

An Information-Theoretic Approach to Adaptive Entanglement Witnessing

Lev Gruber

Professor Theresa W. Lynn, Advisor

May, 2026



Department of Physics

Copyright © 2026 Lev Gruber.

The author grants Harvey Mudd College the nonexclusive right to make this work available for noncommercial, educational purposes, provided that this copyright statement appears on the reproduced materials and notice is given that the copying is by permission of the author. To disseminate otherwise or to republish requires written permission from the author.

Abstract

Quantum entanglement is a vital resource in quantum computation, communication, and cryptography. Entanglement witnesses (EW) provide a method for certifying entanglement from few measurements. I present an adaptive protocol that uses local measurements on many copies of an unknown bipartite quantum state to certify entanglement by measuring EWs. As a given EW only certifies a subset of all entangled states, the protocol adaptively chooses subsequent measurements using the results of prior measurements. While the second-stage measurements overall certify 79.1% of computationally generated random entangled states, selecting a particular set of second-state measurements via our current best choice model leads to a 72.3% success rate. Seeking an upper bound on the success rate achievable by any adaptive choice model, I develop a Bayes classifier model to lower bound this upper bound. I estimate its performance by binning the aforementioned data set using a k-dimensional tree split on parameters informed by our specific witness sets; I find that its success rate lives very close to the success rate of our current best choice model and surpasses it for states with concurrence greater than 0.05. I then provide an existence proof of an information-theoretic bound on any adaptive choice method's performance. I later present a derivation that shows that our entanglement witnesses are equivalent to checking the PPT criterion for separability of a 2-qubit state in a restricted set of density matrix basis states, and use this interpretation to determine the exact set of entangled states our protocol is unable to detect.

Acknowledgments

First and foremost, thank you Professor Lynn for your support and mentorship over the past few years. You have been an amazing advisor who has helped me grow as a scientist and human being while I explored experiment, theory, and the horror of graduate applications. You have taught me how to find the simple solution in a world of complexity, how to poke holes in new theories to better understand them, how to turn half-baked but exciting ideas into new scientific understandings, and so much more. Thank you as well to the rest of Lynn Lab–Brayden Mendoza, Iz Godoy, Elio Thadhani, and Ria Haapala. You all have been such a joy to work with and learn from, and I look forward to seeing what each of you do in the future. Thank you also to past Lynn Lab members Stuart Kerr and Paco Navarro for making my first summer with the lab a great experience.

I would also like to acknowledge the various support systems that helped me as I embarked on the treacherous fight to maintain work-life balance while at HMC. Thank you to the many physics majors who struggled through problem sets with me (Henrik Barck (30+ PSETs), Ananya Venkatachalam (20+), Charlie Schofield (15+), Korin Aldam-Tajima (10+), Daniel Voyles (10+)). Thank you to the HMC ResLife deans–Dean Chris, Dean Vanessa, Dean Karen, and Dean Alyssa–and my fellow Orientation Directors–Julia Manese, Carter Meyer, and Kasey Chung–for being amazing collaborators while planning and running HMC's Orientation and Sophomore Retreats. Thank you to my three-year Atwood ResLife teammate Abby Samson for being one of the funniest and kindest human beings I have the privilege of calling my friend—you're a GOAT. Thank you to each of my Atwood mentees from the past three years for pushing me to be the best version of myself. Thank you to the Claremont Roos for being such an amazing competitive environment for volleyball and for welcoming me back after every single one of my countless number of injuries to countless parts of my body. Thank you to Dr. Mark Dong and Will Yzaguirre for both of your support through our entanglement harvesting project and for believing in each of my ideas. Finally, thank you to everyone who I did not mention but is thinking reading this, "I thought we were tight?"—we are, and I appreciate you too.

Contents

1 Introduction **7**

2 Background **11**
2.1 Notation . . . 11
2.1.1 Quantum States . . . 11
2.1.2 The Bloch Sphere . . . 12
2.1.3 Tensor Product . . . 13
2.1.4 Density Matrices . . . 14
2.1.5 Pauli Matrices . . . 14
2.1.6 Full Tomography . . . 15
2.1.7 The Partial Transpose . . . 15
2.2 Entanglement Witnesses (EW) . . . 16
2.3 Mutual Information and Entropy . . . 17

3 Adaptive Entanglement Witnessing **19**
3.1 Witness Sets . . . 19
3.1.1 $\{W^3\}$ Witness Set . . . 19
3.1.2 $\{W^5\}$ Witness Set . . . 21
3.2 Two Step Adaptive Witnessing Protocol . . . 26
3.3 Choice Models . . . 26
3.3.1 Population Model . . . 28
3.3.2 Five Layer Neural Network . . . 29
3.3.3 Cross Correlation Model . . . 29
3.3.4 A k-dimensional Tree . . . 30
3.3.5 Choice Model Comparison . . . 30

4 Bounding the Protocol's Success Rate **32**
4.1 Bayes Classifier Choice Method . . . 32
4.2 An Information-Theoretic Upper Bound . . . 34
4.2.1 States no Adaptive Choice Model can guess Correctly . . . 34
4.2.2 A Bound from the Gap . . . 36
4.3 Estimating P_{Bayes} . . . 40
4.4 Binning the Dataset: A k-d Tree . . . 41

4.4.1 An Overview of Binning Parameters 42
4.4.2 An Overview of Methods for Splitting the Dataset 43
4.5 Bayes Lower Bound Results 44
4.5.1 A Naive Binning: Stokes Parameters 44
4.5.2 A Smarter Binning with Analytic Minimization 45

5 A General Witnessing Framework from the PPT Criterion 51
5.1 General Framework Summary 51
5.2 A Geometric Understanding 54
5.2.1 Why the Population Model Wins 56
5.3 Witnesses as Density Matrix Elements 59
5.4 Connections to PPT and the most general witness 62

6 Conclusion 65

List of Figures

1.1 Motivating physical scenario for needing efficient entanglement detection of an unknown quantum state: some company is sending Alice (A) and Bob (B) many copies of an entangled two-photon state via satellite. If acting in bad faith, the company sends separable (non-entangled) states and reads their communication; at the same time, the atmosphere may be modifying the state, requiring Alice and Bob certify the entanglement of an unknown quantum state. 8

2.1 Horizontal ($|H\rangle$), vertical ($|V\rangle$), diagonal ($|D\rangle$), and anti-diagonal polarizations ($|A\rangle$). Note that right and left circularly polarized are defined as $|R\rangle = |H\rangle + i|V\rangle$ and $|R\rangle = |H\rangle - i|V\rangle$, respectively. 12
2.2 Bloch sphere in the $|H\rangle$, $|V\rangle$ basis, adapted from [29]. 13

3.1 Measurement bases required for the $\{W^3\}$ (blue diagonal) and $\{W^5\}$ witnesses (color-coded by additional measurement per triplet). . . 21
3.2 Flow chart of the adaptive witnessing protocol. The witness sets are given in Sec. 3.1 and the choice methods in Sec. 3.3. 27
3.3 Protocol success rate for different choice models and witness sets for all states with concurrence greater than or equal to some value on the x-axis. See Sec. 3.3 for a description of each choice model. . 31

4.1 Γ from Eq. 4.2.16 as a function of ϵ (top) and an estimate of P_{Bayes} from $P_{\text{oracle}} = 79.1\%$, and Eq. 4.2.14. Here, ϵ controls how tightly the data set is binned. 39
4.2 k-d tree Bayes classifier performance when given all $\{W^3\}$ measurements. At $c > 0.0$, meaning for all entangled states in the data set, this estimates $P_{\text{Bayes}} = 71.3\%$. 45

4.3 Top 12 features when the k-d tree is given Δ_i and the nine Stokes parameters needed for the $\{W^3\}$ set. I provide both to compare which the tree picks. The asymmetry between S_{zz}, S_{xx}, and S_{yy} is due to each feature being standardized to mean 0 and variance 1 before the first split, such that S_{zz} gets randomly chosen to start off. 46
4.4 Feature usage of a variance-based tree given Stokes parameters and the Δ_i parameters, plotted for the 12 most-used features. 46
4.5 The k-dimensional tree split on the Δ_i parameters of the cross correlation model, evaluated via information gain as discussed in Sec. 4.4. 47
4.6 Bayes lower bound as a function of tree depth for a balanced tree split on information gain and given $\Delta_i(x)$ derived from Eq. 3.3.4. The lower bound is evaluated on two data sets of 10 million (blue) and 16 million (orange) states each. 48
4.7 Information gain-dependent k-d tree lower bound as a function of tree depth, provided only the Δ_i parameters. This is a zoomed in version of Fig. 4.5. 48
4.8 Choice model success rates as a function of concurrence with the k-d tree based lower bound shown both at a depth of 16 and as the result of an exponential fit on a lower bound versus depth curve. Only the Δ_i splitting parameters are provided. 50

5.1 A comparison of the population and cross correlation choice models, evaluated only on states that are witnessed by the first and second W^5 of each triplet (top) or third (bottom). 58

List of Tables

3.1 Population model; the maximum value of the three populations corresponding to which $\{W^5\}$ triplet is chosen. 28

Chapter 1

Introduction

Secure communication is paramount in our increasingly digital world. One way to ensure secure communication rests on a basic facet of quantum information theory: the no-cloning theorem. This states that an arbitrary pure quantum state cannot be exactly copied without disturbing the original state; in other words, no unitary transformation exists that brings the state $|\psi\rangle_A \otimes |\phi\rangle_B$ to $|\phi\rangle_A \otimes |\phi\rangle_B$ for all quantum states $|\phi\rangle$ and any quantum state $|\psi\rangle$, associated with subsystems B and A respectively. This implies that an outside antagonistic observer would be unable to intercept and duplicate a message that is encoded as a quantum state chosen from a set of non-orthogonal possible states without disturbing the message itself.

To illustrate this, consider the scenario as follows: a company which I name Qummunicate sends Alice and Bob–two spatially separated observers who hope to communicate securely–a stream of many identical photons that are entangled in their polarization; see Fig 1.1. If the quantum states of these two streams of photons are entangled, Qummunicate is unable to read any of Alice and Bob's communications despite being the provider of the states. If Qummunicate is instead acting in bad faith and sending non-entangled states, the company can read Alice and Bob's communication. Thus, Alice and Bob need a way to certify the entanglement shared by their photon streams. While doing so, factors such as atmospheric interference may change the state the two observers share, requiring the two to certify entanglement without knowing exactly what the state they are receiving is.

One method to detect entanglement is for Alice and Bob to fully determine the state itself through a full tomography—this is a process in which they take a measurement in every polarization basis and use that to reconstruct the state's density matrix. This can be thought of similar to taking multiple head-on 2D images of a shape to determine that it is a cube because every 2D image is found to be a square. The primary issue with a full tomography is that it is too time-intensive for use in a real-world communication scenario; some current long-

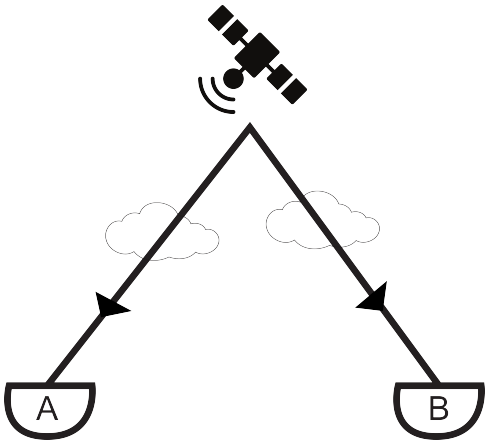


Figure 1.1: Motivating physical scenario for needing efficient entanglement detection of an unknown quantum state: some company is sending Alice (A) and Bob (B) many copies of an entangled two-photon state via satellite. If acting in bad faith, the company sends separable (non-entangled) states and reads their communication; at the same time, the atmosphere may be modifying the state, requiring Alice and Bob certify the entanglement of an unknown quantum state.

range methods take up to 160 hours [15, 16]. Many other current approaches to certifying entanglement rely on prior knowledge of the state.

One method to detect the entanglement with fewer measurements than a full tomography uses entanglement witnesses: block-positive but not positive operators that have negative expectation value on a subset of all entangled states and positive expectation value on all separable (non-entangled) states. Different entanglement witnesses are able to detect different subsets of the space of entangled states, necessitating the measurement of multiple different entanglement witnesses given no prior knowledge of the state you are attempting to detect the entanglement of. A motivating question is what fraction of all entangled states are detectable given you measure some fraction of all measurement bases, or equivalently, measure some set of entanglement witnesses. We address this question by testing different witnesses on a computationally generated random mixed state set, where each random state is created by randomly sampling the four diagonal probabilities of the density matrix as in [17], then rotating the density matrix by a randomly generated unitary matrix as in [23]. Of note, this data set contains 16,831,944 entangled states.

In this thesis, I will frequently refer to two sets of entanglement witnesses. The first is the $\{W^3\}$ set that was originally proposed by Riccardi et al [22] in 2020, which rely on a third of the measurements required for a full tomography and in turn witness a third of all computationally generated random entangled

states. The second is the $\{W^5\}$ set which was derived by previous Quantum Optics Group members (Ben Hartley originally, using the methods of [22], and Paco Navarro later in a reformulated geometric way [18]). This set is split into three triplets, $\{W_i^5\}$, $i \in [1,2,3]$ which rely on different measurements to determine, but each require 55% of a full tomography and detect approximately 79% of all computationally generated random entangled states. If all three triplets were measured simultaneously, that would constitute a full tomography.

Previous Quantum Optics Group members proposed a two-step adaptive entanglement detection protocol that utilizes successive measurements on the many-copy stream of photons to measure the qubit pair in the bases required to test entanglement witnesses. The protocol is as follows: first, measure the $\{W^3\}$ witness set; if any of the corresponding witnesses are negative, you have witnesses entanglement and the protocol ends. If you have not yet witnesses entanglement, use the previous measurement outcomes to make an informed decision of which of the three $\{W^5\}$ triplets to measure next. If any of those are negative, you have witnesses entanglement. If not, your state is either not entangled, you made the incorrect choice of witness triplet to measure, or the state is one of the approximately 21% of entangled states missed entirely by both witness sets.

In this thesis, I focus on bounding the performance of the adaptive choice method. To do so, I derive a Bayes classifier model for the adaptive choice, which is an optimal classification model that minimizes the error rate. I then use a k-dimensional tree to estimate a lower bound on the Bayes classifier success rate, inspired by the fact that methods for random entangled state generation are varied and we wish to determine how our adaptive protocol performs on our data set of entangled states specifically. To ensure that the k-d tree's binning is done in a way that pushes states more likely to be witnessed by a given entanglement witness into the same bin, I derive a new form of the $\{W^3\}$ and $\{W^5\}$ witnesses that give a set of inequalities on which states may be split. To improve the binning process further, I split subsequent tree nodes via information gain, inspired by the fact that our protocol uses subsequent measurements to gain progressively more information about a quantum state.

The Bayes classifier estimation becomes untrustworthy when there are too few states per leaf (lowest node in the k-d tree) because the distribution of the likelihood of each witness to detect states in a given bin becomes unrepresentative of the data set overall. To determine what points are trustworthy and which are not, I look at the performance of the Bayes classifier as a choice model on a set-aside testing set of the 16,831,944 state data set, as it is representative of the success rate if this model were applied in a real-world test case. I find that the success rate peaks at a tree depth of 16—I then treat all points at depth 16 or lower as trustworthy and those above as not. By fitting an exponential curve to the trustworthy points, I estimate a lower bound protocol success rate

of 72.3%, which matches the group's current best choice model–a five layer neural network–to two hundredths of a percent. I proceed to interpret these results and give a simple story for why this method works.

The remainder of this thesis is organized as follows: in Chapter 2 I provide necessary background on quantum information theory and entanglement witnesses. In Chapter 3 I introduce the $\{W^3\}$ and $\{W^5\}$ witness sets which in general depend on minimizable parameters and provide analytic minimizations when possible. I also summarize the group's current choice methods—an analytic method denoted the *population method* and a five-layer neural network, and present two new choice models: the cross correlation model that is based on analytic minimizations of the $\{W^3\}$ and $\{W^5\}$ witnesses, and a k-d tree model. In Chapter 4 I derive the Bayes classifier choice method and estimate its performance using a k-dimensional tree. Finally, in Chapter 5, I demonstrate that the $\{W^3\}$ and $\{W^5\}$ witness sets are equivalent to the positive partial transpose (PPT) criterion—which certifies whether a bipartite two-qubit density matrix is separable—when the density matrix is written in one of a limited set of basis states. I then interpret this result geometrically and use it to determine exactly what quantum states our protocol is able and unable to certify the entanglement of, assuming a perfect choice method.

Chapter 2

Background

2.1 Notation

2.1.1 Quantum States

In the following work, I will represent a quantum state using a ket symbol $|\psi\rangle$ and take it to be a linear combination of basis states, each multiplied by a probability amplitude. The probability amplitude is a complex number whose magnitude squared is the probability of measuring that basis state. We will often use basis states corresponding to the polarization of a photon, namely, $|H\rangle, |V\rangle, |D\rangle, |A\rangle, |R\rangle, |L\rangle$ for horizontal, vertical, diagonal, anti-diagonal, right circular, and left circular polarization—see Fig 2.1. For a state with a 50/50 chance of being horizontally or vertically polarized, we write

$$|\psi\rangle = \frac{1}{\sqrt{2}}|H\rangle + e^{i\phi}\frac{1}{\sqrt{2}}|V\rangle = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ e^{i\phi}\frac{1}{\sqrt{2}} \end{pmatrix} \tag{2.1.1}$$

where $e^{i\phi}$ is a complex phase and $\phi \in [0, 2\pi]$. The column vector is an equivalent way to write a ket given a particular choice of basis. In this thesis, I focus solely on the two particle case, which may be written in the horizontal-vertical polarization basis as

$$|\varphi\rangle = a|HH\rangle + b|HV\rangle + c|VH\rangle + d|VV\rangle = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}, \tag{2.1.2}$$

where the first letter in the ket corresponds to particle A's state, the second letter to particle B's state, and $a, b, c, d \in \mathbb{C}$ are probability amplitudes. To conserve probability, $|a|^2 + |b|^2 + |c|^2 + |d|^2 = 1$. We write the dual (row vector) of the state $|\varphi\rangle$ as

$$\langle\varphi| = a^*\langle HH| + b^*\langle HV| + c^*\langle VH| + d^*\langle VV| = \begin{pmatrix} a^* & b^* & c^* & d^* \end{pmatrix}, \tag{2.1.3}$$

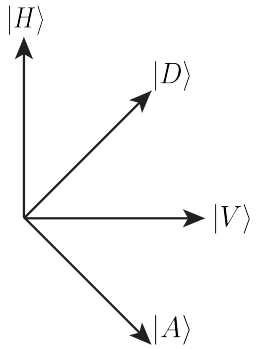


Figure 2.1: Horizontal ($|H\rangle$), vertical ($|V\rangle$), diagonal ($|D\rangle$), and anti-diagonal polarizations ($|A\rangle$). Note that right and left circularly polarized are defined as $|R\rangle = |H\rangle + i\,|V\rangle$ and $|R\rangle = |H\rangle - i\,|V\rangle$, respectively.

where $\langle\varphi|$ is denoted a bra. All quantum states live in a Hilbert space $\mathscr{H}$, which is a complete vector space equipped with an inner product. A bipartite Hilbert space hosts the joint state of particles A and B and is written $\mathscr{H}_A \otimes \mathscr{H}_B$. The tensor product operator $\otimes$ is defined in the next section.

