uncertainty.

Second, I would like to thank my senior colleagues in the Howell Lab Group Curtis Broadbent, and Greg Howland. We've worked on many projects together, and had many fascinating discussions of new ideas, the foundations of our work, and keeping them well grounded in experimental reality. With their help, I've grown from an uncertain grad student into a confident scientist, and I cannot thank them enough for this.

I also would like to thank my junior colleagues, Dan Lum, Sam Knarr, and Justin Winkler for our deep discussions and creative collaborations. Through projects long and short, it's been fun, and I hope for many more.

Finally, I would like to thank my adviser, John Howell for his unwavering optimism, patience, continuing guidance, and his most understanding work philosophy that "great cheese comes from happy cows". Not everyone can say that grad school was fun, but in your group, there wasn't a single day that I didn't enjoy coming to work.

Abstract

This thesis covers much of the research that I have worked on since the summer of 2012, when I started vigorously studying information theory and how it can be used to make quantum entanglement in optical experiments both easier to demonstrate and to understand. In particular, I will discuss my research into demonstrating Continuous-variable EPR-steering with a finite number of discrete measurements; simultaneous imaging of both the position and momentum spectra of an optical field with partial projective measurements; and violating a Bell inequality with the position and momentum statistics of entangled photon pairs.

As a whole, this thesis is intended to show how the techniques of standard information theory can be used to make both the understanding and the demonstration of quantum correlations considerably easier. Most existing notions of uncertainty are those of localization (i.e., how closely or accurately a measurement outcome can be predicted to be near a particular value). Here, we discuss an alternative notion of uncertainty as that of information. In particular, while a localization-based uncertainty relation is a limit on the precision of simultaneous measurements of complementary observables (say, of position and momentum), an information-based uncertainty relation is a lower limit to the amount of information needed to describe the statistics of the measurement outcomes of complementary observables. Thus, while it is understood to be impossible to have a particle whose state describes both an arbitrarily narrow range in position and in momentum, it is also impossible to have a particle whose state allows the descrip-

tion of both position and momentum statistics with arbitrarily small amounts of information (i.e., predicted with certainty).

The first three chapters of this thesis provide a background for the research that follows, discussing the essentials of information theory (Chapter 1), EPR steering (Chapter 2), and the physics underlying the creation of the spatially entangled photon pairs discussed throughout this work (Chapter 3). In the next chapter, we go over our research into how to demonstrate continuous-variable EPR-steering with discrete measurements (Section 4.1), and how our new continuous-variable Fano inequality allows us to demonstrate position-momentum EPR-steering in spite of not having access to the complete position-momentum probability distribution (Section 4.2).

In the closing arc of this thesis, we delve into our more recent research into strategies for accurately measuring the statistics of complementary observables using partially projective measurements (Chapter 5), as well as our strategy for demonstrating position-momentum Bell nonlocality in the transverse spatial statistics of entangled photon pairs (Chapter 6).

Contributors and Funding Sources

This work was supervised by a dissertation committee consisting of Professors John C. Howell [my research advisor], Douglas Cline, and Joseph H. Eberly of the Department of Physics and Astronomy, as well as Professor Miguel Alonso of the Institute of Optics, and chaired by Professor James R. Fienup of the Institute of Optics. My research would not have been possible without the actual funding that paid me. Over the past few years, I have been funded by DARPA DSO InPho Grant No. W911NF-10-1-0404, DARPA DSO Grant No. W31P4Q-12-1-0015, and AFOSR Grant No. FA9550-13-1-0019. I am especially indebted to John Howell for securing all this funding on the behalf of myself and others.

When I started in John Howell's lab group, the first project I was put on was to create an ultra-bright source of entangled photon pairs. This was in the summer of 2010. For the next two years, I worked on many different projects including: trying to observe Hong-Ou Mandel interference in the transverse positions and momenta of entangled photon pairs; quantum ghost imaging through turbulence; exploring quantum discord with the two-qubit correlation matrix; experimental measurement of position and momentum correlations in photon pairs from SPDC; Simulating the photon number phase transition in compressive sensing; error analysis of image acquisition by raster scanning; and measuring the statistics and entropy of "natural" images.

My primary contributions started in the summer of 2012 when I was tasked with trying to figure out how our discrete measurements of position and momen-

tum from an experiment earlier could be used to violate an existing inequality that demonstrates the EPR-paradox. From that point on, I was working on the research contained in this thesis.

My fellow contributors helped significantly in the research presented in this thesis. In a collaboration, multiple people work on the same projects, and it should be understood that although we list the primary contributions from co-authors and collaborators, their contributions extend to multiple shared efforts.

In my 2013 PRL, "Violation of Continuous-Variable Einstein-Podolsky-Rosen Steering with Discrete Measurements", I worked on this paper closely with Curtis Broadbent, while receiving experimental data and advice from Gregory Howland and Ben Dixon. I figured out the useful relationship between discrete and continuous entropies, developed the steering inequality for conditional entropies, provided the figures, and wrote the paper, while Curtis showed how one could remodel the inequality in terms of the actual position and momentum correlations via the mutual information. Curtis also provided extensive guidance in scientific writing, provided simpler ways of expressing the concepts we wanted to convey, and developed new ideas to expand the scope of the paper.

In my 2013 PRA, "Einstein-Podolsky-Rosen Steering Inequalities from Entropic Uncertainty Relations", I worked closely with Curtis Broadbent, while receiving significant contributions from both Stephen Walborn and Eric Cavalcanti. In particular, Curtis and I worked together on the main result of developing entropic steering inequalities from entropic uncertainty relations (though I wrote the paper, and came up with the figures). In addition, Curtis also looked into the possibility of exclusive one-way steerable states. Stephen Walborn showed how these steering inequalities can give rise to secure key rates in quantum key distribution, while Eric Cavalcanti expanded upon the novelty of symmetric steering inequalities, and explored how different steering inequalities compared in their ability to demonstrate steering in Werner states.

In my 2014 PLA, "Improving Einstein-Podolsky-Rosen Steering with State Information", I worked to see whether improved uncertainty relations result in improved steering inequalities, wrote the paper, and made the figures, while Curtis showed how to simplify much of my discussion and arguments.

In the 2014 PRL, "Simultaneous Measurement of Complementary Observables with Compressive Sensing", I developed a model showing how a random partial projective measurement in position affects the momentum distribution of an optical field, while Gregory Howland came up with the general idea, wrote the paper and led the experiment. Daniel Lum assisted in the experimental efforts, and showed how to substantially simplify data acquisition and analysis.

In my 2015 JOSA B, "Demonstrating Continuous-Variable Einstein-Podolsky-Rosen Steering in spite of Finite Experimental Capabilities using Fano Steering Bounds", I developed a Fano inequality for continuous variables, and showed how it could be used to easily demonstrate steering in highly correlated continuous-variable systems. Though I wrote the paper, and provided the figures, Samuel Knarr and Gregory Howland provided the experimental data that I used to show that the inequality was sensitive enough to be of practical use.

In my 2015 arXiv article, "Introduction to the Transverse Spatial Correlations in Spontaneous Parametric Down-Conversion thorough the Biphoton Birth Zone", I reviewed the foundations of spontaneous parametric down-converion (SPDC), calculated the joint position and momentum intensity distributions of down-converted photon pairs, reviewed its agreement with experimental data, and explored consequences and treatments of these correlations. The writing, figures, ideas, and calculations are my own, though I wrote this with advice and input from multiple members of the Howell Lab Group.

Beyond the research contained in this thesis, there were a few standout projects I particularly enjoyed working on. My work on directly measuring the Dirac distribution with Sam Knarr was especially rewarding, as well as my experimental and

theoretical contributions to John Howell's, Gerardo Viza's, and Julian Martinez's project on measuring universal gravitation with a simple pendulum coupled to a displaced Sagnac interferometer. In addition, my work with John Howell on estimating the entropy of "natural images" was a lot of fun; even though it didn't lead to a publication, we could use the empirical probability distribution from a large database of photographs to create a random natural image generator. This not only showed how good our estimation of the entropy of natural images was, but also gave us amusement with an endlessly renewable digital Rorschach test.

Table of Contents

List of Figures

List of Acronyms and Abbreviations

AEP	Asymptotic Equipartition Property
APD	Avalanche Photo-Diode
BBO	Beta-Barium Borate
CHSH	Clauser-Horne-Shimony-Holt
DMD	Digital Micro-mirror Device
EPR	Einstein-Podolsky-Rosen
FWHM	Full-Width at Half-Maximum
i.i.d.	Independent and Identically Distributed
IPR	Inverse Participation Ratio
LED	Light Emitting Diode
LHS	Local Hidden State
LHV	Local Hidden Variable
LOCC	Local Operations and Classical Communication
QKD	Quantum Key Distribution
SLM	Spatial Light Modulator
SPDC	Spontaneous Parametric Down-Conversion

1 Preliminaries on Information Theory, Quantum Entanglement, and Entropic Uncertainty Relations

In this chapter, we discuss the fundamental concepts and relations in classical and quantum information theory forming the background for this thesis. With these concepts understood, we discuss entropic uncertainty relations, information exclusion relations, and how they both inform our understanding of quantum measurement.

1.1 Introducing Shannon Entropy

From fire signals, to optical fibers, the problem of communication over long distances is unchanged in its essence; what is the minimum amount of resources needed to send a message?

The standard meter of information is the bit: the binary digit, or the answer to a two-state (e.g., yes or no) question. The standard measure of how much information a message has is in the smallest number of bits it takes to communicate

it to someone who has no knowledge of it. Information theory, the mathematical theory of communication, examines this question in the following context: how many bits do you need to communicate the outcome of a random variable X with probability distribution $P(X)$?

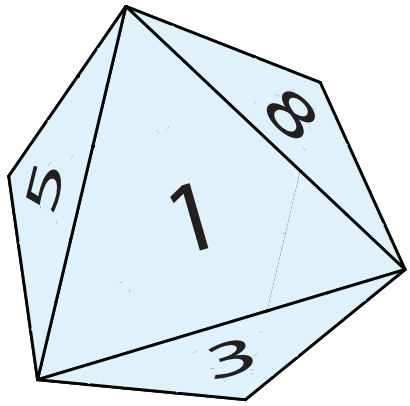

Figure 1.1: An eight-sided die, D8, or octahedral die.

Consider an eight-sided loaded die, with outcomes

$$X = \left\{ \begin{array}{cccc} \text{“side 1”} & \text{“side 2”} & \text{“side 3”} & \text{“side 4”} \\ \text{“side 5”} & \text{“side 6”} & \text{“side 7”} & \text{“side 8”} \end{array} \right\} \tag{1.1}$$

and corresponding probabilities

$$P(X) = \left\{ \begin{array}{cccc} \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} \\ \frac{1}{32} & \frac{1}{64} & \frac{1}{128} & \frac{1}{128} \end{array} \right\}. \tag{1.2}$$

The minimum number of bits needed to communicate a particular outcome of X is the number of digits you need to uniquely label all outcomes (i.e., the base-2 logarithm of the number of possible outcomes rounded up to the nearest integer). The eight-sided die has eight possible outcomes, which take three bits to label them all. With only one bit, one can communicate whether the outcome is in one of two sets (e.g., in outcomes 1-4, or in outcomes 5-8). With this information (say, that it is within outcomes 1-4), the second bit can communicate whether the outcome is in one of two subsets of the known set (within 1-2 or within 3-4).

With this additional information (say, that it is within outcomes 3-4), the third bit can distinguish between the last two possible outcomes. Though it might seem that one always needs three bits to uniquely describe the outcome of X, one can actually do significantly better on average.

Let us say someone rolls the eight sided loaded die many times, and we ask, "How many bits does it take *on average* to know what the outcome of X is?" For the probability distribution $P(X)$, if we ask the following questions in order; "Is it $X =$'side 1'?";"Is it $X =$'side 2'?";"Is it $X =$'side 3'?", and so forth, we only need one bit half of the time, two bits a quarter of the time, three bits an eighth of the time, and so on, which surprisingly adds to an average number of $\frac{127}{64}$ or about 1.98 bits per roll. This minimum average number of bits needed to communicate the outcome of X is such an important quantity in information theory that it is given its own name, the discrete Shannon entropy, $H(X)$:

$$H(X) \equiv -\sum_i P(X_i)\log(P(X_i)) \;=\; \left\langle \log\left(\frac{1}{P(X)}\right)\right\rangle_{P(X)}. \tag{1.3}$$

The discrete Shannon entropy of X is equal to the number of bits it takes on average to communicate the outcome of X. Note that here and throughout this thesis, all logarithms are assumed to be base two unless otherwise specified.

For probability distributions less contrived than the eight-sided loaded die, it is still proven (in Shannon's original paper [1]) that in the limit of many trials, one needs on average $H(X)$ bits per trial to communicate the outcome of X. The proof of this is beyond the scope of this thesis, but the idea is well captured by the Asymptotic Equipartition Property (AEP) theorem [2] (coming from the weak law of large numbers). What the AEP theorem tells us is that given long, independent and identically distributed (i.i.d.) sequences of outcomes of a random variable X (say, the outcomes of many rolls of the loaded die), the fraction of possible outcome sequences that are "typical" (i.e., those whose relative frequencies of outcomes differ from the true probabilities by no more than some small fixed

value ϵ) approaches unity (exponentially) as the sequence grows large [2]. In addition, these "typical" sequences are very nearly equiprobable, and the number of bits needed to uniquely label these "typical" outcomes is very nearly equal to $nH(X)$, where n is the number of trials in the sequence, and $H(X)$ is the entropy of X. What this means for us is that for sufficiently long sequences, it becomes overwhelmingly likely that we need only on average $H(X)$ bits to communicate the outcome of X for each trial.

The discrete Shannon entropy is also a well-qualified measure of the intrinsic uncertainty of a random variable. If the entropy is low, it takes fewer bits to determine the outcome than if it is high. A uniform probability distribution has the highest possible entropy, being the logarithm of the number of outcomes. A probability distribution with only one outcome of nonzero probability has zero entropy; there is no uncertainty in its outcome since it will be the same outcome every time over and over again. The Shannon entropy improves upon variance-based measures of uncertainty in that a sharply peaked bimodal distribution has much less intrinsic uncertainty than a singly-peaked distribution of the same variance (see Fig. 1.2).

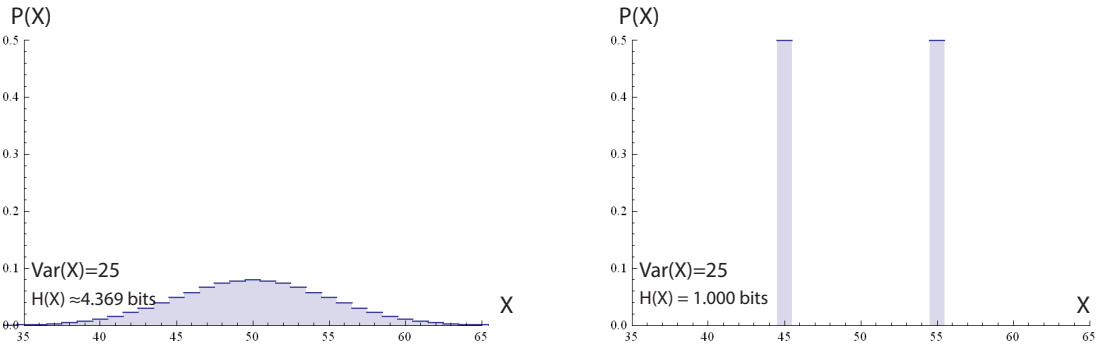

Figure 1.2: Diagram comparing the Shannon entropy of two probability distributions with the same variance $Var(X)$. On the left is the binomial distribution of 100 fair coin tosses. On the right is the minimum entropy distribution of the same mean and variance.

On the other hand, the Shannon entropy is a poor measure of localization (i.e.,

how tightly a distribution of outcomes is clustered about a single point) for the same reason that it is such a good measure of uncertainty. Since the Shannon entropy $H(X)$ is a function of the *probabilities* of the outcomes of X, and not on the outcomes themselves, it is invariant under any permutation or reshuffling of the outcomes of X.

One particularly useful property of the Shannon entropy $H(X)$ is that it is a concave function of the probability distribution $P(X)$. A concave function is a function that is "concave-down" over the domain of the function. What makes this concavity useful comes from Jensen's inequality [2]. Jensen's inequality states that for any concave function $f(x)$, and any weighted average $\langle\cdot\rangle$ with non negative weights $\{w_i\}$ such that they sum to unity, the average of the function never exceeds the function of the average. In particular, where

$$\langle x\rangle \equiv \sum_i w_i x_i, \tag{1.4}$$

Jensen's inequality for concave $f(x)$ is,

$$\langle f(x)\rangle \leq f(\langle x\rangle). \tag{1.5}$$

Since the Shannon entropy is concave in the probability, we know that the entropy of a mixture of probability distributions is never less than the corresponding mixture of the entropies of each distribution. The concavity of the entropy can be used to prove many relations in information theory. In particular, since we can write $P(X)$ as a weighted sum of conditional probabilities,

$$P(X_i) = \sum_j P(X_i, Y_j) = \sum_j P(Y_j)P(X_i|Y_j), \tag{1.6}$$

we get the useful relation

$$H(X) \geq H(X|Y); \tag{1.7}$$

i.e., that conditioning on extra information (the outcomes of another random variable Y), cannot increase the average uncertainty of X.

The discrete Shannon entropy is not the only entropic function discussed in classical information theory. For every kind of probability distribution, there is an associated entropy. Accordingly, we define the joint entropy of random variables X and Y, $H(X,Y)$, as

$$H(X,Y) \equiv -\sum_{i,j} P(X_i,Y_j)\log(P(X_i,Y_j)) \;=\; \left\langle \log\left(\frac{1}{P(X,Y)}\right)\right\rangle_{P(X,Y)}, \quad (1.8)$$

and define the conditional entropy of Y given X, $H(Y|X)$, as

$$H(Y|X) \equiv -\sum_{i,j} P(X_i,Y_j)\log(P(Y_j|X_i)) \;=\; \left\langle \log\left(\frac{1}{P(Y|X)}\right)\right\rangle_{P(X,Y)}. \quad (1.9)$$

With Bayes' rule,

$$P(X,Y) = P(X)P(Y|X) = P(X|Y)P(Y), \quad (1.10)$$

it is straightforward to show that

$$H(X,Y) = H(X) + H(Y|X). \quad (1.11)$$

This leads us to the intuition that the number of bits it takes to learn both X and Y is equal to the bits needed to learn X plus the remaining number of bits needed to learn Y, given that the outcome of X is already known. Consequently, the difference between $H(X)$ and $H(X|Y)$ is equal to the number of bits learned about X by learning the outcome of Y. This difference is called the mutual information $H(X:Y)$,

$$H(X:Y) \equiv H(X) - H(X|Y) = H(X) + H(Y) - H(X,Y). \quad (1.12)$$

The mutual information is a highly important entropic function because it describes the capacity for sending information between a sender whose input symbols may be described by the outcomes X and a receiver whose output symbols may be described by the outcomes of Y, and vise versa. Indeed, when optimizing over all possible inputs X, Shannon showed [1] that the mutual information $H(X:Y)$

is the maximum rate (on average) of error free communication on an input-output channel, whose transmission probabilities are described by $P(Y|X)$. To see how this might be the case, we can consider long series of inputs X and outputs Y of length n, and invoke the AEP theorem. For large n, there are approximately $2^{nH(Y)}$ typical output sequences, and $2^{nH(Y|X)}$ typical outcome sequences for each typical input sequence. The ratio of these two, $2^{nH(X:Y)}$ is the number of sets of output sequences that can be distinguished by knowing which typical input sequence is given. Since an input sequence of length n narrows down the typical output sequence to be within one of $2^{nH(X:Y)}$ sets, one can send on average at most $2^{nH(X:Y)}$ inputs that can be received with low probability of error. Thus, $H(X:Y)$ is the average number of bits that can be communicated in a single use of a channel whose inputs and outputs are described by $P(X,Y)$.

In addition, the mutual information $H(X:Y)$ is an extremely general measure of correlation, capturing any manner of statistical dependence between X and Y, and not just linear dependence as the covariance does. This works because (as mentioned previously) the entropy of a random variable depends only on its set of probabilities (and not on the outcomes to which they refer); the entropy of a probability distribution is invariant to permutations of its outcomes.

The entropic functions previously defined are all defined with respect to a single probability distribution (whether that be of one or many variables). There is one additional entropic function used when comparing two different probability probability distributions, say, $P(X)$ and $Q(X)$. This entropic function comparing $P(X)$ and $Q(X)$ is known as the relative entropy $D(P||Q)$ [3], and is defined as:

$$D(P||Q) \equiv \sum_i P(X_i) \log\left(\frac{P(X_i)}{Q(X_i)}\right) = \left\langle \log\left(\frac{P(X)}{Q(X)}\right)\right\rangle_{P(X)} \tag{1.13}$$

This relative entropy $D(P||Q)$ is asymmetric between P and Q, and is also known as the Kullback-Leibler divergence of Q from P. It is especially useful because: all the previous entropies can be expressed in terms of relative entropies (where

the odd distribution Q would be taken to be a uniform distribution), e.g.,

$$H(X:Y) = D(P(X,Y)||P(X)P(Y)). \tag{1.14}$$

In addition, the relative entropy is non-negative, i.e., $D(P||Q) \geq 0$, and the maximum likelihood fit of a distribution to a raw data set is the one with minimum relative entropy (with P represented by the fitting distribution, and Q representing the data).

1.1.1 The Continuous (Differential) Shannon Entropy

The discrete Shannon entropy, $H(X)$, describes the uncertainty of discrete random variables very well. To characterize the uncertainty of continuous random variables, a differential or continuous Shannon entropy was developed (also in Shannon's original paper [1]) by essentially taking a high-resolution limit of the discrete entropy, and neglecting a constant offset.

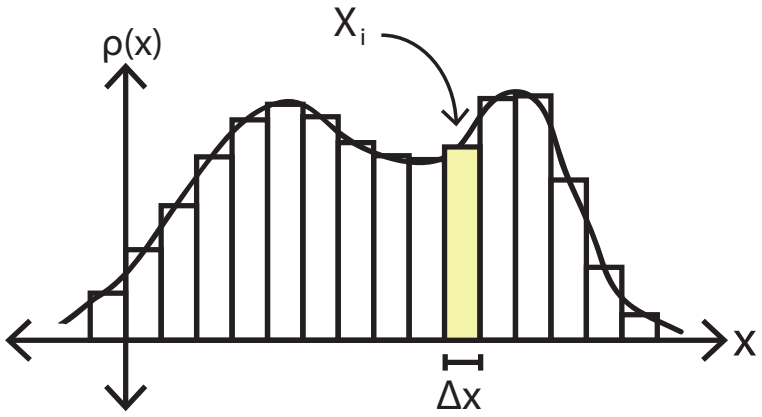

Figure 1.3: Diagram representing the quantization of a continuous probability density $\rho(x)$ into windows X_i of width Δx.

For any continuous random variable x described by density function $\rho(x)$, whose outcome may be any real number, we define the continuous Shannon en-

tropy $h(x)$:

$$h(x) \equiv -\int dx\ \rho(x)\log(\rho(x)). \tag{1.15}$$

This continuous entropy can be defined in analogy to the discrete Shannon entropy by using the AEP theorem and the weak law of large numbers, but we consider the density function $\rho(x)$ quantized into regular windows X_i of size Δx (see Fig. 1.3).

Where

$$\rho(x)dx = \lim_{\Delta x \to 0} P(X_i), \tag{1.16}$$

and $P(X_i)$ is the density function, integrated over all values of x within window X_i, we have that [1]

$$h(x) = \lim_{\Delta x \to 0} \Big(H(X) + \log(\Delta x) \Big). \tag{1.17}$$

The final term is an offset that is independent of the density function; when comparing uncertainties of continuous variables, it suffices only to compare their continuous entropies, $h(x)$.

The continuous Shannon entropy follows rules similar to the discrete Shannon entropy. For example, we have the identities:

$$h(x,y) = h(x) + h(y|x), \tag{1.18}$$

$$h(x:y) = h(x) - h(x|y) = h(x) + h(y) - h(x,y), \tag{1.19}$$

$$D(\rho||\eta) = \int dx \rho(x) \log\left(\frac{\rho(x)}{\eta(x)}\right), \tag{1.20}$$

and

$$h(x) \geq h(x|y), \tag{1.21}$$

in analogy to discrete entropies.

[1] Note: where Δx has a dimension of length, the value of $\log(\Delta x)$ depends on one's length scale. Indeed, the Shannon entropy of a probability density shifts by a constant amount over a uniform rescaling of the probability density. In order to compare the entropies of two different probability densities, one must be sure to use the same length scale for both.

In addition, we have that the maximum continuous entropy is that of a uniform distribution, though this limit is arbitrarily large when x has no upper or lower bounds. However, the continuous entropy is markedly different from the discrete entropy in that it can take negative values because the density function $\rho(x)$ can have values above unity (so long as it still integrates to unity). This means, for example, that $h(x, y)$ can be smaller than either $h(x)$ or $h(y)$ in spite of its being a measure of uncertainty for a pair of random variables.

1.1.2 Relating Continuous and Discrete Shannon Entropy

In order to make sense of continuous-variable degrees of freedom in physical experiments, we must be able to relate discrete measurement data to continuous-variable statistics. Though an arrow on a dial may have any angular position between two limits, recording its position requires that we quantize its position up to a given resolution.

In the research discussed in later chapters, we make use of relating the continuous entropy $h(x)$ to the discrete entropy $H(X)$ of a quantization of x up to a resolution Δx. To express the continuous entropy in terms of the discrete entropy, we cut up the continuous density function $\rho(x)$ into equally spaced regions of size Δx. The continuous density function is then expressed as:

$$\rho(x) = \sum_i P(X_i)\rho(x|X_i) \tag{1.22}$$

where $\rho(x|X_i)$ is the probability density of the measurement outcome of x, given that it is within window X_i. Using this quantization, we can express the continuous entropy in a similar way:

$$h(x) = \sum_i P(X_i)h(x|X_i) + H(X) \tag{1.23}$$

where $h(x|X_i)$ is the continuous entropy of x, conditioned on x being within window X_i. Since the largest possible entropy for a continuous random variable bounded between two limits is that of the uniform distribution, we find:

$$h(x|X_i) \leq \log(\Delta x), \tag{1.24}$$

and using this in our relation between discrete and continuous entropies (1.23), we prove that

$$h(x) \leq H(X) + \log(\Delta x). \tag{1.25}$$

This relation was already known in [4, 5], and it agrees with the intuition that the entropy of the discrete approximation should not be smaller than the entropy of the thing itself. In Chapter 4, we prove that a similar relation holds true for arbitrary conditional entropies:

$$h(x_B|x_A) \leq H(X_B|X_A) + \log(\Delta x_B), \tag{1.26}$$

and then use this result to demonstrate the EPR-paradox in the position-momentum statistics of entangled photon pairs.

1.2 Quantum Entropy and Information

While classical information theory is a mathematical theory of communication and data manipulation, quantum information theory is a mathematical theory of quantum processing, and of the manipulation of quantum systems. While the elementary unit of classical information can be expressed in any number of ways from etchings on clay tablets, to etchings on plastic discs, quantum information is stored exclusively in quantum systems. While classical information is measured in bits, quantum information is measured in *qubits*, i.e., elementary two-state quantum systems. While the Shannon entropy $H(X)$ characterizes the uncertainty in a random variable X, the quantum von Neumann entropy [6], $S(A) = S(\hat{\rho}_A)$:

characterizes the intrinsic uncertainty in a quantum system A described by density matrix $\hat{\rho}_A$. Specifically, the von Neumann entropy of system A is defined as:

$$S(A) \equiv -\mathrm{Tr}(\hat{\rho}_A \log(\hat{\rho}_A)) = -\sum_i \lambda_i \log\big(\lambda_i\big), \tag{1.27}$$

i.e., the Shannon entropy of the eigenvalues λ_i of its density matrix $\hat{\rho}_A$.

1.2.1 Pure and Mixed States

Beyond capturing the resources needed to encode $\hat{\rho}_A$, the von Neumann entropy captures the irreducible mixedness of quantum systems. If a quantum system is described by a single state vector $|\psi\rangle$, we say it is in a pure state. The corresponding density operator for this state would be $\hat{\rho} = |\psi\rangle\langle\psi|$. If instead a quantum system is described as a mixture of state vectors $|\psi_i\rangle$ with non-negative weights p_i (adding to unity), we say the system is in a mixed state with density operator $\hat{\rho} = \sum_i p_i |\psi_i\rangle\langle\psi_i|$. If a quantum state $\hat{\rho}_1$ can be expressed as a mixture of a second quantum state $\hat{\rho}_2$, and the maximally mixed state $\hat{\rho}_{MM}$, then $\hat{\rho}_1$ is more mixed than $\hat{\rho}_2$. To understand the irreducible mixedness of a quantum system, we need to point out that any mixed state can be expressed as one of an infinite number of possible pure state decompositions. Indeed, if we consider a spin-1/2 system which has been prepared to be a 50/50 mixture of spin-up and spin-down, we may also express it at a 50/50 mixture of spin-left and spin-right or indeed a 50/50 mixture of any pair of opposite spin directions, and get the same density matrix as a result. To see this geometrically, it helps to consider the Bloch representation of the spin-1/2 system [7]. Any density matrix of a two-state quantum system (or qubit) can be represented as a weighted sum of the identity matrix and the three Pauli spin matrices:

$$\hat{\rho} = \frac{1}{2}\big(\hat{\mathbf{I}} + P_x\hat{\sigma}_x + P_y\hat{\sigma}_y + P_z\hat{\sigma}_z\big), \tag{1.28}$$

where $\vec{\mathbf{P}} = (P_x, P_y, P_z)$ is the Bloch vector encoding all the information about the qubit.

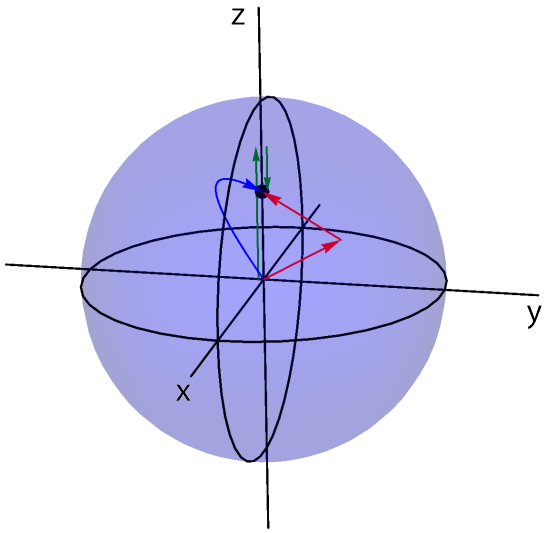

Figure 1.4: Diagram of Bloch sphere representation of pure state decomposition of the state (black point) that is a mixture of 3/4 spin-up, and 1/4 spin-down (green arrows). The red jointed arrow represents a decomposition in terms of two non-orthogonal pure states, while the blue curved arrow represents a decomposition in terms of an infinite number of pure states. In both cases, the arc length of the arrows is one, as the total sum of the weights adds to unity.

This Bloch vector can have any value within the unit sphere in three-dimensional Cartesian space. Pure states have Bloch vectors of unit magnitude, while mixed states have Bloch vectors of smaller magnitudes. If we consider the Bloch vector of a mixed state, a pure state decomposition can be expressed as the weighted sum of the Bloch vectors of the pure states. In that way, we see there are as many pure state decompositions as there are curves of unit arc length connecting the origin of the Bloch sphere to the point giving the mixed state (see Fig. 1.4).

Though there are many pure state decompositions of any mixed state, with the Shannon entropy of the weightings $\{p_i\}$ having no upper limit (one can mix a nearly infinite number of states almost uniformly), there is a unique lower limit to the Shannon entropy of the pure state decomposition, and that is the von Neumann entropy, or the Shannon entropy of the eigenstate decomposition of the

mixed state:

$$H(p_i) \geq S(\hat{\rho}). \tag{1.29}$$

Before going further, it is important to make clear what is meant by the mixedness of a quantum state, and how the von Neumann entropy measures it. To do this, we consider the mixedness of a quantum state to be the inverse of its purity. The purity of a quantum state is generally a measure of how closely the eigenvalue spectrum of powers of a density matrix agree with the eigenvalues of the original density matrix. Since only for pure states can $\mathrm{Tr}[\hat{\rho}^{\alpha}] = \mathrm{Tr}[\hat{\rho}]$ for all non-negative powers of α, we can use the trace of $\hat{\rho}$ to some power α as a measure of the state's purity, and from that, get a measure of its mixedness.