2.1.2 The Bloch Sphere

The Bloch sphere is a way to represent a single-particle quantum state geometrically. For basis vectors $|H\rangle$ and $|V\rangle$, the Bloch sphere is a unit sphere with $|H\rangle$ and $|V\rangle$ pointing in the positive $\hat{z}$ and negative $\hat{z}$ directions, respectively. This is shown in Fig. 2.2. Points on the Bloch sphere correspond to pure states, which can be expressed as

$$|\psi\rangle = \cos(\theta/2)\,|H\rangle + e^{i\phi}\sin(\theta/2)\,|V\rangle\,, \tag{2.1.4}$$

with θ, ϕ representing the typical spherical coordinate representation of a vector of length 1. Thus, on the Bloch sphere, a pure state as in Eq. 2.1.4 may be written in vector form as

$$|\psi\rangle = \begin{pmatrix}\sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta.\end{pmatrix} \tag{2.1.5}$$

Mixed states may be represented as points within a solid Bloch sphere—called the Bloch ball—but this is irrelevant for the rest of this thesis.

A rotation around an axis $\hat{n}$ of the Bloch sphere can be written mathematically as

$$R_{\hat{n}}(\theta) = e^{-i\frac{\theta}{2}\sigma_n}, \tag{2.1.6}$$

where σ_n is a Pauli operator as defined in Sec. 2.1.5.

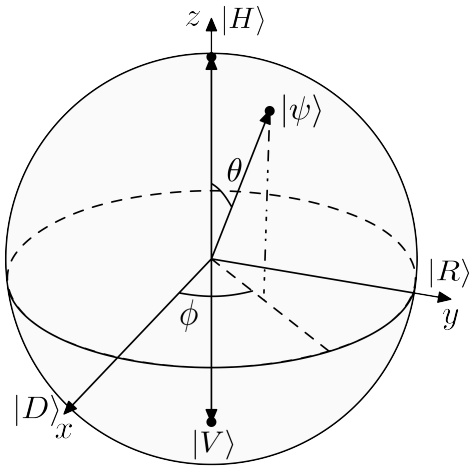


Figure 2.2: Bloch sphere in the $|H\rangle$, $|V\rangle$ basis, adapted from [29].

2.1.3 Tensor Product

A tensor product creates a representation of two lower-dimensional matrices in a higher dimensional space. Given two matrices

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}, \; B = \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{n1} & \cdots & b_{nn}, \end{pmatrix} \tag{2.1.7}$$

their tensor product is defined as

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & & \vdots \\ a_{n1}B & \cdots & a_{nn}B \end{pmatrix}. \tag{2.1.8}$$

An outer product is a tensor product but applied to vectors. Given to vectors

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}, \; \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}, \; n, m \in \mathbb{Z}_+, \tag{2.1.9}$$

their outer product is defined as

$$\mathbf{a} \otimes \mathbf{b} = \begin{pmatrix} a_1 b_1 & a_1 b_2 & \dots & a_1 b_n \\ a_2 b_1 & a_2 b_2 & \dots & a_2 b_n \\ \vdots & \vdots & \ddots & \vdots \\ a_m b_1 & a_m b_2 & \dots & a_m b_n \end{pmatrix}. \tag{2.1.10}$$

2.1.4 Density Matrices

The outer product of a state's ket with its bra forms a density matrix

$$\rho = |\varphi\rangle\langle\varphi| = \begin{pmatrix} |a|^2 & ab^* & ac^* & ad^* \\ ba^* & |b|^2 & bc^* & bd^* \\ ca^* & cb^* & |c|^2 & cd^* \\ da^* & db^* & dc^* & |d|^2 \end{pmatrix}. \tag{2.1.11}$$

Of note, the density matrix representation of a quantum state allows us to differentiate between pure states and mixed states. Pure states are the type we have seen thus far and can be written as a vector $|\varphi\rangle$. A mixed state, however, is the probabilistic sum of pure states and can only be written in density matrix form as

$$\rho = \sum p_i |\psi_i\rangle\langle\psi_i|, \tag{2.1.12}$$

for pure states $|\psi_i\rangle$ and $p_i \in \mathbb{R}$, $\sum p_i = 1$. A two particle quantum state is separable if it may be written as the tensor product of two states, e.g.

$$\rho = |\psi_1\rangle\langle\psi_1| \otimes |\psi_2\rangle\langle\psi_2|. \tag{2.1.13}$$

This implies that each single particle state in a separable bipartite state lives in only its respective Hilbert space. A quantum state is entangled if and only if it is not separable. This means that entangled states live in the joint Hilbert space $\mathscr{H}_A \otimes \mathscr{H}_B$.

2.1.5 Pauli Matrices

The Pauli matrices are a set of $2x2$ complex matrices that with the identity matrix span the space of 2×2 Hermitian matrices. They are given as

$$\sigma_1 = \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ \sigma_2 = \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \ \sigma_3 = \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.1.14}$$

The identity matrix is given as $\mathbb{I}$ and denoted σ_0. Tensor products of Pauli matrices span the space of density matrices, meaning all two-particle density matrices admit the following decomposition

$$\rho = \sum_{i=0,j=0}^{3} S_{ij}\, \sigma_i \otimes \sigma_j. \tag{2.1.15}$$

Here, $S_{ij} = \mathrm{Tr}(\rho(\sigma_i \otimes \sigma_j)) = \langle \sigma_i \otimes \sigma_j \rangle \in \mathbb{R}$ are denoted Stokes parameters. I will write $\mathbb{I}$ versus σ_0 when possible, but write expressions such as $\langle \mathbb{I} \otimes \sigma_i \rangle$ as S_{0i}. To determine $\langle \sigma_i \otimes \sigma_j \rangle$, you may measure the two particles separately in various bases then correlate the results via classical communication (these are called correlated local measurements). In the photon-polarization case considered in

this thesis, there are 36 projective probabilities corresponding to measuring each particle in one of the $|H\rangle, |V\rangle, |A\rangle$, or $|D\rangle$ bases. Obtaining correlated local measurements from a set of mutually orthogonal basis vectors allows you to determine the probabilities of each of the four eigenstates of $\sigma_i \otimes \sigma_j$, in turn allowing you to determine $\langle \sigma_i \otimes \sigma_j \rangle$, as well as $\langle \sigma_i \otimes \mathbb{I} \rangle$ and $\langle \mathbb{I} \otimes \sigma_j \rangle$.

2.1.6 Full Tomography

> "Here it is. The most important process in quantum chocolate factory – the tomography!" - Willy Wonky-Action-at-a-Distance

Recall that any two-particle density matrix can be written as

$$\rho = \sum_{i=0,j=0}^{3} S_{ij}\, \sigma_i \otimes \sigma_j. \tag{2.1.16}$$

Measuring in all 9 bases, where I call a basis one of the $\sigma_i \otimes \sigma_j$ pairs, is called a full quantum tomography. Of note for this thesis, a full tomography requires $\mathcal{O}(N^2)$ measurements on a state of dimension N and can take a long time to complete as mentioned in Chap. 1. The primary factor that influences the completion time of a full tomography is the rate of measurement. For example, in quantum communication schemes where photons are sent from a satellite to a ground-based detector, the atmosphere can interfere and cause a very low measurement rate.

2.1.7 The Partial Transpose

A partial transpose is an operation that applies the transpose to only one subsystem of a quantum system. Given a bipartite state defined on the state space of particles A and B,

$$\rho = \sum_{i,j,k,l} p_{i,j,k,l} (|i\rangle \langle j|)_A \otimes (|k\rangle \langle l|)_B, \tag{2.1.17}$$

its partial transpose with respect to particle B is

$$\rho^{\Gamma_B} = \sum_{i,j,k,l} p_{i,j,k,l} (|i\rangle \langle j|)_A \otimes (|l\rangle \langle k|)_B. \tag{2.1.18}$$

When written as a block matrix (see Eq. 2.1.8), the partial transpose with respect to subsystem B is

$$\rho^{\Gamma_B} = \begin{pmatrix} a_{11}B^\Gamma & \cdots & a_{1n}B^\Gamma \\ \vdots & & \vdots \\ a_{n1}B^\Gamma & \cdots & a_{nn}B^\Gamma \end{pmatrix}. \tag{2.1.19}$$

When written without a subscript on Γ, the partial transpose is understood to be with respect to subsystem B.

2.2 Entanglement Witnesses (EW)

> "Hello. My name is EWnigo Quantoya. You are entangled. Prepare to be witnessed." - EWnigo Quantoya

Entanglement witnesses are operators that can certify the entanglement of a quantum state. They are block-positive meaning that, for an entanglement witness W acting on a bipartite Hilbert space $\mathscr{H}_A \otimes \mathscr{H}_B$,

$$\langle\psi| \otimes \langle\varphi| W |\psi\rangle \otimes |\varphi\rangle > 0, \quad \forall\ |\psi\rangle \in \mathscr{H}_A,\ |\varphi\rangle \in \mathscr{H}_B. \tag{2.2.1}$$

This implies that they have positive expectation value for all separable states. Importantly, they also have positive expectation value for all separable mixed states. At the same time, entanglement witnesses are not positive, meaning that there exists some entangled state–which lives in the joint space $\mathscr{H}_A \otimes \mathscr{H}_B$ because of block-positivity–for which W has a negative expectation value. In other words,

$$\mathrm{Tr}(W\rho) < 0, \text{ for some entangled state } \rho. \tag{2.2.2}$$

An important fact is that for every entangled state, there exists at least one entanglement witness with negative expectation for that state [12]. Conversely, each entanglement witness only witnesses a subset of the space of all entangled states. We continue to characterize optimality in entanglement witnessing.

Let $\{\rho_W\}$ be the set of entangled states witnessed by W. W is optimal only if there exists no other witness W' whose set of witnessed states $\{\rho_{W'}\}$ is a super-set of $\{\rho_W\}$ [6], i.e.

$$W \text{ is optimal only if } \nexists \text{ any witness } W' \text{ for which } \{\rho_W\} \subset \{\rho_{W'}\}. \tag{2.2.3}$$

An important takeaway is there there exists no *universal* entanglement witness, meaning not all entangled states are detected by a single entanglement witness. This requires us to construct sets of entanglement witnesses that are each capable of detecting different sets of entangled states.

One such class of entanglement witnesses are extremal decomposable entanglement witnesses. An entanglement witness is extremal decomposable if it can be written as [6]

$$W = |\phi\rangle\langle\phi|^{\Gamma} \text{ for an entangled state } |\phi\rangle. \tag{2.2.4}$$

In Sec. 2.2, I present two sets of extremal decomposable entanglement witnesses. The sets we consider in this work can be constructed from limited local measurements on many copies of ρ and do not require a full tomography. This is important for reducing the amount of time it takes to certify entanglement.

2.3 Mutual Information and Entropy

> "Entropy, darling, as usual, perfectly increasing", One Hundred and One Bits

In later sections, a refresh of the basics of information theory will be helpful. Here, I present some definitions that should kept in mind.

Consider a discrete random variable X with support $\mathscr{X}$. The probability of some $x \in \mathscr{X}$ being chosen is denoted $p(x)$. The Shannon entropy of X determines the minimum amount of information needed to specify one possible outcome of X, averaged across all outcomes:

$$H(X) = -\sum_{x \in \mathscr{X}} p(x) \log_2 p(x). \tag{2.3.1}$$

An alternative understanding of Shannon entropy is as the minimum number of bits I would need to send a friend to tell them an outcome of X, averaged across all outcomes. As a brief example, if X is a unweighted six-sided die, $\mathscr{X} = \{1,2,3,4,5,6\}$, $p(x) = \frac{1}{6}$ for all x, and $H(X) = -\sum_{i=1}^{6} \frac{1}{6} \log \frac{1}{6} = 2.58$. In the communication sense, this measures the average minimum number of bits I'd need to send a friend to specify an outcome of the roll.

Given two random variables X and Y, the conditional entropy measures the amount of information needed to describe the outcome of X given the outcome of Y. Let $\mathscr{X}$ and $\mathscr{Y}$ be the supports of these two random variables. If $p(x,y)$ is the probability of obtaining value $x \in \mathscr{X}$ given Y takes on $y \in \mathscr{Y}$, then the conditional entropy is

$$H(X|Y) = -\sum_{x \in \mathscr{X}, y = \mathscr{Y}} p(y,x) \log \frac{p(y,x)}{p(y)}, \tag{2.3.2}$$

where $H(X|Y)$ can be read as the conditional entropy of X given Y. In the language of bits, conditional entropy can be thought of as the number of bits still needed to specify X in a message if the outcome of Y is known. If $H(X|Y) = 0$, then the outcome of Y fully specifies X, while if $H(X|Y) = H(X)$, Y does not help predict X.

Given two random variables, X and Y, the mutual information measures the amount of information that knowing the outcome of Y provides about X, on average. We can think of it as the difference between the average original length of a message I would have to send to specify X, $H(X)$, minus the new length of a message I would have to send if Y is known, $H(X|Y)$. Or, as the uncertainty of a variable X minus the uncertainty about X given I know the outcome of Y. Therefore, the mutual information is defined by

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X). \tag{2.3.3}$$

Despite similar colloquial definitions, this identifies conditional entropy and mutual information as two distinct quantities: the first tells us the new uncertainty of X given Y, while the second tells us how much the uncertainty—or average length of a message to specify X—changed once Y was known. This explains why, when Y provides no information about X, $H(X|Y) = H(X)$ as the mutual information is zero (no information is gained and the length of the message to specify X is the same). Conversely, if Y fully specifies X, $H(Y|X) = 0$ and the mutual information is just the information it took to specify X in the first place, $H(X)$.

Chapter 3

Adaptive Entanglement Witnessing

3.1 Witness Sets

> "Toto, I've a feeling we aren't taking full tomographies anymore", Dorothy after being quantum teleported to Oz

3.1.1 $\{W^3\}$ Witness Set

The W^3 witnesses are a set of six extremal decomposable–hence optimal– entanglement witnesses that require 3 of the 9 bases used in a full tomography. They were derived in Riccardi et al. in 2020 [22] and have the form

$$
\begin{aligned}
W_1^3 &= \frac{1}{4}[1+\sigma_z\otimes\sigma_z+\cos(2\theta)(\sigma_x\otimes\sigma_x+\sigma_y\otimes\sigma_y)+\sin(2\theta)(\sigma_z\otimes\mathbb{I}+\mathbb{I}\otimes\sigma_z)],\\
W_2^3 &= \frac{1}{4}[1-\sigma_z\otimes\sigma_z+\cos(2\theta)(\sigma_x\otimes\sigma_x-\sigma_y\otimes\sigma_y)+\sin(2\theta)(\sigma_z\otimes\mathbb{I}-\mathbb{I}\otimes\sigma_z)],\\
W_3^3 &= \frac{1}{4}[1+\sigma_x\otimes\sigma_x+\cos(2\theta)(\sigma_z\otimes\sigma_z+\sigma_y\otimes\sigma_y)+\sin(2\theta)(\sigma_x\otimes\mathbb{I}+\mathbb{I}\otimes\sigma_x)],\\
W_4^3 &= \frac{1}{4}[1-\sigma_x\otimes\sigma_x+\cos(2\theta)(\sigma_z\otimes\sigma_z-\sigma_y\otimes\sigma_y)-\sin(2\theta)(\sigma_x\otimes\mathbb{I}-\mathbb{I}\otimes\sigma_x)],\\
W_5^3 &= \frac{1}{4}[1+\sigma_y\otimes\sigma_y+\cos(2\theta)(\sigma_z\otimes\sigma_z+\sigma_x\otimes\sigma_x)+\sin(2\theta)(\sigma_y\otimes\mathbb{I}+\mathbb{I}\otimes\sigma_y)],\\
W_6^3 &= \frac{1}{4}[1-\sigma_y\otimes\sigma_y+\cos(2\theta)(\sigma_z\otimes\sigma_z-\sigma_x\otimes\sigma_x)-\sin(2\theta)(\sigma_y\otimes\mathbb{I}-\mathbb{I}\otimes\sigma_y)].
\end{aligned}
\tag{3.1.1}
$$

These correspond to the bases colored blue on the diagonal of Fig. 3.1. To search for the most negative values of $\{W_i^3\}$, these witnesses must be minimized with respect to θ once the expectation values of the above Pauli operators are known.

I analytically minimize them using the identity

$$c\cos\alpha \pm d\sin\alpha = \sqrt{c^2+d^2}\cos(\alpha \mp x), \tag{3.1.2}$$

which has minimum $-\sqrt{c^2+d^2}$. To provide a short proof, begin with the left hand side:

$$c\cos\alpha + d\sin\alpha = \sqrt{c^2+d^2}\left[\frac{c}{\sqrt{c^2+d^2}}\cos\alpha + \frac{d}{\sqrt{c^2+d^2}}\sin\alpha\right], \tag{3.1.3}$$

$$= \sqrt{c^2+d^2}\left[\cos x\cos\alpha \pm \sin x\sin\alpha\right] \tag{3.1.4}$$

$$= \sqrt{c^2+d^2}\left[\cos(\alpha \mp x)\right], \tag{3.1.5}$$

where I have defined $\cos x = \frac{c}{\sqrt{c^2+d^2}}$, $\sin x = \frac{d}{\sqrt{c^2+d^2}}$, and $\tan(x) = \frac{d}{c}$. I compute $\langle W_1^3\rangle$ as an example. Begin with

$$\langle W_1^3\rangle_{\min} = \min_\theta\left[\frac{1}{4}[1+\langle\sigma_z\otimes\sigma_z\rangle + \cos(2\theta)(\langle\sigma_x\otimes\sigma_x\rangle + \langle\sigma_y\otimes\sigma_y\rangle) + \sin(2\theta)(\langle\sigma_z\otimes\mathbb{I}\rangle + \langle\mathbb{I}\otimes\sigma_z\rangle)\right]. \tag{3.1.6}$$

Let $\alpha = 2\theta$, $c = (\sigma_x\otimes\sigma_x + \sigma_y\otimes\sigma_y)$, and $d = (\sigma_z\otimes\mathbb{I} + \mathbb{I}\otimes\sigma_z)$ to obtain

$$\langle W_1^3\rangle_{\min} = \frac{1}{4}\left[1+\langle\sigma_z\otimes\sigma_z\rangle + \min_\alpha\left(c\cos(\alpha) + d\sin(\alpha)\right)\right] \tag{3.1.7}$$

Now, apply Eq. 3.1.2 to get with the minimum expectation value

$$\langle W_1^3\rangle_{\min} = \frac{1}{4}[1+S_{zz} - \sqrt{(S_{xx}+S_{yy})^2 + (S_{z0}+S_{0z})^2}], \tag{3.1.8}$$

where I have used that $S_{ij} = \langle\sigma_i\otimes\sigma_j\rangle$ and $\sigma_0 \equiv \mathbb{I}$. The derivations for the rest of the witnesses in Eq. 3.1.1 are equivalent and yield

$$\begin{aligned}
\langle W_1^3\rangle_{\min} &= \frac{1}{4}[1+S_{zz} - \sqrt{(S_{xx}+S_{yy})^2 + (S_{z0}+S_{0z})^2}],\\
\langle W_2^3\rangle_{\min} &= \frac{1}{4}[1-S_{zz} - \sqrt{(S_{xx}-S_{yy})^2 + (S_{z0}-S_{0z})^2}],\\
\langle W_3^3\rangle_{\min} &= \frac{1}{4}[1+S_{xx} - \sqrt{(S_{zz}+S_{yy})^2 + (S_{x0}+S_{0x})^2},\\
\langle W_4^3\rangle_{\min} &= \frac{1}{4}[1-S_{xx} - \sqrt{(S_{zz}-S_{yy})^2 + (S_{x0}-S_{0x})^2}],\\
\langle W_5^3\rangle_{\min} &= \frac{1}{4}[1+S_{yy} - \sqrt{(S_{zz}+S_{xx})^2 + (S_{y0}+S_{0y})^2}],\\
\langle W_6^3\rangle_{\min} &= \frac{1}{4}[1-S_{yy} - \sqrt{(S_{zz}-S_{xx})^2 + (S_{y0}-S_{0y})^2}].
\end{aligned} \tag{3.1.9}$$

I find a $2285\times \pm 258\times$ speed up between the time it takes to minimize all 6 one-parameter $\{W^3\}$ witnesses in Eq. 3.1.1 via SciPy versus using the analytic equations given in Eq 3.1.9.