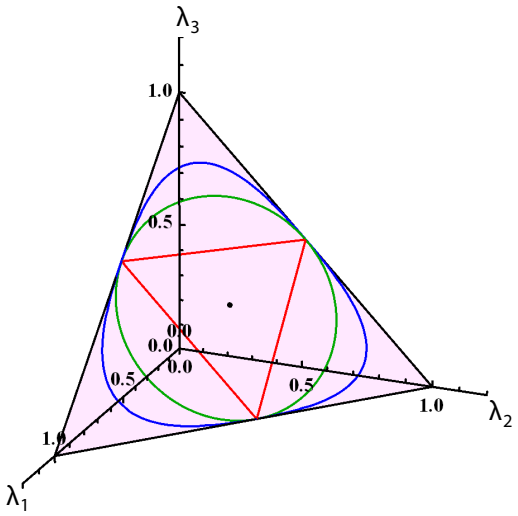

Figure 1.5: Diagram of the plane, $\lambda_1 + \lambda_2 + \lambda_3 = 1$, for the eigenvalues of 3-dimensional quantum systems. The black dot in the middle gives the maximally mixed state. Here, I have plotted four measures of mixedness. The black boundary of this hyperplane gives states with a Hartley entropy of 1 bit (i.e., $H_0(\lambda) = 1$). The outermost (blue) curve gives states with a Shannon entropy of 1 bit (i.e, $H_1(\lambda) = 1$). The circular (green) curve gives states with a Collision entropy of 1 bit (i.e., $H_2(\lambda) = 1$), and the triangular (red) curve gives states with a Min entropy of 1 bit (i.e., $H_{\infty}(\lambda) = 1$).

Though for different powers of α, we get different measures of mixedness, we can group them into a general class of entropies known as Renyi entropies (of the

eigenvalue spectrum) [8]. For an n-dimensional density matrix, the Renyi entropy of order α of its eigenvalue spectrum $\{\lambda_i\}$ is defined as:

$$H_\alpha(\lambda) = \frac{1}{1-\alpha} \log \left(\sum_{i=1}^{n} \lambda_i^\alpha \right). \tag{1.30}$$

With this definition, we note that since pure states have only one nonzero eigenvalue (and that eigenvalue is equal to unity), the Renyi entropy of their eigenvalue spectrum is zero (the minimum value) for all orders α. In addition, The Renyi entropy attains a maximum value for the maximally mixed state for all orders α, and is a monotonically increasing function under any series of mixing operations (i.e., where two or more permutations of the given state are mixed to create a third state). Since this makes the Renyi entropy a satisfactory measure of mixedness for all orders α, the Von Neumann entropy must be as well, since in the limit as $\alpha \to 1$, the Renyi entropy of the eigenvalue spectrum approaches the von Neumann entropy.

1.2.2 Quantum Entropic Functions and Identities

Just as there are many classical entropic functions based on probability distributions, there are similarly useful quantum entropic functions defined for density matrices. For a bipartite quantum state $\hat{\rho}_{AB}$, there is the joint quantum entropy

$$S(A,B) = S(\hat{\rho}_{AB}) \equiv -\mathrm{Tr}(\hat{\rho}_{AB} \log(\hat{\rho}_{AB})), \tag{1.31}$$

and the marginal quantum entropy,

$$S(A) = S(\hat{\rho}_A) \equiv -\mathrm{Tr}(\hat{\rho}_A \log(\hat{\rho}_A)), \tag{1.32}$$

where $\hat{\rho}_A$ is defined as the partial trace of $\hat{\rho}_{AB}$ over the states of B. In addition, there is also the conditional quantum entropy,

$$S(A|B) \equiv S(A,B) - S(B), \tag{1.33}$$

and the quantum mutual information,

$$I(A:B) \equiv S(A) + S(B) - S(A,B). \tag{1.34}$$

Note that here, and throughout the paper, H will refer to classical entropies and mutual informations, while S and I will refer to quantum entropies and quantum mutual informations.

Since $S(A)$ represents the Shannon entropy of the eigenvalue spectrum of a system A (i.e., its Von Neumann entropy), it is also the smallest possible measurement uncertainty of any observable $\hat{Q}_A$ on A;

$$S(A) = \min_{\hat{Q}_A} H(Q_A). \tag{1.35}$$

This is because mixing operations increase the entropy of a probability distribution, and the measurement probabilities of the outcomes of Q_A are mixtures of the probability eigenvalues λ_i (see (1.27)) of the density operator $\hat{\rho}_A$ of system A: e.g.,

$$P(Q_A = 1) = \sum_i \lambda_i |\langle q_{A,(1)} | \lambda_i \rangle|^2, \tag{1.36}$$

where $|q_{A,(1)}\rangle$ is the eigenstate corresponding to the outcome $Q_A = 1$. That is, the measurement probability distribution can be expressed as a result of a mixing operation on the eigenvalue probability distribution, and so will have to be more mixed, and have higher (or at least not lower) entropy.

Unlike classical information theory, there is no formal distinction between the quantum entropies of continuous and discrete variable systems. Since a wavefunction of a particle may always be written as a discrete (though possibly infinite) sum of eigenfunctions (or orthogonal modes), one can always express a continuous-variable quantum system in a discrete basis, and use the same formalism as one would for finite-dimensional systems.

Because pure joint quantum states can be entangled, quantum entropies obey slightly different constraints than classical entropy. For starters, if the joint state

of AB, $\hat{\rho}^{AB}$ is a pure state, then the quantum joint entropy $S(A,B) = 0$, while the quantum marginal entropies $S(A)$ and $S(B)$ may have some positive value. In this case, the quantum conditional entropy $S(B|A)$ is such that

$$S(B|A) \leq 0, \tag{1.37}$$

which is not possible for classical entropies, where the joint entropy $H(X,Y)$ is an upper bound for the marginal entropies $H(X)$ and $H(Y)$. When $\hat{\rho}^{AB}$ is a separable state (1.38), the quantum conditional entropy $S(A|B)$ is non-negative [6], in accord with our classical intuitions. As a result, whenever $S(B|A) < 0$, we know the state $\hat{\rho}_{AB}$ must be entangled.

1.3 Quantum Entanglement: Measures and Witnesses

As a large portion of this thesis deals in one form or another with quantum entanglement, it is important to explain: what we talk about when referring to quantum entanglement; what we are measuring when we measure the entanglement in a quantum state; and what we are witnessing when we witness entanglement.

Consider the following. A pair of sufficiently well-prepared quantum particles A and B may be described by a single pure quantum state $|\Psi\rangle_{AB}$. If these particles were prepared independently, then we may describe A and B by individual pure quantum states, such that their joint state $|\Psi\rangle_{AB}$ factors into the tensor product $|\psi\rangle_A \otimes |\psi\rangle_B$. This is the simplest kind of separable state known as a product state.

If these particles A and B are evolving in time according to a Hamiltonian $\hat{H}$ that acts on each particle independently (so that $\hat{H} = \hat{H}_A \otimes \hat{H}_B$), then we may describe the evolution of each particle by its own Schrödinger equation, and be assured that (within the context of quantum mechanics) both A and B will

have well-defined pure states at any time. Such factorizeable Hamiltonians are non-interacting in that the evolution of the state of B is independent of the state of A (and the "environment"), and vise versa.

If the pair of particles A and B are instead interacting with one another, so that its evolution is not described by a factorizeable Hamiltonian, then the evolution of the state of A may actually depend on the state of B. Over time, though the pair AB may always be described by a joint pure quantum state, the joint state $|\Psi\rangle_{AB}$ may no longer factor into a product of pure states $|\psi\rangle_A \otimes |\psi\rangle_B$. When the joint state $|\Psi\rangle_{AB}$ is no longer separable in this way, we say the state is *entangled*.

1.3.1 Separability and Entanglement

Nonseparability is synonymous with entanglement. Previously, we considered a pair of quantum systems A and B described by a single pure quantum state $|\Psi\rangle_{AB}$, but a complete picture of quantum entanglement includes the entanglement of mixed states as well. In the process we considered, we understood that if a pair of particles is prepared independently, the state describing the pair must be separable. The most general case of a separable state [9, 10] is a mixture of independently prepared pairs of particles whose preparation procedure may be mediated by local operations on each member of each pair, and classical communication. This criterion of local operations and classical communications is abbreviated as LOCC for simplicity, and is used in defining non-entangled, or separable states. The density operator for a separable state $\hat{\rho}_{AB(sep)}$ is generally expressed as:

$$\hat{\rho}_{AB(sep)} = \sum_i p_i \big(\hat{\rho}_{A,i} \otimes \hat{\rho}_{B,i}\big), \tag{1.38}$$

where p_i are non-negative weights that sum to unity. Note that with a suitable choice of weightings, any separable state may be expressed as a weighted mixture of products of pure states as well. Any quantum state for a pair of quantum

systems not expressible in the form (1.38) must also not have been derived exclusively from LOCC operations on independent pairs of systems. In this sense, such a state is entangled. Using LOCC operations, we may also define varieties of separable states for multi-partite systems, and from that, varieties of multi-partite entanglement [11], though a discussion of this is beyond the scope of this thesis.

1.3.2 Witnessing Quantum Entanglement

Now that we know precisely what it means for a state $\hat{\rho}_{AB}$ to be separable, we can consider what tests there are that rule out a given state from being separable. Broadly speaking, entanglement witnesses [12] are criteria that all separable states satisfy. When a state violates an entanglement witness, it is definitely entangled, though the converse is not necessarily true, since there may be entangled states that do not violate a particular entanglement witness. However, for every entangled state, there is some entanglement witness that will succeed in witnessing its entanglement.

To see how we might witness the entanglement of any entangled state, we consider that the set of separable states is convex; all mixtures of separable states are separable states. In Euclidean 3D space, a convex set is some simply connected (i.e., contiguous and un-holey) blob whose surface has everywhere non-negative curvature, like an egg, sphere, or any other surface where the plane tangent to every point on the boundary of that set lies outside that set. For every point outside that convex set, there always exists some other plane where the point will be on one side, and the convex set will be on the other. Since such a plane always exists, there is always some vector defining that plane, whose inner product with every vector in the convex set will be non-negative, while for the point outside the set, will be negative (see Fig. 1.6 for a diagram of the 2D version of this). Similarly, in the vector space of density matrices, there also always exists some

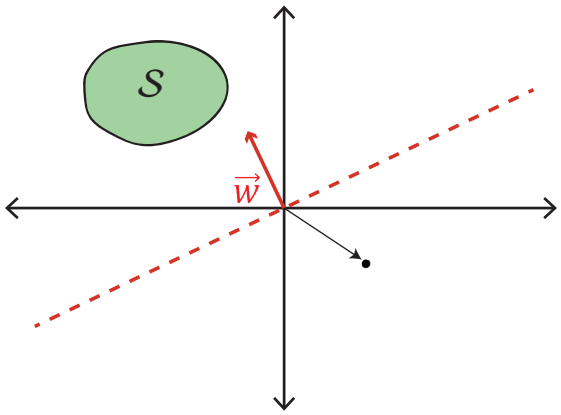

Figure 1.6: Diagram of a convex set S in $2D$ Euclidean space. The black dot is a vector outside that convex set, while the vector $\vec{W}$ defines a plane (dotted line) separating S from the point outside that set.

observable $\hat{W}$ of the joint system AB whose expectation value will be non-negative for all separable states, but be negative for the entangled state in question [2].

Although it is in principle possible to tell whether or not a given state $\hat{\rho}_{AB}$ is entangled, the actual determination of this is currently understood to be an NP-hard problem [13]. Fortunately, there are entanglement witnesses that witness entanglement in a wide variety of states. As one example [14], the expectation value of the observable $\hat{\sigma}_x \otimes \hat{\sigma}_x + \hat{\sigma}_y \otimes \hat{\sigma}_y + \hat{\sigma}_z \otimes \hat{\sigma}_z$, must be between -1 and 1 for all separable (two-qubit) states, while its value is 3 for the maximally entangled Bell singlet state. However, its value is 1 for the other three Bell states; even maximally entangled states are not necessarily detected by a particular entanglement witness.

Beyond single witness observables, there are also a wide variety of inequalities witnessing the entanglement using other statistics of observables. EPR-steering inequalities and Bell inequalities (discussed in Chapter 2), are examples of these, though there are still multiple other inequalities exclusively witnessing the non-

[2]Indeed, the observable $\hat{W}$ need not have a negative expectation value to witness entanglement; it must, however, have an expectation value outside the range of values achievable for separable states.

separability of systems, such as is seen in [15–17]. As one final example, we previously mentioned that a negative quantum conditional entropy (1.37) witnesses entanglement. This particular witness requires knowledge of the density matrix, as opposed to knowledge of the outcomes of measurements of some set of observables. If one knows the joint density matrix, one can test whatever entanglement witness one likes. Indeed, with the joint density matrix, one has complete knowledge of the quantum state measured, and one can in some sense measure how entangled the state is (as discussed in the following section).

1.3.3 Measuring Quantum Entanglement

While it is straightforward to understand the distinction between separable and entangled states, there is also a manner of speaking in which it makes sense to say that some states are more entangled than others (i.e., that there are degrees of entanglement beyond "yes" and "no").

Mathematically, sources generally agree [11, 14] that a measure of entanglement (also known as an entanglement monotone) has the following properties. First, it is a continuous function of the joint density matrix $\hat{\rho}_{AB}$ because infinitesimal perturbations of the density matrix ought to infinitesimally affect the amount of entanglement it contains. Second, it has a minimum (usually zero) value for (and only for) separable states [3]. Third, it is invariant to local unitary transformations, because the degree of entanglement between an isolated pair of particles should not change if the particles are forbidden from (further) interaction. Fourth, it is non-increasing under LOCC operations. This last criterion makes sense in that it allows us to rank any set of entangled states based on whether we can obtain one from another through LOCC operations. Given any pair of joint states

[3]There are some sources that do not require that an entanglement measure of zero implies separability. In these cases, there are relatively easy-to-compute measures of entanglement, such as the negativity, though these fail to capture the entanglement of bound-entangled states.

$\hat{\rho}$ and $\hat{\eta}$, it is never the case that neither state can be converted into the other though LOCC operations [14].

Practically, it is interesting to consider that although one can construct measures of entanglement, it is harder to agree on precisely what is being measured. A large school of thought considers entanglement to be a resource consumed in quantum information protocols. Indeed, as a qubit is an elementary two-state system in quantum information, an *ebit* [6] is an elementary two-qubit maximally entangled Bell state. One can characterize entanglement-based quantum information protocols in terms of the average number of ebits consumed in a channel use. In fact, among measures of entanglement, the entanglement of formation [18] characterizes the number of ebits on average needed to create a given quantum state through LOCC operations, while the distillable entanglement [19, 20] characterizes the mean number of ebits that can be distilled out of many copies of a given quantum state. An alternative school of thought considers entanglement to be a geometric property of quantum states. One measure of entanglement along this line would be the distance (according to some metric) to the nearest separable state [21]. In either school of thought, all the essential properties for a measure of entanglement are satisfied, though information-theoretic measures find more immediate application.

Although measuring entanglement is generally an NP-hard problem (as it is for even determining if an arbitrary state is separable), the problem of measuring entanglement is remarkably straightforward when the joint state is pure. Indeed, there are many measures of entanglement worked out, depending only on the eigenvalues of reduced density matrix $\hat{\rho}_A$ of a joint system described by $|\Psi\rangle_{AB}$. In particular, the reduced von Neumann entropy, $S(A)$ is a measure of the entanglement of $|\Psi\rangle_{AB}$. It is a continuous function of $|\Psi\rangle_{AB}$; it is zero for separable states; it is invariant to local unitary transformations (since it depends only on the eigenvalues of $\hat{\rho}_A$); and it is non-increasing under LOCC operations. Indeed,

there are many entropies of entanglement to choose from since the Renyi entropy of the eigenvalues of $\hat{\rho}_A$ (of which $S(A)$ is a particular case) is also a viable measure of entanglement. One measure of entanglement particularly useful for pure continuous-variable states is known as the Schmidt number κ [22, 23] (also known as the Inverse Participation Ratio);

$$\kappa = \frac{1}{\mathrm{Tr}[(\hat{\rho}_A)^2]} = \frac{1}{\mathrm{Tr}[(\hat{\rho}_B)^2]} \tag{1.39}$$

Though the Schmidt number [4] is not zero for separable states, a simple scaling and shift will suffice to bring it into accord with the standard conventions. What is particularly nice about the Schmidt number is that one does not need to calculate the eigenvalues of e.g., $\hat{\rho}_A$, in order to take a trace of it raised to an integer power. $S(A)$ does not have this convenience, but has been more useful in information-theoretic applications.

1.4 Entropic Uncertainty Relations and Their Consequent Relations

In this section, we explore the uncertainty principle as expressed in entropic terms. We discuss the continuous variable uncertainty relations of Bialynicki-Birula [25], the discrete uncertainty relations of Deutsch [26], later improved by Maassen and Uffink [27], and later again by Berta *et al.* [28], and more still by others. In addition, we discuss how these entropic uncertainty relations inform our understanding of quantum uncertainty. In particular, we discuss information exclusion relations as a consequence of entropic uncertainty relations.

[4] The Schmidt number κ is not to be confused with the Schmidt rank [24], which is the number of nonzero eigenvalues of $\hat{\rho}_A$ or $\hat{\rho}_B$. The Schmidt number and Schmidt rank are related in that the logarithms of these are the quantum Renyi entropy (1.30) of order 2 and 0, respectively. Recall, the quantum Renyi entropy of order 1 is the von Neumann entropy $S(A)$.

The uncertainty principle was originally expressed by Heisenberg [29] in terms of the variances σ^2 of the outcomes of measuring the position x and momentum k of a single quantum particle, i.e.,

$$\sigma_x^2 \sigma_k^2 \geq \frac{1}{4}. \tag{1.40}$$

This uncertainty principle applies not just to quantum wavefunctions, but indeed any waveform (such as classically described sound and light) admitting a Fourier decomposition. Shortly thereafter, this uncertainty relation was generalized to apply to any pair of observables [30].

$$\sigma_Q^2 \sigma_R^2 \geq \frac{1}{4} |\langle [\hat{Q}, \hat{R}] \rangle|^2, \tag{1.41}$$

where the right-hand side is one-fourth of the modulus square of the expectation of the commutator of the pair of observables. Heisenberg uncertainty relations are still ubiquitous to this day, since the variance is a solid, statistical measure of localization. Entropy is an insensitive undiscerning measure of localization since the entropy of a random variable is invariant to permutations of the outcomes in the probability space. However, uncertainty relations based on entropic "uncertainty" are much more prevalent in information-theoretic applications because they directly express one's measurement limitations in terms of the actual information that can be obtained. With entropic uncertainty relations, we can learn how much information we can send using quantum systems (e.g., photons) [31], how much information an adversary can gain with a wiretap [32, 33], how well we can preserve nonclassical correlations in a quantum channel, and so forth.

1.4.1 Historical Development of Entropic Uncertainty Relations

The first entropic uncertainty relation was proved for position and momentum in 1957 by Hirschmann [34]

$$h(\vec{x}) + h(\vec{k}) \geq 0, \tag{1.42}$$

and improved independently in 1975 by Białynicki-Birula and Mycielski [25], and Beckner [35] to;

$$h(\vec{x}) + h(\vec{k}) \geq n \log(\pi e), \tag{1.43}$$

where n is the number of spatial degrees of freedom (e.g., 3 for 3-D space)[5].

Shortly thereafter, entropic uncertainty relations were developed for discrete observables. In 1983, David Deutsch [26] developed the first entropic uncertainty relation for discrete observables;

$$H(\hat{Q}) + H(\hat{R}) \geq 2 \log \left(\frac{2}{1+c} \right) \tag{1.44}$$

where

$$c = max_{i,j} |\langle q_i | r_j \rangle| \tag{1.45}$$

Here, $|\langle q_i | r_j \rangle|$ is the magnitude of the inner product between the i^{th} eigenstate of $\hat{Q}$ and the j^{th} eigenstate of $\hat{R}$, and $H(\hat{Q})$ is the Shannon entropy of the measurement probability distribution of the eigenstates of $\hat{Q}$ for the quantum state being measured. When $\hat{Q}$ and $\hat{R}$ are mutually unbiased, a system prepared in an eigenstate of $\hat{Q}$ is equally likely to be measured in any of the eigenstates of $\hat{R}$, and vise versa. When $\hat{Q}$ and $\hat{R}$ are N-dimensional mutually unbiased observables, c can be no larger than $1/\sqrt{N}$. Deutsch's uncertainty relation is not tight when observables $\hat{Q}$ and $\hat{R}$ are mutually unbiased. By tight, we mean that there are

[5]Note: Although one could imagine choosing position and momentum length scales so as to make $h(\vec{x})$ and $h(\vec{k})$ arbitrarily small, the Fourier transform imposes an intrinsic length scale on momentum related inversely to the length scale in position.

no quantum states which saturate this inequality for an optimal choice of observables $\hat{Q}$ and $\hat{R}$. In 1988, Maassen and Uffink [27] significantly improved Deutsch's uncertainty relation, coming up with

$$H(\hat{Q}) + H(\hat{R}) \geq \log\left(\frac{1}{c^2}\right) \equiv \log(\Omega), \tag{1.46}$$

where we implicitly define Ω for later convenience. This uncertainty relation is saturated when $\hat{Q}$ and $\hat{R}$ are mutually unbiased, provided the system being measured is in a pure state.

In 2010, Berta *et al.* [28] improved upon Maassen and Uffink's entropic uncertainty relation by accounting for the intrinsic uncertainty of the quantum system itself given by its von Neumann entropy. They formulated the uncertainty principle in the presence of quantum memory:

$$S(\hat{Q}_A|B) + S(\hat{R}_A|B) \geq \log(\Omega) + S(A|B). \tag{1.47}$$

Here, $S(\hat{Q}_A|B)$ is the conditional quantum entropy of the system AB, after observable $\hat{Q}$ of A has been measured. If $\hat{Q}$ of A is measured, the state $\hat{\rho}_{AB}$ becomes $\hat{\rho}_{AB'} = \sum_i (\hat{\Pi}^i_Q \otimes \hat{I})\hat{\rho}_{AB}(\hat{\Pi}^i_Q \otimes \hat{I})$, where $\hat{\Pi}^i_Q$ is the projection onto the i^{th} eigenstate of $\hat{Q}$, (i.e., $|q_i\rangle\langle q_i|$), and the index i runs over all the eigenstates of $\hat{Q}$. Then, $S(A|B)$ for $\hat{\rho}_{AB'}$ is equal to $S(\hat{Q}_A|B)$ for $\hat{\rho}_{AB}$. In the scenario leading to the development of (1.47), the experimenter wishes to determine the observables $\hat{Q}$ and $\hat{R}$ of system A, but has the advantage of possessing an additional system B entangled, or simply correlated with A. In this case, she possesses additional side information since she can measure an observable of B to infer the measurement outcome of the corresponding observable of system A. When systems A and B are uncorrelated, the uncertainty relation reduces to

$$H(\hat{Q}_A) + H(\hat{R}_A) \geq \log(\Omega) + S(A), \tag{1.48}$$

in which case, B holds no information about A. Other improvements have been made since then for different pairs of observables [36], but this uncertainty relation remains the state-of-the-art for mutually unbiased pairs of observables.

These entropic uncertainty relations describe how well a pair of noncommuting observables can simultaneously be determined. That the lower limit is not zero (except when they do commute) means that there is no way that a system can be prepared so as to have well-defined (i.e., predictable) values for both $\hat{Q}$ and $\hat{R}$. What makes these entropic uncertainty relations especially useful is that we can use them to develop other relations describing tasks involving quantum systems to place our intuitions on solid ground. In the next subsection, we will develop information exclusion relations [37, 38], showing how measurements that gain information about one observable must limit the information that can be gained about complementary observables of the same system.

1.4.2 Consequence: Information Exclusion Relations

Let θ be a random variable representing the outcome of one's measurement device. Then for some observable $\hat{Q}$ of the N-dimensional system,

$$H(\hat{Q}|\theta) = \sum_i P(\theta_i) H(\hat{Q}|\theta_i). \tag{1.49}$$

With an additional observable $\hat{R}$, related to $\hat{Q}$ by the uncertainty relation:

$$2\log(N) \geq H(\hat{Q}) + H(\hat{R}) \geq \log(\Omega), \tag{1.50}$$

we have that:

$$H(\hat{Q}|\theta) + H(\hat{R}|\theta) \geq \log(\Omega). \tag{1.51}$$

In other words, the more a measurement θ reduces one's uncertainty about the outcome of an observable $\hat{Q}$, the less that same measurement can also reduce the measurement uncertainty of another observable $\hat{R}$, and that limit depends on how uncertain they are with respect to each other (as captured by Ω). This tradeoff is more easily seen when one expresses the information exclusion relation in terms of the mutual information $H(\hat{Q} : \theta)$, such that:

$$H(\hat{Q} : \theta) = H(\hat{Q}) - \sum_i P(\theta_i) H(\hat{Q}|\theta_i). \tag{1.52}$$

By using both the upper and lower bounds for the uncertainty relation (1.50), we can get a weaker, but more transparent information exclusion relation:

$$H(\hat{Q}:\theta) + H(\hat{R}:\theta) \leq \log\left(\frac{N^2}{\Omega}\right). \tag{1.53}$$

Here we see directly that the more information a measurement θ obtains about $\hat{Q}$, the less information it can also obtain about $\hat{R}$. Similarly, for continuous obervables $\hat{x}$ and $\hat{k}$ related by the uncertainty relation:

$$\log(2\pi e\sigma_x\sigma_k) \geq h(\hat{x}) + h(\hat{k}) \geq \log(\pi e), \tag{1.54}$$

we may obtain the information exclusion relations:

$$h(\hat{x}|\theta) + h(\hat{k}|\theta) \geq \log(\pi e), \tag{1.55}$$

and

$$h(\hat{x}:\theta) + h(\hat{k}:\theta) \leq \log(2\sigma_x\sigma_k) \tag{1.56}$$

These information exclusion relations illustrate how any measurement strategy is limited in the information it can obtain about both $\hat{x}$ and $\hat{k}$. While most measurement strategies focus on obtaining a maximum amount of one observable at the expense of the other, there are other measurement strategies that obtain only a small amount of information about one observable, minimizing the disturbance on the statistics of the other observable. Partial projective measurements, discussed later in this thesis, can be used to obtain simultaneously the position and momentum probability distributions of photons without violating the uncertainty principle. In particular, we will see that since these measurements only obtain a small amount of information about the position of a photon, the maximum information about the momentum of a photon that may yet be obtained is reduced only by that small amount. Using an ensemble of these measurements allows us to measure both the position and momentum statistics of a field of photons with high accuracy with the same set of measurements.

2 Preliminaries on Einstein-Podolsky-Rosen Steering, and Bell Nonlocality

In this chapter, we discuss the fundamentals of the original Einstein-Podolsky-Rosen (EPR) paradox, and show how it leads into the development of Local Hidden Variable (LHV) models, and their renunciation by violating Bell inequalities. In addition, we will discuss how the EPR-scenario led directly into the concept of quantum steering, and how this has recently taken off in the development of the new class of inequalities known as EPR-steering inequalities.

2.1 The Eintein-Podolsky-Rosen Scenario

Eighty years ago, three scientists: Albert Einstein, Boris Podolsky, and Nathan Rosen [39] sought to show that although quantum mechanics was a "correct" model of reality, it was not a complete one. To be a correct theory was to be a theory that agreed with experimental results. To be a complete theory was to be a theory where every "element of reality" had a corresponding element in the theory. Fortunately, they did not have to define what an element of reality was in order to make their point. Instead, they gave a sufficient criterion for something

to have an element of reality. In particular, they said (on the first page of [39]):

> "*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.*"

In other words, if a physical (i.e., measurable) quantity can be predicted without interacting with it, then its value is independent of one's measurement strategy, and can be assumed to be "real".

If quantum mechanics is a complete theory, then it must give values for every measurable quantity that has an associated measurement of reality. However, the uncertainty principle assures us that it is not possible to predict both the position and momentum of a particle with certainty. Thus, if quantum mechanics is indeed a "complete" theory, then we cannot assign elements of reality to both the position and the momentum of a quantum particle.

What Einstein, Podolsky, and Rosen considered next, was a pair of particles maximally entangled in position and momentum. Given a pair of particles A and B, whose joint state $|\psi_{AB}\rangle$ is

$$|\psi_{AB}\rangle = \int dx_A dx_B f(x_A, x_B)|x_A\rangle \otimes |x_B\rangle, \tag{2.1}$$

where $f(x_A, x_B)$ is the (non-normalizeable) maximally entangled wavefunction in position space:

$$f(x_A, x_B) \propto \delta(x_A - x_B), \tag{2.2}$$

and its Fourier transform (into momentum space) is:

$$\tilde{f}(k_A, k_B) \propto \delta(k_A + k_B), \tag{2.3}$$

we find that this maximally entangled state has perfect position correlations, and perfect momentum anti-correlations. Note that $\delta(x)$ is the Dirac delta function.

For the purpose of this thesis, what is physically relevant is that there are well-behaved normalizeable entangled quantum states whose position and momentum correlations can both be arbitrarily strong (though finite). The uncertainty principle imposes no limit on the strength of measurement correlations between separate pairs of particles; the difference $\hat{x}_A - \hat{x}_B$ commutes with the sum $\hat{k}_A + \hat{k}_B$, giving the Heisenberg uncertainty relation,

$$\sigma^2(x_A - x_B)\sigma^2(k_A + k_B) \geq 0. \tag{2.4}$$

Note that as the uncertainty in the sum (or difference) of two random variables approaches zero, the outcome of the second random variable becomes determined by the outcome of the first, since their sum (or difference) must be constant.

At this point, Einstein, Podolsky, and Rosen considered isolating the highly entangled pair of particles A and B from each other so that they can no longer interact, or so that nothing done to A can affect the elements of reality associated with B. If the pair of now isolated particles is described by the entangled state $|\psi_{AB}\rangle$, then by measuring the position or momentum of A, one can predict with certainty the corresponding position or momentum of B, without interacting with B. Since the elements of reality associated to B are independent of any measurement on A, we must be able to assign elements of reality to both the position and momentum of B. However, if quantum mechanics is a complete theory, then we cannot assign elements of reality to both the position and momentum of B. This situation is often referred to as the EPR "paradox". On the one hand, a quantum mechanical description of highly entangled pairs of particles allows one to predict either the position or momentum of one particle to arbitrary precision by measuring the corresponding observable of the other particle. On the other hand, using the additional assumption that non-interacting particles impart no disturbances on each other, Einstein, Podolsky, and Rosen concluded that quantum mechanics must be incomplete, though there may be some future theory that may complete the description of reality.

2.1.1 Locality in the EPR Scenario

In the original EPR paper [39], their argument on assigning elements of reality to both the position and momentum of a particle is contingent on the pair of particles being forbidden from interacting with each other. Later [40], Einstein clarified that considering a pair of highly entangled particles spatially separated by a large enough distance (i.e., to be space-like separated[1]), we may consider them to be isolated from one another. As a result, the conclusion of the EPR paper can be rephrased as seen in the following excerpt (from page 682 of [40]):

> "*By this way of looking at the matter it becomes evident that the paradox forces us to relinquish one of the following two assumptions:*
> *(1) the description by means of the ψ-function is complete*
> *(2) the real states of spatially separated systems are independent of each other*"

Alternatively, we can say that it cannot both be the case that quantum theory offers a complete description of reality, and that reality obeys the principle of locality.

Locality is an assumption that all effects on physical systems travel no faster than some finite speed, taken to be the speed of light. What locality implies (and where its name ostensibly originates) is that any event in spacetime is determined by the physical situation within its immediate (spatial) neighborhood in the immediate past (i.e., is locally determined). If the Universe obeys locality, then the only causes of an event in spacetime exist in the past light cone of that event [2].

[1] Two points (i.e., events) in spacetime are space-like separated if the (spatial) distance between them is large enough so that no signal traveling at or below the speed of light can traverse that distance in the time interval between them.

[2] The past light cone of an event in spacetime is the set of all points in spacetime where signals traveling at or below the speed of light could originate that would reach the event in question.

In a local universe, two events are causally separated from one another if they are space-like separated; nothing at one point can effect the physical situation at the other point, and vise versa.

2.2 Local Hidden Variables and Bell-Nonlocality

Since it cannot both be the case that the effect of measurement travels no faster than light, and that quantum mechanics offers a complete description of reality, assuming the Universe is local requires that quantum mechanics is incomplete. At the end of the EPR paper [39], they proposed that in spite of this, there may yet be a future theory that can complete the description of reality. These theories (i.e., theories that both obey locality and accurately represent experimental outcomes) are broadly classified as models of Local Hidden Variables (LHV).

Without extra variables, quantum mechanics allows one to predict the probabilities of measurement outcomes, and how those probabilities change in different circumstances. A hidden variable model is one where knowledge of some extra set of variables λ in addition to the quantum state allows one to predict not only the relative probabilities, but the measurement outcomes themselves. As shown in Bell's seminal paper [41] (whose results we reproduce here), one can make a hidden variable model for spin measurements on a qubit (e.g., a spin-1/2 particle). In particular, one can express a measurement outcome of the qubit along any axis $\vec{n}$ as a function of its quantum state given by its Bloch vector $\vec{P}$ and a hidden random vector $\vec{\lambda}$. If we let $\vec{\lambda}$ have a uniform probability distribution covering the hemisphere $\vec{\lambda} \cdot \vec{P} \geq 0$, then any measurement outcome along $\vec{n}$ may be given by $\text{sgn}(\vec{n} \cdot \vec{\lambda})$ for some $\vec{\lambda}$. In this hidden variable model, we could predict our measurement outcome if we only knew what $\vec{\lambda}$ was.

A *local* hidden variable (LHV) model is a hidden variable model where we also require that the only correlations between a pair of separated systems, A and B,

are due to information given by λ in the intersection of their past light cones (see Fig. 2.1). In order to understand why imposing locality is sufficient to isolate

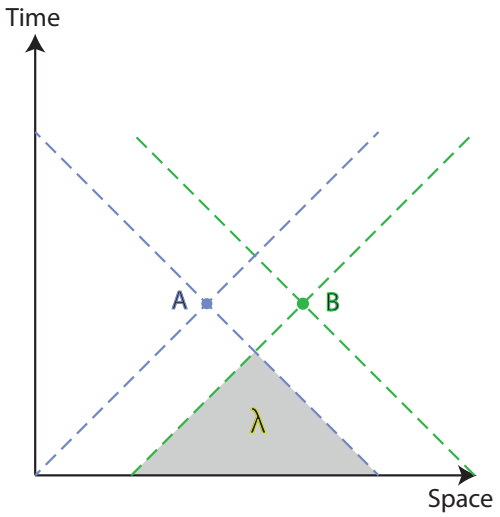

Figure 2.1: Space-time diagram of a pair of measurement events A and B. The diagonal lines emanating from A and B are light-speed trajectories. The region containing λ is the intersection of the past light cones of A and B.

the correlations between A and B to the intersection of their past light cones, it is useful to consider that in a local universe, the only correlations that can be produced as a consequence of both events A and B would be those within the intersection of their future light cones. The intersection of the past light cones on A and B is then the set of points whose future light cones contain both A and B.

In 1964 [41], Bell showed that there were mathematical consequences of LHV models. He showed that if the correlations between a pair of observables $\hat{x}_A$ and $\hat{x}_B$, admit an LHV model, then their joint probability $\rho(x_A, x_B)$ must factor in the following way:

$$\rho(x_A, x_B) = \int d\lambda \rho(\lambda)\rho(x_A|\lambda)\rho(x_B|\lambda) \tag{2.5}$$

Here, $\rho(\lambda)$ is some unknown arbitrary probability density over hidden variables λ. By conditioning on a particular value of λ, the sub-ensemble of outcomes of

$\hat{x}_A$ and $\hat{x}_B$ are statistically independent. As a consequence of this criterion, Bell derived an inequality that measurement statistics must satisfy in an LHV model:

$$1 + E(\vec{b}, \vec{c}) \geq |E(\vec{a}, \vec{b}) - E(\vec{a}, \vec{c})|. \tag{2.6}$$

Here, $E(\vec{a}, \vec{b})$ is the expectation value of the product of the outcomes of measuring a pair of entangled qubits, one along direction $\vec{a}$, and the other along direction $\vec{b}$. More important than this, he showed that the maximally entangled Bell singlet state $|\Psi_-\rangle = \frac{1}{\sqrt{2}}(|+,-\rangle - |-,+\rangle)$ was capable of violating his inequality. As a result, he showed LHV models were fundamentally incompatible with quantum mechanical predictions.

Though Bell's result was profoundly significant, an experimental violation of his particular inequality (2.6) was unrealistic because its derivation relied on the additional assumption that identical measurements on the entangled pair of particles should give perfectly correlated outcomes. This would only be true for a perfect singlet state, which is not something that can be realized experimentally. Still, all inequalities capable of ruling out Bell's LHV criterion (2.5) are known as Bell inequalities in his honor. Five years after Bell's paper, Clauser, Horne, Shimony, and Holt (CHSH) [42] developed an alternative inequality more amenable to experimental investigations. While still starting from the original LHV criterion (2.5), the inequality they derive does away with the additional assumption of perfect correlations for identical measurement settings.

2.2.1 The CHSH Bell Inequality

To explore where the CHSH inequality comes from, let us consider a pair of particles A and B. We let α and α' refer to two different measurement settings on particle A, and let β and β' refer to two different settings on particle B. We let $\langle x_A \rangle_\alpha$ be the expectation value of the random variable associated to the

set of outcomes from observing particle A with measurement setting α, and let the corresponding random variables for other particles and settings be similarly defined. Next, let $f(x)$ be a function where $x \in (-\infty, \infty)$, but where $|f(x)| \leq 1$. If measurements on A and B admit an LHV model, then the expectation product $\langle f(x_A)f(x_B)\rangle_{(\alpha,\beta)}$ would have to be expressible as the following form:

$$\langle f(x_A)f(x_B)\rangle_{(\alpha,\beta)} = \int d\lambda \rho(\lambda)\langle f(x_A)\rangle_{(\alpha,\lambda)}\langle f(x_B)\rangle_{(\beta,\lambda)} \tag{2.7}$$

where $\langle f(x_A)\rangle_{(\alpha,\lambda)}$ is the expectation value of $f(x_A)$ given measurement setting α, and hidden variable(s) λ. Since $f(x)$, must be between -1 and 1, the expectations $\langle f(x_A)\rangle_{(\alpha,\lambda)}$ and $\langle f(x_B)\rangle_{(\beta,\lambda)}$ are bounded between -1 and 1 as well as the expectation product $\langle f(x_A)f(x_B)\rangle_{(\alpha,\beta)}$.

To obtain the CHSH inequality, we bound the difference between $\langle f(x_A)f(x_B)\rangle_{(\alpha,\beta)}$, and $\langle f(x_A)f(x_B)\rangle_{(\alpha,\beta')}$. In particular, we find that

$$\begin{aligned}
\langle f(x_A)f(x_B)\rangle_{(\alpha,\beta)} - \langle f(x_A)f(x_B)\rangle_{(\alpha,\beta')} &= \int d\lambda\rho(\lambda)\Big(\langle f(x_A)\rangle_{(\alpha,\lambda)}\langle f(x_B)\rangle_{(\beta,\lambda)} \\
&\qquad - \langle f(x_A)\rangle_{(\alpha,\lambda)}\langle f(x_B)\rangle_{(\beta',\lambda)}\Big) \\
= \int d\lambda\rho(\lambda)\Big(&\langle f(x_A)\rangle_{(\alpha,\lambda)}\langle f(x_B)\rangle_{(\beta,\lambda)} \pm \langle f(x_A)\rangle_{(\alpha,\lambda)}\langle f(x_A)\rangle_{(\alpha',\lambda)}\langle f(x_B)\rangle_{(\beta,\lambda)}\langle f(x_B)\rangle_{(\beta',\lambda)} \\
&- \langle f(x_A)\rangle_{(\alpha,\lambda)}\langle f(x_B)\rangle_{(\beta',\lambda)} \mp \langle f(x_A)\rangle_{(\alpha,\lambda)}\langle f(x_A)\rangle_{(\alpha',\lambda)}\langle f(x_B)\rangle_{(\beta,\lambda)}\langle f(x_B)\rangle_{(\beta',\lambda)}\Big) \\
= \int d\lambda\rho(\lambda)\Big[&\langle f(x_A)\rangle_{(\alpha,\lambda)}\langle f(x_B)\rangle_{(\beta,\lambda)}\Big(1 \pm \langle f(x_A)\rangle_{(\alpha',\lambda)}\langle f(x_B)\rangle_{(\beta',\lambda)}\Big) \\
&- \langle f(x_A)\rangle_{(\alpha,\lambda)}\langle f(x_B)\rangle_{(\beta',\lambda)}\Big(1 \mp \langle f(x_A)\rangle_{(\alpha',\lambda)}\langle f(x_B)\rangle_{(\beta,\lambda)}\Big)\Big]
\end{aligned} \tag{2.8}$$

Since the expectation product $\langle f(x_A)f(x_B)\rangle_{(\alpha,\beta)}$ is bounded between -1 and 1 for all settings α, and β, we are led to the inequality

$$\begin{aligned}
&|\langle f(x_A)f(x_B)\rangle_{(\alpha,\beta)} - \langle f(x_A)f(x_B)\rangle_{(\alpha,\beta')}| \\
&\qquad \leq \int d\lambda\rho(\lambda)\Big[2 \pm \big(\langle f(x_A)\rangle_{(\alpha',\lambda)}\langle f(x_B)\rangle_{(\beta',\lambda)} + \langle f(x_A)\rangle_{(\alpha',\lambda)}\langle f(x_B)\rangle_{(\beta,\lambda)}\big)\Big],
\end{aligned} \tag{2.9}$$

and since, by the LHV criterion, we may express other expectation products, we arrive at the CHSH inequality:

$$|\langle f(x_A)f(x_B)\rangle_{(\alpha,\beta)} - \langle f(x_A)f(x_B)\rangle_{(\alpha,\beta')}| \leq 2 \pm \big(\langle f(x_A)f(x_B)\rangle_{(\alpha',\beta')} + \langle f(x_A)f(x_B)\rangle_{(\alpha',\beta)}\big) \tag{2.10}$$

This particular formulation of the CHSH inequality will be especially useful in Chapter 6, when we discuss how to demonstrate Bell-nonlocality in the transverse position-momentum statistics of entangled photon pairs.

In the original CHSH paper, they show how to violate their inequality (2.10) using the statistics of polarization-entangled photons as might be generated from an electric dipole cascade[3], in which case $f(x)$ has two values (1 and -1 for different polarization outcomes). In the language of entangled qubits (2.6), the CHSH inequality has the simple form:

$$|E(\vec{a},\vec{b}) - E(\vec{a},\vec{b}')| \leq 2 \pm \big(E(\vec{a}',\vec{b}) + E(\vec{a}',\vec{b}')\big) \tag{2.11}$$

Indeed, Bell's original inequality results when we set both $\vec{a}'$ and $\vec{b}'$ equal to $\vec{c}$.

2.2.2 Consequences of Violating a Bell Inequality

When we violate a Bell inequality, we rule out the possibility of an LHV model describing the correlations between a pair of separate quantum systems. What this implies is that the correlations between a pair of space-like separated measurements A and B are not reducible to information existing in their shared past

[3]An electric dipole cascade is a series of decay transitions from an atomic orbital of zero angular momentum to another lower energy orbital of zero angular momentum. Dipole selection rules forbid a direct transition, so two photons of unit angular momentum each are emitted whose total is zero. As the two decay modes ((one mode being the first photon with $S = +1$, and the second with $S = -1$), and second mode with opposite angular momenta) become indistinguishable, the pair of photons becomes more entangled. See [43] for details.

light cones. Though a full discussion of the physical and metaphysical implications of the violations of Bell inequalities is beyond the scope of this thesis, there are at least two possible consequences of ruling out LHV models. We may say either that the measurement correlations are explained by information existing outside their shared past light cones (a nonlocal hidden variable theory), or that there simply is no information explaining their correlations (i.e., that there are no hidden variables).

2.3 Schrödinger's Response to the EPR Scenario and Quantum Steering

As a response to the original EPR paper, Schrödinger [44] elaborated on the nature of entanglement, and showed that the quantum state of one particle entangled with a second really does depend on the measurement performed on that second particle. In particular, he considered that (from page 556 of [44]):

> *"It is rather discomforting that the theory should allow a system to be steered or piloted into one or the other type of state at the experimenter's mercy, in spite of his having no access to it."*

Indeed, if we consider a pair of particles A and B maximally entangled position-momentum EPR state, measuring the position of A results in the state of B being one of well-defined position, the particular value being highly correlated to the measurement outcome of A. If instead, the momentum of A were measured, then the state of B would be well-defined in momentum, its value being correlated to the momentum outcome of A. Thus, by one's choice of measurement of A, one can *steer* the ensemble of states of B to be an ensemble of either position eigenstates, or momentum eigenstates. This general idea of quantum steering led to the specific formulation of EPR-steering by Wiseman *et al.* in 2007 [45].

However, it is important to understand that this entanglement-mediated manipulation of a quantum state does not allow for faster-than-light communication due to the no-signaling theorem [46]. In short the no signaling theorem shows that given two quantum systems A and B, there is no difference between the density matrix of A before any measurement of B, and the density matrix of A that results from a complete ensemble (including null results) of generalized measurements of B. Simply put, there is no way in principle of determining whether a single particle is half of an entangled pair without additional information outside that single particle. There is no difference between the reduced density matrix of one of an entangled pair of particles, and the density matrix of a statistical mixture of single particles.

2.3.1 Local Hidden State Models and the Hierarchy of Nonlocality

With the general idea of quantum steering in mind, Wiseman, Jones, and Doherty [45] formulated the concept of EPR-steering as a variety of nonlocality intermediate between Bell-nonlocality, and mere entanglement. Specifically, they introduced the idea of a model of Local Hidden *States* (LHS). In an LHV model, the joint probability between two separated observables $\hat{x}_A$ and $\hat{x}_B$ factors as seen previously (2.5). In an LHS model for system B, the joint probability factors similarly:

$$\rho(x_A, x_B) = \int d\lambda \rho(\lambda)\rho(x_A|\lambda)\mathrm{Tr}[\hat{\Pi}^x_B \hat{\rho}_{B,\lambda}]. \tag{2.12}$$

Here, $\mathrm{Tr}[\hat{\Pi}^x_B \hat{\rho}_{B,\lambda}]$ is the probability of measuring a particular outcome of $\hat{x}_B$ where that probability is obtained from a quantum state $\hat{\rho}_{B,\lambda}$, which is determined by local hidden variable(s) λ, and $\hat{\Pi}^x_B = |x_B\rangle\langle x_B|$. Indeed, an LHS model for B is a particular kind of LHV model, but with the added assumption that the measurement statistics of B are determined quantum mechanically. Thus, the set

of states satisfying LHS models for either A or B are subsets of the set of states satisfying LHV models. Note that here, and throughout this thesis, we will refer to states that admit an LHV model as Bell-local states, and states that admit an LHS model as non-steerable states. Having formulated LHS models, Wiseman *et al.* finished a hierarchy of locality by giving a similar model for separable states. In particular, if the quantum state of a pair of particles A and B is separable, then the joint probability between $\hat{x}_A$ and $\hat{x}_B$ factors in the following way:

$$\rho(x_A, x_B) = \int d\lambda \rho(\lambda) \mathrm{Tr}[\hat{\Pi}^x_A \hat{\rho}_{A,\lambda}] \mathrm{Tr}[\hat{\Pi}^x_B \hat{\rho}_{B,\lambda}]. \tag{2.13}$$

Indeed, this conforms to the standard definition of a separable state [10, 47]:

$$\hat{\rho}_{AB(sep)} = \sum_i p_i \hat{\rho}_{A,i} \otimes \hat{\rho}_{B,i} \tag{2.14}$$

With the hierarchy established, they then showed that non-steerable states in fact

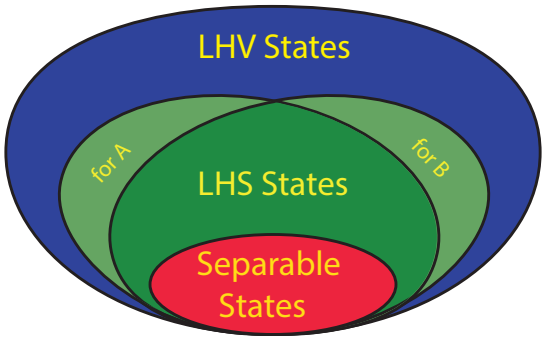

Figure 2.2: A Venn diagram illustrating the hierarchy of nonlocality for a pair of quantum systems A and B. Separable states form a subset of the non-steerable states, which in turn forms a subset of Bell-local states. Note that there are states admitting an LHS model for A, but not for B, and vise versa.

form a strict subset of Bell-local states (i.e., that there were (two-qubit Werner) states [4] that rule out an LHS model, but satisfy an LHV model), and a strict

[4]Two-qubit Werner states are mixtures of the two-qubit Bell singlet state and the two-qubit maximally mixed state.

superset of separable states (see Fig. 2.2 for diagram). Among other questions left open was whether there are states that are exclusively one-way steerable. Recently [48] this was shown to be the case.

Although one can perform some semblance of quantum steering with any pair of entangled particles, EPR-steering has a more stringent operational definition. A state for pair AB is EPR-steerable if, and only if, there is no LHS model for either A or B, or both that can describe their joint measurement statistics. To determine whether or not a given quantum state is steerable may be at least as difficult a problem as determining whether or not a state is separable, which is currently [13, 49], an NP-hard problem [5] (Research into necessary and sufficient demonstrations of steering is ongoing). As such, the exhaustive determination of steerability for high dimensional (say, continuous-variable) systems is computationally unfeasible (if not explicitly impossible) for the time being.

2.3.2 Witnessing EPR-Steering, and the Development of EPR-Steering Inequalities

While it may be impractical or possibly unfeasible to tell whether or not a state admits an LHS model (for one or another party), there are ways to witness EPR-steering in analogy to witnessing entanglement. Any mathematical criterion that all separable states must satisfy is an entanglement witness, for when a state violates that criterion, it cannot be separable.

One especially useful property of Bell-local states, non-steerable states, and separable states, besides that they form a nested series of sets, is that each set of states is *convex*. Mixtures of separable states are other separable states; mixtures

[5] In computational complexity theory, an NP-hard problem is one where the number of steps needed to solve it is not known to scale favorably (i.e., be solvable in an amount of time proportional to some polynomial in the number of bits of the input), but where the number of steps needed to check a candidate solution is known to scale favorably.

of non-steerable states are other non-steerable states; and mixtures of Bell-local states are also Bell-local states. Because these sets of states are each convex, we are assured [6] that for any steerable state, there is always [50] some witness that will show it to be so.

Just as witnesses of Bell-nonlocality are called Bell inequalities, witnesses of EPR steering are called EPR-steering inequalities. Shortly after the inception of EPR-steering it was shown [51] that demonstrations of the EPR paradox are a special case of EPR-steering. As mentioned previously, an LHS model for, say, A of the pair AB is an LHV model, but with the extra assumption that A's measurement statistics are quantum-mechanically generated. What this implies is that in an LHS model for A, any measurement statistics of A, conditional or otherwise, must be constrained by uncertainty relations. Then violation of conditional uncertainty relations, which demonstrate the EPR paradox, also rule out LHS models, and so demonstrate EPR-steering.

Reid's EPR-Steering Inequality

The first EPR-steering inequality was Developed by Margaret Reid in 1989 [52] as a proposed demonstration of the EPR paradox in the field quadratures of entangled light. She considered inferred variances (as seen in [51, 53]), and developed a steering inequality which may be seen as the Heisenberg uncertainty relation, but with conditioned variances. Since it is instructive to see where this steering inequality comes from, we re-derive it for position x and momentum k.

Given a pair of particles A and B entangled in position x and momentum k, an LHS model for B exists for the joint distribution $\rho(x_A, x_B)$ if and only if it

[6] The Hahn-Banach theorem tells us (among other things) that for any vector outside a convex set of vectors, there is always a plane separating that point from every point in the convex set. One can then tell that the vector is outside the convex set by seeing it is on the opposite side of the plane as every vector in the convex set.

factors as in (2.12). In particular, an LHS model for Bob implies that there is some set of hidden variables λ where:

$$\rho(x_B|x_A) = \int d\lambda \rho(\lambda|x_A)\rho_q(x_B|\lambda) \tag{2.15}$$

where for simplicity, $\rho_q(x_B|\lambda) = \mathrm{Tr}[\hat{\Pi}^x_B \hat{\rho}_{B,\lambda}]$, and $\rho(x_B|x_A) = \rho(x_A, x_B)/\rho(x_A)$. The subscript q is a label implying the density it is attached to is quantum-mechanical in origin. Since the variance, like the entropy, is a concave function, any mixture of probability distributions will have a variance no smaller than the corresponding mixture of variances. Using this fact, we find (as in [51]):

$$\sigma^2(x_B|x_A) \geq \int d\lambda \rho(\lambda|x_A)\sigma^2_q(x_B|\lambda). \tag{2.16}$$

Here, $\sigma^2(x_B|x_A)$ is the variance of x_B conditioned *on a particular outcome* of x_A, and $\sigma^2_q(x_B|\lambda)$ is variance of x_B conditioned on a particular value of λ, where this probability distribution is quantum-mechanical in origin. By averaging $\sigma^2(x_B|x_A)$ over all outcomes of x_A, we get the mean inferred variance, which we call $\sigma^2_{\mathrm{inf}}(x_B|x_A)$, and obtain the LHS constraint:

$$\sigma^2_{\mathrm{inf}}(x_B|x_A) \equiv \int dx_A \rho(x_A)\sigma^2(x_B|x_A) \geq \int d\lambda \rho(\lambda)\sigma^2_q(x_B|\lambda). \tag{2.17}$$

Since taking the logarithm of both sides of an inequality does not change its direction (and all terms are positive), we combine the LHS constraints for position and momentum to get:

$$\log(\sigma^2_{\mathrm{inf}}(x_B|x_A)\sigma^2_{\mathrm{inf}}(k_B|k_A)) \geq \int d\lambda \rho(\lambda) \log(\sigma^2_q(x_B|\lambda)\sigma^2_q(k_B|\lambda)). \tag{2.18}$$

Finally, since conditioning on a single set of random variables does not change the fact that measurement statistics of B are quantum-mechanically constrained by the Heisenberg relation, i.e.,

$$\sigma^2_q(x_B|\lambda)\sigma^2_q(k_B|\lambda) \geq \frac{1}{4}, \tag{2.19}$$

we arrive at Reid's EPR-steering inequality:

$$\sigma^2_{\text{inf}}(x_B|x_A)\sigma^2_{\text{inf}}(k_B|k_A) \geq \frac{1}{4}. \tag{2.20}$$

This steering inequality was used to successfully demonstrate the EPR-paradox with the transverse positions and momenta of entangled photon pairs generated via spontaneous parametric down-conversion (SPDC) [54]. In 2011 [55], Steven Walborn and others showed how one could develop a tighter steering inequality by using entropies instead of variances.

Walborn *et al.*'s Entropic Steering Inequality

Just as the concavity of the variances can be used to develop an LHS constraint on measuring variances (2.17), the concavity of the continuous entropy can also be used to develop an LHS constraint on measuring entropies. Indeed, from the very same LHS model for Bob (2.12), Walborn *et al.* showed:

$$h(x_B|x_A) \geq \int d\lambda \rho(\lambda) h_q(x_B|\lambda), \tag{2.21}$$

though they proved this using the fact that the relative entropy (1.20) between distributions $\rho(x_B, \lambda|x_A)$ and $\rho(\lambda|x_A)\rho(x_B|x_A)$ is non-negative (as it is between any pair of distributions. Using the fact, again, that conditioning measurements on a quantum system on a single ensemble of events λ does not affect that those statistics are constrained by uncertainty relations, we can use the entropic uncertainty relation (1.43), to find:

$$h(x_B|\lambda) + h(k_B|\lambda) \geq \log(\pi e), \tag{2.22}$$

and combine this with the entropic LHS criterion (2.21) to get Walborn *et al.*'s steering inequality:

$$h(x_B|x_A) + h(k_B|k_A) \geq \log(\pi e). \tag{2.23}$$

Using the steps leading to Walborn *et al*'s steering inequality (2.23), my collaborators and I showed [56] that *any* entropic uncertainty relation is easily adaptable into an EPR-steering inequality. This is of particular use, as it leads to inequalities capable of witnessing entanglement across disparate degrees of freedom (i.e., hybrid entanglement).

Walborn *et al.*'s steering inequality is an improvement over Reid's steering inequality (2.20) in that one can actually derive Reid's inequality as a special case of Walborn *et al.*'s inequality, but not the other way around; there are states violating Walborn *et al.*'s inequality that do not violate Reid's inequality. The reason that this is so is because the maximum entropy probability distribution for a constant variance is that of a Gaussian:

$$h(x) \leq \frac{1}{2}\log(2\pi e\sigma_x^2). \tag{2.24}$$

Using this, we can place upper limits on the entropic uncertainty relation (1.43), to get a weaker uncertainty relation:

$$\frac{1}{2}(\log(2\pi e\sigma_x^2) + \log(2\pi e\sigma_k^2)) \geq \log(\pi e). \tag{2.25}$$

Simplifying, we actually obtain the Heisenberg uncertainty relation

$$\sigma_x^2\sigma_k^2 \geq \frac{1}{4}. \tag{2.26}$$

Indeed, Walborn *et al.*'s steering inequality is one of the most sensitive continuous-variable steering inequalities to date; many simpler steering inequalities are derived from it, but not the other way around.

2.3.3 EPR-Steering and Quantum Cryptography

In Wiseman *et al.*'s seminal paper on EPR-steering, they defined EPR-steering in terms of an operational task similar to the demonstration of the EPR paradox. This task concerns two parties, named Alice and Bob.

Alice prepares a pair of particles AB, and sends B to Bob, storing her particle in a quantum memory for later measurement. For the sake of simplicity, we will assume AB is a state highly entangled in position and momentum, and concern ourselves with EPR-steering in that degree of freedom. Alice repeats this many times, creating many identically prepared particle pairs. Alice then announces what ensembles she can steer Bob's states into. For example, with AB highly entangled in position or momentum, Alice could steer Bob's state to be in an ensemble of localized position states, or an ensemble of localized momentum states.

Next, Bob randomly picks an ensemble (in this case, either position or momentum), and asks Alice to steer his state into that ensemble. Alice does so, by measuring either $\hat{x}_A$ or $\hat{k}_A$, and announces her measurement outcome, which corresponds to remotely preparing a particular state of Bob's particle. Finally, over many runs, Bob can show that that state is correct through tomography on the sub-ensembles of his particles corresponding to each outcome of Alice's measurements.

There is one complication in this set of events. In order to truly be EPR-steering, the correlations between the results Alice reports and the states Bob obtains must not be explainable by an LHS model. Indeed, if Alice were instead just preparing and sending individual particles to Bob, and using her knowledge of them to announce "results" correlated to Bob's measurement outcomes, then Bob would be receiving a local hidden state. In this way, EPR-steering can be used to verify entanglement between two parties, even when one party's measurements are untrusted. To see how this might work in practice, we consider a simplified version of Wiseman *et al.*'s thought experiment (see Fig. 2.3).

Alice again, controls a source of identically prepared entangled pairs of particles AB, sends B to Bob, and keeps A. Alice and Bob are separated by a large enough distance that their measurements on each pair are space-like separated from one another. Here, Bob randomly decides to measure either $\hat{x}_B$ or $\hat{k}_B$,

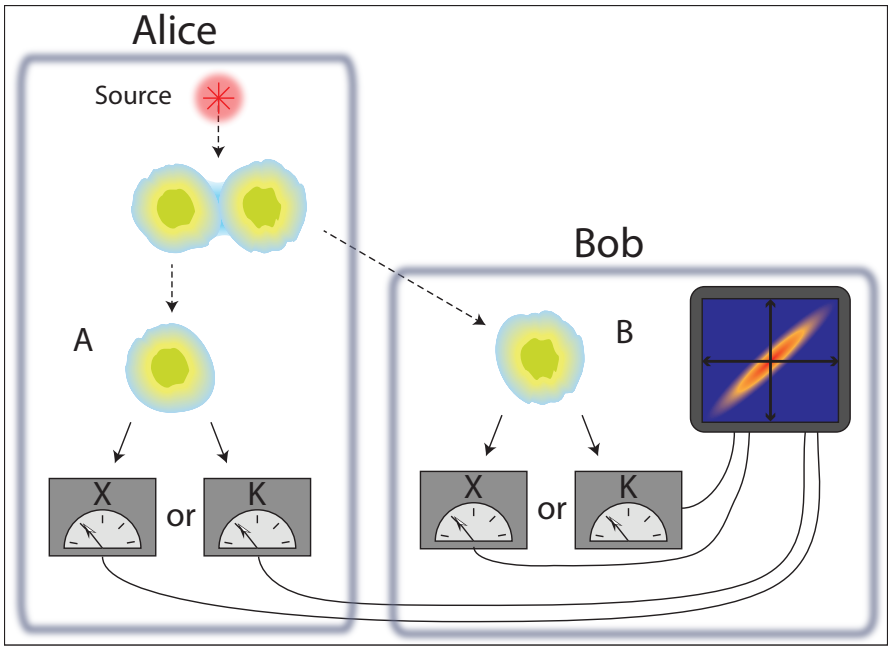

Figure 2.3: A diagram of our simplified version of Wiseman *et al.*'s thought experiment for demonstrating position-momentum entanglement in spite of Bob's skepticism of everything outside his experimental setup.

and communicates his choice of observable to Alice. Alice then measures that observable on her particle A, and tells Bob her measurement outcomes. Bob uses that information along with his own measurements to determine their position and momentum correlations.

The question to consider here is whether Alice can prove to Bob that they share entanglement, even when Bob does not trust the origin of any particle or piece of information entering his own setup. Ordinarily, Alice and Bob could show that they share entanglement if the correlations they share violate a separability inequality (i.e., one that all separable states satisfy). Because Bob does not trust what is outside his setup, he also cannot trust that the information he receives about Alice's measurements outcomes are quantum mechanical in origin. Indeed, in the absence of knowledge beyond his own system, it could be that Alice is actually preparing ensembles of particles well-defined in either $\hat{x}$ or $\hat{k}$, so that

when Alice and Bob both measure $\hat{x}$, Alice can announce her "measurement" result to be highly correlated with Bob's outcome, simply because she knows what state she sent. However, since Bob receives his particle before he dictates Alice's measurement basis, half of the time, Alice would have sent a particle well-defined in a different, complementary basis to the one Bob measures. In that case, Alice can do no better than randomly guessing measurement outcomes, which will on average, be uncorrelated to Bob's outcome. In other words, if Alice were preparing and sending ensembles of particles to Bob, their measurement correlations could only be so strong. If Alice and Bob's measurement correlations are so strong that they violate an EPR-steering inequality, then Bob can be assured that he and Alice must have shared entanglement, since there is no way they could have had such correlations otherwise.

To connect EPR-steering to quantum cryptography, we must understand what else is ruled out when Alice and Bob's measurement correlations violate an EPR-steering inequality. In particular, we rule out the possibility that any third party was intercepting, measuring, and resending the particles en route to Bob a significant portion of the time, since those resent particles are described by a local quantum state only classically correlated to Alice's measurements. While entanglement-based quantum cryptography does not require EPR-steering, EPR steering allows one to verify secure correlations in entanglement-based quantum cryptography, even when one party's measurement device is untrusted. In Sec. 4.2.4, we discuss exactly how this can be accomplished using the position-momentum correlations of entangled photon pairs generated via spontaneous parametric down-conversion (SPDC).

3 Preliminaries on the Statistics of Photon Pairs Generated in Spontaneous Parametric Down-Conversion

First, a disclaimer:
In numerous places, I use the terms "wavefunction", and "transverse spatial amplitude" interchangeably. The state $|\psi\rangle$ of a single photon with spatial amplitude $\psi(x)$ can be expressed as:

$$|\psi\rangle = \int dx\ \psi(x)\ \hat{a}^{\dagger}(x)|0\rangle, \tag{3.1}$$

where $|0\rangle$ is the vacuum state, and $\hat{a}^{\dagger}(x)$ is the creation operator for a photon at position x. Though the time evolution of the spatial amplitude $\psi(x)$ is not governed by a Schrödinger equation, it is governed by a wave equation, and the amplitude is still a square-integrable function having all other necessary properties of a wavefunction. It is only in the sense that we use the term "wavefunction". We avoid this distinction where we expect it will cause confusion.