Figure 3.1: Measurement bases required for the $\{W^3\}$ (blue diagonal) and $\{W^5\}$ witnesses (color-coded by additional measurement per triplet).

3.1.2 $\{W^5\}$ Witness Set

The $\{W^5\}$ witnesses are constructed similarly to the $\{W^3\}$ set but require measurements in additional bases. They may either be derived by varying the Stokes parameters one sets to zero as in [22], or by applying rotation operators to either or both of particles A or B, some angle about x, y, or z; see Paco Navarro's thesis for this geometric derivation [18]. Dissimilarly to the $\{W^3\}$ set, the $\{W^5\}$ set are broken into three triplets, each of which depend on a different set of measurement bases, along with those already required for W^5 measurement. The three triplets are written as

$$\{W^5_{t1}\} = \{W^5_1, W^5_2, W^5_3\} \rightarrow \sigma_x \otimes \sigma_y, \sigma_y \otimes \sigma_x, \tag{3.1.10}$$

$$\{W^5_{t2}\} = \{W^5_4, W^5_5, W^5_6\} \rightarrow \sigma_x \otimes \sigma_z, \sigma_z \otimes \sigma_x, \tag{3.1.11}$$

$$\{W^5_{t3}\} = \{W^5_7, W^5_8, W^5_9\} \rightarrow \sigma_y \otimes \sigma_z, \sigma_z \otimes \sigma_y. \tag{3.1.12}$$

In Fig. 3.1, each $\{W^5\}$ triplet necessitates the three blue diagonal terms as well as one of the color coded off diagonals; $\{W^5_{t1}\}$ is purple, $\{W^5_{t2}\}$ green, and $\{W^5_{t3}\}$ red. Measuring the entire $\{W^5\}$ set constitutes a full tomography, so we choose to measure only one triplet at a time. The expressions are given in Eq. 3.1.13.

$$
\begin{aligned}
W_1^5 = W_{1,z,\mathbb{I}}^3 = \frac{1}{4}[1 &+ \sigma_z \otimes \sigma_z + \cos(2\theta)(\cos(\alpha)(\sigma_x \otimes \sigma_x + \sigma_y \otimes \sigma_y) \\
&- \sin(\alpha)(\sigma_x \otimes \sigma_y - \sigma_y \otimes \sigma_x)) + \sin(2\theta)(\sigma_z \otimes \mathbb{I} + \mathbb{I} \otimes \sigma_z)] \\
W_2^5 = W_{2,z,\mathbb{I}}^3 = \frac{1}{4}[1 &- \sigma_z \otimes \sigma_z + \cos(2\theta)(\cos(\alpha)(\sigma_x \otimes \sigma_x - \sigma_y \otimes \sigma_y) \\
&+ \sin(\alpha)(\sigma_x \otimes \sigma_y + \sigma_y \otimes \sigma_x)) + \sin(2\theta)(\sigma_z \otimes \mathbb{I} - \mathbb{I} \otimes \sigma_z)] \\
W_3^5 = W_{3,z,z}^3 = \frac{1}{4}[1 &+ \cos(\alpha)\cos(\beta)\sigma_x \otimes \sigma_x + \sin(\alpha)\sin(\beta)\sigma_y \otimes \sigma_y \\
&+ \sin(\alpha)\cos(\beta)\sigma_y \otimes \sigma_x + \cos(\alpha)\sin(\beta)\sigma_x \otimes \sigma_y + \cos(2\theta)(\sigma_z \otimes \sigma_z \\
&+ \cos(\alpha)\cos(\beta)\sigma_y \otimes \sigma_y + \sin(\alpha)\sin(\beta)\sigma_x \otimes \sigma_x - \sin(\alpha)\cos(\beta)\sigma_x \otimes \sigma_y \\
&- \cos(\alpha)\sin(\beta)\sigma_y \otimes \sigma_x) + \sin(2\theta)(\cos(\alpha)\sigma_x \otimes \mathbb{I} + \sin(\alpha)\sigma_y \otimes \mathbb{I} \\
&+ \cos(\beta)\mathbb{I} \otimes \sigma_x + \sin(\beta)\mathbb{I} \otimes \sigma_y)] \\
W_4^5 = W_{3,x,\mathbb{I}}^3 = \frac{1}{4}[1 &+ \sigma_x \otimes \sigma_x + \cos(2\theta)(\cos(\alpha)(\sigma_z \otimes \sigma_z + \sigma_y \otimes \sigma_y) \\
&- \sin(\alpha)(\sigma_y \otimes \sigma_z - \sigma_z \otimes \sigma_y)) + \sin(2\theta)(\sigma_x \otimes \mathbb{I} + \mathbb{I} \otimes \sigma_x)] \\
W_5^5 = W_{4,x,\mathbb{I}}^3 = \frac{1}{4}[1 &- \sigma_x \otimes \sigma_x + \cos(2\theta)(\cos(\alpha)(\sigma_z \otimes \sigma_z - \sigma_y \otimes \sigma_y) \\
&- \sin(\alpha)(\sigma_y \otimes \sigma_z + \sigma_z \otimes \sigma_y)) - \sin(2\theta)(\sigma_x \otimes \mathbb{I} - \mathbb{I} \otimes \sigma_x)] \\
W_6^5 = W_{1,x,x}^3 = \frac{1}{4}[1 &+ \cos(\alpha)\cos(\beta)\sigma_z \otimes \sigma_z + \sin(\alpha)\sin(\beta)\sigma_y \otimes \sigma_y \\
&- \sin(\alpha)\cos(\beta)\sigma_y \otimes \sigma_z - \cos(\alpha)\sin(\beta)\sigma_z \otimes \sigma_y + \cos(2\theta)(\sigma_x \otimes \sigma_x \\
&+ \cos(\alpha)\cos(\beta)\sigma_y \otimes \sigma_y + \sin(\alpha)\cos(\beta)\sigma_z \otimes \sigma_y + \cos(\alpha)\sin(\beta)\sigma_y \otimes \sigma_z \\
&+ \sin(\alpha)\sin(\beta)\sigma_z \otimes \sigma_z) + \sin(2\theta)(\cos(\alpha)\sigma_z \otimes \mathbb{I} \\
&- \sin(\alpha)\sigma_y \otimes \mathbb{I} + \cos(\beta)\mathbb{I} \otimes \sigma_z - \sin(\beta)\mathbb{I} \otimes \sigma_y)] \\
W_7^5 = W_{5,y,\mathbb{I}}^3 = \frac{1}{4}[1 &+ \sigma_y \otimes \sigma_y + \cos(2\theta)(\cos(\alpha)(\sigma_z \otimes \sigma_z + \sigma_x \otimes \sigma_x) \\
&+ \sin(\alpha)(\sigma_x \otimes \sigma_z - \sigma_z \otimes \sigma_x)) + \sin(2\theta)(\sigma_y \otimes \mathbb{I} + \mathbb{I} \otimes \sigma_y)] \\
W_8^5 = W_{6,y,\mathbb{I}}^3 = \frac{1}{4}[1 &- \sigma_y \otimes \sigma_y + \cos(2\theta)(\cos(\alpha)(\sigma_z \otimes \sigma_z - \sigma_x \otimes \sigma_x) \\
&+ \sin(\alpha)(\sigma_x \otimes \sigma_z + \sigma_z \otimes \sigma_x)) - \sin(2\theta)(\sigma_y \otimes \mathbb{I} - \mathbb{I} \otimes \sigma_y)] \\
W_9^5 = W_{1,y,y}^3 = \frac{1}{4}[1 &+ \cos(\alpha)\cos(\beta)\sigma_z \otimes \sigma_z + \sin(\alpha)\sin(\beta)\sigma_x \otimes \sigma_x \\
&+ \sin(\alpha)\cos(\beta)\sigma_x \otimes \sigma_z + \cos(\alpha)\sin(\beta)\sigma_z \otimes \sigma_x \\
&+ \cos(2\theta)(\sigma_y \otimes \sigma_y + \cos(\alpha)\cos(\beta)\sigma_x \otimes \sigma_x - \sin(\alpha)\cos(\beta)\sigma_z \otimes \sigma_x \\
&- \cos(\alpha)\sin(\beta)\sigma_x \otimes \sigma_z + \sin(\alpha)\sin(\beta)\sigma_z \otimes \sigma_z) + \sin(2\theta)(\cos(\alpha)\sigma_z \otimes \mathbb{I} \\
&+ \sin(\alpha)\sigma_x \otimes \mathbb{I} + \cos(\beta)\mathbb{I} \otimes \sigma_z + \sin(\beta)\mathbb{I} \otimes \sigma_x)], \qquad (3.1.13)
\end{aligned}
$$

with $\theta \in [0, \frac{\pi}{2}]$, $\alpha \in [0, 2\pi]$, and $\beta \in [0, 2\pi]$. However, we may also derive each $\{W^5\}$ witness by applying a local rotation matrix to either or both particles associated with the $\{W^3\}$ [18]; in Eq. 3.1.13, the notation $W^3_{1,z,\mathbb{I}}$ means to rotate particle 1 about the $\hat{z}$ axis by an angle α:

$$W^3_{1,z,\mathbb{I}} = (\hat{R}_z(\alpha) \otimes \mathbb{I}) W^3_1 ((\hat{R}^\dagger_z(\alpha) \otimes \mathbb{I}). \tag{3.1.14}$$

For reference, these rotations follow the rules:

$$\begin{aligned}
\sigma_x \text{ about } z: &\quad \cos(\alpha)\sigma_x + \sin(\alpha)\sigma_y. \\
\sigma_y \text{ about } z: &\quad \cos(\alpha)\sigma_y - \sin(\alpha)\sigma_x. \\
\sigma_x \text{ about } y: &\quad \cos(\alpha)\sigma_x - \sin(\alpha)\sigma_z. \\
\sigma_z \text{ about } y: &\quad \cos(\alpha)\sigma_z + \sin(\alpha)\sigma_x. \\
\sigma_y \text{ about } x: &\quad \cos(\alpha)\sigma_y + \sin(\alpha)\sigma_z. \\
\sigma_z \text{ about } x: &\quad \cos(\alpha)\sigma_z - \sin(\alpha)\sigma_y.
\end{aligned} \tag{3.1.15}$$

The first two witnesses in each triplet admit an analytic minimization:

$$\begin{aligned}
\min W^5_1 &= \frac{1}{4}[1 + S_{zz} - \sqrt{(S_{xx} + S_{yy})^2 + (S_{z0} + S_{0z})^2 + (S_{xy} - S_{yx})^2}] \\
\min W^5_2 &= \frac{1}{4}[1 - S_{zz} - \sqrt{(S_{xx} - S_{yy})^2 + (S_{z0} - S_{0z})^2 + (S_{xy} + S_{yx})^2}] \\
\min W^5_4 &= \frac{1}{4}[1 + S_{xx} - \sqrt{(S_{zz} + S_{yy})^2 + (S_{0x} + S_{x0})^2 + (S_{yz} - S_{zy})^2}] \\
\min W^5_5 &= \frac{1}{4}[1 - S_{xx} - \sqrt{(S_{zz} - S_{yy})^2 + (S_{0x} - S_{x0})^2 + (S_{yz} + S_{zy})^2}] \\
\min W^5_7 &= \frac{1}{4}[1 + S_{yy} - \sqrt{(S_{zz} + S_{xx})^2 + (S_{y0} + S_{0y})^2 + (S_{zx} - S_{xz})^2}] \\
\min W^5_8 &= \frac{1}{4}[1 - S_{yy} - \sqrt{(S_{zz} - S_{xx})^2 + (S_{y0} - S_{0y})^2 + (S_{zx} + S_{xz})^2}].
\end{aligned} \tag{3.1.16}$$

I derive this in two ways and will demonstrate both with W^5_1. The first method begins with the general expression:

$$\begin{aligned}
\min_{\alpha,\theta} \langle W^5_1 \rangle = \min_{\alpha,\theta} \frac{1}{4} &[1 + \langle \sigma_z \otimes \sigma_z \rangle + \cos(2\theta)(\cos(\alpha)(\langle \sigma_x \otimes \sigma_x \rangle + \langle \sigma_y \otimes \sigma_y \rangle) \\
&- \sin(\alpha)(\langle \sigma_x \otimes \sigma_y \rangle - \langle \sigma_y \otimes \sigma_x \rangle)) + \sin(2\theta)(\langle \sigma_z \otimes \mathbb{I} \rangle + \langle \mathbb{I} \otimes \sigma_z \rangle)] \\
= \min_{\alpha,\theta} \frac{1}{4} &\left[1 + S_{zz} + \cos(2\theta)(\cos(\alpha)(S_{xx} + S_{yy}) - \sin(\alpha)(S_{xy} - S_{yx}) + \sin(2\theta)(S_{z0} + S_{0z})\right].
\end{aligned} \tag{3.1.17}$$

Let

$$\begin{aligned}
a &= 1 + S_{zz}, \\
b &= S_{xx} + S_{yy}, \\
c &= S_{z0} + S_{0z}, \\
d &= S_{xy} - S_{yx}.
\end{aligned} \tag{3.1.18}$$

Then, this expression becomes

$$\min_{\alpha,\theta}\langle W_1^5\rangle = \min_{\alpha,\theta}\frac{1}{4}[a+\cos(2\theta)(b\cos(\alpha)-d\sin(\alpha))+c\sin(2\theta)]. \tag{3.1.19}$$

Recall the identity

$$c\cos\alpha \pm d\sin\alpha = \sqrt{c^2+d^2}\cos(\alpha \mp x), \tag{3.1.20}$$

from Eq. 3.1.2 which has a minimum value of $-\sqrt{c^2+d^2}$.

Complete the α minimization:

$$\min_{\alpha,\theta}\langle W_1^5\rangle = \min_{\theta}\frac{1}{4}[a+\cos(2\theta)(-\sqrt{b^2+d^2})+c\sin(2\theta)] \tag{3.1.21}$$

I play the same trick but with the 2θ terms to kill the θ minimization:

$$\begin{aligned}\langle W_1^5\rangle_{\min} &= \frac{1}{4}[a-\sqrt{b^2+c^2+d^2}] \\ &= \frac{1}{4}[1+S_{zz}-\sqrt{(S_{xx}+S_{yy})^2+(S_{z0}+S_{0z})^2+(S_{xy}-S_{yx})^2}].\end{aligned} \tag{3.1.22}$$

The other expressions in Eq. 3.1.16 follow from this method. I can also derive these equations by applying the necessary rotation to the minimized $\{W^3\}$ expressions in Eq. 3.1.9, then minimizing over the angles brought by the rotation:

$$\langle W_1^5\rangle_{\min} = \min_{\alpha}\langle(\hat{R}_z(\alpha)\otimes\mathbb{I})W_1^3(\hat{R}_z^\dagger(\alpha)\otimes\mathbb{I})\rangle \tag{3.1.23}$$

$$= \min_{\alpha}\langle(\hat{R}_z(\alpha)\otimes\mathbb{I})\frac{1}{4}[1+\sigma_z\otimes\sigma_z-\sqrt{(\sigma_x\otimes\sigma_x+\sigma_y\otimes\sigma_y)^2+(\sigma_z\otimes\mathbb{I}+\mathbb{I}\otimes\sigma_z)^2}](\hat{R}_z^\dagger(\alpha)\otimes\mathbb{I})\rangle \tag{3.1.24}$$

$$= \min_{\alpha}\frac{1}{4}[1+S_{zz}-\sqrt{(S_{xx}\cos\alpha+S_{yx}\sin\alpha+S_{yy}\cos\alpha-S_{xy}\sin\alpha)^2+(S_{z0}+S_{0z})^2}] \tag{3.1.25}$$

$$= \min_{\alpha}\frac{1}{4}[1+S_{zz}-\sqrt{((S_{xx}+S_{yy})\cos\alpha+(S_{yx}-S_{xy})\sin\alpha)^2+(S_{z0}+S_{0z})^2}] \tag{3.1.26}$$

I then minimize over α using the same identity as before in Eq. 3.1.20, yielding

$$\langle W_1^5\rangle_{\min} = \frac{1}{4}[1+S_{zz}-\sqrt{\left(\sqrt{(S_{xx}+S_{yy})^2+(S_{xy}-S_{yx})^2}\right)^2+(S_{z0}+S_{0z})^2}] \tag{3.1.27}$$

$$= \frac{1}{4}[1+S_{zz}-\sqrt{(S_{xx}+S_{yy})^2+(S_{xy}-S_{yx})^2+(S_{z0}+S_{0z})^2}], \tag{3.1.28}$$

which matches Eq. 3.1.16.

I also simplify the expressions for the third witness in each triplet—W_3^5, W_6^5, and W_9^5—by analytically completing the θ minimization. Begin with W_1^5 in Eq. 3.1.13

and resolve the θ minimization via the identity given in Eq. 3.1.20; each witness then reads

$$
\begin{aligned}
W_3^5 = \frac{1}{4}[&1 + \cos(\alpha)\cos(\beta)\sigma_x \otimes \sigma_x + \sin(\alpha)\sin(\beta)\sigma_y \otimes \sigma_y \\
&+ \sin(\alpha)\cos(\beta)\sigma_y \otimes \sigma_x + \cos(\alpha)\sin(\beta)\sigma_x \otimes \sigma_y \\
&- \Big[(\sigma_z \otimes \sigma_z + \cos(\alpha)\cos(\beta)\sigma_y \otimes \sigma_y + \sin(\alpha)\sin(\beta)\sigma_x \otimes \sigma_x \\
&\quad - \sin(\alpha)\cos(\beta)\sigma_x \otimes \sigma_y - \cos(\alpha)\sin(\beta)\sigma_y \otimes \sigma_x)^2 \\
&\quad + \big(\cos(\alpha)\sigma_x \otimes \mathbb{I} + \sin(\alpha)\sigma_y \otimes \mathbb{I} + \cos(\beta)\mathbb{I} \otimes \sigma_x + \sin(\beta)\mathbb{I} \otimes \sigma_y\big)^2\Big]^{\frac{1}{2}}] \\
W_6^5 = \frac{1}{4}[&1 + \cos(\alpha)\cos(\beta)\sigma_z \otimes \sigma_z + \sin(\alpha)\sin(\beta)\sigma_y \otimes \sigma_y \\
&- \sin(\alpha)\cos(\beta)\sigma_y \otimes \sigma_z - \cos(\alpha)\sin(\beta)\sigma_z \otimes \sigma_y \\
&- \Big[(\sigma_x \otimes \sigma_x + \cos(\alpha)\cos(\beta)\sigma_y \otimes \sigma_y + \sin(\alpha)\cos(\beta)\sigma_z \otimes \sigma_y \\
&\quad + \cos(\alpha)\sin(\beta)\sigma_y \otimes \sigma_z + \sin(\alpha)\sin(\beta)\sigma_z \otimes \sigma_z)^2 \\
&\quad + \big(\cos(\alpha)\sigma_z \otimes \mathbb{I} - \sin(\alpha)\sigma_y \otimes \mathbb{I} + \cos(\beta)\mathbb{I} \otimes \sigma_z - \sin(\beta)\mathbb{I} \otimes \sigma_y\big)^2\Big]^{\frac{1}{2}}] \\
W_9^5 = \frac{1}{4}[&1 + \cos(\alpha)\cos(\beta)\sigma_z \otimes \sigma_z + \sin(\alpha)\sin(\beta)\sigma_x \otimes \sigma_x \\
&+ \sin(\alpha)\cos(\beta)\sigma_x \otimes \sigma_z + \cos(\alpha)\sin(\beta)\sigma_z \otimes \sigma_x \\
&- \Big[(\sigma_y \otimes \sigma_y + \cos(\alpha)\cos(\beta)\sigma_x \otimes \sigma_x - \sin(\alpha)\cos(\beta)\sigma_z \otimes \sigma_x \\
&\quad - \cos(\alpha)\sin(\beta)\sigma_x \otimes \sigma_z + \sin(\alpha)\sin(\beta)\sigma_z \otimes \sigma_z)^2 \\
&\quad + (\cos(\alpha)\sigma_z \otimes \mathbb{I} + \sin(\alpha)\sigma_x \otimes \mathbb{I} + \cos(\beta)\mathbb{I} \otimes \sigma_z + \sin(\beta)\mathbb{I} \otimes \sigma_x)^2\Big]^{\frac{1}{2}}]
\end{aligned}
\tag{3.1.29, 3.1.30}
$$

I identically reach Eq. 3.1.29 by rotating the corresponding $\{W^3\}$ witnesses from Eq. 3.1.9 as described in detail in Paco Navarro's thesis [18]. However, I am unable to eliminate the α or β parameters—even if I attempt to split the two-particle rotation then minimization step into two separate one-particle rotations with two minimizations—because of the presence of rotated terms outside of each square root in the minimized $\{W^3\}$ expressions. In Sec. 5.2, I describe what these three witnesses test geometrically and explain why having two parameters is expected.