In this chapter, we briefly discuss the theoretical foundation behind the experimental source of position-momentum entangled photon pairs used in our research, namely, type-I degenerate (nearly collinear) Spontaneous Parametric

Down-Conversion (SPDC). In particular, we show what approximations are taken to derive the transverse spatial biphoton wavefunction (and what one gets as a result), and how it departs from idealized models. In addition, we discuss how to witness the entanglement, EPR-steerability, and Bell-nonlocality of a simplified biphoton state. A more expansive discussion of these foundations can be found in my article [57].

3.1 Essential Concepts and Development of the Biphoton Wavefunction

In the study of light and how it interacts with matter, an electromagnetic wave, passing through a dielectric material induces a polarization within the material that affects the wave's subsequent propagation. For most treatments of optics [58], we consider the polarization response of a material to be directly proportional to the magnitude of the electric field component of the electromagnetic wave:

$$\vec{\mathbf{P}}^{(1)} = \epsilon_0 \chi^{(1)} \vec{\mathbf{E}} \tag{3.2}$$

where here, $\vec{\mathbf{P}}^{(1)}$ is the (linear) polarization field within the material, ϵ_0 is the permittivity of free space, $\vec{\mathbf{E}}$ is the incident electric field, and $\chi^{(1)}$ is the first-order dielectric susceptibility tensor, (i.e., a constant matrix relating each component of $\vec{\mathbf{E}}$ to each component of $\vec{\mathbf{P}}$). Indeed, (3.2) forms the basis of standard optics, and knowing the real and imaginary components [1] of $\chi^{(1)}$ as a function of position and frequency gives you all the necessary information needed to describe how the light interacts and propagates within a material, provided that its response is linear.

Within optics, there is a sub-discipline dealing with situations where a material's response to an optical field is no longer linear. In this subfield, known

[1] The real and imaginary components of $\chi^{(1)}$ give one the index of refraction and the absorption coefficient respectively, for the location and frequency that $\chi^{(1)}$ is given.

as nonlinear optics, processes are organized according to the highest-order power (e.g., square, cube, etc) of the optical field on which the polarization response $\vec{\mathbf{P}}$ effectively depends [59]. Provided that the electric field strength of a light pulse incident on a material is not so strong as to be comparable to the electric field strength between electrons and nuclei within a material (so that the light does not ionize the material), we can express each component of the polarization response $\vec{\mathbf{P}}$ as a series depending on various powers of the electric field:

$$P_i = \epsilon_0\big(\chi^{(1)}_{ij}E_j + \chi^{(2)}_{ijk}E_jE_k + \chi^{(3)}_{ijk\ell}E_jE_kE_\ell + ...\big). \tag{3.3}$$

Here, we use Einstein notation, so that repeated indices imply a summation over components. In keeping with this series expansion, we define $\chi^{(2)}$ as a (rank-3) tensor, called the second order nonlinear susceptibility.

SPDC is a second-order nonlinear-optical process occurring in birefringent[2] materials in which a photon (called a pump photon) is in one instant converted into a pair of daughter photons (termed signal and idler), whose energy and momenta add up to those of the pump photon. While many nonlinear-optical processes can be modeled without an appeal to quantum mechanics, (say, by including the nonlinear polarization as an extra term in the optical wave equation), SPDC requires a quantum-mechanical treatment because the process does not occur classically without there already being significant fields oscillating at the signal/idler frequencies. Quantum mechanically, the process is stimulated by spontaneous fluctuations in the electromagnetic quantum vacuum oscillating at the signal/idler frequencies [60, 61].

To describe SPDC quantum mechanically, we can use the electromagnetic field Hamiltonian for a $\chi^{(2)}$-nonlinear material. In particular, we will describe

[2]In order to produce SPDC, one does not necessarily need a *birefringent* crystal. Though there are other means of ensuring a constant phase relationship between the pump photon and the signal/idler photon pair (known as phase matching), birefringent phase matching is a popular method, and the one that we use here.

a collimated pump beam propagating in the z-direction incident on a nonlinear crystal cut in the shape of an L_x by L_y by L_z rectangular prism (as first discussed in [62])(see Fig. 3.1).

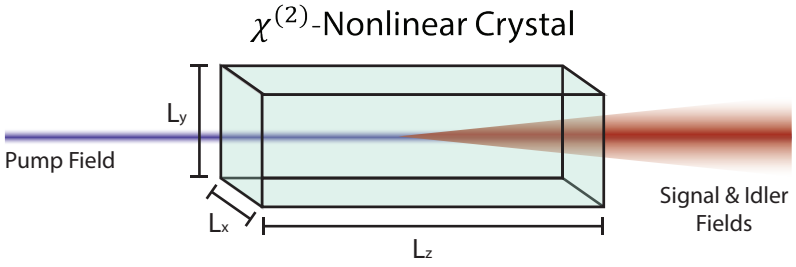

Figure 3.1: Diagram of an L_x by L_y by L_z birefringent $\chi^{(2)}$-nonlinear crystal facilitating SPDC. The pump field is collimated, while the SPDC process creates signal/idler fields with a significant transverse momentum spread.

Since only the nonlinear term in the electromagnetic Hamiltonian will contribute to the SPDC process, we can use only this term, as seen in [60, 62]. The nonlinear part of the electromagnetic Hamiltonian is given by:

$$\mathcal{H}_{NL} = \frac{1}{2}\epsilon_o \int d^3r\ \tilde{\chi}^{(2)}_{ijl} E_i(\vec{\mathbf{r}},t) E_j(\vec{\mathbf{r}},t) E_l(\vec{\mathbf{r}},t). \tag{3.4}$$

To quantize this field, we use the standard procedure in quantum optics [61]: we decompose the field into a sum of cavity modes, find the canonically conjugate coordinates and momenta of each mode (i.e., those that satisfy Hamilton's equations of motion), and substitute for those coordinates, canonically conjugate observables (i.e., whose commutator is $i\hbar$). In doing so, we eventually arrive at the following expression for the quantum Hamiltonian:

$$\hat{H}_{NL} = \frac{1}{2}\epsilon_o \int d^3r\ \tilde{\chi}^{(2)}_{ijl} \hat{E}_i(\vec{\mathbf{r}},t) \hat{E}_j(\vec{\mathbf{r}},t) \hat{E}_l(\vec{\mathbf{r}},t), \tag{3.5}$$

where

$$\hat{E}_i(\vec{\mathbf{r}},t) = \frac{1}{V^{\frac{1}{2}}} \sum_{\vec{\mathbf{k}},s} i\sqrt{\frac{\hbar\omega(\vec{\mathbf{k}})}{2\epsilon_0}} \hat{a}_{\vec{\mathbf{k}},s}(t)\ \vec{\epsilon}_{\vec{\mathbf{k}},s}\ e^{i\vec{\mathbf{k}}\cdot\vec{\mathbf{r}}} + \text{H.C.}, \tag{3.6}$$

and H.C. denotes Hermitian Conjugate, $\hat{a}_{\vec{\mathbf{k}},s}$ is the photon annihilation operator for a photon with polarization indexed by s and momentum $\vec{\mathbf{k}}$, and V is the quantization volume of a hypothetical optical cavity. We take the limit as V becomes arbitrarily large to get the quantum treatment of photons in free space.

In order to go from this very general Hamiltonian (which may describe any second-order nonlinear optical process), to the Hamiltonian driving type-I, degenerate, nearly collinear SPDC, we make multiple approximations[3]. First, we assume that the pump field is bright enough that it can be treated classically, and is not significantly attenuated in the crystal. Second, we assume the crystal is embedded in a material of the same refractive index, but with no nonlinearity so that we may ignore the effect of internal reflections, and that $\chi^{(2)}$ is isotropic throughout the crystal. Third, we assume the pump in nearly monochromatic so that we can separate its time dependence from its spatial dependence. Fourth, we assume the pump is sufficiently collimated so that its the longitudinal component of its momentum dominates over all others, allowing us to separate the longitudinal and transverse dependencies of the pump field.

Once we have a usable Hamiltonian regarding type-1 nearly collinear degenerate SPDC, we can use first-order time-dependent perturbation theory in the interaction picture to obtain the state of the down-converted photon pair, which gives:

$$|\Psi(t)\rangle_{SPDC} \approx \left(1 - \frac{i}{\hbar}\int_0^t dt' \hat{H}_{NL}(t')\right)|0\rangle. \tag{3.7}$$

Note that we integrate over the time it takes for light at the pump wavelength to travel through the crystal. Since the nonlinear Hamiltonian is slowly varying over that time scale, we may just as well use the time it takes for light at the signal/idler wavelengths to propagate through the crystal as well. The reason we only need first-order perturbation theory to get an accurate picture of the state of the downconverted fields is that the energy of the nonlinear contribution to

[3]To see a step-by-step implementations of these approximations, see [57].

the electromagnetic field is minuscule in comparison to the energy of the linear contribution. Alternatively, the energy of the pump field is much much larger (by more than a factor of 10^8 for typical experimental parameters) than the energy of the signal/idler fields. For such small corrections, first-order perturbation theory works well, and subsequent corrections in higher-order perturbation theory are small enough to be ignored.

Carrying out the perturbation integral (with the aforementioned approximations), we find the downconverted state to be (as shown in [63]):

$$\begin{aligned} |\Psi_{SPDC}\rangle &\approx C_0|0_1,0_2\rangle \\ &+ C_1 \iint d^3k_1 d^3k_2 \; \Phi(\vec{\mathbf{k}}_1,\vec{\mathbf{k}}_2)\hat{a}^\dagger(\vec{\mathbf{k}}_1)\hat{a}^\dagger(\vec{\mathbf{k}}_2)|0_1,0_2\rangle, \end{aligned} \tag{3.8}$$

where C_0 and C_1 are normalization constants giving the relative likelihood of creating a signal-idler photon pair (note: $|C_0| >> |C_1|$), and

$$\Phi(\vec{k}_1,\vec{k}_2) = \iint dk_{xp}dk_{yp}\Big[\prod_{m=1}^{3}\mathrm{sinc}\Big(\frac{\Delta k_m L_m}{2}\Big)\Big]\nu(k_{xp},k_{yp}). \tag{3.9}$$

Here, k_{xp} is the x-component of the momentum of the pump photon, and $\Delta\vec{k}$ is defined as the difference between the pump momentum and the sum of signal and idler momenta, so that $\Delta k_x \equiv k_{x1} + k_{x2} - k_{xp}$. In addition, $\nu(k_{xp},k_{yp})$ is the transverse momentum amplitude of the pump field, which we take to be a Gaussian, and $\mathrm{sinc}(x)$=$\sin(x)/x$.

To get our biphoton wavefunction for degenerate, nearly collinear type-I SPDC, we note that since our crystal dimensions are much larger than the typical wavelengths passing through it (by a factor of 10^3-10^4 for visible wavelengths and a crystal a few millimeters thick), this integral is straightforward, since the sinc function factors depending on transverse coordinates act like delta functions (the only significant contributions being when, e.g., $\Delta k_x = 0$, or when the sum of the transverse signal and idler momenta is very nearly equal to the transverse pump

momentum. Carrying out those integrals, we find:

$$\Phi(\vec{k}_1, \vec{k}_2) = \mathcal{N}\,\mathrm{sinc}\left(\frac{\Delta k_z L_z}{2}\right)\nu(k_{x1}+k_{x2}, k_{y1}+k_{y2}), \tag{3.10}$$

which, when we use the small angle approximation (to express Δk_z in terms of Δk_x and Δk_y), and consider only the x component (for simplicity)[4], becomes our biphoton wavefunction:

$$\Phi(k_{1x}, k_{2x}) = \mathcal{N}\,\mathrm{sinc}\left(\frac{L_z\lambda_p}{8\pi}(k_{x1}-k_{x2})^2\right)e^{-\frac{(k_{x1}+k_{x2})^2}{4\sigma_{kxp}^2}}, \tag{3.11}$$

where σ_{kxp} is the standard deviation in the x-component of the pump momentum.

3.2 Transverse Spatial Statistics of Biphotons in SPDC

Using the biphoton wavefunction discussed previously (3.11), we can calculate the transverse spatial probability density in position and in momentum, and from that, calculate correlation statistics, and especially, find alternative approximations to (3.11) that are easier to work with.

To simplify notation, let $a \equiv \frac{L_z\lambda_p}{4\pi}$, and we relabel k_{x1} as just k_1, and k_{x2} similarly. The biphoton wavefunction is then:

$$\Phi(k_1, k_2) = \mathcal{N}\,\mathrm{sinc}\big(\frac{a}{2}(k_1-k_2)^2\big)e^{-\frac{(k_1+k_2)^2}{4\sigma_{kp}^2}}. \tag{3.12}$$

To simplify our expressions further, we transform to the rotated coordinates:

$$k_+ \equiv \frac{k_1+k_2}{\sqrt{2}} \qquad : \qquad k_- \equiv \frac{k_1-k_2}{\sqrt{2}} \tag{3.13}$$

[4]The biphoton transverse spatial amplitude does not generally factor into a product of horizontal and vertical wavefunctions. However, in the approximations we make (such as, that for small x and y, $\mathrm{sinc}(x+y) \approx \mathrm{sinc}(x)\mathrm{sinc}(y)$), it is accurate to do so.

so that our biphoton wavefunction is a product of a pair of wavefunctions, one of k_+, and one of k_-:

$$\Phi(k_+, k_-) = \mathcal{N}\,\mathrm{sinc}(ak_-^2)e^{-\frac{k_+^2}{2\sigma_{kp}^2}}. \tag{3.14}$$

With the simplified biphoton wavefunction, we can transform to the position-domain with the Fourier transform, and get [57]:

$$\psi(x_+, x_-) = \mathcal{N}' e^{-\frac{x_+^2}{(1/2\sigma_{kp}^2)}}\left[x_-\sqrt{2\pi}\left(\mathcal{S}\left(\frac{x_-}{\sqrt{2\pi a}}\right) - \mathcal{C}\left(\frac{x_-}{\sqrt{2\pi a}}\right)\right) + \right.$$
$$\left. + 2\sqrt{a}\left(\cos\left(\frac{x_-^2}{4a}\right) + \sin\left(\frac{x_-^2}{4a}\right)\right)\right], \tag{3.15}$$

where $\mathcal{C}(x)$ and $\mathcal{S}(x)$ are the Fresnel integrals, integrating over $\cos(\frac{\pi}{2}t^2)$ and $\sin(\frac{\pi}{2}t^2)$, respectively.

Calculating statistics from these biphoton wavefunctions is not trivial, but with some effort, one can obtain a simple expression for the transverse correlation width $\sigma_{(x_1-x_2)}$, here defined as the standard deviation of $x_1 - x_2$. By using position operators, and performing the calculation in momentum space, one can find [57]:

$$\sigma_{(x_1-x_2)} = \sqrt{\frac{18a}{5}} = \sqrt{\frac{9L_z\lambda_p}{10\pi}}. \tag{3.16}$$

3.2.1 The Double-Gaussian Approximation

Although all transverse spatial statistics can in principle be derived from our calculated biphoton wavefunction, it is very useful to consider approximating our wavefunction with a Double-Gaussian wavefunction (i.e., the product of a Gaussian in x_+ and another Gaussian in x_- with widths chosen for good fitting [23, 64–66]. The general double-Gaussian wavefunction $\psi_{(DG)}(x_+, x_-)$ has the form:

$$\psi^{(DG)}(x_+, x_-) \equiv \left(\frac{1}{\sqrt{2\pi\sigma_{x_+}\sigma_{x_-}}}\right)e^{-\frac{x_-^2}{4\sigma_{x_-}^2}}e^{-\frac{x_+^2}{4\sigma_{x_+}^2}}. \tag{3.17}$$

where σ_{x_+} and σ_{x_-} are chosen for good fitting.

Though we considered the transverse pump profile to already be a Gaussian, approximating the sinc-based element of the biphoton wavefunction as a Gaussian can be approached in at least two ways. One can either fit a Double-Gaussian distribution to (3.15) by matching the statistics of one to the other, or one can fit the Double-Gaussian by matching some shape parameter of the other (e.g., the Full Width at Half Maximum (FWHM)). If we consider only the factor of our biphoton wavefunction (3.15) depending on x_-, we can take its modulus square to find a probability density $\rho(x_-)$. We plot our comparisons in Fig. 3.2.

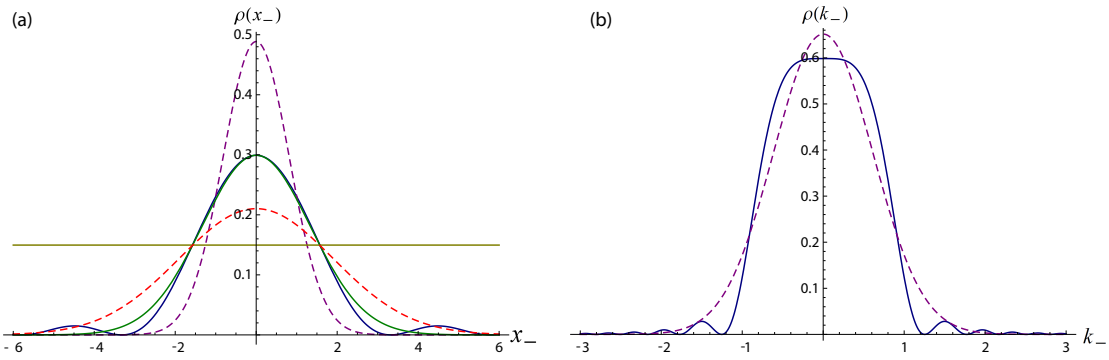

Figure 3.2: (a) Plot of $\rho(x_-)$ (shallow blue wavy curve) comparing it to different Gaussian fits. The tall skinny dotted curve, is the Gaussian approximation, when matching the variance of k_-. The shallow red dotted curve is the Gaussian approximation when matching the variance of x_-. The green medium Gaussian is the peak-matching approximation, which in this case also fits the full width at 48.2% of the maximum, and differs at the half-maximum by only 0.3%. (b) Plot of $\rho(k_-)$ (blunted blue wavy curve) comparing it to a Gaussian fit of matching mean and variance of k_-. Note that in both these plots, we set $a = 2$ for convenience.

In plotting our comparisons of different Gaussian fittings for $\rho(x_-)$ and $\rho(k_-)$, we find that a Double-Gaussian approximation of (3.15) is coarse, and one can at best form a reasonable fit of the central peak in position space. We note that the peak-matching Double-Gaussian (i.e., the Gaussian in Fig. 3.2.a with the same peak value as our calculated $\rho(x_-)$ is the one where probabilities close to the central maximum are most accurately represented. Fitting a Gaussian to $\rho(x_-)$ by matching the mean and variance of x_- does literally give the maximum

likelihood fit, but ironically creates a distribution that is significantly more broad than the central peak of $\rho(x_-)$.

Regardless of which fitting procedure one uses, they both agree with experimental data [54, 66–68] within experimental uncertainty, though the exact value again depends on one's fitting process. When measuring $\rho(x_-)$ through coincidence counting, thresholding away a noise floor gives one an estimate that more conforms to the peak-matching approximation of $\rho(x_-)$ since the small wings may be unresolvable in the experimental data. In this case, the estimated transverse correlation width from a peak-matching Double-Gaussian $\sigma^{PM}_{(x_1-x_2)}$ is:

$$\sigma^{PM}_{(x_1-x_2)} = \sqrt{\frac{4L_z\lambda_p}{9\pi}}, \tag{3.18}$$

which is smaller than the exact correlation width (3.16) by a factor of $\sqrt{40/81}$, or about 70.2%.

The reason many wish to approximate our transverse biphoton spatial amplitude by a Double-Gaussian wavefunction is that the latter is very easy to work with. Many statistics are straightforward to work out for the Double-Gaussian state. It is easy to relate a Double-Gaussian wavefunction to its Fourier transform (which is another Double-Gaussian), making both transverse position and transverse momentum statistics easy to calculate. The Double-Gaussian wavefunction is easy to propagate in the paraxial regime (the same regime in which we calculated our our biphoton state to begin with). In a remarkable mathematical tour de force, it is even possible to diagonalize the partial trace of the Double-Gaussian, giving a Schmidt decomposition of the state, and making it exceptionally convenient in continuous-variable quantum information. An extensive discussion of each of these benefits is beyond the scope of this thesis, but an interested reader may want to look at [57], and its accompanying references.

3.3 Entanglement, EPR-Steerability, and Bell-Nonlocality of the Double-Gaussian Biphoton Wavefunction

The Double-Gaussian biphoton state is ubiquitous in studies of continuous-variable quantum entanglement and quantum information [15, 23, 55, 64–66, 69–75], because, among other reasons, it is one of the few continuous-variable wavefunctions whose partial trace (giving the density operator of a subsystem) can be explicitly diagonalized. In doing so, measures of entanglement can be directly calculated instead of relying on entanglement witnesses. That being said, entanglement witnesses, EPR-steering inequalities, and Bell inequalities are also straightforward to test with the Double-Gaussian state.

3.3.1 Entanglement of the Double-Gaussian State

The Double-Gaussian wavefunction (3.17) describing the position statistics of a signal-idler pair of photons generated in SPDC cannot be factored into a product of a signal position wavefunction, and an idler position wavefunction. By definition, the Double-Gaussian state is entangled, except in the case where there are no position correlations.

Since the double-Gaussian state is a pure state, we can measure its entanglement from its reduced density operator. The density operator associated to a pure joint position wavefunction $\psi(x_1, x_2)$ is expressed as:

$$\hat{\rho}_{AB} = \int dx_1 dx_2 dx_1' dx_2' \psi^*(x_1, x_2)\psi(x_1', x_2')|x_1, x_2\rangle\langle x_1', x_2'|. \tag{3.19}$$

Its reduced density matrix $\hat{\rho}_A$ is expressed as:

$$\hat{\rho}_A = \int dx_1 dx_1' d\lambda \psi^*(x_1, \lambda)\psi(x_1', \lambda)|x_1\rangle\langle x_1'|. \tag{3.20}$$

A simple-to-calculate measure of entanglement of a pure quantum state is the Schmidt number κ, or the reciprocal of the trace of the square of the reduced density matrix:

$$\kappa = \frac{1}{\mathrm{Tr}[(\hat{\rho}_A)^2]} \tag{3.21}$$

Here, the Schmidt number is expressed as:

$$\frac{1}{\kappa} = \int dx_1 dx_1' dx_2 dx_2' \psi(x_1, x_2)\psi^*(x_1', x_2)\psi(x_1', x_2')\psi^*(x_1, x_2') \tag{3.22}$$

For the Double-Gaussian wavefunction (3.17), the Schmidt number has the simple expression:

$$\kappa(\psi^{(DG)}) = \frac{1}{2}\left(\frac{\sigma_{x_+}}{\sigma_{x_-}} + \frac{\sigma_{x_-}}{\sigma_{x_+}}\right). \tag{3.23}$$

Note that when σ_{x_-} is much less than σ_{x_+}, we have strong position correlations, and strong momentum anti-correlations. As this ratio grows, the correlations become stronger, and the state as measured by the Schmidt number becomes more entangled as well.

The Double-Gaussian state is special amongst continuous-variable entangled wavefunctions in that the eigenvalues of the reduced density operator are actually known [64]. The eigenvalues λ_n of $\hat{\rho}_A$ are:

$$\lambda_n = 4\sigma_{x_+}\sigma_{x_-}\frac{(\sigma_{x_+} - \sigma_{x_-})^{2n}}{(\sigma_{x_+} + \sigma_{x_-})^{2n+2}} \tag{3.24}$$

From these eigenvalues, we can calculate $S(A)$ (the von Neumann entropy of the reduced density operator), which is also a viable measure of entanglement. Interestingly, since the eigenvalues λ_n are geometrically distributed, and the geometric distribution is a maximum entropy distribution for a given ratio $\sigma_{x_+}/\sigma_{x_-}$, the Double-Gaussian state is also the maximally entangled state for that ratio.

3.3.2 EPR-Steerability of the Double-Gaussian State

Using Walborn *et al.*'s steering inequality (2.23), we can test the steerability of the Double-Gaussian state by evaluating its position conditional entropy:

$$h(x_1|x_2) = \frac{1}{2}\log\left(4\pi e\frac{\sigma_{x_+}^2\sigma_{x_-}^2}{\sigma_{x_+}^2+\sigma_{x_-}^2}\right), \tag{3.25}$$

and its momentum conditional entropy (expressed in terms of position statistics):

$$h(k_1|k_2) = \frac{1}{2}\log\left(\pi e\frac{1}{\sigma_{x_+}^2+\sigma_{x_-}^2}\right), \tag{3.26}$$

and compare it to the bound of $\log(\pi e)$. With this in mind, all non-steerable Double-Gaussian states satisfy the inequality:

$$\frac{1}{2}\left(\frac{\sigma_{x_+}}{\sigma_{x_-}}+\frac{\sigma_{x_-}}{\sigma_{x_+}}\right) \leq 1. \tag{3.27}$$

As the left hand side of this inequality is the Schmidt number of the Double-Gaussian state, and the Schmidt number has a minimum value of unity, no matter the state, we find that all Double-Gaussian states with nonzero correlations are not only entangled, they are also EPR-steerable (as seen in [55]).

3.3.3 Bell-Nonlocality of the Double-Gaussian State

The Double-Gaussian state is Bell-nonlocal, and yet will not violate a Bell inequality using transverse position statistics. For finite-dimensional pure-entangled states, Bell-nonlocality is assured as a consequence [5] of Gisin's theorem [76–78]. Pure entangled states of continuous-variable systems (or infinite-dimensional discrete systems) have not yet been shown to be Bell-nonlocal for the same reason. However, there are Bell inequalities [79, 80] that the double-Gaussian state will violate (though not explicitly in the position-momentum degree of freedom). On

[5] Gisin's theorem states that for any pure entangled state $|\Psi_{AB}\rangle$ of a pair of d-dimensional quantum systems, there exists some Bell inequality that $|\Psi_{AB}\rangle$ will violate.

the other hand, we also know that the Double-Gaussian state will not violate a Bell inequality with transverse spatial statistics because the Wigner function associated to the Double-Gaussian state is a well-behaved probability density (a Quadruple-Gaussian of sorts), and therefore can be used as an explicit local hidden variable (LHV) model for those correlations. In Chapter 6, we discuss our research into violating a position momentum bell inequality with the sinc-based transverse spatial amplitude, even though the Double-Gaussian state will never do so.

4 Position-Momentum EPR-Steering from Entropic Uncertainty Relations

In this chapter, we discuss our results in demonstrating position-momentum EPR-steering with discrete measurements. In particular, we show how to demonstrate continuous-variable EPR-steering experimentally in spite of a finite resolution, and finite range of detection. In doing so, we develop a new kind of Fano inequality suitable for continuous variables. The research discussed here can be found in my papers [81–83].

4.1 How to Witness Continuous-Variable EPR-Steering with Discrete Measurements

The measurement process may be analog or digital; the information may be in the position of a needle or in an LED readout; but quantifiable data, expressed with a finite number of significant figures, is essentially discrete. In particular, when acquiring information about a continuous degree of freedom, such as position or time, one is forced to do binning, either by the hardware (pixels in a digital camera) or the software (say, the number of bits used to store the arrival time in

photon counting). Although we may want to experimentally determine, say, the position probability distribution of photons on a detector, we may only measure the position statistics up to a given resolution, and often can only make statements about the position statistics subject to this caveat. However this is not always the case.

In this section, I shall discuss the relationship between a continuous probability density $\rho(x)$ and the probability distribution $P(X)$ resulting from discretizing $\rho(x)$ into equally spaced windows of size Δx. In doing so, I shall show I, with help from my colleague Curtis Broadbent, was able to successfully bound the continuous conditional entropy $h(x_B|x_A)$ using the entropy of its discrete approximation. With this bound, we shall show how knowledge only of the discretization of the (joint) position and momentum probability densities is sufficient to demonstrate EPR-steering in the continuous position-momentum degree of freedom.

As mentioned in section 1.1.2, there is a straightforward relationship (1.23) between the continuous entropy $h(x)$ and the entropy of its discrete approximation $H(X)$. This relationship also extends to multiple variables. In particular:

$$h(x,y) = \sum_{i,j} P(X_i, Y_j) h_{ij}(x,y) + H(X,Y), \tag{4.1}$$

where $h_{ij}(x,y)$ is the abbreviated form of $h(x,y|X_i,Y_j)$, which is the continuous entropy of x and y, conditioned on x being within window X_i and y being within window Y_j. Since the continuous conditional entropy $h(x|y)$ is just the difference between $h(x,y)$ and $h(y)$, we get a similar, slightly more cumbersome relationship for conditional continuous entropy:

$$h(x|y) = \sum_{i,j} P(X_i, Y_j)\big(h_{ij}(x,y) - h_j(y)\big) + H(X|Y) \tag{4.2}$$

Since conditioning does not increase the entropy, we arrive at an inequality for the conditional entropy:

$$h(x|y) \leq \sum_{i,j} P(X_i, Y_j)\big(h_{ij}(x|y)\big) + H(X|Y). \tag{4.3}$$

In addition, since the uniform distribution has a maximum entropy, i.e.,

$$h_{ij}(x|y) \leq \log(\Delta x). \tag{4.4}$$

we get our formal bound between discrete and continuous conditional entropies:

$$h(x|y) \leq H(X|Y) + \log(\Delta x). \tag{4.5}$$

With an inequality such as this, finding an upper bound to the continuous conditional entropy is only as hard as finding the discrete probability distribution, where any element can be determined experimentally (in principle). On the other hand, one can also show that in the limit of high resolution, the discrete approximation approaches the continuous distribution, and our inequality (4.6) becomes saturated (i.e., becomes an equivalency). Note that the prior relations are not exclusive to pairs of random variables, but can also be used to relate pairs of vectors of random variables. Indeed, where x and y are instead n-dimensional vectors $\vec{x}$ and $\vec{y}$, we find

$$h(\vec{x}|\vec{y}) \leq H(\vec{X}|\vec{Y}) + \log(\Pi_{i=1}^{n}\Delta x_i). \tag{4.6}$$

4.1.1 Continuous-Variable EPR-Steering with Discrete Entropies

What motivated my research into the conditional entropy bound (4.6) was to see whether it was possible to demonstrate position-momentum EPR-steering via Walborn *et al.*'s [55] inequality:

$$h(\vec{x}_B|\vec{x}_A) + h(\vec{k}_B|\vec{k}_A) \geq n\log(\pi e). \tag{4.7}$$

This steering inequality is very general in that it encompasses other steering inequalities (such as Reid's inequality using conditional variances [52], among others). The unfortunate downside to Walborn *et al.*'s inequality (4.7) is that the continuous entropies require knowledge of the continuous probability densities to calculate. Though it is a very good inequality for theoretical analyses [55, 71, 84, 85]

of entangled states, using it experimentally posed significant challenges. Trying to approximate the continuous entropy from discrete data is far from straightforward, and claims of EPR steering would be subject to the strengths of those approximations. To make a solid claim of EPR-steering, we needed a solid bound on those conditional entropies that experimental data could be used to evaluate.

With the conditional entropy bound (4.6), and Walborn *et al.*'s inequality (4.7), we developed the following steering inequality, as seen in [81, 82]:

$$H(\vec{X}_B|\vec{X}_A) + H(\vec{K}_B|\vec{K}_A) \geq \log\left(\frac{(\pi e)^n}{\Pi_{i=1}^n(\Delta x_{i,B}\Delta k_{i,B})}\right) \tag{4.8}$$

This steering inequality depends only on the discretized probability distributions $P(\vec{X}_A, \vec{X}_B)$ and $P(\vec{K}_A, \vec{K}_B)$, and bin sizes $\Delta x_{i,B}$ and $\Delta k_{i,B}$ (along each direction in space), which are determined by the experimental setup. Following this, we used the assumption that all flux is within the area (range) $L_{xA}L_{yA}$, $L_{xB}L_{yB}$ and $L_{kxA}L_{kyA}$, and $L_{kxB}L_{kyB}$ of our finite detection capabilities to derive the symmetric steering inequality:

$$H(\vec{X}_A : \vec{X}_B) + H(\vec{K}_A : \vec{K}_B) \leq \max_{\ell\in\{A,B\}} \log\left(\frac{\Pi_{i=1}^n L_{i\ell}L_{ki\ell}}{(\pi e)^n}\right), \tag{4.9}$$

where $H(\vec{X}_A : \vec{X}_B)$ is the mutual information between $\vec{X}_A$ and $\vec{X}_B$, and L_{iA} is the total range of positions along the i^{th} dimension with nonzero probability (within the detection range of detector A). With these inequalities relying on experimental parameters, we tested both of them using the transverse position and momentum statistics of entangled photon pairs generated via spontaneous parametric down-conversion (SPDC). Since our measurements were using both the horizontal and vertical degrees of freedom, we tested our inequality for $n = 2$.