When comparing the three-parameter and two-parameter versions of W_3^5 (Eq. 3.1.13 vs. Eq. 3.1.29) I find a $4.7\times \pm 0.8\times$ speed up in minimizing a computationally generated random states, important for efficiency in future applications to quantum communication protocols and for our group computing witnesses on large data sets. Likewise, I find a $3913\times \pm 999\times$ speedup between the three-parameter and analytically minimized versions of W_2^5 (Eq.3.1.13 vs Eq.3.1.16).

3.2 Two Step Adaptive Witnessing Protocol

"E.W. detect entanglement" - E.W.: The Extranglement-Witerrestrial

As a reminder, entanglement witnesses are typically used to detect the entanglement of states that the observer has prior knowledge of [27]. An example of this would be when an experimentalist has a state they expect they created, but which they need to confirm the entanglement of. This thesis focuses on the scenario in which no prior knowledge is known of the state. I now introduce a two-step adaptive witnessing protocol that uses the $\{W^3\}$ and $\{W^5\}$ witness sets to iteratively identify the entanglement of unknown bipartite states from few measurements, this protocol having been developed by previous Quantum Optics Group members. This protocol allows users to take the minimum number of measurements required to detect a quantum state ρ, assuming that ρ is witnessed by at least one witness within $\{W^3\}$ or $\{W^5\}$.

The first step is to determine the expectation values of all $\{W^3\}$ witnesses. This corresponds to measuring in the $\sigma_x \otimes \sigma_x, \sigma_y \otimes \sigma_y$, and $\sigma_z \otimes \sigma_z$ bases. If any of the W^3 witnesses have a negative expectation value, ρ has been witnessed as entangled, and the protocol ends.

If no W^3 witnesses are negative, we proceed to measure **one** of the three W^5 triplets as the second step of this protocol. Each triplet requires the same measurement bases as the W^3 witnesses, with an additional basis set $\sigma_i \otimes \sigma_j, \sigma_j \otimes \sigma_i$, $i, j \in [x, y, z]$ and $i \neq j$. Before this measurement we must make a choice of which of the three triplets to measure. This choice is adaptive, meaning it uses the measurement results from W^3 to inform its decision. We have two choice models, discussed and compared in Sec 3.3. After measuring in our chosen additional bases, we check whether any of the chosen triplet's expectation values are negative—if any are, we have successfully detected entanglement. If not, the state is either separable, witnessed by a different W^5 triplet (i.e. we made the wrong choice), or are undetected by any W^3 or W^5. This protocol is visualized in Fig. 3.2. By testing across a data set of 16,831,944 computationally generated random states via the method described in Roik et al. [23], I find that 33% of these states are witnessed by the $\{W^3\}$s, and 79% by the $\{W^5\}$s if the perfect choice is made.

3.3 Choice Models

The performance of this protocol is dependent on the success rate, or accuracy, of the choice model used. In this thesis, I define the accuracy of a choice model as the number of entangled states that this choice model makes the correct choice for divided by the number of entangled states in the data set. In hindsight, it would have been more pertinent to define the success rate as the number of entangled states that are witnessed by a $\{W^3\}$ witness *or* that the choice

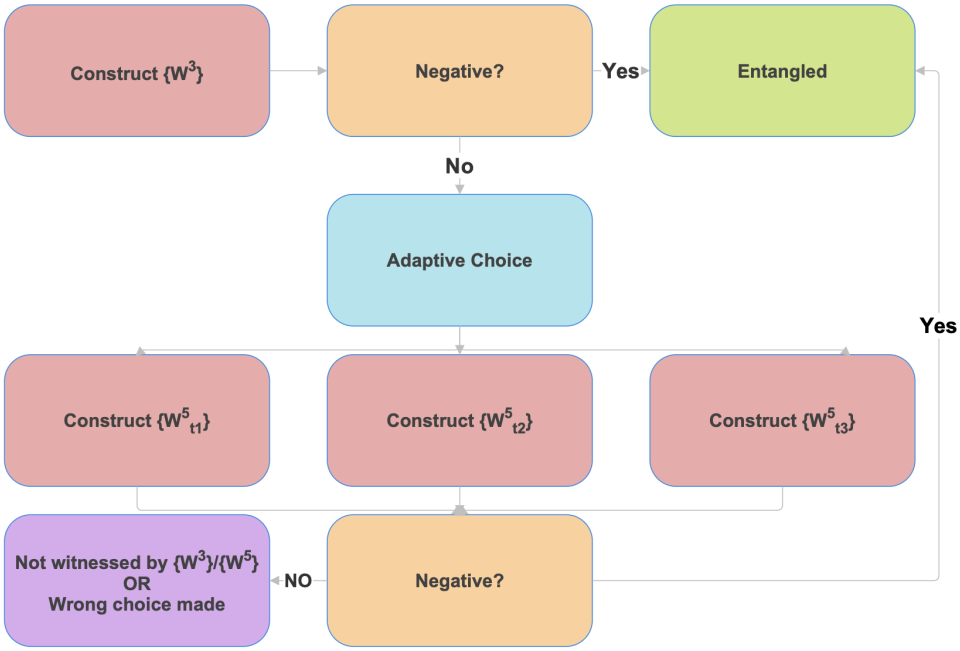


Figure 3.2: Flow chart of the adaptive witnessing protocol. The witness sets are given in Sec. 3.1 and the choice methods in Sec. 3.3.

model chooses correctly for, divided by all entangled states. However, I found that the choice model success rates differ by less than 0.1% between the two definitions, which indicates that our choice methods are already doing a good job of choosing correctly for states witnessed by a $\{W^3\}$.

We define the accuracy of a choice model as the number of states witnessed by the two-step protocol while using this choice model, divided by the number of states witnessed if we make the ideal choice every time. The model that makes the ideal choice every time is called the W^5 oracle, as it has knowledge of which, if any, of the W^5 triplets witness any state ρ.

3.3.1 Population Model

The population model is an analytic choice model that chooses to measure the measurement bases orthogonal to the W^3 basis along which the maximum information resides. It does so by calculating the magnitude of three W^3 Stokes parameters $|\langle \sigma_i \otimes \sigma_i \rangle| = |S_{ii}|$, $i \in [x, y, z]$. The maximum magnitude represents the axis on which we have the most information about the state. The population method then chooses to gain information about the axes orthogonal to $\sigma_i \otimes \sigma_i$. See Table 3.1 for the specific choice mapping.

Greatest $\lvert S_{ii}\rvert$, $i \in [x, y, z]$	$\sigma_i \otimes \sigma_j$ pair
$\lvert S_{zz}\rvert$	$\sigma_x \otimes \sigma_y$
$\lvert S_{xx}\rvert$	$\sigma_y \otimes \sigma_z$
$\lvert S_{yy}\rvert$	$\sigma_x \otimes \sigma_z$

Table 3.1: Population model; the maximum value of the three populations corresponding to which $\{W^5\}$ triplet is chosen.

Observe that the positive portions of the first two witnesses in each triplet are dependent on $1 \pm S_{ii}$, $i \in [x, y, z]$. By choosing the largest $|S_{ii}|$ we are guaranteed to minimize the positive portion of one of these two witnesses, therefore making it more likely that it has an overall negative expectation value. As an example, see the first two witnesses in $\{W^5_{t1}\}$:

$$\begin{aligned} \min W_1^5 &= \frac{1}{4}[1 + S_{zz} - \sqrt{(S_{xx} + S_{yy})^2 + (S_{z0} + S_{0z})^2 + (S_{xy} - S_{yx})^2}] \\ \min W_2^5 &= \frac{1}{4}[1 - S_{zz} - \sqrt{(S_{xx} - S_{yy})^2 + (S_{z0} - S_{0z})^2 + (S_{xy} + S_{yx})^2}] \end{aligned} \tag{3.3.1}$$

If $|S_{zz}|$ is the largest of $|S_{ii}|$, $i \in [x, y, z]$, then one of the two witnesses in Eq. 3.3.1 has a minimal positive portion and is therefore more likely to be negative. The population model was originally derived heuristically by a previous group member [30]; my work was to give the mathematical interpretation presented in this section.

3.3.2 Five Layer Neural Network

The five layer neural network is a computational choice model trained using WandB [4], which is a platform for optimizing and tracking machine learning models. It takes in the nine of the projective probability measurements obtained when calculating the $\{W^3\}$ witness set

$$HH, HV, VV, DD, DA, DD, RR, RL, LL, \tag{3.3.2}$$

with projective probabilities discussed briefly in Sec. 2.1.5. The model returns a goal vector $\hat{y} = [W^5_{t1}, W^5_{t2}, W^5_{t3}]$, which is a multi-hot encoded vector with each index being assigned value 1 if $W^5_{ti} < 0$, or 0 if not. Hyperparameter sweeps with a 10^{-3} learning rate and batch size of 2000 resulted in the ideal model containing 5 hidden layers of 200 nodes each, ReLu activation functions for all hidden layers, and a Sigmoid activation function for the final output layer. This model was conceived of and trained by previous group members [25].

3.3.3 Cross Correlation Model

The cross correlation model is an abstraction of the population model. From the analytic minimizations in Eq. 3.1.16, I set a bound on what Stokes parameters a state must have to be detected by a given witness:

$$\begin{aligned}
\langle W^5_1\rangle < 0 &\iff (1+S_{zz})^2 < (S_{xy}-S_{yx})^2 + (S_{xx}+S_{yy})^2 + (S_{z0}+S_{0z})^2,\\
\langle W^5_2\rangle < 0 &\iff (1-S_{zz})^2 < (S_{xy}+S_{yx})^2 + (S_{xx}-S_{yy})^2 + (S_{z0}-S_{0z})^2,\\
\langle W^5_4\rangle < 0 &\iff (1+S_{xx})^2 < (S_{yz}-S_{zy})^2 + (S_{zz}+S_{yy})^2 + (S_{x0}+S_{0x})^2,\\
\langle W^5_5\rangle < 0 &\iff (1-S_{xx})^2 < (S_{yz}+S_{zy})^2 + (S_{zz}-S_{yy})^2 + (S_{0x}-S_{x0})^2,\\
\langle W^5_7\rangle < 0 &\iff (1+S_{yy})^2 < (S_{zx}-S_{xz})^2 + (S_{zz}+S_{xx})^2 + (S_{y0}+S_{0y})^2,\\
\langle W^5_8\rangle < 0 &\iff (1-S_{yy})^2 < (S_{zx}+S_{xz})^2 + (S_{zz}-S_{xx})^2 + (S_{y0}-S_{0y})^2.
\end{aligned} \tag{3.3.3}$$

Then, collect terms we know from measuring the $\{W^3\}$ witness set on one side and the unknown cross-correlation term on the other:

$$\begin{aligned}
\langle W^5_1\rangle < 0 &\iff (1+S_{zz})^2 - [(S_{xx}+S_{yy})^2 + (S_{z0}+S_{0z})^2] < (S_{xy}-S_{yx})^2,\\
\langle W^5_2\rangle < 0 &\iff (1-S_{zz})^2 - [(S_{xx}-S_{yy})^2 + (S_{z0}-S_{0z})^2] < (S_{xy}+S_{yx})^2,\\
\langle W^5_4\rangle < 0 &\iff (1+S_{xx})^2 - [(S_{zz}+S_{yy})^2 + (S_{x0}+S_{0x})^2] < (S_{yz}-S_{zy})^2,\\
\langle W^5_5\rangle < 0 &\iff (1-S_{xx})^2 - [(S_{zz}-S_{yy})^2 + (S_{0x}-S_{x0})^2] < (S_{yz}+S_{zy})^2,\\
\langle W^5_7\rangle < 0 &\iff (1+S_{yy})^2 - [(S_{zz}+S_{xx})^2 + (S_{y0}+S_{0y})^2] < (S_{zx}-S_{xz})^2,\\
\langle W^5_8\rangle < 0 &\iff (1-S_{yy})^2 - [(S_{zz}-S_{xx})^2 + (S_{y0}-S_{0y})^2] < (S_{zx}+S_{xz})^2.
\end{aligned} \tag{3.3.4}$$

From Eq. 3.3.4, I define six quantities, $\Delta_i(X)$, as the left hand side's quantity for each $\langle W^5_i\rangle$. X is the set of 9 Stokes parameters necessary to measure $\{W^3\}$. The

cross correlation model seeks to minimize these terms, as the triplet is then more likely to witness a given state because it needs a smaller exchange symmetric cross correlation term (of the form $(S_{ij} \pm S_{ji})^2$) to satisfy one of the inequalities.

3.3.4 A k-dimensional Tree

In Chapter 4, I present a k-dimensional tree approach to lower bound the protocol's success rate. I also use the k-d tree as a choice method—this means that I train, or develop a set of rules to place states into bins, using a large training set taken out of 16,831,944 computationally generated random states. Then, for each state that was not used for training, I place it into a bin using the rules from training and check whether the witness triplet that detects it is the same as the witness triplet that detects the majority of the training states in the bin. The success probability is then the number of test states for which their witness triplet matches their bin's, divided by the number of test states.

3.3.5 Choice Model Comparison

> "Is that a bird? Is that a plane?? No! It's a gap between the success rates of the W^5 oracle and neural network!" - An observer of Super-operator-man

I evaluate the performance of the W^3 witnesses, W^5 witnesses with perfect choice (W^5 oracle), population method, cross-correlation method, k-dimensional tree model, and neural network in Fig. 3.3 as a function of concurrence, which is a monotone measure of a state's degree of entanglement that is 0 for all separable states and 1 for maximally entangled states [11]. More specifically, I use a concurrence threshold, so evaluate success rates on all states with concurrence greater than or equal to some $c \in [0,1]$. I evaluate success rates using a computationally generated set of 16,831,944 random entangled states that we generated using the method from Roik et al. [23]. For models that require training—the neural network and k-dimensional tree—I use 5-fold cross validation. This means that I split the data set into 5 subsets, train on four of them, test on the remaining one, and then repeat the splitting five times to form an average. From Fig. 3.3, we immediately observe that the neural network chooses the incorrect $\{W^5\}$ triplet for 7% of states. The first question we can ask is whether the optimal choice model can saturate the 79% bound set by the W^5 oracle, or if there is instead some information-theoretic bound based on the limited measurements being provided to the choice model. I approach this question in the following chapter. A second question is why the population model is more successful than the cross correlation model if the cross correlation model uses more information about the state than the population model. The short answer is that because the

cross correlation model is based on the first two witnesses in each triplet, it mischaracterizes the third witness in each triple easily. The long answer is provided in Sec. 5.2.1, after I present a geometric interpretation of the model necessary to understand my explanation. This also answers why the k-d tree model does worse than the neural network.

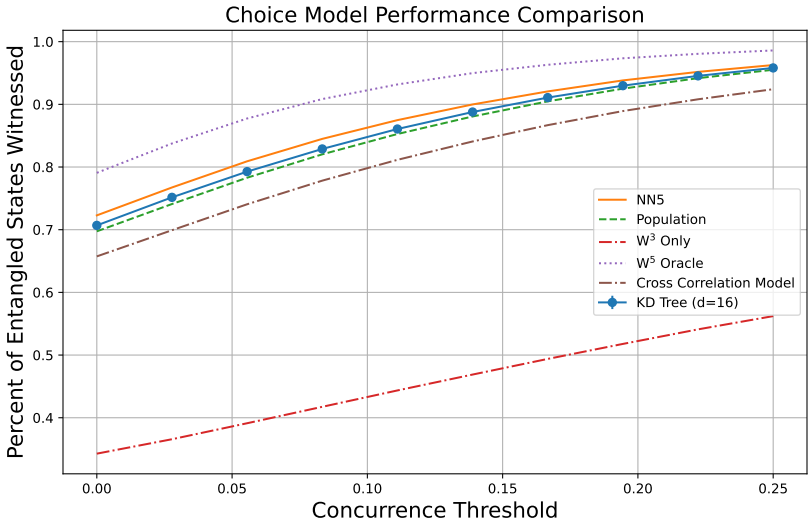


Figure 3.3: Protocol success rate for different choice models and witness sets for all states with concurrence greater than or equal to some value on the x-axis. See Sec. 3.3 for a description of each choice model.

Chapter 4

Bounding the Protocol's Success Rate

The goal of this thesis is to derive a bound on the success rate of the two step adaptive entanglement witnessing protocol. First, I show that there exists an information-theoretic upper bound on the protocol's success rate by deriving two states that have identical $\{W^3\}$-measured Stokes parameters but are detected by two different $\{W^5\}$ witness triplets. I then estimate this bound on the large data set of 16,831,944 random computationally generated entangled states by estimating the percent of entangled states that have the same $\{W^3\}$-measured Stokes parameters but are detected by two different $\{W^5\}$ witness triplets. I then estimate a separate upper bound by directly estimating the success rate of the Bayes classifier.

4.1 Bayes Classifier Choice Method

"Behold!!! My optimal-choice-makerinator!!!" - Dr. Qu-finshmirtz.

In this section, I derive an optimal adaptive choice model. Let ρ be a random bipartite density matrix, i.e. a random variable, drawn from a distribution $\pi(\rho)$. Define

$$G : \{M^3\} \to \{W^5_{t1}, W^5_{t2}, W^5_{t3}\} \tag{4.1.1}$$

to be a classifier that takes in the nine Stokes parameters calculated for W^3 measurement, which I denote $\{M^3\}$, and returns a classification choice of which W^5_i triple to measure next. I represent these measurement outcomes as a random vector $x \in \mathbb{R}^9$; note that the values of x are constrained such that they have a non-trivial distribution which I denote f_{stokes}. Importantly, there is no clean closed-form expression for f_{stokes} that I found in the literature or was able to

derive. Define the event that the i-th $\{W^5\}$ witness triplet, $\{W^5_{ti}\}$, contains a negative witness for the random variable ρ as

$$E_i(\rho) \equiv \{\min_j \mathrm{Tr}(\rho W^5_j) < 0 \mid W^5_j \in \{W^5_{ti}\}\}. \tag{4.1.2}$$

Then, the probability of this event occurring given the random outcome vector x is written as

$$g_i(x) = \Pr(E_i = 1|X = x) = \mathbb{E}(E_i|X = x), \tag{4.1.3}$$

where $\mathbb{E}$ is the expectation value. The best possible hypothetical classification model—as in the one that minimizes its error—is the Bayes classifier [7] which is defined by

$$G_{\text{bayes}}(x) = \text{argmax}_{i\in\{1,2,3\}}\, g_i(x). \tag{4.1.4}$$

This model chooses the witness triplet with the maximum probability of witnessing a state ρ based on its W^3 measurement outcome vector x. It is optimal because it knows the full probability distribution relationship between any quantum state and $\{W^5\}$, without having knowledge of which witness will detect that state. This is a hypothetical choice model in that it represents the highest success rate any actual choice model could have without telling us exactly how to make that choice.