The experimental setup is laid out as follows. First we used a 325nm HeCd pump laser to illuminate a BBO (β-Barium Borate) nonlinear crystal cut for type-I SPDC. We placed a filter blocking out light at the pump wavelength after the exiting face of the crystal, so that only down-converted light would pass through

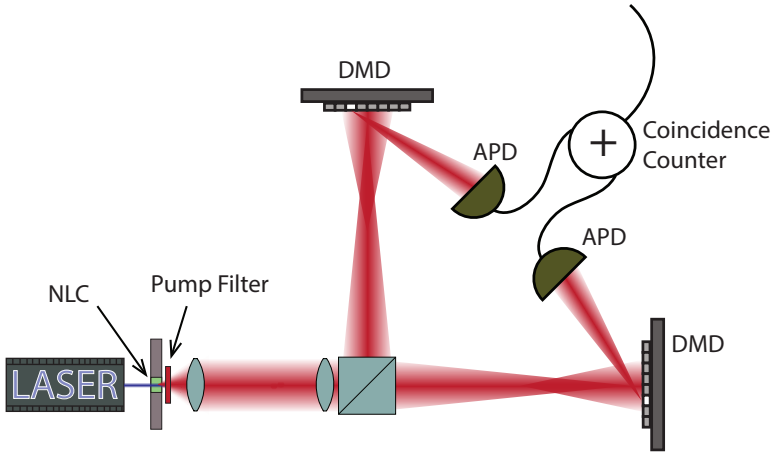

Figure 4.1: Diagram of experiment to witness position-momentum EPR steering. NLC is an abbreviation for nonlinear crystal. The lenses before the beamsplitter may be chosen to image either the exiting face of the NLC, or its Fourier transform.

the rest of the setup. Next, the downconverted light was separated[1] by a 50/50 beamsplitter into signal and idler arms, and DMD (Digital Micro-mirror Device) arrays are placed at the end of each arm. Lenses were placed in the signal and idler arms to image either the exit face of the nonlinear crystal (to measure position correlations), or the Fourier transform of the exit face (to measure momentum correlations). Much as in their name, DMD arrays are arrays of small mirrors arranged in a screen, where each mirror can be placed into one of two settings, reflecting incident light in one of two directions. Each DMD array (we call them A and B) was calibrated to reflect light either toward or away from a detector (in our case, an Avalanche Photo-Diode (APD). With an APD after the DMD array in each arm, we measured the transverse position (and momentum) statistics through coincidence counting. Since the photon pairs are highly correlated in time, we can measure the difference in arrival times between detectors, and see a significant peak of coincidence counts at the time difference related to the path

[1]The separation of the signal-idler photon pairs by a beamsplitter is statistical in nature. Half of the time, the signal photon goes down one arm, and the idler photon goes down the other. We can restrict our measurements to just these separated pairs through coincidence counting.

length difference in the signal and idler arms (including the path length from the DMD arrays to the detectors). By raster scanning a single pixel on each DMD array, one can obtain a coincidence histogram for the transverse positions (or momenta) of the signal and idler photons up to the resolution of the pixels on the DMD arrays. In order to estimate errors in experimental data, we noted that the coincidence counts are Poisson-distributed in time, so that we could take the square root of the number of coincidence counts as our uncertainty in them.

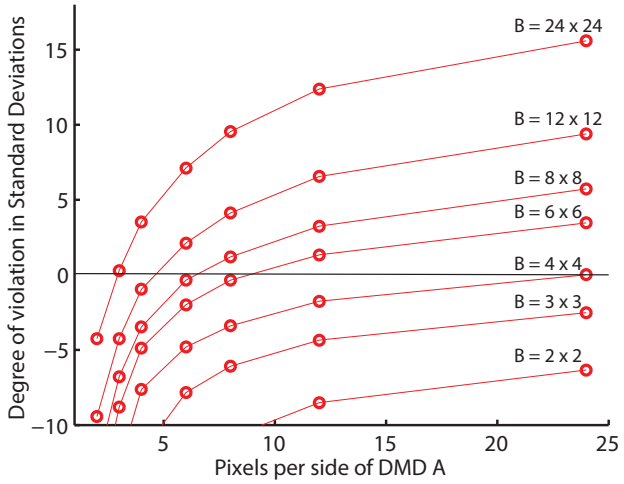

Figure 4.2: Plot of violation of conditional entropy steering inequality as a function of resolution of signal DMD (A). Each curve gives the violation for a constant resolution of the idler DMD (B). The degree of violation is in the number of standard deviations of the difference between the conditional entropy and the bound.

Using the joint coincidence histograms, properly normalized, as joint probability distributions, we tested our inequality, and found successful violation at multiple resolutions (where we may take groups of pixels to perform a lower resolution measurement). The reason that we looked at lower resolution measurements was two-fold. First, lower resolution measurements are easier to perform, making the claim about EPR-steering less cumbersome to test. Second is that we wished to probe the lower limit of resolution below which the inequality would be impossible to violate in principle. When the product of pixel sizes $\Pi_{i=1}^{n}(\Delta x_{i,B}\Delta k_{i,B})$ is

larger than $(\pi e)^n$, the bound on the right hand side of our inequality (4.8) becomes negative. Even with perfect position and momentum correlations, the left hand side could be no less than zero, so the inequality could never be violated at such resolutions. As seen in the plot in Fig. 4.2, the threshold where we no longer violate our inequality is larger than the theoretical one since our measured correlations are imperfect.

Our successful claim to demonstrating position-momentum EPR-steering is subject to a number of caveats or loopholes. First, the beamsplitter separating the photon pairs into signal/idler arms only does so 50% of the time; the rest of the time both photons go down one arm, or both down the other. This means that only 50% of the photon pairs are being recorded as coincidence counts, amounting to a substantial detection or fair sampling loophole. However, since a single beamsplitter is not expected to alter the transverse position or momentum correlations, the measurements we obtained should be accurate representations of the underlying joint position and momentum probability distributions. This particular complication could be avoided by using photon pairs from type-II SPDC, since photon pairs from type-II SPDC have orthogonal polarizations, and can therefore be more cleanly separated by a polarizing beamsplitter. The reason we did not use type-II SPDC was that it has a smaller conversion efficiency (resulting in fewer photon pairs and coincidence counts), and that the experimental setup would be more difficult to align (the signal and idler beams are not parallel). The second major caveat is also centered on accurate sampling. In order to normalize our coincidence histograms and convert them to empirical probability distributions, we had to assume that there were no photon pairs incident on the image planes of the DMD arrays outside their range, or ability to reflect that light into the APDs. Alternatively, we had to assume that the probability of getting a photon pair with positions outside the range set by the DMD arrays was small enough to be truly negligible. The problem here is that small probability events can con-

tribute significantly to the total entropy. Though this effect is not significant if the probability density decreases exponentially (or faster than any power law), the issue warranted further study, and lead to our even stronger (though less encompassing) steering inequality that allowed us to compensate for certain amounts of technical limitations such as these.

4.2 Position-Momentum EPR-Steering in spite of Finite Experimental Limitations

In our experimental demonstrations of position-momentum EPR-steering, we were forced to assume that the set of transverse positions and momenta imposed by the experiment (i.e., imposed by the DMD arrays) are the only ones that the photon pairs actually exhibit. Alternatively, we say that there are no significant counts outside the detection range of the experiment. However, even with the DMD arrays imposing specific detection ranges, there are actually gaps in between the mirrors where light is not being reflected. This fill factor is between 94% and values approaching 100% when using spatial light modulators (SLMs) in place of DMD arrays. The problem here is that small probabilities can contribute significantly to the total entropy. For Gaussian, exponential, and other probability distributions decaying faster than any power law, the effect on the total entropy may be insignificant, but being able to deal with these small probabilities is paramount if we are to make claims of position-momentum EPR steering in spite of experimental limitations. In particular, if we can use the probabilities we do measure to bound the statistics based on probabilities we cannot measure, we will have a powerful method indeed.

In [83], I show just how this may be done by developing a new kind of Fano inequality. While other papers [86] have dealt with the efficiency of detectors

(for polarization entanglement), our methods allow one to compensate for limited detector efficiency, dead space between pixels, and a finite viewing area. With this in mind, it is useful to understand just what the Fano inequality is.

4.2.1 The Fano Inequality

The Fano inequality is an inequality that bounds the conditional entropy $H(X|Y)$ between two discrete random variables X and Y with the probability that X and Y have identical outcomes. For X and Y having the same set of N outcomes, Fano's inequality [2] has the form:

$$H(X|Y) \leq h_2(\eta) + (1-\eta)\log(N-1). \tag{4.10}$$

where $h_2(\eta)$ is the binary entropy function $-\eta\log(\eta) - (1-\eta)\log(1-\eta)$, and the agreement probability η is defined as

$$\eta \equiv \sum_i P(X_i, Y_i) = P(X = Y), \tag{4.11}$$

and the sum is the maximum over all permutations of the outcomes of X and Y.

To understand our research into adapting Fano's inequality for continuous variables, it is useful to see where it comes from, as its derivation is straightforward [2]. To begin, we assume we have three random variables: X, Y, and G (see Fig. 4.3). The variables X and Y are the N-outcome random variables we discussed previously. The variable G is defined to have two outcomes. If $G = 0$, then we have the event $X \neq Y$. If $G = 1$, then we have the opposite event $X = Y$. We call G the "agreement" variable.

To start bounding $H(X|Y)$, we use the fact that adding a variable never reduces the entropy:

$$H(X|Y) \leq H(X, G|Y). \tag{4.12}$$

Next, we use Bayes' rule (also known as the chain rule for entropies) to get:

$$H(X|Y) \leq H(G|Y) + H(X|G, Y), \tag{4.13}$$

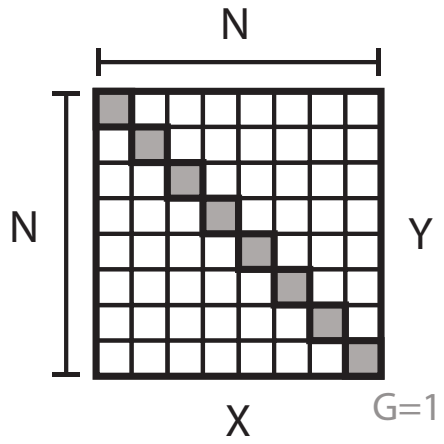

Figure 4.3: Diagram of discrete joint probability distribution $P(X, Y)$. The area in gray corresponds to those outcomes where $X = Y$, or the event $G = 1$.

and then the fact that removing a conditioning variable never reduces the entropy to get what starts to resemble the Fano inequality:

$$H(X|Y) \leq H(G) + H(X|G, Y). \tag{4.14}$$

Note that from our definition of the agreement probability η, we have that $H(G) = h_2(\eta)$.

By using the definition of the conditional entropy, we can expand $H(X|G, Y)$ as a sum over the outcomes of G (of which there are just two):

$$H(X|Y) \leq H(G) + P(G = 0)H(X|G = 0, Y) + P(G = 1)H(X|G = 1, Y). \tag{4.15}$$

Now, if $G = 1$, then $X = Y$, which means the conditional entropy $H(X|G = 1, Y)$ must be zero, leading us closer to the Fano inequality:

$$H(X|Y) \leq H(G) + P(G = 0)H(X|G = 0, Y). \tag{4.16}$$

To get Fano's inequality, we finally place an upper bound on $H(X|G = 0, Y)$. If $G = 0$, then $X \neq Y$, and the largest entropy $H(X|G = 0, Y)$ could have would be $\log(N - 1)$, since there are $N - 1$ possible outcomes of X, excluding the outcome of Y, and the uniform distribution has maximum entropy. As a result, we find:

$$H(X|Y) \leq H(G) + P(G = 0) \log(N - 1), \tag{4.17}$$

which is identical to our original statement of Fano's inequality, where $H(G) = h_2(\eta)$, and $P(G = 0) = (1 - \eta)$.

Using Fano's Inequality to Witness Entanglement

The Fano inequality has two principal applications. The first is in using the conditional entropy $H(X|Y)$ to bound the agreement (alternatively error) probability. In a communications channel whose output is given by Y, and input given by X, Fano's inequality gives a lower bound to the error probability $(1 - \eta)$ when the conditional entropy $H(X|Y)$ is known. This is useful in cases where we wish to prove, say, that sending messages through this channel in a particular way results in an error probability of at most so-and-so value. Indeed, Fano's inequality is used in this way to prove the inverse of the Shannon coding theorem [2] (i.e., that we cannot send information with an arbitrarily low probability of error at a rate larger than the channel capacity).

The second principal application of Fano's inequality is in using the agreement probability η to place an upper bound on the conditional entropy $H(X|Y)$. This was known to be usable to witness entanglement using the semi-classical analogue of the uncertainty principle in the presence of quantum memory (see (1.47) and neighboring text for definitions):

$$H(Q_A|Q_B) + H(R_A|R_B) \geq \log(\Omega) + S(A|B). \tag{4.18}$$

When the conditional entropies $H(Q_A|Q_B)$ and $H(R_A|R_B)$ are sufficiently small, the quantum conditional entropy $S(A|B)$ will have to be negative, and the state will have to be entangled. When $\hat{Q}_A$, $\hat{Q}_B$, $\hat{R}_A$, and $\hat{R}_B$ are N-dimensional observables, we can use Fano's inequality to place an upper bound of the conditional entropies $H(Q_A|Q_B)$ and $H(R_A|R_B)$ using their respective agreement probabilities, which we shall call η_q and η_r. As η_q and η_r become large (approaching unity), they make the upper bounds on $H(Q_A|Q_B)$ and $H(R_A|R_B)$ very small— small

enough to show that $S(A|B)$ is negative and that the pair AB is entangled. Indeed, using the entropic steering inequality that my collaborators and I developed in [56], one can show that under these same conditions (i.e., that the large agreement probabilities imply the entropy sum is less than $\log(\Omega)$), the quantum state describing AB must be EPR-steerable as well. In both these cases, Fano's inequality is useful because it serves as a witness of entanglement (and EPR-steering) using only enough information to estimate the agreement probabilities (a small subset of the total joint probability distribution).

4.2.2 Extending Fano's Inequality to Continuous Variables: Problem and Solution

While Fano's inequality is very useful for discrete random variables with a finite number of outcomes, the bound depends not just on the agreement probability η, but also on the logarithm of the number of outcomes N. Indeed, as the number of outcomes N grows, the bound grows on the order of $\log(N)$, even for a constant large agreement probability η.

In order to adapt Fano's inequality for continuous variables, we must be able to make sense of random variables with an infinite number of outcomes. However, as $N \to \infty$, the bound in Fano's inequality ceases to be useful, as we already know that $H(X|Y)$ is finite. Without addressing this problem, we cannot witness EPR-steering in continuous-variable systems, without being subject to significant sampling loopholes.

The way we address this problem, as seen in [83], is as follows. Note that although we only consider one position dimension here for simplicity, these arguments also hold for multiple dimensions. First, we consider a pair of continuous random variables x_A and x_B, discretized into windows of size Δx_A and Δx_B, respectively. These discretized random variables X_A and X_B may have a (count-

ably) infinite number of possible outcomes. If we were to try to bound $H(X_A|X_B)$ using Fano's inequality, the bound would cease to be useful. However, if we consider an additional random variable W, that subdivides the outcomes of X_A and X_B into $\bar{N}$-pixel "windows" (see diagram), we can find a useful upper bound to $H(X_A|X_B)$.

In particular, we define W to be a random variable with outcomes $\{0, 1, 2, 3, ...\}$. We consider "$W = 0$" to be the event that the outcomes of X_A and X_B are both within the $\bar{N}$-pixel range of the experiment. When $W \neq 0$, we have that X_A and X_B are not both within the $\bar{N}$-pixel range of the experiment, but are within other $\bar{N}$-pixel ranges (see diagram). If we consider our prior derivation of Fano's

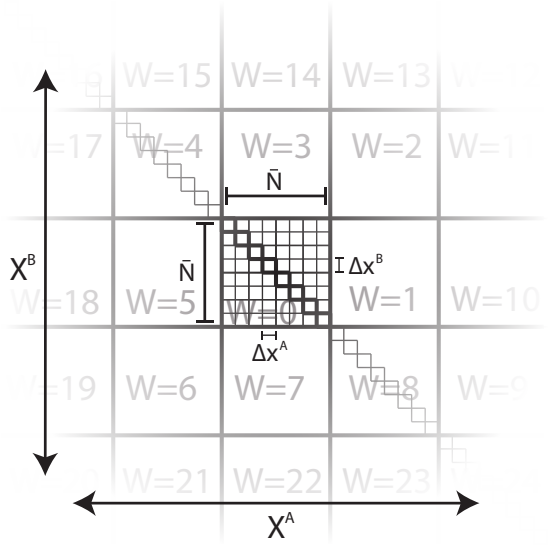

Figure 4.4: Diagram of the discretized joint probability distribution $P(X_A, X_B)$. Beyond the quantization into pixels of widths Δx_A and Δx_B, we cut up the space also into windows of $\bar{N}$ pixels to a side. Different windows correspond to different values of W, though we define $W = 0$ to be the event that X_A and X_B are within their $\bar{N}$-pixel detection ranges.

inequality, we can arrive at a similar step for $H(X_A|X_B)$, though now we have an additional term dependent on W, and we also have that the conditional entropy also depends on W, i.e.:

$$H(X_A|X_B) \leq H(W) + H(G) + H(X_A|G, W, X_B) \tag{4.19}$$

Next, we can break up $H(X_A|G,W,X_B)$ according to the outcomes of G, and simplify as before:

$$H(X_A|G,W,X_B) = P(G=0)H(X_A|G=0,W,X_B) \tag{4.20}$$

In addition, we break up $H(X_A|G=0,W,X_B)$ according to the outcomes of W:

$$H(X_A|G,W,X_B) = P(G=0)\sum_{i=0}^{\infty} P(W=i)H(X_A|G=0,W=i,X_B) \tag{4.21}$$

Furthermore, since each outcome of W narrows down X_A and X_B to be within $\bar{N}$-pixel domains (excluding the case that $X_A = X_B$), we know that $H(X_A|G=0,W=i,X_B)$ has an upper limit of $\log(\bar{N}-1)$. Putting this all together, we find:

$$H(X_A|X_B) \leq H(W) + H(G) + P(G=0)\log(\bar{N}-1). \tag{4.22}$$

If 100% of the probability is within the experimental range $W=0$, then $H(W)=0$, and (4.22) reduces to the original Fano inequality (4.17).

In order to finish our adaptation of Fano's inequality for continuous variables, we must deal with the matter that $H(W)$ has no upper bound. With an infinite number of possible outcomes, $H(W)$ can be arbitrarily large. However, if we include an additional very broad assumption, that W has a finite expectation value, we can get a rather tight upper bound. Since W has values $\{0,1,2,3,...\}$, the maximum entropy probability distribution for W given that it has a finite expectation value is the entropy of the geometric distribution. Second, since the entropy of the geometric distribution is a function only of its maximum probability (and a monotonically decreasing one at that), we can place an upper limit on $H(W)$ just using $P(W=0)$. For later simplicity, we define the domain probability $\mu_x = P(W=0)$ to be the probability that X_A and X_B are both within the experimental $\bar{N}$-pixel range. With this in mind, we have that:

$$H(W) \leq \frac{h_2(\mu_x)}{\mu_x}. \tag{4.23}$$

To finish this argument, we note that even though μ_x may not be the largest probability among all the outcomes of W, the entropy of a geometric distribution where μ_x is the maximum probability is larger than the entropy of a geometric distribution with a larger maximum probability.

As one additional complication, measuring the agreement probability ($P(G = 1) = \eta_x$) requires being able to normalize the probabilities obtained within the experiment by the total number of trials, which cannot be known if the outcome is undetectable. However, we can even deal with this by breaking up the agreement probability $P(G = 1)$ into two terms;

$$P(G = 1) = P(G = 1|W = 0)P(W = 0) + P(G = 1|W \neq 0)P(W \neq 0). \quad (4.24)$$

Here, we define the *measured* agreement probability $\bar{\eta}_x$ as $P(G = 1|W = 0)$. From this, we know that:

$$\eta_x \geq \bar{\eta}_x \mu_x. \quad (4.25)$$

When $\bar{\eta}_x \mu_x \geq 1/2$, we can just substitute $\bar{\eta}_x \mu_x$ for η_x, and finally arrive at a Fano inequality useful when using random variables with an infinite number of outcomes:

$$H(X_A|X_B) \leq \frac{h_2(\mu_x)}{\mu_x} + h_2(\bar{\eta}_x \mu_x) + (1 - \bar{\eta}_x \mu_x) \log(\bar{N} - 1). \quad (4.26)$$

To obtain a Fano inequality that can bound *continuous* conditional entropies, we use our earlier relation between discrete and continuous conditional entropies (4.6), and obtain:

$$h(x_A|x_B) \leq \log(\Delta x_A) + \frac{h_2(\mu_x)}{\mu_x} + h_2(\bar{\eta}_x \mu_x) + (1 - \bar{\eta}_x \mu_x) \log(\bar{N} - 1). \quad (4.27)$$

To recap, the assumptions going into our continuous-variable Fano inequality, beyond what went into the original Fano inequality are that x_A and x_B have finite expectation values, and that $\bar{\eta}_x \mu_x \geq 1/2$. We can do away with this second assumption by substituting $h_2(\bar{\eta}_x \mu_x)$ with its maximum value of unity, but this

would result in a looser (though perhaps still useful) bound. In addition, to use this Fano inequality for higher-dimensional continuous random variables, we need only consider that μ_x be the probability of both $\vec{x}_A$ and $\vec{x}_B$ being within the $\bar{N}$-pixel area or volume detectable by the experiment, and that $\bar{\eta}_x$ is the probability that $\vec{X}_A = \vec{X}_B$ given that they are both within that detectable area.

4.2.3 Improving Position-Momentum EPR-Steering with the Continuous-Variable Fano Inequality

With our continuous-variable Fano inequality (4.27), we can bound the continuous conditional entropies in Walborn *et al.*'s steering inequality (4.7), and demonstrate position-momentum entanglement and EPR-steering as a result. Considering the work that went into developing the continuous-variable Fano inequality, it is important to understand exactly what advantages demonstrating EPR-steering this way has over conventional methods.

The advantages of our Fano inequality stem from the relative ease of characterizing the bound, rather than the conditional entropy itself. The bound is a function of two probabilities, $\bar{\eta}_x$ and μ_x, and experimentally determined resolution $\bar{N}$ and pixel size Δx. Indeed, it takes substantially fewer resources to measure where one expects to see correlations in order to get a value for $\bar{\eta}_x$. Though $\bar{\eta}_x$ is technically determined by the entire joint probability distribution (since it is a maximum over all permutations of outcomes), one only needs to determine a sufficiently large value of $\bar{\eta}_x$, as any sum over correlated outcomes gives a lower limit to $\bar{\eta}_x$, and underestimating $\bar{\eta}_x$ only gives a bound larger than what the true bound would be (so long as $\bar{\eta}_x \mu_x \geq 1/2$). In addition, estimating the domain probability μ_x can be done by fitting, and comparing the probability of x_A and x_B within the experimental range to unity within that fitting distribution. Furthermore, underestimating the domain probability also results in a bound larger

than it would be otherwise (again, so long as $\bar{\eta}_x \mu_x \geq 1/2$).

In order to show that our Fano inequality is sufficiently sensitive for practical use, we used it to demonstrate position-momentum EPR-steering in photon pairs generated by SPDC. The experimental setup we used is similar to the one in Fig. 4.1, though it has minor differences, since the data was taken for a separate experiment in our group whose results are being written up at the present time. In particular, we measured $\bar{\eta}_x$, $\bar{\eta}_k$, and estimated μ_x and μ_k by fitting. If these four probabilities are known to be sufficiently large, then the conditional entropies in Walborn *et al.*'s inequality (4.7) will be known to be sufficiently small to produce a violation, demonstrating entanglement and EPR-steering.

In our experiments, we performed measurements at a resolution where $\bar{N} = 256$ (or 16×16), and found a maximum position agreement probability $\bar{\eta}_x$ of 69.4%, and a maximum momentum agreement probability of 75.1% (see Fig. 4.5). In addition, we estimated a minimum position domain probability μ_x of 99.7%, and a minimum momentum domain probability μ_k of 95.2% for momentum by fitting Gaussians to the empirical probability distributions.

The second major advantage of using our Fano inequality to demonstrate steering is that we can use good correlations (i.e., high agreement probabilities) to demonstrate EPR-steering for a whole range of possible domain probabilities. In this way, we can be assured that even if we do not know what the domain probabilities μ_x and μ_k are to a high precision, we can understand that if they are above a certain range of values, we may still claim to demonstrate EPR-steering. In Fig. 4.5, we plotted the violation of Walborn *et al.*'s steering inequality using our Fano inequality. The threshold for demonstrating EPR-steering is determined by the domain probabilities, and the threshold gets harder to violate as the domain probabilities get smaller.

If we could assume that 100% of the probability is within the $\bar{N}$ pixel domains we detected (i.e., that $\mu_x = \mu_k = 1$), then the threshold coming from those domain

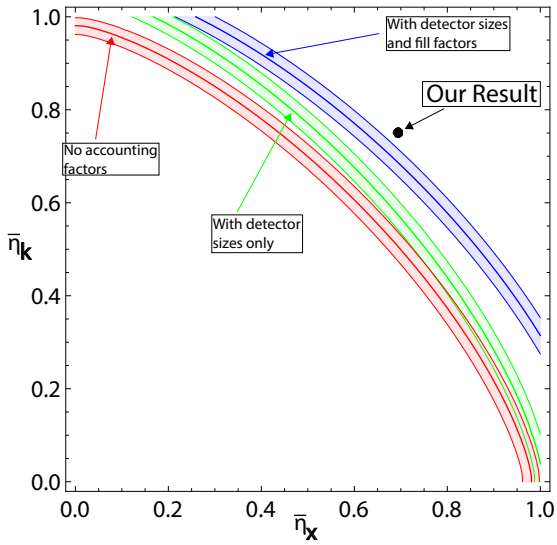

Figure 4.5: Contour plot of the violation of Walborn *et al.*'s steering inequality via our new Fano inequality as a function of the measured position and momentum agreement probabilities $\bar{\eta}_x$ and $\bar{\eta}_k$. The sub-contours on either side of the central contours in each band give values five standard deviations in the bound above and below the threshold for EPR-steering. Each band gives the threshold for EPR-steering for the domain probabilities based on different assumptions (as seen in the plot).

probabilities is given by the red band in Fig. 4.5. With that rather restrictive assumption, we do demonstrate EPR-steering, just as we did before when we directly measured the conditional entropies. However, with our relatively large agreement probabilities, there is a significant range of domain probabilities where the measurements will still demonstrate steering.

Considering that these domain probabilities can be interpreted as the probability of joint detection, there are other factors that affect their values than the wings of the probability distributions hanging outside the range of the experiment. Indeed, the DMD arrays we use to measure the position correlations have fill factors of about 92% (i.e., that only 92% of the surface of the DMD arrays were reflecting light; the rest of the space being gaps between mirrors). To measure the momentum correlations, we used spatial light modulators (SLMs), which had

effectively a fill factor of 100%. The SLMs serve the same purpose as the DMD arrays; their inclusion was to mitigate losses due to the $\approx 30\%$ reflectivity of the DMD arrays. The SLMs were included as a part of another experiment, where we (successfully) attempted to simultaneously measure the position and momentum correlations of downconverted light with a single set of partially projective measurements (see Ch. 5 for a description of partial projective measurements).

By accommodating the domain probabilities from Gaussian fitting, the threshold for demonstrating EPR-steering (as seen in Fig. 4.5) moves from the inner (red) band, to the middle (green) band. By accommodating both a finite detection range and the position and momentum fill factors into our domain probabilities, we move further away toward the outermost (blue) band. Even in this case, our agreement probabilities are still large enough to successfully demonstrate EPR-steering and entanglement in the position-momentum degree of freedom.

4.2.4 The Continuous-Variable Fano Inequality in Quantum Cryptography

Demonstrating position-momentum EPR-steering using our continuous-variable Fano inequality is useful because it allows us to successfully demonstrate EPR-steering with many fewer measurements [2], and it allows us to credibly demonstrate EPR-steering with sufficiently good data in spite of not having access to the entire joint probability distribution. Though we have improved how one might demonstrate position momentum EPR-steering, there is a useful application to doing so as well.

[2]While it takes of the order $\bar{N}^2$ measurements to determine the position and momentum probabilities (and therefore the conditional entropies) within the experiment, it takes only of the order $\bar{N}$ measurements to get a lower bound on the agreement probability through measuring where one expects to see correlations.

In entanglement-based quantum key distribution (QKD) [87], two parties, Alice and Bob, use measurements on pairs of separate, entangled particles in order to generate a secret key (i.e., two identical strings of random bits) known only to them. As an example, one way to do this in position and momentum (in one dimension for simplicity) would be as follows. First, Alice generates entangled pairs of photons AB, and sends each B to Bob. Next, Alice and Bob measure their photons in independently random measurement bases (either position of momentum) up to a pre-determined resolution. After the measurements are finished, Alice communicates to Bob her measurement bases, and they discard any measurements where they did not both measure in the same basis. The series of measurement outcomes Alice and Bob share is correlated. Indeed, the mutual information $H(X_A : X_B)$ characterizes the correlations between their outcomes when they both measure in position, and $H(K_A : K_B)$ characterizes the same correlations for momentum. Alice and Bob then have two correlated (but not perfectly) raw keys defined by their measurement outcomes.

If a third party, say Eve, were gathering information about their measurement correlations by intercepting B en route to Bob, and either measuring or coupling it to some other system she could measure later (say, once Alice and Bob communicate their measurement bases), then Eve could have a string of measurement outcomes correlated to both Alice's and Bob's outcomes. However, if Alice and Bob's measurement correlations are strong enough to violate an entropic EPR-steering inequality, they can guarantee that their measurement correlations are larger than any measurement correlations Eve could have with either of them. This is due in part to the monogamy of entanglement [88] and of EPR-steering [89]. More importantly, if Alice and Bob's measurement correlations are larger than Eve's correlations with either Alice or Bob, then Alice and Bob can create a secret key from their raw keys at a rate equal to the difference.

To understand how this monogamy works, we note that Alice, Bob's and Eve's

measurement correlations are limited by the complementary information tradeoffs [90]:

$$H(X_A|X_B) + H(K_A|K_E) \geq \log\left(\frac{\pi e}{\Delta x_A \Delta k_A}\right) \tag{4.28}$$

$$H(X_A|X_E) + H(K_A|K_B) \geq \log\left(\frac{\pi e}{\Delta x_A \Delta k_A}\right) \tag{4.29}$$

These tradeoffs can be derived from the information exclusion relation (see Sec. 1.4.2):

$$H(X_A|\theta) + H(K_A|\theta) \geq \log\left(\frac{\pi e}{\Delta x_A \Delta k_A}\right), \tag{4.30}$$

where we let the measurement variable θ refer to the pair of variables X_B and K_E, or the pair X_E and K_B. Since removing a conditioning variable never reduces the entropy, removing one of the pair from each entropy in the exclusion relation gives us the preceding complementary information tradeoffs.

By adding the complementary information tradeoffs (4.28) and (4.29) together, and rearranging the terms, we can get the following relation:

$$\left(H(X_A|X_B) + H(K_A|K_B) - \log\left(\frac{\pi e}{\Delta x_A \Delta k_A}\right)\right) + \\ + \left(H(X_A|X_E) + H(K_A|K_E) - \log\left(\frac{\pi e}{\Delta x_A \Delta k_A}\right)\right) \geq 0 \tag{4.31}$$

This relation is essentially the sum of two position-momentum EPR-steering inequalities. If the total of the first group of terms in parentheses is negative, then the correlations Alice and Bob share exhibit EPR-steering. On the other hand, if the total of the second group of terms in parentheses is negative, then the correlations between Alice's and Eve's measurements exhibit EPR-steering. What is remarkable about this inequality, is that *if* Alice and Bob's measurements violate their entropic steering inequality by R bits, then we know without having to consult Eve, that her measurements fail to violate an entropic steering inequality by at least that same amount.

To get a secure key rate, we may use the standard formula for three random variables (i.e., the wiretap channel [32, 91]). If X_A, X_B, and X_E are described

by a joint probability distribution $P(X_A, X_B, X_E)$ (and it is), then in the limit of long key strings, Alice and Bob may communicate secretly at a rate of R bits per channel use, where:

$$R \geq H(X_A : X_B) - H(X_A : X_E) = H(X_A|X_E) - H(X_A|X_B). \tag{4.32}$$

Combining this asymptotic secret key rate formula with the complementary information tradeoff (4.29), we find the rate formula:

$$R \geq \log\left(\frac{\pi e}{\Delta x_A \Delta k_A}\right) - H(X_A|X_B) - H(K_A|K_B). \tag{4.33}$$

Thus, if Alice and Bob can violate an entropic steering inequality by R bits, they can be assured, at least in the asymptotic limit, that they can communicate secretly at a rate of at least R bits per pair.

What makes our continuous-variable Fano inequality relevant to quantum cryptography is that it allows us to verify a violation of an EPR-steering inequality in spite of having incomplete access to the joint probability distribution. If Alice and Bob can only measure X and K within certain ranges, we can take the pessimistic viewpoint that Eve has access to every piece of information about Alice and Bob's measurement outcomes that Alice and Bob cannot access experimentally. If we can successfully demonstrate EPR-steering with our continuous-variable Fano inequality, we can show that Alice and Bob's correlations do violate the entropic steering inequality by at least the amount given by using our Fano inequality, guaranteeing a secure key rate even assuming all lost counts are in Eve's possession.

5 Partially Projective Measurements, and Their Effects on the Statistics of Complementary Observables

In this chapter, we discuss our research into the recent phenomenon of partially projective measurements. In particular, while a standard strong measurement in position projects the quantum state to be within a very narrow band of eigenvalues, a partially projective measurement in position instead projects the quantum state to be within a large randomly selected space of eigenvalues. Here, we discuss the nature of these partial projective measurements in position, and show that their effect on the momentum statistics of a quantum system is fundamentally different from a standard strong measurement. The research in this chapter constitutes my contributions to Howland *et al.*'s paper [92].

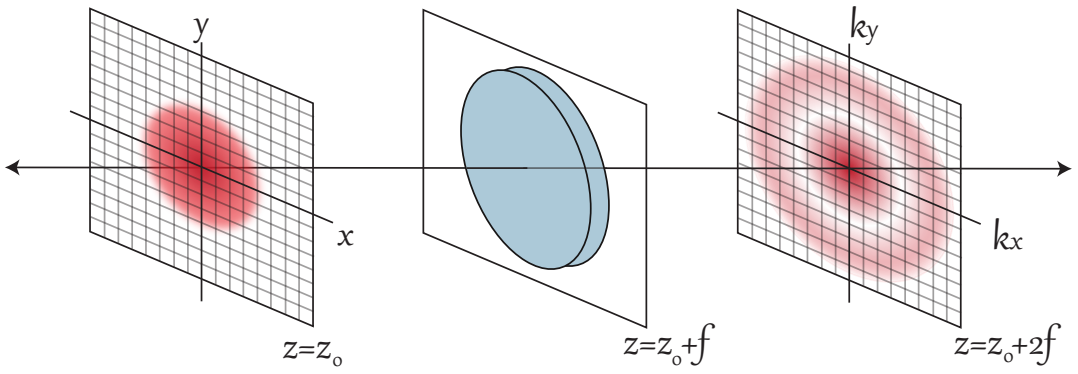

Figure 5.1: Depiction of $2-f$ optical setup with transverse position and momentum planes.

5.1 Strong Projective Measurements and the Uncertainty Principle

Consider the following. Let us have an optic axis defined by coordinate z, and a beam of light traveling along this optic axis with transverse coordinates (x, y), whose origin intersects the optic axis. What we want to do is to measure both the transverse position and transverse momentum statistics of the light in that plane as precisely as possible. The uncertainly principle places a fundamental limit on how well we may do this. In order to measure the transverse position statistics of a beam of light (one photon at a time) in some plane (say, $z = z_0$) perpendicular to the optic axis, one needs to measure the relative intensity of light as a function of transverse position. To measure the transverse momentum statistics of the light at $z = z_0$, one can use a Fourier transforming lens (of focal length f at transverse plane $z = z_0 + f$), and image the Fourier transform of the transverse position amplitude (i.e., the transverse momentum amplitude) onto a spatially resolving photodetector in the Fourier plane $z = z_0+2f$. Technical issues aside, the position measurement process amounts to counting the flux of photons incident on different squares on a fine grid laid out at $z = z_0$. By normalizing with the total flux on the grid, one can get an estimate of the transverse position

probability distribution at $z = z_0$ (see Fig. 5.1).

To measure the intensity of light passing through a single pixel at $z = z_0$, one can either place a spatially resolving photodetector at the plane $z = z_0$, or one can place a pinhole in that plane so that the opening is centered on that pixel. In either case, measuring the light passing through that pixel amounts to excluding, filtering, or *projecting* out all other transverse positions.

These position measurements, in which the outcome is narrowed down to a very narrow band of values are called *strong projective measurements*. By performing one of a series of strong projective measurement in position, the state after the measurement is localized (i.e., projected) to one of a number of small position eigenvalue ranges. In the ideal picture of quantum measurement, a strong projective measurement of an observable narrows down the statistics of that observable completely to a single eigenvalue. Since in that sense a true strong measurement of position is not realizable, we consider the limiting case of tight localization instead.

If we wish to measure both the transverse position and momentum statistics of the light at $z = z_0$ (using only one experimental setup), we can scan a pinhole in the plane $z = z_0$ while having a spatially resolving detector in the Fourier plane $z = 2f$. However, we are only detecting the light that has passed through the pinhole. Though we narrow down the possible transverse positions to a narrow band as a result of the pinhole, the uncertainty principle informs us that such a measurement cannot also give us significant information about the transverse momentum statistics at $z = z_0$.

As mentioned in Ch.1, the uncertainty principle can be understood in two contexts: that of localization, and that of information. In terms of localization, the Heisenberg uncertainty relation readily expresses the fundamental limits of

strong projective measurements in position or momentum:

$$\sigma_x \sigma_k \geq \frac{1}{2}. \tag{5.1}$$

Where the standard deviation σ_x is a measure of how tightly the position statistics are clustered about a central point, we understand that the more a position measurement localizes the position statistics, the more the momentum statistics must be spread out as a result. Strongly measuring the position statistics at high resolution limits the resolution at which we can also measure the momentum statistics. In particular, the effect of a strong projective position measurement is to blur the momentum distribution;

To understand how this blurring works, we consider the following. Let $\psi(x, y)$ represent the position transverse amplitude function of the light in the plane $z = z_0$. Let $h(x, y)$ be a filter (or aperture) function (i.e., zero everywhere except at the positions defining the pinhole, and unity otherwise). As a result of this pinhole, the subsequent transverse amplitude function $\bar{\psi}(x, y)$ (neglecting normalization) is equal to the product of the unperturbed transverse amplitude and the filter function: (i.e., $\bar{\psi}(x, y) = h(x, y)\psi(x, y)$). Thus the filter functions are unity in regions where light is transmitted, and zero where light is not transmitted. In momentum space, the perturbed momentum transverse amplitude $\bar{\psi}(k_x, k_y)$ is the convolution of the Fourier transform of the filter function $h(k_x, k_y)$, and the Fourier transform of the position transverse amplitude function (i.e., the unperturbed momentum amplitude function) $\psi(k_x, k_y)$ [93]. In other words, the effect of a pinhole on the transverse momentum amplitude is expressed as the convolution of the unperturbed momentum amplitude and the Fourier transform of the filter function;

$$\bar{\psi}(k_x, k_y) = \mathcal{N}\, h(k_x, k_y) * \psi(k_x, k_y), \tag{5.2}$$

where $\mathcal{N}$ is some normalization constant, and $(*)$ denotes convolution. The filter function $h(x, y)$ is a narrow function in position space. Its Fourier transform

$h(k_x, k_y)$ is therefore a broad function in momentum space. Since the convolution of a pair of functions is a measure of their overlap as one function is translated relative to the second, convolving $\psi(k_x, k_y)$ with a broad function in momentum space results in a blurring; each point (k_x, k_y) is in a sense a windowed average of $\psi(k_x, k_y)$ over a wide range of neighboring points [1]. The fine details of $\psi(k_x, k_y)$ are smeared out, since the effect of the narrow pinhole is to effectively block out all quickly varying components of $\psi(k_x, k_y)$ (i.e., as a conjugate low-pass filter). See Fig. 5.2 for an example of this blurring.

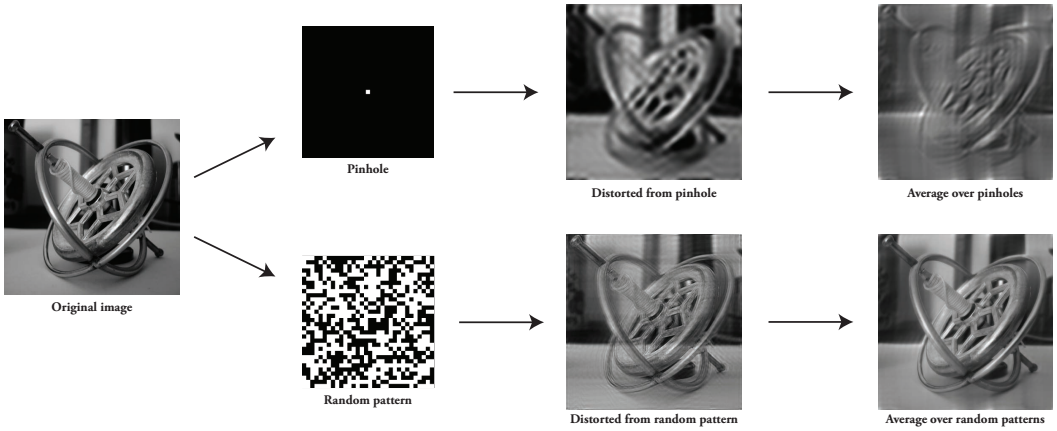

Figure 5.2: A photograph of a gyroscope distorted by filtering its discrete Fourier transform either with a single pinhole, or with a random binary pattern. The single pinhole centered at the origin acts as a low-pass filter, blurring all the fine structure in the image, while the random pattern (normalized for viewing) samples a wide variety of spatial frequencies randomly, preserving much of the original structure of the image. Averaging the distorted images over all pinhole positions gives some improvement, but there is still a fundamental resolution limit. On the other hand, averaging over multiple random patterns gives a nearly perfect copy of the original image.

[1] Indeed, if we convolve a rectangular function $\Pi(k_x, k_y)$ with $\psi(k_x, k_y)$, the result is a function where its value at every point (k_x, k_y) is precisely (constant factors aside) the mean of $\psi(k_x, k_y)$ over the range of values defined by the rectangular function centered at (k_x, k_y).

5.2 Partial Projective Measurements

While strong projective measurements in position are straightforward to implement, we have seen that if one wants to strongly measure both the position and momentum statistics of a quantum system, there is an immutable tradeoff in resolution. One is tempted to consider this tradeoff fundamental, as Heisenberg's uncertainty relation (5.1) supports this intuition. However, there are other measurement strategies known as *partially projective* measurements that avoid this tradeoff in favor of another.

As discussed in Ch.1, there is a more general notion of uncertainty beyond localization. For any measurement strategy whose outcomes are given by the random variable(s) θ, the entropic uncertainty relation applies:

$$h(x|\theta) + h(k|\theta) \geq \log(\pi e). \tag{5.3}$$

No matter one's measurement strategy θ, the remaining position and momentum uncertainty is bounded below by the same limit. This entropic uncertainty relation can be understood as an information-based uncertainty relation (as opposed to localization-based). What this also implies is that the more information a measurement strategy obtains about position (lowering $h(x|\theta)$), the less it can also obtain about momentum. However, this leaves open the possibility that a measurement obtaining a small amount of position information may yet also obtain a large amount of momentum information.

While a single strong projective measurement of the position of a particle examines whether or not the particle is within a particular narrow band of position eigenvalues, a *random partially projective measurement* of position examines whether or not the particle is in one of a very large number of random (narrow) bands of position eigenvalues. Just as a series of strong position projective measurements (i.e., scanning the pinhole) allows one to obtain the position statistics

of a quantum system, a series of random partial position projective measurements also allows one (with some effort) to obtain the same statistics.

Recall our example of a beam of light. A single strong position measurement is carried out by setting all pixels in our filter function to zero except for the position we wish to measure, and recording the flux that passes through. A single random partially projective position measurement is carried out by randomly setting each pixel of our filter function $h(x, y)$ to be zero or unity. Roughly half of the light passes through the random pinhole filter (also referred to as a random pattern). However, the amount of light that passes through is proportional to how well the filter function overlaps with the transverse position amplitude. With a series of position projective measurements, implemented by an ensemble of random filter functions $h_i(x, y)$, where i is an index running over the different random patterns, one can accurately obtain the transverse position statistics.

To see how one can do this, we can consider the following. Let $\vec{\mathbf{w}}$ be an M-dimensional vector, whose components are the intensities of the light passing through each (of M) random patterns. Let $\vec{\mathbf{v}}$ be an N-dimensional vector whose components are the mean intensities of light incident on each pixel (so that the total number of pixels is N), giving the transverse position probability distribution (note: for simplicity, the vector would be reshaped to be a single column.). Then, the measurement vector, $\vec{\mathbf{w}}$, is related to the signal vector $\vec{\mathbf{v}}$, by the matrix equation:

$$\vec{\mathbf{w}} = \mathbf{A}\vec{\mathbf{v}} + \vec{\mathbf{\Phi}}. \tag{5.4}$$

Here, each row of the $M \times N$ matrix $\mathbf{A}$ is a sequence of random zeroes and ones corresponding to the respective transmitting and blocking pixels of a particular random pattern, and $\vec{\mathbf{\Phi}}$ is a random M-dimensional vector accounting for the random variation of the intensity of the detected light. With a sufficiently large number of random patterns (i.e., for $M > N$) [94], one can reconstruct $\vec{\mathbf{v}}$ from $\vec{\mathbf{w}}$ by inverting $\mathbf{A}$. Indeed, one can even come up with a simple noisy estimate

of $\vec{\mathbf{v}}$ by taking the weighted average of the random patterns with weightings proportional to the transmitted intensities $\vec{\mathbf{w}}$. In the limit as the number of random patterns M becomes large (i.e., as $M \gg N$), this noisy estimate approaches $\vec{\mathbf{v}}$ exactly. However, both of these strategies become cumbersome for high-resolution measurements either because of the sheer number of random patterns required or because of the difficulty of inverting high-dimensional random matrices. Fortunately, such signal reconstructions can be carried out efficiently (i.e., with $M < N$) using compressive sensing techniques by essentially treating these random partial projective measurements in position as an implementation of a single-pixel camera [95].

Though the details of compressive sensing are beyond the scope of this thesis, it is useful to cover the basic idea behind how it works. Equation (5.4) is an under-determined system with M equations and N unknowns; there are an infinite number of possible signals $\vec{\mathbf{v}}$ that could correspond to our measurement vector $\vec{\mathbf{w}}$. However, if we know that the signal $\vec{\mathbf{v}}$ is sparse in some basis (i.e., has a low number of significant or nonzero elements), and if the random sensing matrix is a restricted isometry [96] for all sparse vectors $\vec{\mathbf{v}}$, then $\vec{\mathbf{v}}$ is uniquely determined from $\mathbf{A}$ and $\vec{\mathbf{w}}$. In practice, $\vec{\mathbf{v}}$ may be found as a solution of a variety of optimization problems. Random binary sensing matrices satisfy this property with overwhelming probability when M is larger than approximately $s \log(N/s)$, where s is the sparsity [2] of $\vec{\mathbf{v}}$.

[2] A vector $\vec{v}$ is S-sparse in a given basis $\{\mathcal{B}\}$ if it has S or fewer nonzero components when expressed in that basis. A signal represented by vector $\vec{v}$ may be approximately S-sparse in $\{\mathcal{B}\}$ if it has S or fewer *significant* components in that basis.

5.2.1 The Effect of Partial Projective Measurements of Position on Transverse Momentum Statistics

Previously, we showed how it is possible to obtain the transverse position statistics of light passing through the plane $z = z_0$ (recall Fig. 5.1) either by raster-scanning a pinhole (strong projections), or through a series of random patterns (partial projections). Knowing this, it is important to understand what the effect of these random partial projective measurements in position has on the transverse momentum amplitude, and how it differs from the blurring effect of strong position projections.

Just as with the strong measurement, the effect of each random filter function $h_i(x, y)$ on the transverse momentum amplitude is expressed as a convolution:

$$\bar{\psi}(k_x, k_y) = \mathcal{N}\, h_i(k_x, k_y) * \psi(k_x, k_y), \tag{5.5}$$

Unlike our strong position measurement, $h_i(x, y)$ is a sum of many pinhole filter functions placed at random locations:

$$h_i(x, y) = \sum_{\ell,m} a^{(i)}_{\ell,m} \prod_{\Delta_x,\Delta_y} (x - g\ell, y - gm). \tag{5.6}$$

Here we consider $\Pi_{\Delta_x,\Delta_y}(x, y)$ to be a top-hat function (with top equal to unity) centered at $(x, y) = (0, 0)$ with widths $\Delta_x = \Delta_y = g$. The random filter function $h_i(x, y)$ is then a square lattice of pixels, whose values are given by $a^{(i)}_{\ell,m}$, which is zero or unity with equal probability.

In order to get $h_i(k_x, k_y)$ so that we can examine its convolution with $\psi(k_x, k_y)$, and therefore its effect on it, we take the Fourier transform of our random filter function, which is just the sum of the transforms of the translated top-hats:

$$h_i(k_x, k_y) = \frac{\Delta_x \Delta_y}{2\pi} \mathrm{sinc}\big(\frac{\Delta_x k_x}{2}\big) \mathrm{sinc}\big(\frac{\Delta_y k_y}{2}\big) \sum_{\ell,m} a^{(i)}_{\ell,m} e^{-i(k_x g\ell + k_y gm)}, \tag{5.7}$$

Here we see that as a result, we get the Fourier transform of a single top hat multiplied by the sum of translation exponentials (as seen in the shift theorem

[93]). In addition, this model of the random filter function also allows us a direct comparison between partial and strong projective measurements.

The random filter function (5.7) is characterized by two terms: a Sinc-based envelope function we shall denote as $\mathrm{Env}(k_x, k_y)$, and a weighted sum of complex exponentials. If we were performing a strong position measurement with a single pinhole, all elements $a_{\ell,m}^{(i)}$ would be zero except for one corresponding to the position of the pinhole. In this case, the effect of a such a strong position measurement is a convolution with the broad shallow sinc-based envelope function and the exponential due to the shift, amounting to a blurring as expected. However, the sum of multiple random complex exponentials has an unusual effect on the transverse momentum amplitude $\psi(k_x, k_y)$ as we shall see.

To understand the behavior of the exponential sum in (5.7), we make the following assumptions. First, we assume the number of pixels N is large enough, so that nearly half of them will be zero (and the other half unity) for any given random pattern. Because of this, we can say that for $(k_x, k_y) = (0, 0)$, there are $N/2$ phasors adding constructively, so that $h_i(0, 0) \approx \frac{N\Delta_x\Delta_y}{4\pi}$. Second, we assume that for $(k_x, k_y) \neq (0, 0)$, the sum of exponentials can be modeled as a sum of $N/2$ uniformly distributed *random* phasors (of unit length). This means that the exponential sum is considered to be a random walk in the complex plane of unit step size, arbitrary direction, and $N/2$ steps. This random-phasor approximation is limited in that the phases are not each independent and identically distributed; there is a finite pool of N phasors to choose from for any point. However, we will see that this approximation gives us a conservative estimate of the disturbance to $\psi(k_x, k_y)$; a more accurate treatment would result in a smaller disturbance.

The Central Limit Theorem tells us that the net displacement of a random walk is Gaussian distributed (see Fig. 5.3 for an example). This allows us to model our random filter function $h_i(k_x, k_y)$ as a weighted sum of two functions. One is a very narrow sinc function of unit height that we may approximate as the ratio of

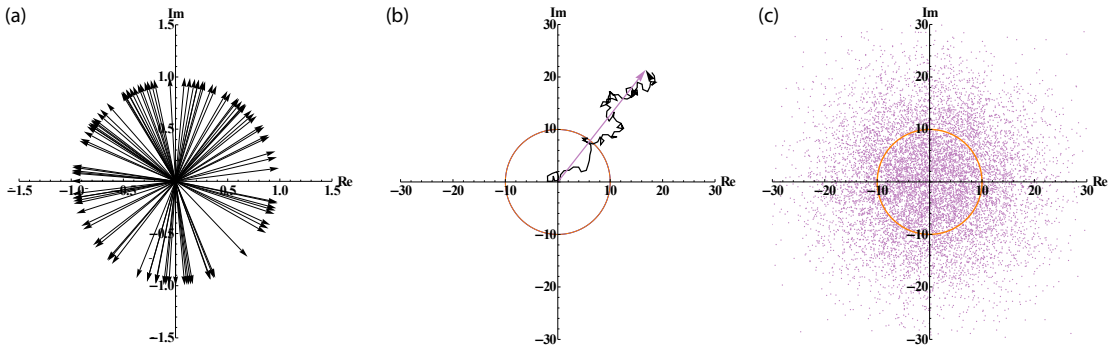

Figure 5.3: (a) A depiction of 100 randomly drawn phasors. (b) A depiction of the sum of 100 randomly drawn phasors. This represents a random walk of unit step size on the complex plane with 100 steps. The orange circle gives the the radius of one standard deviation for a random walk with 100 steps. (c) Scatterplot of 10,000 random walks of 100 steps each. The density of points approaches a Gaussian.

a Dirac delta function $\delta(k_x, k_y)$, and the area A of our grid of pixels [3], while the second is a random function $\phi_i(k_x, k_y)$ for every other value of k_x and k_y so that:

$$h_i(k_x, k_y) = \mathrm{Env}(k_x, k_y)\left(\alpha\frac{\delta(k_x, k_y)}{A} + \beta\phi_i(k_x, k_y)\right), \tag{5.8}$$

where again

$$\mathrm{Env}(k_x, k_y) \equiv \frac{\Delta_x\Delta_y}{2\pi}\mathrm{sinc}\left(\frac{\Delta_x k_x}{2}\right)\mathrm{sinc}\left(\frac{\Delta_y k_y}{2}\right), \tag{5.9}$$

and

$$\alpha = \frac{N}{2} \qquad : \qquad \beta = \sqrt{\frac{N}{2}}. \tag{5.10}$$

[3] If one supposes a random pixel array of infinite extent, then the Fourier transform of that random filter function would be the sum of a Dirac delta function and an insignificantly small Gaussian random noise floor. A random pixel array of finite extent is expressible as the product of an infinite random array with a pinhole of finite extent. In momentum space, the random filter function is the convolution of the transformed infinite array, with the narrow sinc function corresponding to the broad pinhole. The narrow sinc function of unit height has an area surrounding the origin $(k_x, k_y) = (0, 0)$ of approximately $1/A$, where all of the exponentials in the sum (5.7) will interfere constructively. In the limit where the area A of our grid in position space is large enough to cover the significant elements of the transverse position amplitude $\psi(x, y)$, $\psi(k_x, k_y)$ will be slowly varying enough that we may use our delta function approximation.

Here, α and β are weightings, and $\phi_i(k_x, k_y)$ is a function, where for each value of (k_x, k_y), the function takes a random complex value distributed according to a Gaussian centered at the origin of the complex plane with real and imaginary variances set equal to unity.

In order to find values for the weightings α and β, we consider what the mean and mean-square values of $h_i(k_x, k_y)$ when averaging over many random patterns. The mean value of $h_i(k_x, k_y)$ when averaging over many patterns (indexed by i) is a function that is zero where $(k_x, k_y) \neq (0, 0)$, and $\frac{N\Delta_x\Delta_y}{4\pi}$ where $(k_x, k_y) = (0, 0)$. From this, we can infer that $\alpha = N/2$. When averaging the magnitude squares of $h_i(k_x, k_y)$ over many random patterns, we find that for $(k_x, k_y) \neq (0, 0)$ we get $N/2$ times the square of the envelope function. From this, we can infer that $\beta = \sqrt{N/2}$.

With our model for $h_i(k_x, k_y)$ established, we find there is a significant difference between $h(k_x, k_y)$ for a single pinhole, and $h_i(k_x, k_y)$ for a random set of pinholes. In particular, for a single pinhole, the qualitative behavior of the filter function is dominated by the sinc-based envelope $\mathrm{Env}(k_x, k_y)$. When we instead use a random array of pinholes, the random filter function is a product of the envelope function, and a sharply peaked function with a random shallow noise floor. The consequence of this is that while $h(k_x, k_y)$ for a single pinhole gets broader as the pinhole size gets smaller (and N gets larger to cover the same space), $h_i(k_x, k_y)$ for a random array of pinholes remains sharply peaked (see Fig. 5.4 for example).

As mentioned previously, the effect of a filter function on the transverse momentum amplitude is expressed as the convolution (5.5) of the filter function in momentum space, and the unperturbed momentum amplitude. When $h_i(k_x, k_y)$ is a narrow function, its effect on $\psi(k_x, k_y)$ will not be a blurring, but rather a modulation with a noise floor. With our model of the random filter function (5.8), we can directly calculate its effect of the transverse momentum intensity

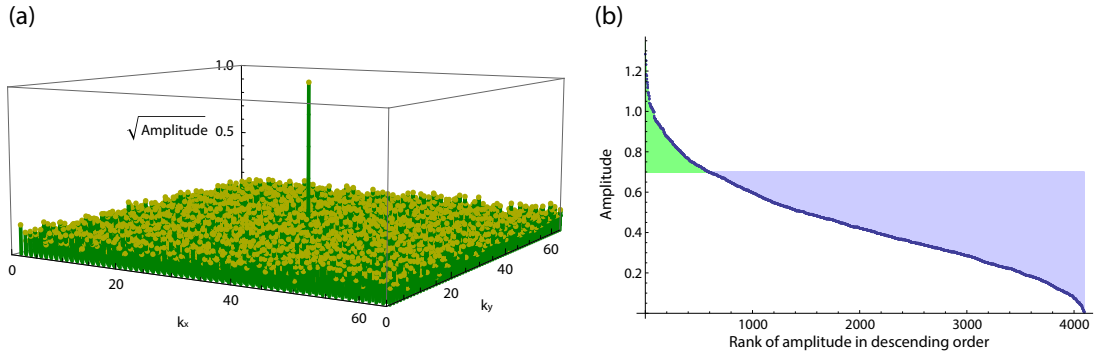

Figure 5.4: (a) Plot of the square root of the amplitudes of the discrete Fourier transform of a random 64×64 binary pattern, normalized so that the maximum amplitude is unity. (b) Plot of the amplitudes of the same Fourier transform in descending order. As one can see, most of the components are of the order $\sqrt{2/N}$ of the peak intensity (i.e, below the threshold defined by the shading).

distribution, as given by $|\psi(k_x, k_y)|^2$.

$$|\bar{\psi}(k_x, k_y)|^2 = \left|\psi(k_x, k_y) * \mathrm{Env}(k_x, k_y)\left(\frac{N}{2A}\delta(k_x, k_y) + \sqrt{\frac{N}{2}}\phi_i(k_x, k_y)\right)\right|^2 \quad (5.11)$$

For simplicity, let us consider the case, where the pinhole size Δ_x (and Δ_y) is small enough that $\mathrm{Env}(k_x, k_y)$ can be considered constant over $\psi(k_x, k_y)$, such as when Δ_x and Δ_y are smaller than the reciprocal of the horizontal and vertical momentum bandwidths [4] in k_x and k_y. Noting that convolution with a Dirac delta function does not change the function convolved, we find:

$$|\bar{\psi}(k_x, k_y)|^2 \approx \mathcal{N}\left|\psi(k_x, k_y) + A\sqrt{\frac{2}{N}}\psi(k_x, k_y) * \phi_i(k_x, k_y)\right|^2 \quad (5.12)$$

where $\mathcal{N}$ is some normalization constant (absorbing factors of 2, N, and A). Here we see for a constant L, the perturbed momentum amplitude $\bar{\psi}(k_x, k_y)$ is approximately the unperturbed momentum amplitude $\psi(k_x, k_y)$ plus a correction function seeming of the order $1/\sqrt{N}$.

[4] The horizontal momentum bandwidth is meant to be read here as range of values k_x enclosing the significant elements of $\psi(k_x, k_y)$. The vertical momentum bandwidth is the corresponding range of values of k_y.

In order to understand the relative significance of this correction function, we need to understand the relative magnitude of $\psi(k_x, k_y) * \phi_i(k_x, k_y)$. Where the convolution can be understood as the area of overlap between $\psi(k_x, k_y)$ and $\phi_i(k_x, k_y)$, it is also of the order of the integral of $\psi(k_x, k_y)$ over all space, which is 2π times the position space amplitude $\psi(x, y)$ at the origin. With this, we see that $\psi(k_x, k_y) * \phi_i(k_x, k_y)$ is of the order unity, and does not scale with N.

While a single random partial projective measurement in position may disturb the momentum distribution with a noise floor, performing a series of partial projective measurements in position mitigates this disturbance. Indeed, by averaging the momentum intensity distributions at $z = z_0 + 2f$ (see Fig. 5.1) obtained from each of a series of random patterns, the resulting intensity distribution is much closer to the unperturbed momentum distribution than a single measurement can so. To see how this works, we expand our approximation to the perturbed probability $|\bar{\psi}(k_x, k_y)|^2$, as a sum of three terms:

$$\begin{aligned} |\bar{\psi}(k_x, k_y)|^2 \approx \mathcal{N}\Bigg(& |\psi(k_x, k_y)|^2 \\ & + 2A\sqrt{\frac{2}{N}}\mathrm{Re}\big[\psi^*(k_x, k_y)(\psi(k_x, k_y) * \phi_i(k_x, k_y))\big] \\ & + \frac{2A^2}{N}|\psi(k_x, k_y) * \phi_i(k_x, k_y)|^2\Bigg). \end{aligned} \tag{5.13}$$

If we take $|\bar{\psi}(k_x, k_y)|^2$ from each of, say M random patterns (indexed by i), add them together, and divide by M (finding their mean), we find that this mean value scales not as $1/\sqrt{N}$, but as $1/N$. Where this mean can be calculated as the mean of each term, we see that the second term vanishes with increasing M; the convolution is linear, and the mean of $\phi_i(k_x, k_y)$ diminishes as M increases. Thus for large M and N, $|\bar{\psi}(k_x, k_y)|^2$ is very nearly $|\psi(k_x, k_y)|^2$, but with an additional low-level noise floor.

5.3 Measuring Complementary Correlations with Partial Projections

In the previous section, we discussed the effect of a random filter function on a single optical field. We found that while a single pinhole in position substantially limited our resolution in momentum measurements (and this limitation only grew worse with smaller pixel size), random arrays of pinholes in position instead had the effect of adding noise to our momentum measurements, an effect that grows less prevalent as the number of pixels N grows large. This advantage of random partial projective measurements is not limited to single optical fields. Indeed, it has been shown [97] that partial projective measurements offers a similar advantage when measuring the transverse position and momentum amplitudes of entangled optical fields as generated by spontaneous parametric down-conversion (SPDC).

The transverse spatial amplitude of the signal-idler photon pair generated in SPDC can be represented (in position space) by the wavefunction $\psi(x_1, y_1, x_2, y_2)$, where subscripts 1 and 2 refer to signal and idler fields respectively. To measure the transverse spatial correlations using partial projective measurements, we could use an experimental setup similar to Fig. 4.1. We would separate the signal and idler fields using a 50/50 beamsplitter and post-selecting on coincidence counts. We place separate random pinhole arrays in the image planes of the nonlinear crystal (in the signal and idler arms, respectively). Then we place the DMD (digital micro-mirror device) arrays in the Fourier planes of the nonlinear crystal to serve as a spatially resolving detector setup.