Consider the probability that the event E_i occurs with i chosen so that $\mathrm{P}(E_i = 1|X = x)$ is maximized for a Stokes vector x,

$$g_{G_{\text{bayes}}(X)}(X). \tag{4.1.5}$$

This is the highest probability that one of the witness triplets detects the state associated with x, which is equivalent to the probability that the choice model that chooses this highest-probability triplet succeeds for that state. Averaging over all states and their associated x vectors gives the average value (expectation value) of this highest probability, and therefore gives the success rate of this hypothetical optimal choice method,

$$\begin{aligned} P_{\text{Bayes}} &= \mathbb{E}[g_{G_{\text{bayes}}(X)}(X)], && (4.1.6)\\ &= \mathbb{E}[\max_i g_i(X)], && (4.1.7)\\ &= \int \max_{i\in\{1,2,3\}} g_i(x) f_{\text{stokes}}(x)\, dx, && (4.1.8) \end{aligned}$$

where $f(x)$ is the distribution of the $\{W^3\}$ Stokes parameters in $\mathbb{R}^9$. Of note, $\mathbb{E}[G_{\text{bayes}}(x)] < 0.791$ as we are bounded above by the $\{W^5\}$ oracle. This is in turn equivalent to the success rate of the protocol itself if using this choice method. This is because whether a state is witnessed by a $\{W^3\}$ witness is fully encoded in x already, so $g_{G_{\text{bayes}}(X)}(X) = 1$ if this is so. P_{Bayes} as written is an **upper bound** on how well the protocol could perform with *any* choice model because it is the probability of success for the hypothetical best possible choice model.

Recognize that the distribution $f_{\text{stokes}}(x)$ of $\{W^3\}$-measured Stokes parameters, $\{M^3\} \in \mathbb{R}^9$, in Eq. 4.1.8 is dependent on the distribution from which we draw our quantum states. However, there is no single distribution of mixed quantum states agreed upon in the literature [32], and there are therefore multiple methods by which to generate random mixed quantum states. The goal of this thesis is to bound the performance of the entanglement witnessing protocol, which means that we must choose the distribution $f(x)$ to match the entangled states we use to characterize each choice model's performance in Fig. 3.3. As such, it makes the most sense to derive a bounding method that determines a bound on data generated by the same method as that used to create Fig. 3.3.

4.2 An Information-Theoretic Upper Bound

Consider two states, ρ_1 and ρ_2, that have the same 9 Stokes parameters measured to calculate the $\{W^3\}$ witness set in step one of our adaptive witnessing protocol. Imagine that these two states have different cross correlation Stokes parameters of the form S_{ij} and S_{ji}, $i, j \in x, y, z$, and are witnessed by different $\{W^5\}$ triplets. Then, any conceivable adaptive choice method would fail to choose the correct $\{W^5\}$ witness triplet for one of these two states. This implies that for some states, the 9 Stokes parameters provided to the adaptive choice method do not contain sufficient information for a perfect choice. In this section, I begin by deriving two states for which this scenario is true. Then, I estimate an upper bound on any adaptive choice method's success rate using the percent of states in our data set of computationally random for which this scenario applies.

4.2.1 States no Adaptive Choice Model can guess Correctly

Here, I construct two states that have identical $\{W^3\}$-necessary Stokes parameters but which are witnessed by two different triplets. In Sec. 3.3.3, I derived that W_1^5 and W_4^5—which are in triplets 1 and 2 respectively—detect a given quantum state if the following inequalities hold true with respect to that state's Stokes parameters:

$$\begin{aligned}
\langle W_1^5 \rangle < 0 &\iff (1+S_{zz})^2 - [(S_{xx}+S_{yy})^2 + (S_{z0}+S_{0z})^2] < (S_{xy}-S_{yx})^2, \\
\langle W_4^5 \rangle < 0 &\iff (1+S_{xx})^2 - [(S_{zz}+S_{yy})^2 + (S_{x0}+S_{0x})^2] < (S_{yz}-S_{zy})^2.
\end{aligned} \tag{4.2.1}$$

To simplify this construction, I set $S_{i0} = S_{0i} = S_{yy} = 0$, $i \in x, y, z$, for both states. With this, the W_1^5 and W_4^5 witnessing conditions in Eq. 4.2.1 become

$$\begin{aligned}
\langle W_1^5 \rangle < 0 &\iff (1+S_{zz})^2 - (S_{xx})^2 < (S_{xy}-S_{yx})^2, \\
\langle W_4^5 \rangle < 0 &\iff (1+S_{xx})^2 - (S_{zz})^2 < (S_{yz}-S_{zy})^2.
\end{aligned} \tag{4.2.2}$$

Then, let

$$\begin{aligned} S_{xx} &= a, \\ S_{zz} &= b, \\ S_{xy} &= -S_{yx} = c, \\ S_{yz} &= -S_{zy} = d, \\ S_{xz} &= S_{zx} = 0, \end{aligned}$$

where $S_{xz} = S_{zx} = 0$ is chosen so that neither are witnessed by the third triplet. For no $\{W^3\}$ witness to detect these states, the left hand side of both inequalities in Eq. 4.2.2 need to be non-zero, but for the inequalities to hold, they must also be small. With this in mind, I choose

$$a = -\frac{2}{3},\ b = -\frac{1}{6}. \tag{4.2.3}$$

Eq. 4.2.2 then becomes

$$\begin{aligned} \langle W_1^5 \rangle < 0 &\iff \frac{1}{4} < 4c^2, \\ \langle W_4^5 \rangle < 0 &\iff \frac{1}{12} < 4d^2. \end{aligned} \tag{4.2.4}$$

Solving, I obtain

$$\begin{aligned} \langle W_1^5 \rangle < 0 &\iff \frac{1}{4} < c, \\ \langle W_4^5 \rangle < 0 &\iff \frac{1}{6} < d. \end{aligned} \tag{4.2.5}$$

Choose $c = d = \frac{1}{3}$ so that these inequalities hold true. Set $S_{yz} = S_{zy} = 0$ for ρ_1, which is the state I will derive that is only witnessed by the first triplet, and set $S_{xy} = S_{yz} = 0$ for ρ_2, the state witnessed by the second triplet. This is so that ρ_1 is not witnessed by the second triplet, and vice versa for ρ_2. This yields the states

$$\begin{aligned} \rho_1 &= \frac{1}{4}\left[\mathbb{I}\otimes\mathbb{I} - \frac{2}{3}\sigma_x\otimes\sigma_x - \frac{1}{6}\sigma_z\otimes\sigma_z + \frac{1}{3}\sigma_x\otimes\sigma_y - \frac{1}{3}\sigma_y\otimes\sigma_x\right] \\ &= \begin{pmatrix} \frac{5}{24} & 0 & 0 & -\frac{1}{6} \\ 0 & \frac{7}{24} & -\frac{1}{6}(1-i) & 0 \\ 0 & -\frac{1}{6}(i+1) & \frac{7}{24} & 0 \\ -\frac{1}{6} & 0 & 0 & \frac{5}{24} \end{pmatrix}, \end{aligned} \tag{4.2.6}$$

and

$$\begin{aligned} \rho_1 &= \frac{1}{4}\left[\mathbb{I}\otimes\mathbb{I} - \frac{2}{3}\sigma_x\otimes\sigma_x - \frac{1}{6}\sigma_z\otimes\sigma_z + \frac{1}{3}\sigma_y\otimes\sigma_z - \frac{1}{3}\sigma_z\otimes\sigma_y\right] \\ &= \begin{pmatrix} \frac{5}{24} & \frac{i}{12} & -\frac{i}{12} & -\frac{1}{6} \\ -\frac{i}{12} & \frac{7}{24} & -\frac{1}{6} & \frac{i}{12} \\ \frac{i}{12} & -\frac{1}{6} & \frac{7}{24} & -\frac{i}{12} \\ -\frac{1}{6} & -\frac{i}{12} & \frac{i}{12} & \frac{5}{24} \end{pmatrix}. \end{aligned} \tag{4.2.7}$$

These are trace 1, hermitian, semi-positive definite matrices and are thus valid density matrices. The first is witnessed by W_1^5 and W_3^5, so only by witnesses in the first triplet. The second is witnessed by W_4^5 and W_6^5, so only by witnesses in the second triplet. Thus, because they share the 9 measured $\{W^3\}$-necessary Stokes parameters, any conceivable adaptive choice model only given those 9 parameters would fail to detect one of them. I confirm this using the neural network choice model discussed in Sec. 3.3; I find that it predicts the second triplet, meaning that it guesses correctly for ρ_2 but fails for ρ_1.

4.2.2 A Bound from the Gap

I showed in Sec. 4.2.1 that there are some sets of entangled states for which the information provided by the 9 $\{W^3\}$-determined Stokes parameters is insufficient to correctly classify them. More specifically, there exists states that have the same 9 $\{W^3\}$-determined Stokes parameters but are nevertheless detected by different $\{W^5\}$ triplets. This implies a gap between the oracle and Bayes choice model. In this section, I estimate an upper bound on the classification success rate of any possible adaptive choice model provided these 9 parameters by estimating the number of states that the optimal choice model will always get wrong.

Let $\mathbf{I}_i(\rho_x)$ to be an indicator variable that is 1 when the i-th $\{W^5\}$ triplet detects the state ρ with 9 measured Stokes parameters x. Then, the oracle success rate is the average probability that a state with associated x is detected by any three triplet,

$$P_{\text{oracle}} = \mathbb{E}_x \left[P(\mathbf{I}_1(x) \cup \mathbf{I}_2(x) \cup \mathbf{I}_3(x)) | X = x) \right], \tag{4.2.8}$$

where the expectation value $\mathbb{E}$ is taken over all states. Let

$$g_i(x) = \Pr(\mathbf{I}_i(x) | X = x) \tag{4.2.9}$$

be the probability that the i-th witness detects a state given it has 9 measured Stokes parameters x. Here, I denote P to be a success rate averaged over all x, and Pr a probability for a single x. Then, if $G(x)$ is the triplet that any deterministic choice model given only the 9 measured Stokes parameters chooses, the success rate of the model is

$$P_G = \mathbb{E}_x[g_{G(x)}(x)]. \tag{4.2.10}$$

I have already shown that there necessarily exists a gap between the oracle success rate and the success rate of any 9 Stokes parameter-provided adaptive choice model. I then quantify this gap as

$$\begin{aligned} P_{\text{oracle}} - P_G &= \mathbb{E}_x \left[\Pr(\mathbf{I}_1(x) \cup \mathbf{I}_2(x) \cup \mathbf{I}_3(x)) | X = x) \right] - \mathbb{E}_x[g_{G(x)}(x)] \\ &= \mathbb{E}_x \left[\Pr(\mathbf{I}_1(x) \cup \mathbf{I}_2(x) \cup \mathbf{I}_3(x)) - g_{G(x)}(x)) | X = x \right] \\ &= \mathbb{E}_x \left[\Pr(A | X = x) - \Pr(B | X = x) \right], \end{aligned} \tag{4.2.11}$$

where $A = \mathbf{I}_1(x) \cup \mathbf{I}_2(x) \cup \mathbf{I}_3(x)$ and $B = \mathbf{I}_i(x)$. See that because $B \subseteq A$, A can be decomposed as $A = B \cup (A \cap \overline{B})$, where $\overline{B}$ is the complement of B. Conditioning on $X = x$ yields

$$\Pr(A|X=x) = \Pr(B|X=x) + \Pr\Big(A \cap \overline{B}|X=x\Big), \tag{4.2.12}$$

or

$$\Pr(A|X=x) - \Pr(B|X=x) = \Pr\Big(A \cap \overline{B}|X=x\Big). \tag{4.2.13}$$

Substituting this in to Eq. 4.2.11 gives

$$\begin{aligned} P_{\text{oracle}} - P_G &= \mathbb{E}_x\left[\Pr\Big(A \cap \overline{B}|X=x\Big)\right] \\ &= \mathbb{E}_x\left[\Pr\Big((\mathbf{I}_1(x) \cup \mathbf{I}_2(x) \cup \mathbf{I}_3(x)) \cap \overline{\mathbf{I}_{G(x)}(x)}\Big)\right]. \end{aligned} \tag{4.2.14}$$

$\Pr\Big((\mathbf{I}_1(x) \cup \mathbf{I}_2(x) \cup \mathbf{I}_3(x)) \cap \overline{\mathbf{I}_{G(x)}(x)}\Big)$ represents the probability that the state is detectable by one of the $\{W^5\}$ triplets but that the model $G(x)$ chose incorrectly. Its expectation value is then the average of this probability over all states, which is the gap between the oracle success rate and the success rate of the optimal choice model provided only the 9 measured Stokes parameters. By estimating this gap, which I call Γ, I estimate the success rate of the Bayes classifier and therefore estimate an upper bound on the success rate of any possible adaptive choice model.

To estimate

$$\Gamma = \mathbb{E}_x\left[\Pr\Big((\mathbf{I}_1(x) \cup \mathbf{I}_2(x) \cup \mathbf{I}_3(x)) \cap \overline{\mathbf{I}_{G(x)}(x)}\Big)\right], \tag{4.2.15}$$

I bin the large data set $\mathscr{D}$ of 16,831,944 computationally generated random entangled states according to their 9 $\{W^3\}$-determined Stokes parameters. A state with measured Stokes parameters x_1 is placed in bin $B(x)$ if its 9 measured Stokes parameters are equal to x's within some tolerance ϵ. I then choose $G(x)$ via the optimal classification rule of the Bayes classifier discussed in Sec. 4.1. Then,

$$\begin{aligned} \Gamma &= \sum_B \frac{|B|}{|\mathscr{D}|} \frac{|\{\{W^5_{t\,\text{oracle on }\rho_x}\} < 0,\ \{W^5_{t\,G(x)}\} > 0\}|x \in B\}|}{|B|} \\ &= \sum_B \frac{|\{\{W^5_{t\,\text{oracle on }\rho_x}\} < 0,\ \{W^5_{t\,G(x)}\} > 0\}|x \in B\}|}{|\mathscr{D}|} \end{aligned} \tag{4.2.16}$$

In words, this is the average number of states associated with every x that are witnessed by one of the $\{W^5\}$ witnesses but not the triplet chosen by the Bayes classifier, estimated via binning. The top figure in Fig. 4.1 shows the gap Γ from Eq. 4.2.16 across different ϵ values. As $\epsilon \to 0$, Γ converges to zero as well, indicating that the 16 million state data set is not fine enough to have Γ converge to the

non-zero value we expect. From Eq. 4.2.14 with Eq. 4.2.16 and $P_{\text{oracle}} = 79.1\%$, I calculate the estimated Bayes classifier success rate and plot it in the bottom figure of Fig. 4.1. I also find that at $\epsilon = 0.24$, the estimate of P_{Bayes} becomes 72.1% which is below the neural network success rate of 72.3%; as ϵ is increased further to 4.0, the estimated P_{Bayes} falls further to 69.9%. Thus, this estimate is untrustworthy until a specific ϵ is justified—I leave this to future work.

In Fig. 4.1, there is no comparison metric to determine what ϵ value leads to a valid estimate of the bound or gap. In the next section, I focus instead on directly estimating a lower bound on the success rate of a hypothetical optimal choice model.

It would be interesting in future work to apply this gap-based bound to an experimental system—one could choose ϵ based on how finely their experimental apparatus can determine Stokes parameters. Then, this model provides the percent of states missed due to the overlap of measured Stokes parameters, with the measurement error from the experimental apparatus informing a choice of ϵ.

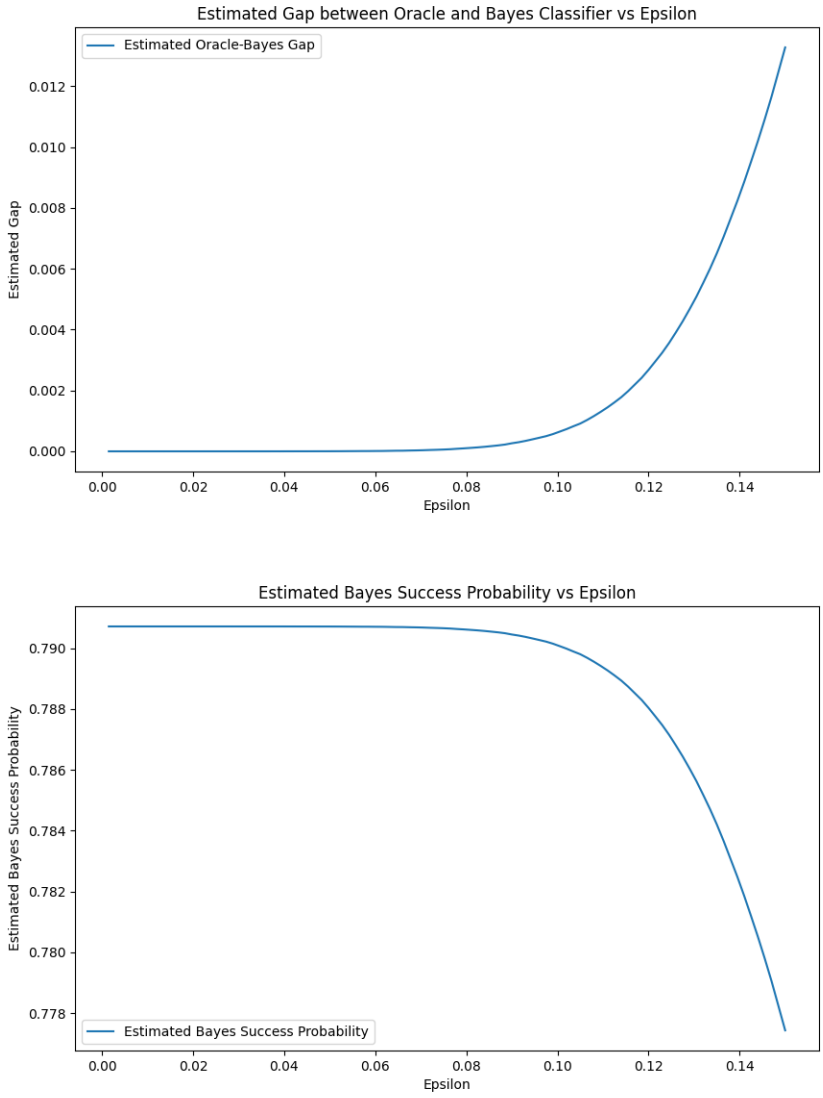


Figure 4.1: Γ from Eq. 4.2.16 as a function of ϵ (top) and an estimate of P_{Bayes} from $P_{\text{oracle}} = 79.1\%$, and Eq. 4.2.14. Here, ϵ controls how tightly the data set is binned.

4.3 Estimating P_{Bayes}

In this section, I estimate P_{Bayes} directly to estimate an upper bound on any adaptive choice model's success rate. Recall that P_{Bayes} is an upper bound on the success probability of our protocol using the hypothetical best possible choice model, a Bayes classifier. I choose to do this, rather than continue finetuning the gap-based bound in Sec. 4.2.2, because it will also provide me a choice model that indicates when the model's estimated bound is trustworthy or not.

My motivation is to generate a large data set in the same fashion as the data used for Fig. 3.3 and use that data set to estimate the relevant probabilities ($g_i(x)$ from Eq. 4.1.3). I choose to estimate the probability that the state ρ_x associated with Stokes 9-vector x is witnessed by the i-th $\{W^5\}$ triplet by partitioning the data set into many bins and then calculating the probability that the i-th $\{W^5\}$ triplet detects ρ_x given it is partitioned into some bin which I denote $B(x)$. I make this choice because, assuming I can bin the data set in a smart way as I discuss later in Sec. 4.4, I can increase what percentage of states in a bin are witnessed by the same triplet. I was further convinced because computing this probability is easy: divide the number of states in a bin witnessed by triplet i by the total number of states in that bin. A drawback of this approach is that the curse of dimensionality in $\mathbb{R}^9$ and my use of a k-d tree both make it easy to overestimate the success probability; I discuss this in Sec. 4.4 as well.

I now proceed with the derivation of the binned success rate. Split the data set, denoted $\mathscr{D}$, via a partition consisting of N bins, $\{B_1, \ldots, B_N\}$.