The effect of random pinhole arrays in position on an entangled pair of optical fields is much like it is for a single optical field. The pair of filter functions $h_i(x_1, y_1)$ and $h_j(x_2, y_2)$ acts on the joint transverse amplitude by either transmitting or not

transmitting the light incident on the pinholes, giving:

$$\bar{\psi}(x_1, y_1, x_2, y_2) = h_i(x_1, y_1)h_j(x_2, y_2)\psi(x_1, y_1, x_2, y_2) \tag{5.14}$$

Just as with a single optical field, the perturbed momentum amplitude is represented by the convolution:

$$\bar{\psi}(k_{x1}, k_{y1}, k_{x2}, k_{y2}) = \big(h_i(k_{x1}, k_{y1})h_j(k_{x2}, k_{y2})\big) * \psi(k_{x1}, k_{y1}, k_{x2}, k_{y2}) \tag{5.15}$$

If N is sufficiently large, we may use our model for $h_i(k_x, k_y)$ (5.8) to describe our random filter functions. For simplicity, we assume that both the signal and idler filter functions have the same number of pixels N and same total area A. As a result, we find:

$$\begin{aligned}\bar{\psi}(k_{x1}, k_{y1}, k_{x2}, k_{y2}) = \mathcal{N}\Bigg[&\mathrm{Env}(k_{x1}, k_{y1})\mathrm{Env}(k_{x2}, k_{y2})\times \\ &\times \Bigg(\alpha^2\frac{\delta(k_{x1}, k_{y1}, k_{x2}, k_{y2})}{A^2}+ \\ &+ \alpha\beta\frac{\delta(k_{x1}, k_{y1})}{A}\phi_j(k_{x2}, k_{y2}) + \alpha\beta\frac{\delta(k_{x2}, k_{y2})}{A}\phi_i(k_{x1}, k_{y1})+ \\ &+ \beta^2\phi_i(k_{x1}, k_{y1})\phi_j(k_{x2}, k_{y2})\Bigg)\Bigg] * \psi(k_{x1}, k_{y1}, k_{x2}, k_{y2})\end{aligned} \tag{5.16}$$

If we again assume that the pixel size is small enough that our envelope functions are more or less uniform over the significant elements of $\psi(k_{x1}, k_{y1}, k_{x2}, k_{y2})$, we can neglect them in our description of $\bar{\psi}(k_{x1}, k_{y1}, k_{x2}, k_{y2})$. By noting that the convolution with a delta function is an identity operation, we can express the perturbed wavefunction as:

$$|\bar{\psi}(k_{x1}, k_{y1}, k_{x2}, k_{y2})|^2 = \mathcal{N}\big[|\psi(k_{x1}, k_{y1}, k_{x2}, k_{y2})|^2 + \frac{c_1}{\sqrt{N}} + \frac{c_2}{N} + \frac{c_3}{N\sqrt{N}} + \frac{c_4}{N^2}\big], \tag{5.17}$$

where $c_1, ..., c_4$ are functions of $(k_{x1}, k_{y1}, k_{x2}k_{y2})$ of the order unity. By averaging $|\bar{\psi}(k_{x1}, k_{y1}, k_{x2}, k_{y2})|^2$ over more and more random filter functions, the contributions due to functions c_1 and c_3 diminish since they are linear in the random noise function $\phi_i(k_x, k_y)$, which diminishes under averaging. For large M and N,

$|\bar{\psi}(k_{x1}, k_{y1}, k_{x2}, k_{y2})|^2$ closely resembles $|\psi(k_{x1}, k_{y1}, k_{x2}, k_{y2})|^2$ with an additional noise floor. Indeed, using these sorts of measurements, we [97] were also able to show that the transverse spatial correlations were strong enough to demonstrate EPR-steering.

5.4 Applications of Partial Projective Measurements

In [92], we used partial projective measurements to obtain both the transverse position and momentum statistics of a beam of light. Besides being of fundamental interest (i.e., being able to get the salient information/statistics of complementary observables without violating the uncertainty principle), random partial projective measurements have useful applications as well.

To begin, partially projective measurements, combined with optimization via compressive sensing techniques, allow one to acquire the same joint probability distributions one would obtain with strong measurements, but with fewer of them. Indeed, M may be much less than N if the image is sufficiently sparse in a given basis, since the number of measurements M needed to acquire a signal in compressive sensing depends not on its total dimensionality N, but on its compressibility (i.e, sparsity).

Second, the efficient acquisition of both the transverse position and momentum intensity distributions will also improve the resources needed to implement phase retrieval algorithms [98] that would give us information about the complete transverse spatial amplitudes.

6 Position-Momentum Bell-Nonlocality in Spontaneous Parametric Down-Conversion

In this chapter, we discuss our strategy for demonstrating the Bell-nonlocality in the transverse position statistics of entangled photon pairs. In particular, we show that by sign-binning (i.e., measuring the correlations in the signs of the transverse positions of the photon pairs), we can demonstrate position-momentum Bell nonlocality in a manner discussed by Bell [99, 100], even though the ideal-Double-Gaussian model for correlated photon pairs can never do so. Furthermore, we discuss why the transverse spatial statistics of photon pairs from degenerate type-I spontaneous parametric down-conversion (SPDC) may yet demonstrate Bell nonlocality, and how we might tailor the biphoton wavefunction (to make such violations easier) through alterations to the $\chi^{(2)}$ nonlinear crystal from which these pairs originate.

6.1 Bell's Approach to Continuous-Variable Nonlocality

One of the predictions of quantum mechanics that casts it in starkest relief to our classical intuitions is the violation of Bell inequalities. Indeed, the violation of Bell inequalities is synonymous with the ruling out of models of local hidden variables (LHV), or where correlations between particles can be explained by causal influences traveling at or below the speed of light [1]. Currently, all Bell inequalities descend in one form or another from the same LHV model [101]. Given two space-like separated particles A and B, an LHV model for the statistics of observables $\hat{x}_A$ and $\hat{x}_B$ exists if and only if the joint probability density $\rho(x_A, x_B)$ factors in the following way:

$$\rho(x_A, x_B) = \int d\lambda \; \rho(\lambda)\rho(x_A|\lambda)\rho(x_B|\lambda). \tag{6.1}$$

Here, λ is the variable (or variables) encoding the information existing in the intersection of the past light cones of both A and B (see Fig. 2.1 for diagram).

The principal Bell inequality derived as a consequence of the LHV model (6.1) is the CHSH inequality [42], for random variables with a bounded set of outcomes. As shown in Ch. 2, a general form of the CHSH inequality can be expressed as follows. Let the function $f(x)$ be bounded between -1 and 1. Then, if an LHV model exists describing $\rho(x_A, x_B)$, the inequality,

$$|\langle f(x_A)f(x_B)\rangle_{\alpha,\beta} - \langle f(x_A)f(x_B)\rangle_{\alpha,\beta'}| \pm \big(\langle f(x_A)f(x_B)\rangle_{\alpha',\beta} + \langle f(x_A)f(x_B)\rangle_{\alpha',\beta'}\big) \leq 2, \tag{6.2}$$

must hold true, where α and α' are different measurement settings for x_A, and β and β' are different measurement settings for x_B. Examples of different measurement settings could be orientation about a given axis, time translation, or

[1] For a more comprehensive summary of local hidden variables and Bell nonlocality, see Sec. 2.1-2.2

even arbitrary transformations, such as might be implemented by a spatial light modulator (SLM).

The CHSH inequality was first used to demonstrate Bell nonlocality in the polarization of entangled photon pairs [102–104], though later it was also applied to other observables in two-dimensional subspaces of higher dimensional discrete observables [105]. It has even been used to demonstrate Bell-nonlocality in the (continuous) frequency-time degree of freedom through Franson interference [106–108]. However, demonstrating Bell nonlocality in the *position-momentum* degrees of freedom remains a significant challenge.

In 1986 [99], Bell showed how one might use the CHSH inequality to witness position-momentum Bell nonlocality. In particular, he considered measuring the signs of the outcomes of position measurements, and using the probabilities of getting positive and non-positive measurement outcomes to test the CHSH inequality. Currently, we call this technique sign-binning. If the statistics of functions of continuous variables violate a Bell inequality, then the continuous variables themselves are Bell nonlocal as well. In this Chapter, we define the function $f(x)$, where x is a real number such that $f(x) = 1$ for $x \geq 0$, and $f(x) = -1$ for $x < 0$.

In his discussion of continuous-variable Bell-nonlocality [100], Bell showed that if the Wigner function $W(x_A, k_A, x_B, k_B)$ [2] describing the state of a pair of particles A and B is non-negative, then there is no way that *those measurement statistics* [109] can violate a Bell inequality; the Wigner function in this case is a classical probability distribution over (x_A, x_B, k_A, k_B), and can be used as an LHV model for their measurement outcomes. Indeed where $\rho(x_A, x_B)$ is expressible as an integral over the Wigner function, a non-negative Wigner function can always be

[2]Note: we define the transverse Wigner function such that

$$W(x_1, x_2, k_1, k_2) \equiv \frac{1}{(2\pi)^2} \iint dq_1 dq_2 \; e^{i(q_1 x_1 + q_2 x_2)} \tilde{\psi}(k_1 + \frac{q_1}{2}, k_2 + \frac{q_2}{2}) \tilde{\psi}^*(k_1 - \frac{q_1}{2}, k_2 - \frac{q_2}{2}). \quad (6.3)$$

used as the argument of the integral in the LHV model (6.1), where λ corresponds to the arguments of the Wigner function.

Following this, Bell considered the case of a wavefunction $\psi^{(BV)}(x_A, x_B)$ whose Wigner function was not non-negative:

$$\psi^{(BV)}(x_A, x_B) = \mathcal{N}((x_A - x_B)^2 - 8\sigma_-^2) e^{-\frac{(x_A+x_B)^2}{8\sigma_+^2}} e^{-\frac{(x_A-x_B)^2}{8\sigma_-^2}}. \tag{6.4}$$

This wavefunction is similar to the Double-Gaussian wavefunction (3.17) used to model the transverse spatial amplitude of photon pairs in spontaneous parametric down-conversion (SPDC), but with an additional quadratic component in the position difference. Using this entangled wavefunction (6.4), and time-evolving A and B independently (i.e., to separate times for A and for B) according to the one-dimensional free particle Hamiltonian [3], Bell was able to show that this state (in the limit as σ_+ becomes very large) is indeed Bell-nonlocal. The CHSH inequality that was violated uses the probabilities for positive and non-positive position outcomes, and the measurement settings corresponded to times of measurement of x_A and x_B.

Though Bell's results on position-momentum nonlocality are very significant, experimental demonstrations of this phenomenon have been difficult. Although sufficiently strong correlations across complementary observables imply entanglement (through a demonstration of the Einstein-Podolsky-Rosen (EPR) paradox), they do not (as yet) imply Bell nonlocality [45].

In the original EPR paper [39], Einstein, Podolsky, and Rosen considered a pair of particles (we call A and B) in the EPR state: a simultaneous eigenstate of the position difference $\hat{x}_A - \hat{x}_B$ and the momentum sum $\hat{k}_A + \hat{k}_B$, where

$$\psi^{EPR}(x_A, x_B) \to \mathcal{N}\delta(x_A - x_B) \qquad : \qquad \psi^{EPR}(k_A, k_B) \to \mathcal{N}\delta(k_A + k_B). \tag{6.5}$$

[3] The $1D$ free-particle Hamiltonian for a pair of particles A and B is $\hat{H} = \hbar^2 \frac{\hat{k}_A^2}{2m_A} + \hbar^2 \frac{\hat{k}_B^2}{2m_B}$, where m_A and m_B are the masses of particles A and B, respectively.

The EPR state is non-normalizeable, but can be considered as a limit of the Double-Gaussian state:

$$\psi^{DG}(x_A, x_B) = \frac{1}{\sqrt{2\pi\sigma_+\sigma_-}} e^{-\frac{(x_A+x_B)^2}{8\sigma_+^2}} e^{-\frac{(x_A-x_B)^2}{8\sigma_-^2}}, \tag{6.6}$$

as $\sigma_- \to 0$, and $\sigma_+ \to \infty$. The Wigner function for the Double-Gaussian state is:

$$W^{DG}(x_A, k_A, x_B, k_B) = \frac{1}{\pi^2} e^{-\frac{(x_A+x_B)^2}{4\sigma_+^2}} e^{-\frac{(x_A-x_B)^2}{4\sigma_-^2}} e^{-\frac{(k_A+k_B)^2}{4(\frac{1}{2\sigma_+})^2}} e^{-\frac{(k_A-k_B)^2}{4(\frac{1}{2\sigma_-})^2}}, \tag{6.7}$$

which in that limit becomes:

$$W^{EPR}(x_A, k_A, x_B, k_B) = \mathcal{N} 2\pi\delta(x_A - x_B)\delta(k_A + k_B). \tag{6.8}$$

In spite of these states having arbitrarily strong position and momentum correlations, their Wigner functions are positive definite. Even though these states are necessarily highly entangled [4] in transverse position and momentum, no Bell inequality using statistics of position and momentum can be violated. Using alternate observables, such as parity space observables [110], or pseudo-spin observables [111] one can in a sense witness Bell nonlocality in the transverse position-momentum degree of freedom, but the position and momentum statistics themselves will not show this if the Wigner function is non-negative.

6.2 Adapting Bell's approach to Photon Pairs in SPDC

In order to make Bell's strategy of demonstrating continuous-variable Bell nonlocality experimentally realizable, we use the remarkable fact that the equation governing the propagation of a paraxial beam of light (i.e., the paraxial Helmholtz

[4] The EPR state is very highly entangled since its entropy of entanglement (see Sec. 1.3) is arbitrarily large.

equation) and the Schrödinger equation governing the two-dimensional evolution of a free quantum particle are mathematically identical (up to a change in variables):

$$-\frac{\partial^2 A}{\partial x^2} - \frac{\partial^2 A}{\partial y^2} = ik_p \frac{\partial A}{\partial z} \quad \sim \quad -\frac{\partial^2 \Psi}{\partial x^2} - \frac{\partial^2 \Psi}{\partial y^2} = i\frac{2m}{\hbar}\frac{\partial \Psi}{\partial t}. \tag{6.9}$$

In this way, we see that we may simply test the CHSH inequality (6.2) using the signs of the transverse position outcomes in one dimension. Instead of using different measurement times for our measurement settings $(\alpha, \beta, \alpha', \beta')$, we may simply use different propagation distances.

In particular, let us consider the following experimental setup (see Fig. 6.1). We have a laser incident on a $\chi^{(2)}$ nonlinear crystal (NLC), which produces entangled photon pairs with transverse statistics described by a biphoton amplitude. We then image the face of the nonlinear crystal onto pairs of avalanche photodiodes (APDs) with a 4-f imaging system. By translating the APDs along the optic axis, we use coincidence counts to measure the transverse sign correlations as a function of a variable propagation distance in each arm.

Though it is not difficult to imagine an optical experiment that might witness position-momentum Bell nonloclaity if the transverse spatial amplitude has the form of Bell's wavefunction (6.4), the actual transverse state is reasonably well approximated by a Double-Gaussian wavefunction (6.6). In spite of being possibly highly entangled, if the photon pairs are described by (6.6), then they cannot demonstrate position-momentum Bell nonlocality.

However, more precise calculations of the transverse spatial amplitude show that the transverse position amplitude of the down-converted photon pair (3.15) is a good deal different than a double-Gaussian, only really bearing a resemblance near the central peak (see Fig. 3.2(a)). Indeed, as shown in the Fig. 6.2, the Wigner function associated to the biphoton amplitude has many regions dipping below zero, with significant dips in locations similar to where we see them for Bell's wavefunction. In that figure, we plotted the portions of the Wigner function

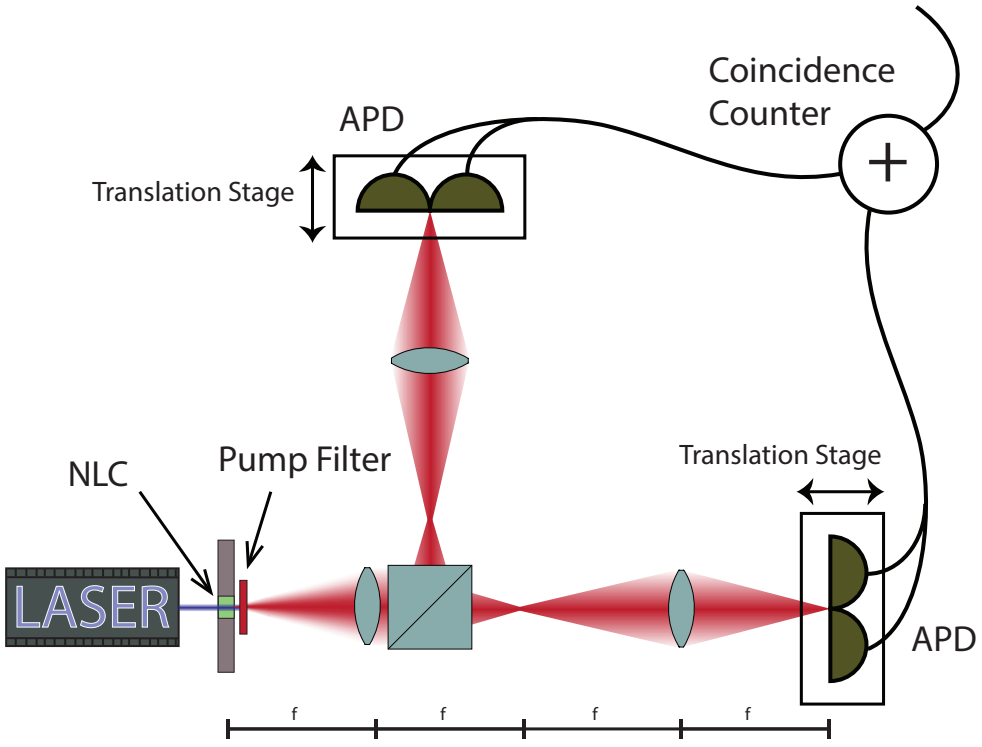

Figure 6.1: Here is an idealized diagram of an experimental setup to violate the CHSH inequality for position and momentum with sign-binning.

depending on the difference coordinates x_- and k_- as defined in (3.13). Since the biphoton wavefunction $\psi(x_A, x_B)$ expressed in terms of the rotated coordinates x_+ and x_- factors as:

$$\psi(x_A, x_B) = \psi_+(x_+)\psi_-(x_-), \tag{6.10}$$

the Wigner function factors in a similar way, and we can plot the Wigner functions' dependence on x_- and k_- in a straightforward manner.

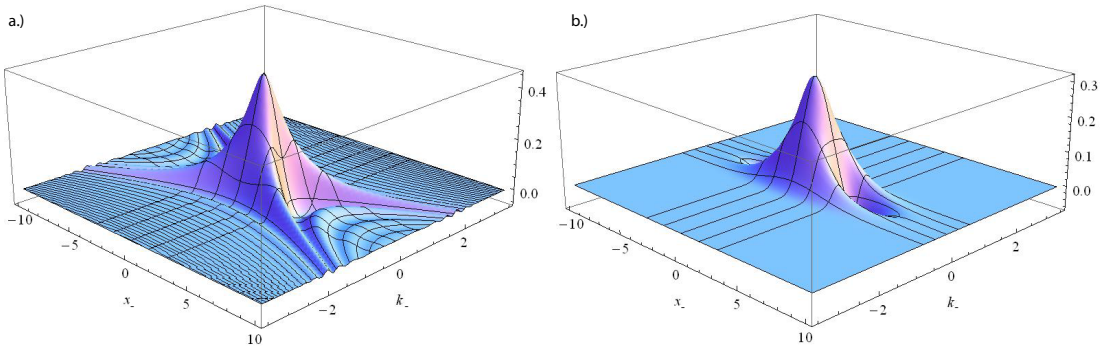

Figure 6.2: On the left (a) is a plot of the Wigner function $W_-^{(SPDC)}(x_-, k_-)$ obtained from direct calculations of the biphoton amplitude in SPDC, with axes scaled to see the central peak. We note that although the biphoton Wigner function can be approximated as a Gaussian, there are significant regions of negativity. On the right (b) is a plot of Bell's Wigner function $W_-^{(BV)}(x_-, k_-)$ for σ_- chosen to match position moments of the SPDC biphoton state. The values below the level contours are negative.

6.3 Testing the Bell Nonlocality of Bell's Transverse Spatial Amplitude

If the transverse position amplitude of the signal/idler photon pair exiting the nonlinear crystal was described by Bell's wavefunction $\psi^{(BV)}(x_A, x_B)$, then we can show its Bell nonlocality without having to take the limit that σ_+ is arbitrarily large. We started by using parameters $\sigma_+ = 1$ mm, and $\sigma_- = 10\ \mu$m, for the initial biphoton wavefunction at planes $z_A = z_B = 0$. We propagated this biphoton field to different optical planes using paraxial free space propagation assuming the wavelength of the downconverted light was 650 nm. In particular, if the joint wavefunction in momentum space at planes z_A and z_B is $\psi(k_A, k_B; z_A, z_B)$, we have that

$$\psi(k_A, k_B; z_A, z_B) = \psi(k_A, k_B; 0, 0) e^{-i\frac{z_A}{2k_p}k_A^2} e^{-i\frac{z_B}{2k_p}k_B^2}. \tag{6.11}$$

Note that here, k_p (the pump momentum) is 2π divided by the wavelength of the downconverted light. In addition, we have neglected constant phase factors, which

do not affect the transverse probability density. In order to more easily perform the calculations needed to get the propagated transverse position probability density, we found it easier to work with the propagated Wigner function. Indeed, paraxial free space propagation takes the simple form of a shear transformation of the Wigner function [112]:

$$W(x_A, k_A, x_B, k_B; z_A, z_B) = W(x_A - \frac{2k_A z_A}{k_p}, k_A, x_B - \frac{2k_B z_B}{k_p}, k_B; 0, 0). \quad (6.12)$$

By propagating the Wigner function in this way, and integrating over k_A and k_B to get the propagated transverse position probability density $\rho(x_A, x_B; z_A, z_B)$, we were able to test the CHSH inequality using sign-binning, where the measurement settings $(\alpha, \beta, \alpha', \beta') = (z_A, z_B, z'_A, z'_B)$ are different propagation distances. In order to get the actual sign correlations used in the CHSH inequality (6.2), we numerically integrated $\rho(x_A, x_B; z_A, z_B)$ over each of the four position quadrants. In doing so, we found a small, but significant violation: the CHSH sum was 2.00046 with the propagation distances $(z_A, z_B, z'_A, z'_B) = (-3, -3, 8, 8)$ in millimeters (see Fig. 6.3). The numerical integral converged with eight digits of precision, a small tolerance for error compared to the violation of 4.6×10^{-4}. Small violations in this sign-binning strategy are also noted in studies on the Bell nonlocality of the field quadratures of entangled pairs of photons as measured with homodyne detection [113–118].

6.4 Experimental Challenges to Position-Momentum Bell Nonlocality

6.4.1 Experimental Restrictions on Bell's Wavefunction

Though it is theoretically possible to demonstrate position-momentum Bell nonlocality if the transverse biphoton amplitude resembles Bell's wavefunction $\psi^{BV}(x_A, x_B)$,

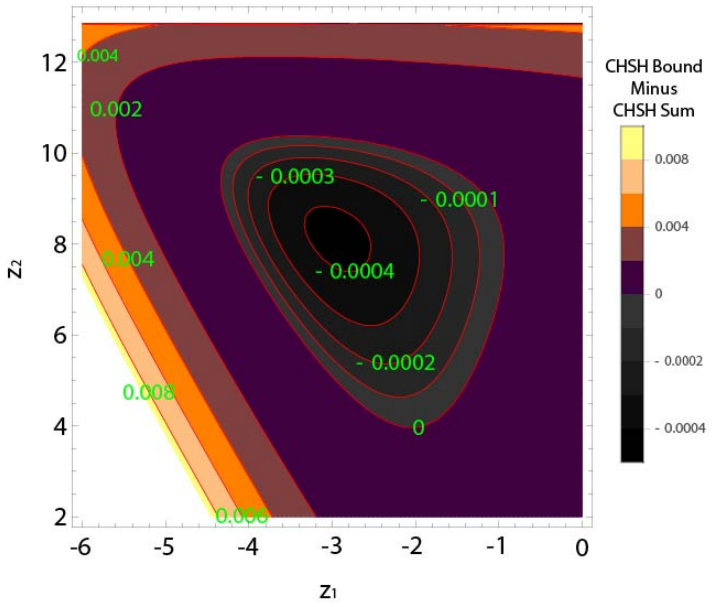

Figure 6.3: Here is a contour plot of the violation of the CHSH inequality of $\psi^{BV}(x_A, x_B; z_A, z_B)$. The violation is plotted against two propagation distances, which we call z_1 and z_2. The set of distances used in the CHSH inequality is then $(z_A, z_B, z'_A, z'_B) = (z_1, z_1, z_2, z_2)$. A negative value indicates a violation of the CHSH inequality.

experimentally realizing this result poses significant difficulties. First, we must consider the number of coincidence counts needed to resolve a violation of only four parts in ten thousand. Since the pump laser is a coherent state of light, the number statistics of the pump photons are Poisson distributed. Since the rate of down-conversion events is proportional to pump photon flux through the non-linear crystal, the coincidence counts are to a good approximation also Poisson distributed. If the mean number of coincidence counts is 10^4 per second, the standard deviation in the coincidence counts will be 10^2 per second, an uncertainty of one part in 10^2. If we wish to be able to resolve differences of intensity of one part in 10^4, we will need at least 10^8 coincidence counts for each probability we wish to measure (16 in total for the four pairs of propagation distances in the CHSH inequality). We can expect a well-aligned SPDC experimental setup to give us a detectable flux of approximately 10^4 coincidence counts per second, requiring on

the order of 27-36 hours for each joint probability.

By themselves, long integration times are not an unreasonable complication here. However, the difficulty in aligning such an experimental setup so as to nearly perfectly center the downconverted light on the pairs of photodetectors in each arm is not a trivial feat. Following our discovery of a Bell violation for $\psi^{BV}(x_A, x_B)$, we considered what the effect of displacing one of the photodetectors by a constant distance would be on our ability to violate the CHSH inequality. As seen in Fig. 6.4, the CHSH inequality is only violated if the displacement is less than 5 microns, and violated by a value close to the maximum for a displacement less than 2 microns. With displacements as small as these, it might seem that beam pointing stability (i.e., the random dynamical fluctuations in the mean direction of the pump beam) becomes a significant issue. Indeed, the 325nm HeCd laser used

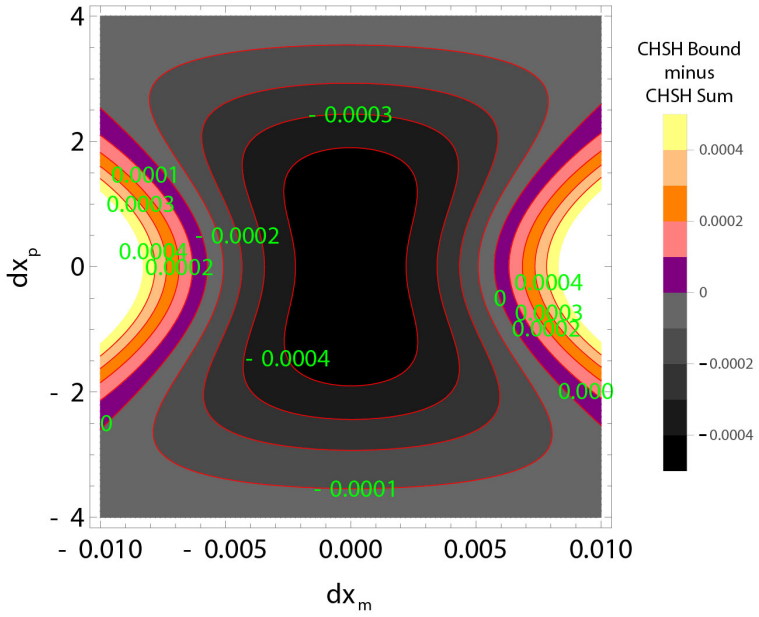

Figure 6.4: Here is a contour plot of the violation of the CHSH inequality of $\psi^{BV}(x_A, x_B; z_A, z_B)$. The violation is plotted against two different transverse shifts, dx_p and dx_m (in millimeters). A negative value indicates a violation of the CHSH inequality.

in previous downconversion setups mentioned in earlier chapters has a pointing

instability [5] of 2.5×10^{-5} radians. If the distance between the pump laser and the nonlinear crystal was 10cm, then that corresponds to a transverse position instability of 2.5μm, on par with our ability to even resolve a violation of the CHSH inequality. However, since the pointing instability of the pump imparts an identical shift on both the signal and idler beam profiles, the effect of pointing stability does not substantially affect our ability to violate the CHSH inequality.

As an example, I have plotted in Fig. 6.4 the violation of the CHSH inequality as a function of a pair of transverse shifts dx_p and dx_m in the signal and idler detector arrays. A shift of dx_p is one where the signal and idler beams are both shifted by the same amount, while a shift of dx_m is one where the signal and idler beams are shifted by opposite amounts. Indeed if dx_A is the misalignment in the signal beam, and dx_B is the misalignment in the idler beam, then $dx_p = dx_A + dx_B$ and $dx_m = dx_A - dx_B$. Though a difference misalignment dx_m of the order a few microns will compromise our ability to violate the CHSH inequality, a sum misalignment of the same amount will not. Instead, the sum misalignment must be over three orders of magnitude larger (a few millimeters) before the CHSH inequality becomes impossible to violate.

6.4.2 Testing the Bell Nonlocality of Real SPDC Biphoton Sources

Though the state of downconverted light exiting a nonlinear crystal may yet be Bell-nonlocal (see Fig. 6.2.a), obtaining theoretical results for that down-converted state remains an open problem. In order to test the position-momentum Bell nonlocality of actual down-converted photon pairs, we need to consider the possibility of tailoring the downconverted biphoton state to resemble $\psi^{BV}(x_A, x_B)$.

[5] The pointing instability of a laser is the standard deviation of the angular position of the centroid of the beam acquired over multiple time intervals.

In [119], it is noted that there is a Fourier-transform relationship between the spatial dependence of the nonlinear susceptibility $\chi^{(2)}$, and the phase matching function (i.e., the component of the biphoton amplitude that depends on the momentum mismatch Δk). Indeed, the rectangular shape of the nonlinear crystal is the chief reason why the biphoton amplitude (3.9) depends on Sinc functions of each component of the momentum mismatch $\Delta\vec{k}$. When the second-order nonlinear susceptibility depends on z, we find:

$$\Phi(\vec{k}_1, \vec{k}_2) = \mathcal{N}\left[\frac{1}{\sqrt{2\pi}} \int dz \bar{\chi^{(2)}}_{\text{eff}}(z) e^{-i\Delta k_z z}\right] \nu(k_{x1} + k_{x2}, k_{y1} + k_{y2}) \tag{6.13}$$

where $\bar{\chi}^{(2)}_{\text{eff}}(z)$ is the scaled effective nonlinear susceptibility as a function of distance along the optic axis z, and $\mathcal{N}$ is a normalization constant. Although one cannot have a continuously varying susceptibility as a function of z (indeed, it must take one of two values (1 or -1) at any given location within the crystal), Dixon *et al.* [119] showed that it was possible to engineer a biphoton wavefunction that more closely resembled the ideal double-Gaussian state by custom-engineering a periodically poled nonlinear crystal to have a constant poling period, but a spatially-varying duty cycle (see Fig. 6.5 for diagram).