Begin with Eq. 4.1.7,

$$P_{\text{Bayes}} = \mathbb{E}_X[\max_i g_i(X)]], \tag{4.3.1}$$

where the expectation value is over all Stokes vectors X associated with entangled states. Partition the data set as described above and apply the law of total expectation, denoting $B(X)$ the bin containing X to obtain

$$P_{\text{Bayes}} = \mathbb{E}_{B(X)}[\mathbb{E}_{X\in B(X)}[\max_i g_i(X)|B(X)]]. \tag{4.3.2}$$

Here, $\mathbb{E}_{X\in B(X)}$ is the expectation value over all X within some bin $B(X)$, and $\mathbb{E}_{B(X)}$ is the expectation value over all bins in the partition. Recognize that max is a convex function. For any convex function f, Jensen's inequality states that

$$\mathbb{E}[f(x)] \geq f(\mathbb{E}[x]). \tag{4.3.3}$$

Therefore, using Jensen's inequality on the max function I rewrite Eq. 4.3.2 as

$$P_{\text{Bayes}} = \mathbb{E}[\mathbb{E}[\max_i g_i(X)|B(X)]] \tag{4.3.4}$$

$$\geq \mathbb{E}[\max_i \mathbb{E}[g_i(X)|B(X)]] \tag{4.3.5}$$

$$= \mathbb{E}[\max_i P(E_i = 1|X \in B(X))] \tag{4.3.6}$$

$$= \sum_{n=1}^{N} P(X \in B_n) \max_{i\in\{1,2,3\}} P(E_i = 1|X \in B_n). \tag{4.3.7}$$

I then estimate $P(X \in B_n)$ as the relative bin size and $P(E_i = 1|X \in B_n)$ as the number of states in bin n that are witnessed by the i-th triplet divided by the total number of states in bin n. This yields

$$P_{\text{Bayes}} \geq \sum_{n=1}^{N} \frac{|B_n|}{|\mathscr{D}|} \max_{i\in\{1,2,3\}} \frac{|\{x \in B_n, E_i(\rho_x) = 1\}|}{|B_n|}. \tag{4.3.8}$$

Of note, this is now an *estimate* of the lower bound and can exceed the true lower bound. I now explore how we may bin this data set such that states likely to be witnessed by the same triplet exist in the same bins.

4.4 Binning the Dataset: A k-d Tree

In this section, I discuss a method for binning the data set that places states witnessed by the same triplet in the same bin, without knowledge of the actual witness value associated with each state. I first describe the general theory behind a k-d tree, then describe specific implementations and their performance in the following subsections.

A k-d tree, or k-dimensional tree, is a binary tree where each node is split in two according to how the data is distributed at that node with respect to one of k parameters. A balanced k-d tree is created by, for each node in the tree, choosing one of k parameters to split on then placing all data with parameter value below its median into one child node and the rest into a different child node. In this thesis, I discuss only a balanced tree, where each child node is half the size of its parent. as I found it performed better than unbalanced trees, which is when the child k-d tree's performance, I am speaking to the success rate of the associated lower bound from Sec. 4.3 or choice model from Sec. 3.3. As one changes how the splitting parameter is chosen, which splitting parameters the tree has access to, the number of leaves (lowest children nodes) allowed, and the depth (longest chain from root to leaf) allowed, the performance of the tree changes. As an example, if I allow the tree to split on a state's Stokes parameters and choose which to split on via each parameter's variance, the k-d tree will choose the Stokes parameter of greatest variance in the dataset then split the data set into two, determined by whether each point is above or below the median of the parameter of

highest variance. The median is chosen so that the tree is balanced. I discuss the provided parameters in Sec. 4.4.1 and potential ways to choose which parameter to split on at a given node in Sec. 4.4.2.

By the end of this algorithm, we have binned our data set in 2^D equally sized bins, where D is the depth of the tree. More importantly, we have constructed a set of rules that determine how to split future data points. In other words, we have created a set of nodes, each associated with a parameter and parameter value to split with respect to, that we may split any data set with—this is a new choice model that approaches the Bayes success rate as the number of data points and depth approaches infinity. In the rest of this thesis, I create this set of *rules* using a training set, and run the lower bound calculation across a separate test set of quantum states.

4.4.1 An Overview of Binning Parameters

In this section, I discuss the different parameter sets I can provide the k-d tree. However, I save a discussion of the results of each parameter set for Sec. 4.5 so that I can talk about both the parameter sets and how a node's splitting parameter is chosen at the same time.

As a first guess, I allow our k-d tree to choose one of the nine Stokes parameters measured for the $\{W^3\}$ witness set,

$$\{M^3\} = \{\langle \sigma_i \otimes \sigma_0 \rangle, \langle \sigma_0 \otimes \sigma_i \rangle, \langle \sigma_i \otimes \sigma_i \rangle\},\ i \in [x, y, z]. \tag{4.4.1}$$

Recall that $\sigma_0 = \mathbb{I}$ is the identity operator. The algorithm then, at each node, chooses a Stokes parameter and splits the data set. I find that this algorithm performs sub-optimally and that its success rate does not surpass nor approach the neural network's performance.

There is, however, no reason to believe that a state's Stokes parameters best determine which $\{W^5\}$s triplet witnesses it. Furthermore, the neural network outperforming the Bayes lower bound indicates that there are correlations between Stokes parameters and negative witness triplet that the Stokes parameters alone fail to capture. Recall from Sec. 3.1.2 that six of the nine $\{W^5\}$ witnesses admit an analytic minimization, and from Sec. 3.3.3 that this leads to six parameters $\Delta_i(x)$ that characterize how large the cross correlation Stokes parameter term obtained by the additional $\{W^5\}$ measurements must be for a triplet to detect a state. Here, x is the vector consisting of the 9 Stokes parameters obtained by measuring the $\{W^3\}$ set. My motivation for using the $\Delta_i(x)$ is that I want parameters that already respect how likely it is for a state to be witnessed by a given $\{W^5\}$ triplet.

4.4.2 An Overview of Methods for Splitting the Dataset

I now present two methods for choosing a binning parameter to split against at each node. The performance of each is discussed in Sec. 4.5.

One way to split is against the parameter of highest variance. This means that at each node, the variance of all provided binning parameters is calculated, the binning parameter of highest variance is chosen, and the data set is split at the mean of that chosen parameter into two child nodes.

We have no reason to believe, however, that variance is the best way to separate states witnessed by different triplets; this was hinted at by the fact that my previous attempts this year to use a nearest neighbor model split on Euclidean distance in the space of the nine $\{W^3\}$ Stokes parameters failed to outperform the population model or neural network.

To inform a different method, I look back at the form of the lower bound, Eq. 4.3.8. The Bayes lower bound looks like a sum over the maximum of probabilities, these probabilities being of a given witness triplet detecting a state in a given leaf. The maximum is over the three triplets and sum over all states. As such, at each step in the training process, I am motivated to increase the probability of the highest-probability witness triplet of a node when creating its children. The quantity I care about here is entropy: when entropy is high, we have very little information on what triplet is most likely to witness states in a given bin, and therefore the probabilities are nearly equal—this decreases the Bayes lower bound. When entropy is low, the probability that one of the triplets witness the states in the node is high, which raises the Bayes lower bound. Decreasing entropy is equivalent to increasing the information known about a system. Inspired by this, I choose how to split a node based on the parameter that results in the highest *information gain*. In machine learning, information gain is defined as

$$\mathrm{IG}(B, a) = H(B) - H(B|a), \tag{4.4.2}$$

where $H(B)$ is the entropy of node B, and $H(B|a)$, the entropy of B given a splitting on parameter a, is the average entropy of its children nodes if you split on that parameter. Recall from Sec. 2.3 that mutual information measures the amount of information that knowing the outcome of some random variable Y provides about a different random variable X, on average. Information gain can be thought of as analogous to mutual information, except that the random variable Y have been replaced by a bin and the random variable X by a splitting parameter. I found that a balanced tree performed better than an unbalanced one, so always choose to split a given node at the median of whatever parameter maximizes the information gain for that node.

4.5 Bayes Lower Bound Results

In this section, I present the results of the Bayes lower bound via binning approach. I find it saturates and in some cases exceeds the neural network success rate. Importantly, note that all features are normalized to mean 0 and variance 1 before the tree is fit to them, meaning that the first node's splitting parameter is arbitrary, while the following are not because I do not re-normalize after the first split. Future work should investigate how choosing different first-node splitting parameters affects the resulting tree performance. The section is organized as follows: Sec. 4.5.1 contains results when using the variance-based splitting on Stokes parameters and Sec. 4.5.2 provides results when either variance or information gain is used on parameters inspired by the analytic-minimization-inspired quantities.

4.5.1 A Naive Binning: Stokes Parameters

Here, I choose to split on the state's 9 measured Stokes parameters and make the choice via highest variance. First, I consider how this bounds the success probability as a function of concurrence. For reference, I include the five-layer neural network and the W^5 oracle choice models as well. Each model is evaluated on the data set of 16,831,944 states discussed in Chap. 1. See Fig. 4.2; of note, the performance stays below the neural network's for all concurrences, indicating that the neural network is finding correlations between Stokes parameters that the Bayes model misses. This figure is generated with a tree depth of 17, which as discussed in Sec. 4.5.2 is likely an overestimate due to a low number of states per leaf. At the safer depth of 15, the lower bound at zero concurrence was approximately 68%.

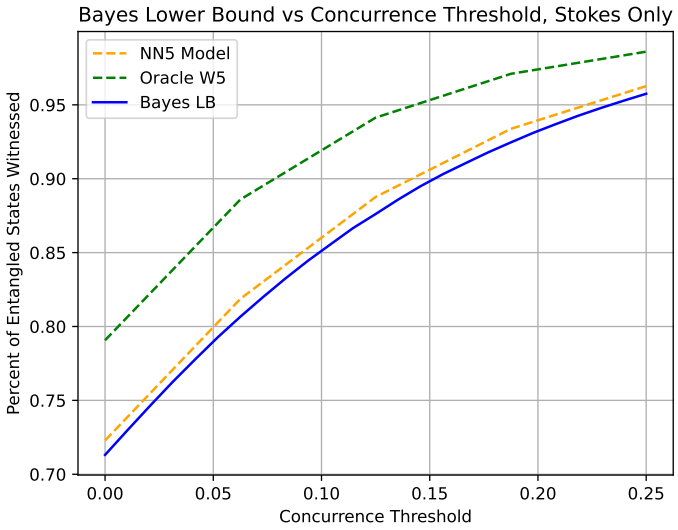


Figure 4.2: k-d tree Bayes classifier performance when given all $\{W^3\}$ measurements. At $c > 0.0$, meaning for all entangled states in the data set, this estimates $P_{\text{Bayes}} = 71.3\%$.

4.5.2 A Smarter Binning with Analytic Minimization

For the k-d tree to be informed by the six analytically minimized $\{W^5\}$ witnesses, I choose to split on the parameters $\Delta_i(x)$, $i \in \{1,2,4,5,7,8\}$, derived from Eq. 3.3.4. Furthermore, I choose to split using information gain, as I found its success rate to be 1% greater than splitting on variance on when given either the six $\Delta_i(x)$ parameters or the 9 measured Stokes parameters.

To confirm that the $\Delta_i(x)$ terms provide better binning than the 9 measured Stokes parameters, and knowing that information gain performs better than variance, I plot the top 12 features used across an information-gain tree when given both $\Delta_i(x)$ and the 9 measured Stokes parameters in Fig. 4.3. I find that the $\Delta_i(x)$ terms are overall better to split on, as only two of the measured Stokes parameters occupy the top 8 features, with the other 6 features being the $\Delta_i(x)$ terms. In contrast with the variance-based model discussed in the next paragraph, the information gain model splits on the S_{ii}, $i \in [x,y,z]$ terms, which is aligned with the fact that the population model outperforms the cross-correlation model.

I still find it interesting to look at the feature usage of a variance-based tree when provided both the $\Delta_i(x)$ temrms and the 9 measured Stokes parameters despite its smaller lower bound—this is plotted in Fig. 4.4. Here, I find that it relies heavily on single-particle $\langle \sigma_i \otimes \sigma_0 \rangle, \langle \sigma_0 \otimes \sigma_i \rangle$, $i \in [x,y,z]$ terms in early splits as well as the $\Delta_j(x)$ terms, $j \in [1,6]$.

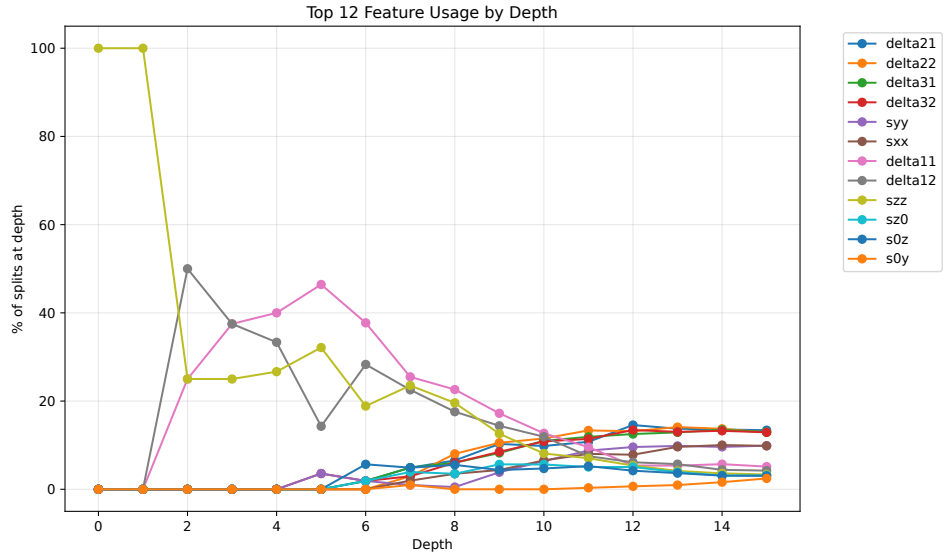


Figure 4.3: Top 12 features when the k-d tree is given Δ_i and the nine Stokes parameters needed for the $\{W^3\}$ set. I provide both to compare which the tree picks. The asymmetry between S_{zz}, S_{xx}, and S_{yy} is due to each feature being standardized to mean 0 and variance 1 before the first split, such that S_{zz} gets randomly chosen to start off.

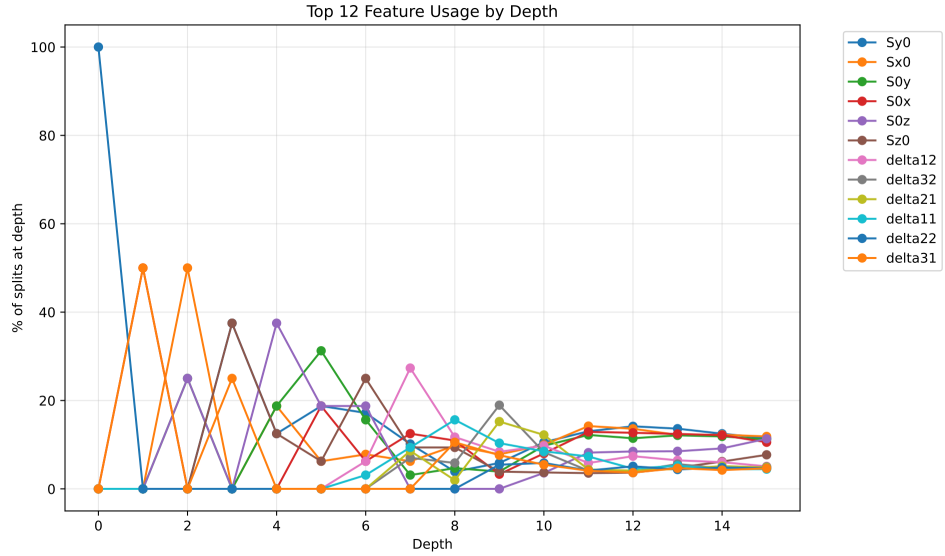


Figure 4.4: Feature usage of a variance-based tree given Stokes parameters and the Δ_i parameters, plotted for the 12 most-used features.

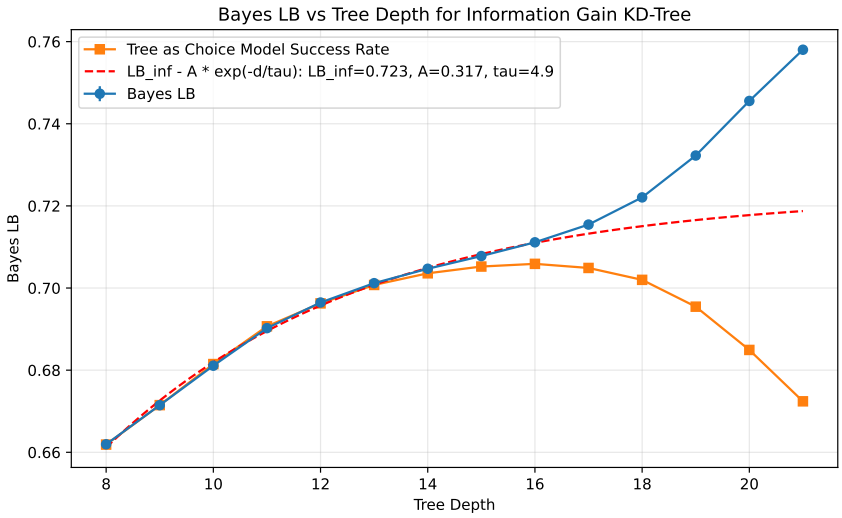


Figure 4.5: The k-dimensional tree split on the Δ_i parameters of the cross correlation model, evaluated via information gain as discussed in Sec. 4.4.

I split on information gain in order for the k-d tree to choose parameters that are more likely to help it separate states detected by different $\{W^5\}$ triplets, shown in Fig. 4.5. In this figure, I use only the $\Delta_i(x)$ parameters, having previously shown them to be superior to the 9 measured Stokes parameters for binning. The three curves represent the lower bound calculated as in Eq. 4.3.8, the performance of this k-d tree as a choice model, and an exponential curve fit to the lower bound points up to and including a depth of 16. I now clarify the difference between these two models. The Bayes lower bound is calculated by fitting the tree on training data, placing a separate test set into the trained tree, then in each bin calculating the probability of every witness triplet detecting states within that bin *on the testing states* and averaging over the maximum probabilities for each bin. Dissimilarly, the choice model success rate is determined by fitting a tree on training data, then calculating the percent of times that a separate test state that is placed into a bin in the trained tree is detected by the same $\{W^5\}$ triplet that detects the most *training* states in the bin. The choice model is more realistic and tells us how the k-d tree would perform if applied in a real-world entanglement witnessing protocol. I use the depth at the peak of the choice model, 16, and all depths below it, to fit an exponential curve to the lower bound and estimate its asymptote. I call depths at or below this peak 'trustworthy'.

I interpret the points leaving the exponential fit in Fig. 4.5 as over-estimation of the lower bound due to having very few states in each leaf; as I take a maximum over probabilities in the lower bound calculation, having few states ar-

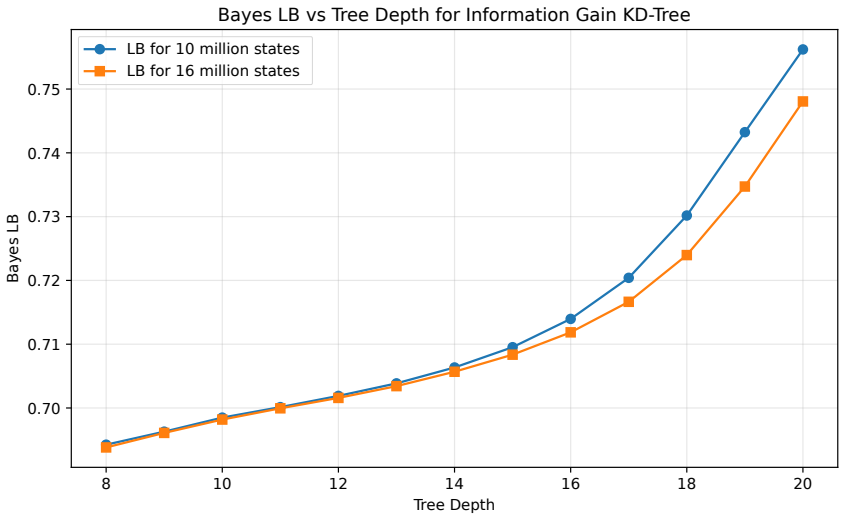


Figure 4.6: Bayes lower bound as a function of tree depth for a balanced tree split on information gain and given $\Delta_i(x)$ derived from Eq. 3.3.4. The lower bound is evaluated on two data sets of 10 million (blue) and 16 million (orange) states each.