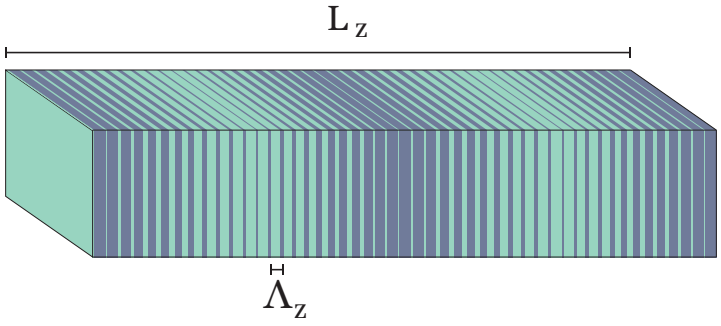

Figure 6.5: Here is a depiction of a periodically poled nonlinear crystal of length L_z and poling period Λ_z. The positive poling is colored a light green, while the negative poling is gray. The duty cycle (i.e, the fraction of positive to negative poling) is smoothly varied from 2/16 to 14/16 and back again twice.

In a periodically-poled nonlinear crystal, the scaled nonlinear susceptibility

$\bar{\chi}^{(2)}_{\text{eff}}(z)$ switches from -1 to 1, and back again over a distance of one poling period Λ_z. Within each poling period, the duty cycle $r(z)$ is defined as a function between 0 and and 1, and is defined within each poling period as the fraction of the poling period where $\bar{\chi}^{(2)}_{\text{eff}}(z) = 1$, while for the rest of the period, it is -1. When the poling period Λ_z is small relative to the crystal length L_z, the scaled nonlinear susceptibility $\bar{\chi}^{(2)}_{\text{eff}}(z)$ is a rapidly varying function whose mean value over a small number of periods is smoothly varying. When the grating period is sufficiently small compared to the crystal length, we may (to a certain approximation) engineer a biphoton wavefunction as though we could smoothly vary $\bar{\chi}^{(2)}_{\text{eff}}(z)$, by varying $r(z)$ instead. It is expected that Bell's wavefunction could be engineered in a similar fashion, but the experimental tolerances for violating the CHSH inequality are small enough that the fidelity of the Bell wavefunction generated would likely need to be large, and poses a significant challenge.

7 Concluding Remarks

The foundations of quantum physics are remarkably accessible in quantum optics. In the past few decades, spontaneous parametric down-conversion (SPDC) has become a very popular tool in quantum optics, since one may generate highly entangled states with relative ease. Using the entangled pairs of photons generated via SPDC, I have made investigations into entanglement, EPR-steering and Bell nonlocality from both a theoretical and experimental perspective.

In this final section, I shall illustrate what avenues remain for future research, based on the work my colleagues and I have accomplished here. In Chapter 3, where I summarize the theoretical foundations of the spatial correlations of photon pairs in SPDC, I developed a pair of compact formulas for the transverse correlation width (3.16)(3.18). I did this because there did not seem to be a compact reference on the subject, and I wanted to make intuitions more concrete. There is much to be researched in SPDC, particularly in quantum engineering custom biphoton wavefunctions for specific applications in quantum information, and in the discovery of more efficient methods of accomplishing SPDC, either with materials with higher nonlinear susceptibilities, or as yet unforeseen methods.

EPR-steering is a field only eight years old by some indications, and there is a substantial amount to be discovered both theoretically and experimentally. Theo-

retical investigations into EPR-steering have primarily been focused between two parties; only very recently has third-party and multi-partite EPR-steering been investigated. While two-party EPR-steering is an intermediate level of nonlocal correlation between non-separability and Bell nonlocality, there are many other possible models to be considered where three or more parties are involved. In addition, the development of a Fano inequality for continuous variables is an innovation only months old. Far from witnessing entanglement, the applications of a continuous-variable Fano inequality are largely unexplored.

The recent idea of partial projective measurements presents new avenues of research as well. While we explored their effect on Fourier-conjugate pairs of observables, it remains to be seen if a similar advantage is obtained using other pairs of mutually unbiased observables. In addition, the efficient acquisition of complementary statistics (as this method affords us) paves the way to an efficient method of estimating the quantum state itself.

Finally, there is much to be done in obtaining a violation of the CHSH inequality using the position-momentum statistics of entangled photon pairs. First and foremost, the calculation of the relevant statistics for the actual biphoton state (3.15) (propagated to signal and idler planes) as opposed to the more convenient Bell's wavefunction (6.4) remains to be accomplished. Though the biphoton state bears a sufficient gross resemblance to Bell's wavefunction that we might expect a violation of the CHSH inequality under similar circumstances, the current difficulties of an experimental demonstration are enough that we should make sure of where and how we may best make our measurements. As a side point, only briefly touched upon, was that Franson interference allows us to demonstrate continuous-variable Bell-nonlocality in the time/frequency degree of freedom. It would be worth investigating if Franson interference could also demonstrate Bell nonlocality in the transverse position degree of freedom.

Bibliography

[1] C. Shannon, "A Mathematical Theory of Communication," Bell System Technical Journal **27**, 379 (1948).

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley and Sons, New York, 2006), 2nd ed.

[3] S. Kullback and R. A. Leibler, "On Information and Sufficiency," Ann. Math. Statist. **22**, 79 (1951), URL http://dx.doi.org/10.1214/aoms/1177729694.

[4] I. Białynicki-Birula, "Entropic Uncertainty Relations," Physics Letters A **103**, 253 (1984), ISSN 0375-9601, URL http://www.sciencedirect.com/science/article/pii/037596018490118X.

[5] I. Białynicki-Birula and J. Madajczyk, "Entropic Uncertainty Relations for Angular Distributions," Physics Letters A **108**, 384 (1985), ISSN 0375-9601, URL http://www.sciencedirect.com/science/article/pii/0375960185902774.

[6] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2000).

[7] U. Fano, "Pairs of Two-Level Systems," Rev. Mod. Phys. **55**, 855 (1983), URL http://link.aps.org/doi/10.1103/RevModPhys.55.855.

[8] A. Rnyi, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics* (University of California Press, Berkeley, Calif., 1961), pp. 547–561, URL http://projecteuclid.org/euclid.bsmsp/1200512181.

[9] R. F. Werner, "Quantum States with Einstein-Podolsky-Rosen Correlations Admitting a Hidden-Variable Model," Phys. Rev. A **40**, 4277 (1989), URL http://link.aps.org/doi/10.1103/PhysRevA.40.4277.

[10] A. Peres, "Separability Criterion for Density Matrices," Phys. Rev. Lett. **77**, 1413 (1996), URL http://link.aps.org/doi/10.1103/PhysRevLett.77.1413.

[11] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, "Quantum Entanglement," Rev. Mod. Phys. **81**, 865 (2009), URL http://link.aps.org/doi/10.1103/RevModPhys.81.865.

[12] B. M. Terhal, "Bell Inequalities and the Separability Criterion," Physics Letters A **271**, 319 (2000), ISSN 0375-9601, URL http://www.sciencedirect.com/science/article/pii/S0375960100004011.

[13] Y. Huang, "Computing Quantum Discord is NP-Complete," New Journal of Physics **16**, 033027 (2014).

[14] V. Vedral, *Introduction to Quantum Information Science* (Oxford University Press, 2006).

[15] S. P. Walborn, B. G. Taketani, A. Salles, F. Toscano, and R. L. de Matos Filho, "Entropic Entanglement Criteria for Continuous Variables," Phys. Rev. Lett. **103**, 160505 (2009), URL http://link.aps.org/doi/10.1103/PhysRevLett.103.160505.

[16] L.-M. Duan, G. Giedke, J. I. Cirac, and P. Zoller, "Inseparability Criterion for Continuous Variable Systems," Phys. Rev. Lett. **84**, 2722 (2000), URL http://link.aps.org/doi/10.1103/PhysRevLett.84.2722.

[17] S. Mancini, V. Giovannetti, D. Vitali, and P. Tombesi, "Entangling Macroscopic Oscillators Exploiting Radiation Pressure," Phys. Rev. Lett. **88**, 120401 (2002), URL http://link.aps.org/doi/10.1103/PhysRevLett.88.120401.

[18] S. Hill and W. K. Wootters, "Entanglement of a Pair of Quantum Bits," Phys. Rev. Lett. **78**, 5022 (1997), URL http://link.aps.org/doi/10.1103/PhysRevLett.78.5022.

[19] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, "Mixed-State Entanglement and Quantum Error Correction," Phys. Rev. A **54**, 3824 (1996), URL http://link.aps.org/doi/10.1103/PhysRevA.54.3824.

[20] E. M. Rains, "Rigorous Treatment of Distillable Entanglement," Phys. Rev. A **60**, 173 (1999), URL http://link.aps.org/doi/10.1103/PhysRevA.60.173.

[21] V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight, "Quantifying Entanglement," Phys. Rev. Lett. **78**, 2275 (1997), URL http://link.aps.org/doi/10.1103/PhysRevLett.78.2275.

[22] R. Grobe, K. Rzazewski, and J. H. Eberly, "Measure of Electron-Electron Correlation in Atomic Physics," Journal of Physics B: Atomic, Molecular and Optical Physics **27**, L503 (1994), URL http://stacks.iop.org/0953-4075/27/i=16/a=001.

[23] C. K. Law and J. H. Eberly, "Analysis and Interpretation of High Transverse Entanglement in Optical Parametric Down Conversion," Phys.

Rev. Lett. **92**, 127903 (2004), URL http://link.aps.org/doi/10.1103/PhysRevLett.92.127903.

[24] B. M. Terhal and P. Horodecki, "Schmidt Number for Density Matrices," Phys. Rev. A **61**, 040301 (2000), URL http://link.aps.org/doi/10.1103/PhysRevA.61.040301.

[25] I. Białynicki-Birula and J. Mycielski, "Uncertainty Relations for Information Entropy in Wave Mechanics," Communications in Mathematical Physics **44**, 129 (1975), ISSN 0010-3616, URL http://dx.doi.org/10.1007/BF01608825.

[26] D. Deutsch, "Uncertainty in Quantum Measurements," Phys. Rev. Lett. **50**, 631 (1983), URL http://link.aps.org/doi/10.1103/PhysRevLett.50.631.

[27] H. Maassen and J. B. M. Uffink, "Generalized Entropic Uncertainty Relations," Phys. Rev. Lett. **60**, 1103 (1988), URL http://link.aps.org/doi/10.1103/PhysRevLett.60.1103.

[28] M. Berta, M. Christandl, R. Colbeck, J. M. Renes, and R. Renner, "The uncertainty Principle in the Presence of Quantum Memory," Nature Physics (2010).

[29] W. Heisenberg, "Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik," Zeitschrift für Physik **43**, 172 (1927), ISSN 0044-3328, URL http://dx.doi.org/10.1007/BF01397280.

[30] H. P. Robertson, "The Uncertainty Principle," Phys. Rev. **34**, 163 (1929), URL http://link.aps.org/doi/10.1103/PhysRev.34.163.

[31] C. M. Caves and P. D. Drummond, "Quantum Limits on Bosonic Communication Rates," Rev. Mod. Phys. **66**, 481 (1994), URL http://link.aps.org/doi/10.1103/RevModPhys.66.481.

[32] M. Christandl and A. Winter, "Uncertainty, Monogamy, and Locking of Quantum Correlations," Information Theory, IEEE Transactions on **51**, 3159 (2005), ISSN 0018-9448.

[33] J. Leach, E. Bolduc, D. J. Gauthier, and R. W. Boyd, "Secure Information Capacity of Photons Entangled in Many Dimensions," Phys. Rev. A **85**, 060304 (2012), URL http://link.aps.org/doi/10.1103/PhysRevA.85.060304.

[34] J. Hirschman, I. I., "A Note on Entropy," American Journal of Mathematics **79**, pp. 152 (1957), ISSN 00029327, URL http://www.jstor.org/stable/2372390.

[35] W. Beckner, "Inequalities in Fourier Analysis," Annals of Mathematics **102**, pp. 159 (1975), ISSN 0003486X, URL http://www.jstor.org/stable/1970980.

[36] A. K. Pati, M. M. Wilde, A. R. U. Devi, A. K. Rajagopal, and Sudha, "Quantum Discord and Classical Correlation can Tighten the Uncertainty Principle in the Presence of Quantum Memory," Phys. Rev. A **86**, 042105 (2012), URL http://link.aps.org/doi/10.1103/PhysRevA.86.042105.

[37] M. J. W. Hall, "Information Exclusion Principle for Complementary Observables," Phys. Rev. Lett. **74**, 3307 (1995), URL http://link.aps.org/doi/10.1103/PhysRevLett.74.3307.

[38] P. J. Coles and M. Piani, "Improved Entropic Uncertainty Relations and Information Exclusion Relations," arXiv preprint arXiv:1307.4265 (2013), URL http://arxiv.org/abs/1307.4265.

[39] A. Einstein, B. Podolsky, and N. Rosen, "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?," Phys. Rev. **47**, 777 (1935), URL http://link.aps.org/doi/10.1103/PhysRev.47.777.

[40] A. Einstein, *Albert Einstein: Philosopher-Scientist*, vol. 7 (Library of Living Philosophers, 1949).

[41] J. S. Bell et al., "On the Einstein-Podolsky-Rosen Paradox," Physics **1**, 195 (1964).

[42] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed Experiment to Test Local Hidden-Variable Theories," Phys. Rev. Lett. **23**, 880 (1969).

[43] S. Das and G. Agarwal, "Nonclassical Correlation of Polarization-Entangled Photons in a Biexciton-Exciton Cascade," Journal of Physics B: Atomic, Molecular and Optical Physics **41**, 1 (2008).

[44] E. Schrödinger, "Discussion of Probability Relations between Separated Systems," Mathematical Proceedings of the Cambridge Philosophical Society **31**, 555 (1935), URL http://dx.doi.org/10.1017/S0305004100013554.

[45] H. M. Wiseman, S. J. Jones, and A. C. Doherty, "Steering, Entanglement, Nonlocality, and the Einstein-Podolsky-Rosen Paradox," Phys. Rev. Lett. **98**, 140402 (2007), URL http://link.aps.org/doi/10.1103/PhysRevLett.98.140402.

[46] A. Peres and D. R. Terno, "Quantum Information and Relativity Theory," Rev. Mod. Phys. **76**, 93 (2004), URL http://link.aps.org/doi/10.1103/RevModPhys.76.93.

[47] R. F. Werner, "Quantum States with Einstein-Podolsky-Rosen Correlations Admitting a Hidden-Variable Model," Phys. Rev. A **40**, 4277 (1989), URL http://link.aps.org/doi/10.1103/PhysRevA.40.4277.

[48] J. Bowles, T. Vértesi, M. T. Quintino, and N. Brunner, "One-Way Einstein-Podolsky-Rosen Steering," Phys. Rev. Lett. **112**, 200402 (2014), URL http://link.aps.org/doi/10.1103/PhysRevLett.112.200402.

[49] S. Gharibian, "Strong NP-Hardness of the Quantum Separability Problem," Quantum Information and Computation **10**, 343 (2010), URL http://arxiv.org/abs/0810.4507.

[50] M. Horodecki, P. Horodecki, and R. Horodecki, "Separability of Mixed States: Necessary and Sufficient Conditions," Physics Letters A **223**, 1 (1996), ISSN 0375-9601, URL http://www.sciencedirect.com/science/article/pii/S0375960196007062.

[51] E. G. Cavalcanti, S. J. Jones, H. M. Wiseman, and M. D. Reid, "Experimental Criteria for Steering and the Einstein-Podolsky-Rosen Paradox," Phys. Rev. A **80**, 032112 (2009), URL http://link.aps.org/doi/10.1103/PhysRevA.80.032112.

[52] M. D. Reid, "Demonstration of the Einstein-Podolsky-Rosen Paradox Using Nondegenerate Parametric Amplification," Phys. Rev. A **40**, 913 (1989), URL http://link.aps.org/doi/10.1103/PhysRevA.40.913.

[53] M. D. Reid, P. D. Drummond, W. P. Bowen, E. G. Cavalcanti, P. K. Lam, H. A. Bachor, U. L. Andersen, and G. Leuchs, "*Colloquium*: The Einstein-Podolsky-Rosen Paradox: From Concepts to Applications," Rev. Mod. Phys. **81**, 1727 (2009), URL http://link.aps.org/doi/10.1103/RevModPhys.81.1727.

[54] J. C. Howell, R. S. Bennink, S. J. Bentley, and R. W. Boyd, "Realization of the Einstein-Podolsky-Rosen Paradox Using Momentum- and Position-Entangled Photons from Spontaneous Parametric Down Conversion," Phys. Rev. Lett. **92**, 210403 (2004), URL http://link.aps.org/doi/10.1103/PhysRevLett.92.210403.

[55] S. P. Walborn, A. Salles, R. M. Gomes, F. Toscano, and P. H. Souto Ribeiro, "Revealing Hidden Einstein-Podolsky-Rosen Nonlocality," Phys. Rev. Lett. **106**, 130402 (2011).

[56] J. Schneeloch, C. J. Broadbent, S. P. Walborn, E. G. Cavalcanti, and J. C. Howell, "Einstein-Podolsky-Rosen Steering Inequalities from Entropic Uncertainty Relations," Phys. Rev. A **87**, 062103 (2013), URL http://link.aps.org/doi/10.1103/PhysRevA.87.062103.

[57] J. Schneeloch and J. C. Howell, "Introduction to the Transverse Spatial Correlations in Spontaneous Parametric Down-Conversion through the Biphoton Birth Zone," arXiv preprint arXiv:1502.06696v5 (2015), URL http://arxiv.org/abs/1502.06996.

[58] J. D. Jackson and J. D. Jackson, *Classical Electrodynamics*, vol. 3 (Wiley New York etc., 1962).

[59] R. W. Boyd, *Nonlinear Optics* (Academic press, 2007), 3rd ed.

[60] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge university press, 1995).

[61] R. Loudon, *The Quantum Theory of Light* (Oxford university press, 2000).

[62] C. K. Hong and L. Mandel, "Theory of Parametric Frequency Down Conversion of Light," Phys. Rev. A **31**, 2409 (1985), URL http://link.aps.org/doi/10.1103/PhysRevA.31.2409.

[63] C. H. Monken, P. H. S. Ribeiro, and S. Pádua, "Transfer of Angular Spectrum and Image Formation in Spontaneous Parametric Down-Conversion," Phys. Rev. A **57**, 3123 (1998), URL http://link.aps.org/doi/10.1103/PhysRevA.57.3123.

[64] M. V. Fedorov, Y. M. Mikhailova, and P. A. Volkov, "Gaussian Modelling and Schmidt Modes of SPDC Biphoton States," Journal of Physics B: Atomic, Molecular and Optical Physics **42**, 175503 (2009), URL http://stacks.iop.org/0953-4075/42/i=17/a=175503.

[65] P. B. Dixon, G. A. Howland, J. Schneeloch, and J. C. Howell, "Quantum Mutual Information Capacity for High-Dimensional Entangled States," Phys. Rev. Lett. **108**, 143603 (2012), URL http://link.aps.org/doi/10.1103/PhysRevLett.108.143603.

[66] M. Edgar, D. Tasca, F. Izdebski, R. Warburton, J. Leach, M. Agnew, G. Buller, R. Boyd, and M. Padgett, "Imaging High-Dimensional Spatial Entanglement with a Camera," Nat. Commun. **3**, 1 (2012), URL http://dx.doi.org/10.1038/ncomms1988.

[67] R. S. Bennink, S. J. Bentley, R. W. Boyd, and J. C. Howell, "Quantum and Classical Coincidence Imaging," Phys. Rev. Lett. **92**, 033601 (2004), URL http://link.aps.org/doi/10.1103/PhysRevLett.92.033601.

[68] I. Ali Khan and J. C. Howell, "Experimental Demonstration of High Two-Photon Time-Energy Entanglement," Phys. Rev. A **73**, 031801 (2006), URL http://link.aps.org/doi/10.1103/PhysRevA.73.031801.

[69] K. W. Chan, J. P. Torres, and J. H. Eberly, "Transverse Entanglement Migration in Hilbert space," Phys. Rev. A **75**, 050101 (2007), URL http://link.aps.org/doi/10.1103/PhysRevA.75.050101.

[70] P. B. Dixon, G. Howland, M. Malik, D. J. Starling, R. W. Boyd, and J. C. Howell, "Heralded Single-Photon Partial Coherence," Phys. Rev. A **82**, 023801 (2010), URL http://link.aps.org/doi/10.1103/PhysRevA.82.023801.

[71] S. P. Walborn and A. H. Pimentel, "Generalized HermiteGauss Decomposition of the Two-Photon State Produced by Spontaneous Parametric Down Conversion," Journal of Physics B: Atomic, Molecular and Optical Physics **45**, 165502 (2012), URL http://stacks.iop.org/0953-4075/45/i=16/a=165502.

[72] S. Walborn, C. Monken, S. Padua, and P. Souto Ribeiro, "Spatial Correlations in Parametric Downconversion," Physics Reports **495**, 87 (2010).

[73] D. S. Tasca, L. Rudnicki, R. M. Gomes, F. Toscano, and S. P. Walborn, "Reliable Entanglement Detection under Coarse-Grained Measurements," Phys. Rev. Lett. **110**, 210502 (2013), URL http://link.aps.org/doi/10.1103/PhysRevLett.110.210502.

[74] G. A. Howland and J. C. Howell, "Efficient High-Dimensional Entanglement Imaging with a Compressive-Sensing Double-Pixel Camera," Phys. Rev. X **3**, 011013 (2013), URL http://link.aps.org/doi/10.1103/PhysRevX.3.011013.

[75] P.-A. Moreau, F. Devaux, and E. Lantz, "Einstein-Podolsky-Rosen Paradox in Twin Images," Phys. Rev. Lett. **113**, 160401 (2014), URL http://link.aps.org/doi/10.1103/PhysRevLett.113.160401.

[76] N. Gisin, "Bell's Inequality Holds for All Non-Product States," Physics Letters A **154**, 201 (1991), URL http://www.sciencedirect.com/science/article/pii/037596019190805I.

[77] N. Gisin and A. Peres, "Maximal Violation of Bell's Inequality for Arbitrarily Large Spin," Physics Letters A **162**, 15 (1992), ISSN 0375-9601, URL http://www.sciencedirect.com/science/article/pii/037596019290949M.

[78] S. Yu, Q. Chen, C. Zhang, C. H. Lai, and C. H. Oh, "All Entangled Pure States Violate a Single Bell's Inequality," Phys. Rev. Lett. **109**, 120402 (2012), URL http://link.aps.org/doi/10.1103/PhysRevLett.109.120402.

[79] T. Yarnall, A. F. Abouraddy, B. E. A. Saleh, and M. C. Teich, "Experimental Violation of Bell's Inequality in Spatial-Parity Space," Phys. Rev. Lett. **99**, 170408 (2007), URL http://link.aps.org/doi/10.1103/PhysRevLett.99.170408.

[80] L. Praxmeyer, B.-G. Englert, and K. Wdkiewicz, "Violation of Bells Inequality for Continuous-Variable EPR States," The European Physical Journal D - Atomic, Molecular, Optical and Plasma Physics **32**, 227 (2005), ISSN 1434-6060, URL http://dx.doi.org/10.1140/epjd/e2005-00021-1.

[81] J. Schneeloch, P. B. Dixon, G. A. Howland, C. J. Broadbent, and J. C. Howell, "Violation of Continuous-Variable Einstein-Podolsky-Rosen Steering with Discrete Measurements," Phys. Rev. Lett. **110**, 130407 (2013), URL http://link.aps.org/doi/10.1103/PhysRevLett.110.130407.

[82] J. Schneeloch, "The Relationship Between Discrete and Continuous Entropy in EPR-Steering Inequalities," arXiv preprint arXiv:1312.2604 (2013), URL http://arxiv.org/abs/1312.2604.

[83] J. Schneeloch, S. H. Knarr, G. A. Howland, and J. C. Howell, "Demonstrating Continuous-Variable Einstein–Podolsky–Rosen Steering in spite of

Finite Experimental Capabilities using Fano Steering Bounds," JOSA B **32**, A8 (2015).

[84] P. Chowdhury, T. Pramanik, A. S. Majumdar, and G. S. Agarwal, "Einstein-Podolsky-Rosen Steering using Quantum Correlations in Non-Gaussian Entangled States," Phys. Rev. A **89**, 012104 (2014), URL http://link.aps.org/doi/10.1103/PhysRevA.89.012104.

[85] C.-W. Lee, S.-W. Ji, and H. Nha, "Quantum Steering for Continuous-Variable States," J. Opt. Soc. Am. B **30**, 2483 (2013), URL http://josab.osa.org/abstract.cfm?URI=josab-30-9-2483.

[86] D. A. Evans, E. G. Cavalcanti, and H. M. Wiseman, "Loss-Tolerant Tests of Einstein-Podolsky-Rosen Steering," Phys. Rev. A **88**, 022106 (2013), URL http://link.aps.org/doi/10.1103/PhysRevA.88.022106.

[87] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lütkenhaus, and M. Peev, "The Security of Practical Quantum Key Distribution," Rev. Mod. Phys. **81**, 1301 (2009), URL http://link.aps.org/doi/10.1103/RevModPhys.81.1301.

[88] V. Coffman, J. Kundu, and W. K. Wootters, "Distributed Entanglement," Phys. Rev. A **61**, 052306 (2000), URL http://link.aps.org/doi/10.1103/PhysRevA.61.052306.

[89] M. Reid, "Monogamy Inequalities for the EPR Paradox and Quantum Steering," arXiv preprint arXiv:1310.2729 (2013), URL http://arxiv.org/abs/1310.2729.

[90] J. M. Renes and J.-C. Boileau, "Conjectured Strong Complementary Information Tradeoff," Phys. Rev. Lett. **103**, 020402 (2009), URL http://link.aps.org/doi/10.1103/PhysRevLett.103.020402.

[91] I. Csiszár and J. Korner, "Broadcast Channels with Confidential Messages," Information Theory, IEEE Transactions on **24**, 339 (1978), URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1055892.

[92] G. A. Howland, J. Schneeloch, D. J. Lum, and J. C. Howell, "Simultaneous Measurement of Complementary Observables with Compressive Sensing," Phys. Rev. Lett. **112**, 253602 (2014), URL http://link.aps.org/doi/10.1103/PhysRevLett.112.253602.

[93] J. W. Goodman et al., *Introduction to Fourier Optics*, vol. 2 (McGraw-hill New York, 1968).

[94] S. S. Welsh, M. P. Edgar, R. Bowman, P. Jonathan, B. Sun, and M. J. Padgett, "Fast Full-Color Computational Imaging with Single-Pixel Detectors," Opt. Express **21**, 23068 (2013), URL http://www.opticsexpress.org/abstract.cfm?URI=oe-21-20-23068.

[95] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk, "Single-Pixel Imaging via Compressive Sampling," Signal Processing Magazine, IEEE **25**, 83 (2008), ISSN 1053-5888.

[96] E. Candes and M. Wakin, "An Introduction To Compressive Sampling," Signal Processing Magazine, IEEE **25**, 21 (2008), ISSN 1053-5888.

[97] G. A. Howland, S. H. Knarr, J. Schneeloch, D. J. Lum, and J. C. Howell, "Compressively Characterizing High-Dimensional Entangled States with Sequential, Random Filtering," In Preparation (2015).

[98] J. R. Fienup, "Phase Retrieval Algorithms: A Comparison," Appl. Opt. **21**, 2758 (1982), URL http://ao.osa.org/abstract.cfm?URI=ao-21-15-2758.

[99] J. S. Bell, "EPR Correlations and EPW Distributions," Annals of the New York Academy of Sciences **480**, 263 (1986).

[100] J. S. Bell, *Speakable and unspeakable in quantum mechanics* (Cambridge University Press, New York, 1987), 1st ed.

[101] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, "Bell Nonlocality," Rev. Mod. Phys. **86**, 419 (2014), URL http://link.aps.org/doi/10.1103/RevModPhys.86.419.

[102] S. J. Freedman and J. F. Clauser, "Experimental Test of Local Hidden-Variable Theories," Phys. Rev. Lett. **28**, 938 (1972), URL http://link.aps.org/doi/10.1103/PhysRevLett.28.938.

[103] A. Aspect, P. Grangier, and G. Roger, "Experimental Tests of Realistic Local Theories via Bell's Theorem," Phys. Rev. Lett. **47**, 460 (1981).

[104] 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. **49**, 91 (1982), URL http://link.aps.org/doi/10.1103/PhysRevLett.49.91.

[105] J. Leach, B. Jack, J. Romero, M. Ritsch-Marte, R. W. Boyd, A. K. Jha, S. M. Barnett, S. Franke-Arnold, and M. J. Padgett, "Violation of a Bell Inequality in Two-Dimensional Orbital Angular Momentum State-Spaces," Opt. Express **17**, 8287 (2009), URL http://www.opticsexpress.org/abstract.cfm?URI=oe-17-10-8287.

[106] J. D. Franson, "Bell Inequality for Position and Time," Phys. Rev. Lett. **62**, 2205 (1989), URL http://link.aps.org/doi/10.1103/PhysRevLett.62.2205.

[107] P. G. Kwiat, A. M. Steinberg, and R. Y. Chiao, "High-Visibility Interference in a Bell-Inequality Experiment for Energy and Time," Phys. Rev. A **47**, R2472 (1993), URL http://link.aps.org/doi/10.1103/PhysRevA.47.R2472.

[108] C. J. Broadbent, R. M. Camacho, R. Xin, and J. C. Howell, "Preservation of Energy-Time Entanglement in a Slow Light Medium," Phys. Rev. Lett. **100**, 133602 (2008), URL http://link.aps.org/doi/10.1103/PhysRevLett.100.133602.

[109] M. Revzen, P. A. Mello, A. Mann, and L. M. Johansen, "Bell's Inequality Violation with Non-Negative Wigner Functions," Phys. Rev. A **71**, 022103 (2005), URL http://link.aps.org/doi/10.1103/PhysRevA.71.022103.

[110] A. F. Abouraddy, T. Yarnall, B. E. A. Saleh, and M. C. Teich, "Violation of Bell's Inequality with Continuous Spatial Variables," Phys. Rev. A **75**, 052114 (2007), URL http://link.aps.org/doi/10.1103/PhysRevA.75.052114.

[111] Z.-B. Chen, J.-W. Pan, G. Hou, and Y.-D. Zhang, "Maximal Violation of Bell's Inequalities for Continuous Variable Systems," Phys. Rev. Lett. **88**, 040406 (2002), URL http://link.aps.org/doi/10.1103/PhysRevLett.88.040406.

[112] M. A. Alonso, "Wigner Functions in Optics: Describing Beams as Ray Bundles and Pulses as Particle Ensembles," Adv. Opt. Photon. **3**, 272 (2011), URL http://aop.osa.org/abstract.cfm?URI=aop-3-4-272.

[113] W. J. Munro, "Optimal States for Bell-Inequality Violations using Quadrature-Phase Homodyne Measurements," Phys. Rev. A **59**, 4197 (1999), URL http://link.aps.org/doi/10.1103/PhysRevA.59.4197.

[114] J. Wenger, M. Hafezi, F. Grosshans, R. Tualle-Brouri, and P. Grangier, "Maximal Violation of Bell Inequalities using Continuous-Variable Measurements," Phys. Rev. A **67**, 012105 (2003), URL http://link.aps.org/doi/10.1103/PhysRevA.67.012105.

[115] H. Nha and H. J. Carmichael, "Proposed Test of Quantum Nonlocality for Continuous Variables," Phys. Rev. Lett. **93**, 020401 (2004), URL http://link.aps.org/doi/10.1103/PhysRevLett.93.020401.

[116] R. García-Patrón, J. Fiurášek, N. J. Cerf, J. Wenger, R. Tualle-Brouri, and P. Grangier, "Proposal for a Loophole-Free Bell Test Using Homodyne Detection," Phys. Rev. Lett. **93**, 130409 (2004), URL http://link.aps.org/doi/10.1103/PhysRevLett.93.130409.

[117] R. García-Patrón, J. Fiurášek, and N. J. Cerf, "Loophole-Free Test of Quantum Nonlocality using High-Efficiency Homodyne Detectors," Phys. Rev. A **71**, 022105 (2005), URL http://link.aps.org/doi/10.1103/PhysRevA.71.022105.

[118] A. Acín, N. J. Cerf, A. Ferraro, and J. Niset, "Tests of Multimode Quantum Nonlocality with Homodyne Measurements," Phys. Rev. A **79**, 012112 (2009), URL http://link.aps.org/doi/10.1103/PhysRevA.79.012112.

[119] P. B. Dixon, J. H. Shapiro, and F. N. C. Wong, "Spectral Engineering by Gaussian Phase-Matching for Quantum Photonics," Opt. Express **21**, 5879 (2013), URL http://www.opticsexpress.org/abstract.cfm?URI=oe-21-5-5879.