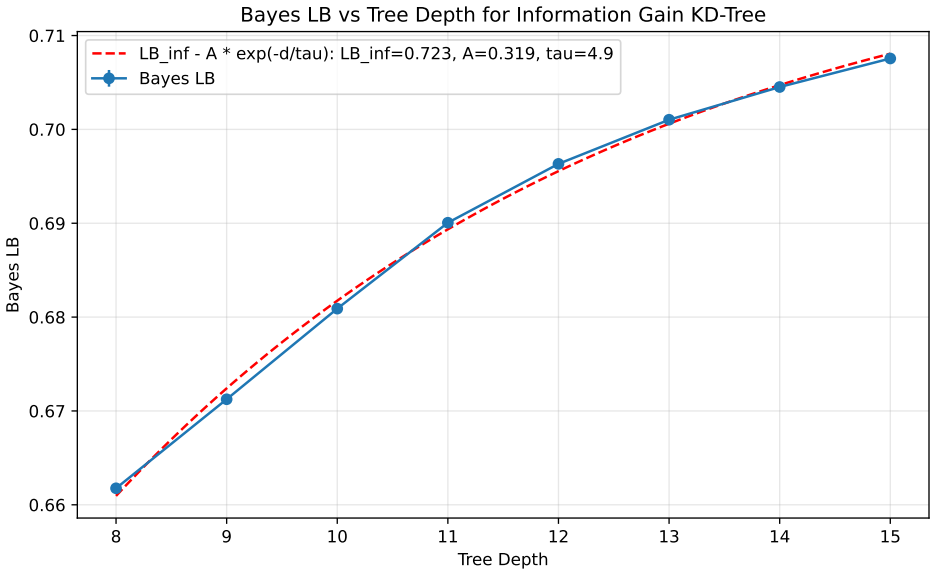


Figure 4.7: Information gain-dependent k-d tree lower bound as a function of tree depth, provided only the Δ_i parameters. This is a zoomed in version of Fig. 4.5.

tificially inflates the associated probabilities. I demonstrate this in Fig. 4.6 by plotting a lower bound calculated from 10 million states versus 16 million states. I find that the 16 million state lower bound is less than the 10 million state lower bound. This is because the 10 million state set has even fewer states per bin than the 16 million state set and supports my reasoning that having two-few-states is the cause of trees at high depths leaving the exponential fit. The trustworthy points of Fig. 4.5 are plotted in Fig. 4.7.

I then investigate the lower bound performance of this tree as a function of concurrence threshold, plotting both the the lower bound at a depth of 16, which is trustworthy due to the choice model reasoning above, and the asymptote of the fitted exponential. A concurrence threshold means that I evaluate success rates on all states with a concurrence greater than to some value c, with c plotted from greater than 0.0 to 0.25. For $c > 0$, meaning all entangled states, the exponential asymptotes lower bound is 72.3%, which matches the five layer neural network performance of 72.3%. I then observe that while the depth of 16 lies below the neural network, the fitted exponential asymptote lies exactly on or above the neural network for all concurrences. A reason for the odd jumping behavior of this fitted curve is currently unknown; I first thought there weren't enough points for an accurate lower bound (due to the overfitting that occurs when there are too few states per leaf) but out of the 16 million states in our data set, $10,277,704$ have a concurrence greater than 0.16. Then, the number of states to avoid overfitting at a depth of 16 is $2^{16} \times 50 = 3,276,800$. This indicates that the concurrence leading to too-few-states per leaf does not lead to the jumping behavior. A possible reason is that high concurrences provide extra information that the k-d tree is able to use—confirmation of this hypothesis is the subject of future work.

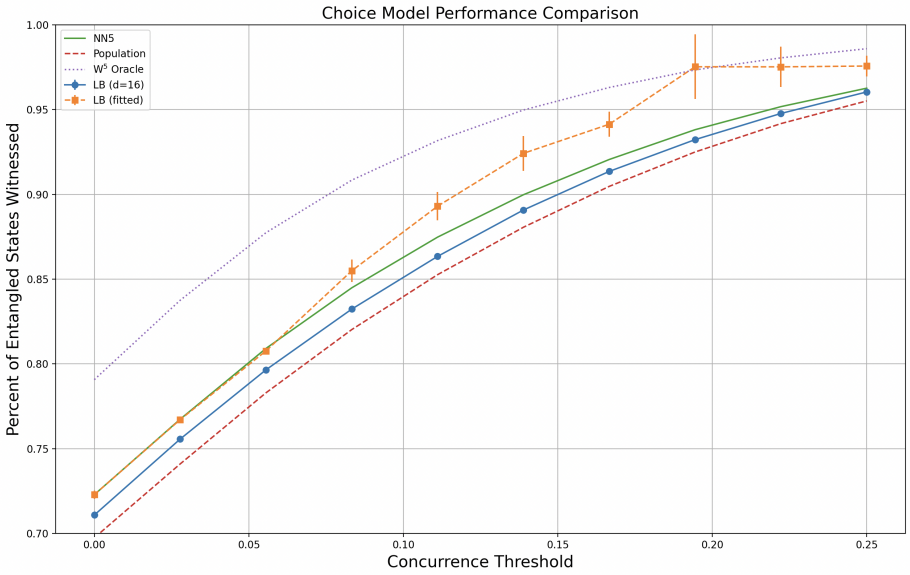


Figure 4.8: Choice model success rates as a function of concurrence with the k-d tree based lower bound shown both at a depth of 16 and as the result of an exponential fit on a lower bound versus depth curve. Only the Δ_i splitting parameters are provided.

Chapter 5

A General Witnessing Framework from the PPT Criterion

"I'll be positive." - The PPTerminator

In this chapter, I relate the $\{W^3\}$ and $\{W^5\}$ witness sets to a well-known separability criterion for bipartite quantum states. This chapter is structured as follows: in Sec. 5.1 I provide a high-level summary of a framework connecting the $\{W^3\}$ and $\{W^5\}$ witnesses to the positive partial transpose (PPT) criterion. In Sec. 5.2 I discuss a geometric interpretation of this framework and use it to explain why the population model has a higher success rate than the cross correlation model, with each model discussed in Sec. 3.3. In Sec. 5.3 I derive each $\{W^3\}$ and $\{W^5\}$ in terms of Stokes parameters. Then, in Sec. 5.4, I prove that the $\{W^3\}$ and $\{W^5\}$ are equivalent to the PPT criterion evaluated in specific density matrix bases and that testing the PPT criterion in all bases is equivalent to a hypothetical general entanglement witness that can be extrapolated from the forms of the $\{W^3\}$ and $\{W^5\}$ using the geometric intuition from Sec. 5.2.

5.1 General Framework Summary

In this section, I provide a summary of the whole chapter—this means that many of the statements in this section are left unjustified until later sections. I refer to density matrix elements for the following density matrix with basis elements along $\hat{z}$:

$$\rho = \begin{pmatrix} \rho_{11} & \rho_{12} & \rho_{13} & \rho_{14} \\ \rho_{21} & \rho_{22} & \rho_{23} & \rho_{24} \\ \rho_{31} & \rho_{32} & \rho_{33} & \rho_{34} \\ \rho_{41} & \rho_{42} & \rho_{43} & \rho_{44} \end{pmatrix}. \tag{5.1.1}$$

The PPT criterion states that any bipartite state for which $\rho^\Gamma \geq 0$ is separable [19]. This is equivalent to all principal minors of ρ^Γ being non-negative [9], in every basis for ρ. The $\{1,4\}$ and $\{2,3\}$ principal minors give the following conditions that are necessary and sufficient to guarantee separability if true in every basis:

$$\rho_{11}\rho_{44} \geq |\rho_{23}|^2 \text{ or } \rho_{22}\rho_{33} \geq |\rho_{14}|^2. \tag{5.1.2}$$

This implies that a bipartite quantum state is entangled if and only if

$$\rho_{11}\rho_{44} < |\rho_{23}|^2 \text{ or } \rho_{22}\rho_{33} < |\rho_{14}|^2, \tag{5.1.3}$$

where must be true in at least *one* basis. Any entangled bipartite quantum state follows Eq. 5.1.3 when its density matrix is rotated into a basis in which the negative eigenvector has support only in the $\{1,4\}$ *or* $\{2,3\}$ principal submatrices. It is always possible to find a set of unitaries $U_A \otimes U_B$ such that this is true. Note that you may always rotate an eigenvector with support only in the $\{1,4\}$ submatrix into the $\{2,3\}$ submatrix and vice versa, meaning that the conditions in Eq. 5.1.3 are equivalent. I present them separately because they do affect how I later compare them to entanglement witnesses. I now relate Eq. 5.1.3 to the $\{W^3\}$ and $\{W^5\}$ entanglement witness sets, using W_1^5 as an example. Write the density matrix in the (z,z) basis as in Eq. 5.1.1 in terms of Stokes parameters as

$$\rho = \frac{1}{4}\begin{pmatrix} (1+S_{zz})+(S_{z0}+S_{0z}) & (S_{0x}+S_{zx})-i(S_{0y}+S_{zy}) & (S_{x0}+S_{xz})-i(S_{y0}+S_{yz}) & (S_{xx}-S_{yy})-i(S_{yx}+S_{xy}) \\ (S_{0x}+S_{zx})+i(S_{0y}+S_{zy}) & (1-S_{zz})+(S_{z0}-S_{0z}) & (S_{xx}+S_{yy})-i(S_{yx}-S_{xy}) & (S_{x0}-S_{xz})-i(S_{y0}-S_{yz}) \\ (S_{x0}+S_{xz})+i(S_{y0}+S_{yz}) & (S_{xx}+S_{yy})+i(S_{yx}-S_{xy}) & (1-S_{zz})-(S_{z0}-S_{0z}) & (S_{0x}-S_{zx})-i(S_{0y}-S_{zy}) \\ (S_{xx}-S_{yy})+i(S_{yx}+S_{xy}) & (S_{x0}-S_{xz})+i(S_{y0}-S_{yz}) & (S_{0x}-S_{zx})+i(S_{0y}-S_{zy}) & (1+S_{zz})-(S_{z0}+S_{0z}) \end{pmatrix}. \tag{5.1.4}$$

Notice that the entanglement condition from the PPT criterion restricted to testing only the real portion of the coherence $\text{Re}(\rho_{23}) = \text{Re}((S_{xx}+S_{yy})-i(S_{yx}-S_{xy})) = (S_{xx}+S_{yy})$ is

$$\begin{aligned} (S_{xx}+S_{yy})^2 &> ((1+S_{zz})+(S_{z0}+S_{0z}))\,((1+S_{zz})-(S_{z0}+S_{0z}))\,, \\ (S_{xx}+S_{yy})^2 &> (1+S_{zz})^2-(S_{z0}+S_{0z})^2, \end{aligned}$$

or

$$(1+S_{zz})^2 - \left[(S_{xx}+S_{yy})^2+(S_{z0}+S_{0z})^2\right] < 0. \tag{5.1.5}$$

This is precisely the detection condition of the analytically minimized W_1^3 expression in Eq. 3.1.9. Furthermore, if the full coherence term is tested, the entanglement condition requires

$$\begin{aligned} |\left((S_{xx}-S_{yy})-i(S_{yx}+S_{xy})\right)|^2 &> ((1+S_{zz})+(S_{z0}+S_{0z}))\,((1+S_{zz})-(S_{z0}+S_{0z}))\,, \\ (S_{xx}+S_{yy})^2+(S_{yx}+S_{xy})^2 &> (1+S_{zz})^2-(S_{z0}+S_{0z})^2, \end{aligned}$$

or

$$(1+S_{zz})^2 - \left[(S_{xx}+S_{yy})^2+(S_{z0}+S_{0z})^2\right] < (S_{yx}+S_{xy})^2. \tag{5.1.6}$$

This is detection condition of the analytically minimized W_1^5 expression in Eq. 3.1.16. I am then able to derive similar mappings between the entanglement condition from the PPT criterion and the analytically minimized $\{W^3\}$ and $\{W^5\}$ witnesses by changing the basis states of ρ. I summarize this now.

When the analytic minimizations of the $\{W^3\}$ witnesses given in Sec. 3.1.1 are written in terms of density matrix elements, they test the entanglement conditions

$$\rho_{11}\rho_{44} < \mathrm{Re}(\rho_{23})^2, \text{ and/or } \rho_{22}\rho_{33} < \mathrm{Re}(\rho_{14})^2, \tag{5.1.7}$$

when ρ is written in the (z,z) basis for W_1^3 and W_2^3, the (x,x) basis for W_3^3 and W_4^3), and the (y,y) basis W_5^3 and W_6^3. Restricting to only test the real portion of ρ_{14} and ρ_{23} allows us to use fewer Stokes parameters as they no longer require the cross correlation terms $(\sigma_i \otimes \sigma_j)$.

The first two witnesses in each $\{W^5\}$ triplet test the full PPT criteria Eq. 5.1.3 with density matrix basis states along the (z,z) basis for W_1^5 and W_2^5, the (x,x) basis for W_4^5 and W_5^5, and the (y,y) basis for W_7^5 and W_8^5.

The third witness in each $\{W^5\}$ triplet tests the same restricted PPT criterion as the $\{W^3\}$ witness set given in Eq. 5.1.7, but rotates the density matrix basis states by a shared-axis rotation $R_i(\alpha) \otimes R_i(\beta)$, where R is the rotation operator about axis $\hat{i}$. W_3^5 rotates a density matrix originally in the (x,x) basis about the z axis, W_6^5 rotates a density matrix originally in the (z,z) basis about the x axis, and W_9^5 rotates a density matrix originally in the (z,z) basis about the y axis.

Define $\{W^{all}\}$ to be the theoretical set of witnesses that obey the PPT-provided criteria in Eq. 5.1.3, and allow for arbitrary rotation of both density matrix subsystems around separate axes. The span of this witness set will detect every entangled state. This is equivalent to PPT.

This framework characterizes what states will never be witnessed by a $\{W^3\}$ or $\{W^5\}$ witness. One of these witness sets is unable to certify the entanglement of ρ if the optimal basis, defined by the basis in which ρ^Γ's negative eigenvector has support only in the $\{1,4\}$ *or* $\{2,3\}$ principal submatrices, is not reachable by our rotations about specific shared axes.

Interestingly, this implies that any bipartite entangled state can be detected using only 7 out of 15 Stokes parameters, if only you know what basis to test in. Generally, finding that basis likely required a full tomography.

We can also think of the different witness sets as a choice of the number of free parameters to search over when choosing your unitaries (when choosing how to rotate ρ with basis states along $\hat{z}$ into an arbitrary basis). The first two witnesses in each triplet have zero free parameters as you fix three bases, the third witness in each triplet has 2 free parameters (α,β), and the most general PPT criteria requires 6 free parameters because it is generally a matrix in $SU(2) \otimes SU(2)$. We can think of the difference between having 2 and 6 parameters as why 21% of states remain unwitnessed–these are the states that require rotations about multiple axes simultaneously on at least one qubit. In other words, they

are the states that cannot be reached by any shared-axis rotation.

Previous work [18] showed that the limited rotations used in the protocol's $\{W^5\}$ set (a subset of all possible $\{W^5\}$-like witnesses as described above), are comprehensive—i.e. they are able to witness all states that the general $\{W^5\}$-like witnesses can detect. Navarro did this by showing that the rotations used in the protocol's specific $\{W^5\}$ set are equivalent to those obtained by a derivation where you set specific Stokes parameters to zero, as in [22]. A remaining question is to better understand why this is. As an idea, could it be due to the specific restriction to exchange symmetric cross correlators (terms like $\langle\sigma_i \otimes \sigma_j\rangle \pm \langle\sigma_j \otimes \sigma_i\rangle$)?

I see three pathways to explore this. First, it may be possible to show that any rotation of both subsystems of a $\{W^3\}$ witness about a shared axis or one subsystem about an axis traces out equivalent great circles to the 9 defined $\{W^5\}$ witnesses or aligns along one of the fixed axes; see Sec.5.2 for more info on this. Second, you may be able to show that the PPT-necessary criterion that is tested by any arbitrary (but limited to single-axis for one subsystem and shared-axis for both subsystems) either matches the PPT-necessary criterion set by the $\{W^5\}$ witnesses as currently defined or is covered by them—for example, if you find a rotated $\{W^3\}$ follows a constraint that does not match any of the first two witnesses in each triplet but can be derived by choosing specific α, β angles for one of the third witness in each of the triplets. Third, show that every separate-axis rotation for two subsystems either introduces a non-exchange symmetric cross correlator pair that was set to zero in the derivation based on Riccardi et al. or detects a subset of entangled states that follow PPT-necessary constraints that are not included in those used in the current $\{W^5\}$ set. This would be equivalent to showing that the shared-axis rotations are the *only* way to form the protocol's triplet without mixing cross correlator terms and losing the exchange symmetry. Follow up questions include whether we may improve the protocol by focusing on terms that are not exchange symmetric, such as $\langle\sigma_i \otimes \sigma_j\rangle \pm \langle\sigma_j \otimes \sigma_k\rangle$.

5.2 A Geometric Understanding

As discussed above, our protocol is able to witness any entangled bipartite quantum state whose density matrix can be rotated by shared-axis fixed rotations about $\hat{x}$, $\hat{y}$, or $\hat{z}$ such that its partial transpose's negative eigenvector has support only in the $\{1,4\}$, or equivalently, $\{2,3\}$ submatrix. Here, negative eigenvector means the eigenvector associated with the negative eigenvalue of ρ^Γ for entangled ρ. Any vector in a bipartite Hilbert space $\mathscr{H}_A \otimes \mathscr{H}_B$ can be expressed as a Schmidt decomposition of the form

$$|\psi\rangle = \sum_i \sqrt{p_i}\,|i\rangle_A \otimes |i'\rangle_B\,, \tag{5.2.1}$$

where $|i\rangle_A \otimes |i'\rangle_B$ is a particular orthonormal basis of $\mathscr{H}_A$ and $\mathscr{H}_B$ [20]. Therefore, this negative eigenvector admits a Schmidt decomposition

$$|\phi\rangle = \cos\theta\,|a_0\rangle\,|b_0\rangle + \sin\theta\,|a_1\rangle\,|b_1\rangle\,, \tag{5.2.2}$$

where $|a_0\rangle$ and $|a_1\rangle$ are orthonormal, as are $|b_0\rangle$ and $|b_1\rangle$. Our protocol witnessing this state is dependent on whether one of the shared-axis rotations is able to rotate $|a_0\rangle$ to $|H\rangle$ and $|a_1\rangle$ to $|V\rangle$, with the same needing to be true for $|b_0\rangle$ and $|b_1\rangle$. Because $|a_0\rangle$ and $|a_1\rangle$ are orthonormal, the direction of $|a_0\rangle$ on the Bloch sphere fully determines the necessary U_A needed–call this unit vector $\hat{n}_A$. Thus, to detect a given state we need to align $\hat{n}_A$ with $\hat{z}_A$ and $\hat{n}_B$ with $\hat{z}_B$. In most geometric talk in this thesis, I refer to the space of all possible Schmidt decomposition basis vectors for ρ^Γ's negative eigenvector, assuming entangled ρ.

The goal now is to demonstrate what rotations of $(\hat{n}_A, \hat{n}_B)$ our protocol allows. From now on, I represent $\hat{n}_A$ and $\hat{n}_B$ as vectors in (x, y, z) notation on the Bloch sphere that are turned into $|H\rangle$ (equivalently $+z$) via rotations. A single-axis rotation has the form $R_k(\alpha) = e^{-i\alpha\sigma_k/2}$. Therefore,

$$R_x(\alpha)\,|0\rangle = \cos(\alpha/2)\,|0\rangle - i\sin(\alpha/2)\,|1\rangle \;\Rightarrow \hat{n} = \begin{pmatrix}0 & -\sin\alpha & \cos\alpha,\end{pmatrix} \tag{5.2.3}$$

$$R_y(\alpha)\,|0\rangle = \cos(\alpha/2)\,|0\rangle - \sin(\alpha/2)\,|1\rangle \;\Rightarrow \hat{n} = \begin{pmatrix}-\sin\alpha & 0 & \cos\alpha,\end{pmatrix} \tag{5.2.4}$$

$$R_z(\alpha)\,|0\rangle = e^{-i\alpha/2}\,|0\rangle \;\Rightarrow \hat{n} = \begin{pmatrix}0 & 0 & 1\end{pmatrix}. \tag{5.2.5}$$

From this, we see that the first two rotations trace out a great circle (largest possible circle drawn on the surface of a sphere) on the Bloch sphere (of the above described space) when α is varied. Of note, the R_z rotation has the effect of changing the off-diagonal real and imaginary terms into each other. I now explore what great circles the three $\{W^5\}$ witnesses generate. I define $\tilde{\rho}_{ij} = \langle i|U\rho U^\dagger|j\rangle = \langle\phi_i|\rho|\phi_j\rangle$, $|\phi_j\rangle = U^\dagger\,|j\rangle$. Recall from Sec. 3.1.2 that each $\{W^5\}$ witness can be derived by applying a rotation to one of the $\{W^3\}$ witnesses.

W_3^5 is derived by rotating W_3^3 about the shared z axis. W_3^3 itself is defined with respect to the x axis, so $\langle W_3^5\rangle$ can be written as

$$\mathrm{Tr}(W_3^5\rho) = \mathrm{Tr}(\big(U(\alpha)\otimes U(\beta)\big)\,W_1^3\,\big(U(\alpha)\otimes U(\beta)\big)^\dagger\,\rho) \tag{5.2.6}$$

$$= \mathrm{Tr}(W_1^3\,\big(U(\alpha)\otimes U(\beta)\big)^\dagger\,\rho\,\big(U(\alpha)\otimes U(\beta)\big)) \tag{5.2.7}$$

with $U(\alpha) = R_z(\alpha)R_y$ and $R_y = R_y(\pi/2)$. Then,

$$U(\alpha)\sigma_z U(\alpha)^\dagger = R_z(\alpha)\sigma_x R(\alpha) = \cos\alpha\sigma_x + \sin\alpha\sigma_y, \tag{5.2.8}$$

meaning that the Bloch sphere coordinates post-rotation look like

$$\hat{n} = \big(\cos\alpha, \sin\alpha, 0\big), \tag{5.2.9}$$

such that the rotation creates the xy great circle. The first two witnesses in this triplet act as a fixed evaluation of the PPT criterion along the $\hat{z}$ axis, which is orthogonal to the great circle swept by the α minimization of the third witness.

W_6^5 uses $R_x(\alpha) \otimes R_x(\beta)$, so optimization over α and β trace out a great circle in the yz plane, independently for subsystems A and B. The first two triplets evaluate the PPT criterion along $\hat{x}$.

W_9^5 uses $R_y(\alpha) \otimes R_y(\beta)$, so optimization over α and β trace out a great circle in the xz plane, independently for subsystems A and B. The first two triplets evaluate the PPT criterion along $\hat{y}$.

To summarize, W_3^5, W_6^5, and W_9^5 trace out the xy, yz, and xz great circles respectively while checking the weakened PPT criterion for $\mathrm{Re}(\rho_{23})^2$ or $\mathrm{Re}(\rho_{14})^2$. Needing to optimize where along the great circle, for each subsystem, each witness tests explains why these three witnesses require minimizing two parameters. It would be interesting for future work to explore whether this geometric framework allows an analytic minimization of the last three witnesses in each $\{W^5\}$ triplet. The other two witnesses in each triplet check the PPT criterion more strongly (with $|\rho_{23}|^2$ or $|\rho_{14}|^2$) along a fixed axis orthogonal to the great circle its third witness generates. For example, $\{W_{t1}^5\}$ measures along the z axis to check $|\rho_{23}|^2$ or $|\rho_{14}|^2$ versus the population terms, then traces a great circle in the xy plane to compare $\mathrm{Re}(\rho_{23})^2$ or $\mathrm{Re}(\rho_{14})^2$ to the population terms.

5.2.1 Why the Population Model Wins

Recall that the population model splits on the $|S_{ii}|$ Stokes parameter magnitudes for $i \in x, y, z$, while the cross correlation model uses more information via the Δ_i terms derived from the analytic minimizations of six $\{W^5\}$ witnessed; refer to Sec. 3.3.1 and Sec. 3.3.3 for a reminder on the population and cross correlation models respectively. Because the cross correlation model contains more information than the population model and is directly informed by the structure of the $\{W^5\}$ witnesses, I would expect that it would have a higher success rate than the population model. Depsite this, as discussed briefly in Sec. 3.3.5, the opposite is true. In this section, I explain why this is the case.

In Fig. 5.1 I plot the two choice models but only on states detected by the first and second W^5s witness in each triplet (top) and third witness (bottom). This figure shows that the cross correlation model is on average $\sim 2\%$ more successful than the population model until a concurrence threshold of $c > 0.166$, where the population model overtakes. On the third witness side, the population model succeeds on average 7% more than the cross correlation model, due to being more general.

I interpret Δ_i, which is the quantity the cross correlation model chooses according to, as a distance (not in a strict mathematical sense) to the fixed axis points in the space of possible Schmidt decomposition basis vector for ρ^Γ's negative eigenvector, assuming ρ is entangled. I discuss in Sec. 5.2 that each triplet

invokes the PPT criterion as written in Eq. 5.1.3 along a fixed axis, then use the weakened version in Eq. 5.1.7 while rotating the Schmidt decomposition basis vector through a great circle in the plane orthogonal to the axis checked by the first two witnesses. I interpret the Δ_i as telling us how close the unknown state's optimal Schmidt decomposition basis vector is to one of the fixed axis directions ($\hat{x}$, $\hat{y}$, or $\hat{z}$). I define the optimal Schmidt decomposition basis set as the basis set for which the PPT criterion as written in Eq. 5.1.3 is satisfied. However, Δ_i provides little information about what great circle is optimal to sweep in. For example, if the optimal Schmidt decomposition basis vector is oriented 45° between two axes, the different Δ values will be similar, as the two are equidistant from two axes. As a second example, consider a state whose optimal basis set is rotated away from the $\hat{x}$ axis along the xy great circle but too far to be witnessed by the first two witnesses in each triplet that test a fixed axis. In this case, the Δ chosen will be for the second triplet as its the closest to that axis, which would then allow it the basis vector to be rotated along the yz plane. However, this won't allow it to be aligned with any of $\hat{x}$, $\hat{y}$, or $\hat{z}$, so the cross correlation model fails. I still hope to better understand why the population method does better, aside from it being more general.

Fascinatingly, at a concurrence threshold of 0.166 in the left side of Fig. 5.1, the population and cross correlation models have the exact same success rate to four digits across 10,000 states. This indicates that at this level of entanglement, the cross correlation model reduces to the population model; this occurs when the terms that look like $(S_{ii} + S_{jj})^2 + (S_{k0} + S_{0k})^2$, with $i, j \in \{x, y, z\}$ and $k \neq i$, $k \neq j$ are zero or close to it. This occurs only for states detected by the first two witnesses in each triplet, meaning states whose optimal Schmidt decomposition basis states are aligned with or close to a fixed axis in $\{x, y, z\}$. In future work, it would be interesting to dive into why the crossover in Fig. 5.1 exists at all.

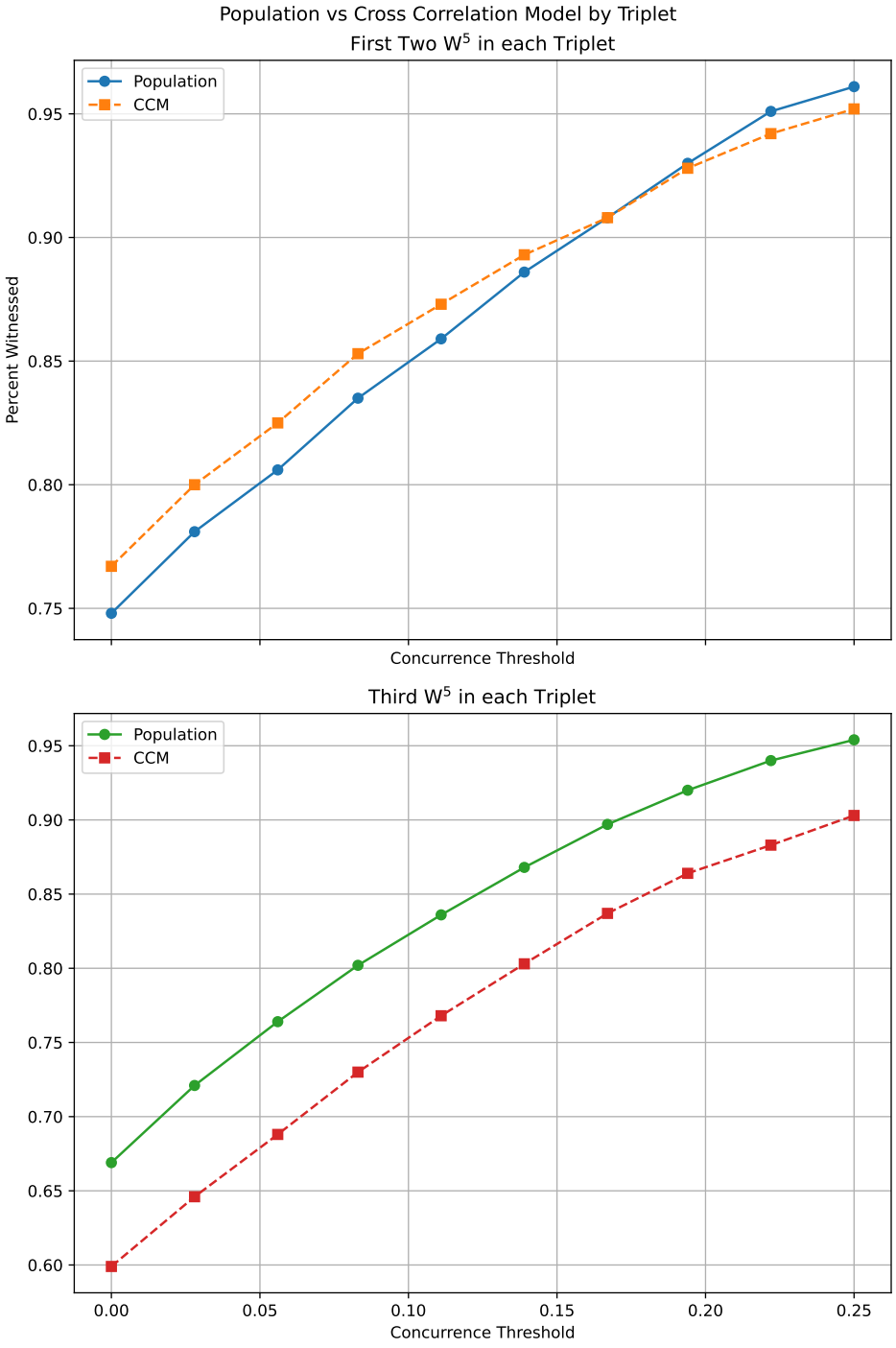


Figure 5.1: A comparison of the population and cross correlation choice models, evaluated only on states that are witnessed by the first and second W^5 of each triplet (top) or third (bottom).

5.3 Witnesses as Density Matrix Elements

Begin with the Stokes parameter constraints on whether a $\{W^5\}$ is negative, derived originally by analytically minimizing six of the nine $\{W^5\}$ witnesses:

$$\langle W_1^5\rangle < 0 \iff 1 + S_{zz} < \sqrt{(S_{xx}+S_{yy})^2 + (S_{z0}+S_{0z})^2 + (S_{xy}-S_{yx})^2}, \tag{5.3.1}$$

$$\langle W_2^5\rangle < 0 \iff 1 - S_{zz} < \sqrt{(S_{xx}-S_{yy})^2 + (S_{z0}-S_{0z})^2 + (S_{xy}+S_{yx})^2}, \tag{5.3.2}$$

$$\langle W_4^5\rangle < 0 \iff 1 + S_{xx} < \sqrt{(S_{zz}+S_{yy})^2 + (S_{x0}+S_{0x})^2 + (S_{yz}-S_{zy})^2}, \tag{5.3.3}$$

$$\langle W_5^5\rangle < 0 \iff 1 - S_{xx} < \sqrt{(S_{zz}-S_{yy})^2 + (S_{0x}-S_{x0})^2 + (S_{yz}+S_{zy})^2}, \tag{5.3.4}$$

$$\langle W_7^5\rangle < 0 \iff 1 + S_{yy} < \sqrt{(S_{zz}+S_{xx})^2 + (S_{y0}+S_{0y})^2 + (S_{zx}-S_{xz})^2}, \tag{5.3.5}$$

$$\langle W_8^5\rangle < 0 \iff 1 - S_{yy} < \sqrt{(S_{zz}-S_{xx})^2 + (S_{y0}-S_{0y})^2 + (S_{zx}+S_{xz})^2}. \tag{5.3.6}$$

Focus on the first two expressions,

$$1 + S_{zz} < \sqrt{(S_{xx}+S_{yy})^2 + (S_{z0}+S_{0z})^2 + (S_{xy}-S_{yx})^2}, \tag{5.3.7}$$

$$1 - S_{zz} < \sqrt{(S_{xx}-S_{yy})^2 + (S_{z0}-S_{0z})^2 + (S_{xy}+S_{yx})^2}, \tag{5.3.8}$$

I compute in Mathematica each Stokes parameter in terms of the elements of the following density matrix with the constraints that $\rho_{11}+\rho_{22}+\rho_{33}+\rho_{44}=1$ to conserve probability and $\rho_{ij}=\rho_{ji}^*$ from hermiticity. This yields

$$\langle W_1^5\rangle < 0 \iff |\rho_{23}|^2 > \rho_{11}\rho_{44}, \tag{5.3.9}$$

$$\langle W_2^5\rangle < 0 \iff |\rho_{14}|^2 > \rho_{22}\rho_{33}. \tag{5.3.10}$$

In words, the witnesses test whether the coherence between one pair of basis states is greater than the product of the populations of the other pair. When determining the expressions in Eq. 5.3.1 for triplets 2 and 3, I found they were much uglier. This is because the nice expressions in Eq. 5.3.9 only hold because the basis states of ρ are in the $\hat{z}$ direction. By rotational symmetry, if I write the density matrix in the (x,x) basis, the Stokes parameter labels transform as $x\to y,\ y\to z,\ z\to x$. Similarly, the density matrix in the (y,y) basis takes $x\to z,\ y\to x,\ z\to y$ if starting from the (z,z) basis. I denote a rotated density matrix

$$\rho^{i,k} = ((R_i\cdot R_k(-\alpha))\otimes(R_i\cdot R_k(-\beta)))\rho(\ldots)^\dagger, \tag{5.3.11}$$

where R_i is a rotation about axis i by $\pi/2$,

$$R_x = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & i\\ i & 1\end{pmatrix},\ R_y = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix}. \tag{5.3.12}$$

If ρ is written with no superscript, it is in the (z,z) basis. If either superscript is a 0, it means that the rotation corresponding to that superscript in not applied (is $\mathbb{I}$). It follows from this symmetry and by comparing with Eq. 5.3.1 that the remaining witnesses (just the first two in each triplet) admit the nice coherence-population expression from Eq. 5.3.9 if we rotate our density matrix to have basis states along the $\hat{x}$ direction for $\{W^5_{t1}\}$ or $\hat{y}$ direction for $\{W^5_{t2}\}$. Computing the rotations in Mathematica, I find

$$\langle W^5_1\rangle < 0 \iff |\rho_{23}|^2 > \rho_{11}\rho_{44}, \tag{5.3.13}$$
$$\langle W^5_2\rangle < 0 \iff |\rho_{14}|^2 > \rho_{22}\rho_{33}, \tag{5.3.14}$$
$$\langle W^5_4\rangle < 0 \iff |\rho^{x,0}_{23}|^2 > \rho^{x,0}_{11}\rho^{x,0}_{44}, \tag{5.3.15}$$
$$\langle W^5_5\rangle < 0 \iff |\rho^{x,0}_{14}|^2 > \rho^{x,0}_{22}\rho^{x,0}_{33}, \tag{5.3.16}$$
$$\langle W^5_7\rangle < 0 \iff |\rho^{y,0}_{23}|^2 > \rho^{y,0}_{11}\rho^{y,0}_{44}, \tag{5.3.17}$$
$$\langle W^5_8\rangle < 0 \iff |\rho^{y,0}_{14}|^2 > \rho^{y,0}_{22}\rho^{y,0}_{33}. \tag{5.3.18}$$

To express this more simply, in general,

$$\langle W^5_{ti,1}\rangle < 0 \iff |\rho^{i,0}_{23}|^2 > \rho^{i,0}_{11}\rho^{i,0}_{44}, \tag{5.3.19}$$
$$\langle W^5_{ti,2}\rangle < 0 \iff |\rho^{i,0}_{14}|^2 > \rho^{i,0}_{22}\rho^{i,0}_{33}, \tag{5.3.20}$$
$$\tag{5.3.21}$$

with $W^5_{ti,1}$ referring to the first witness in triplet i, for $i \in [(1,\hat{z}),(2,\hat{x}),(3,\hat{y})]$. I now turn to the third witness in each triplet. We know that these generalize the preceding witnesses by allowing an arbitrary rotation about a shared axis by α

for particle A and β for particle B. In Stokes parameters, these witnesses read

$$\begin{aligned}
\langle W_3^5\rangle < 0 \iff 0 > \frac{1}{4}[1 &+ \cos(\alpha)\cos(\beta)S_{xx} + \sin(\alpha)\sin(\beta)S_{yy} \\
&+ \sin(\alpha)\cos(\beta)S_{yx} + \cos(\alpha)\sin(\beta)S_{xy} \\
&- \Big[\big(S_{zz} + \cos(\alpha)\cos(\beta)S_{yy} + \sin(\alpha)\sin(\beta)S_{xx} \\
&\quad - \sin(\alpha)\cos(\beta)S_{xy} - \cos(\alpha)\sin(\beta)S_{yx}\big)^2 \\
&\quad + \big(\cos(\alpha)S_{x0} + \sin(\alpha)S_{y0} + \cos(\beta)S_{0x} + \sin(\beta)S_{0y}\big)^2\Big]^{\frac{1}{2}}] \\
\langle W_6^5\rangle < 0 \iff 0 > \frac{1}{4}[1 &+ \cos(\alpha)\cos(\beta)S_{zz} + \sin(\alpha)\sin(\beta)S_{yy} \\
&- \sin(\alpha)\cos(\beta)S_{yz} - \cos(\alpha)\sin(\beta)S_{zy} \\
&- \Big[\big(S_{xx} + \cos(\alpha)\cos(\beta)S_{yy} + \sin(\alpha)\cos(\beta)S_{zy} \\
&\quad + \cos(\alpha)\sin(\beta)S_{yz} + \sin(\alpha)\sin(\beta)S_{zz}\big)^2 \\
&\quad + \big(\cos(\alpha)S_{z0} - \sin(\alpha)S_{y0} + \cos(\beta)\S0z - \sin(\beta)S_{0y}\big)^2\Big]^{\frac{1}{2}}] \\
\langle W_9^5\rangle < 0 \iff 0 > \frac{1}{4}[1 &+ \cos(\alpha)\cos(\beta)S_{zz} + \sin(\alpha)\sin(\beta)S_{xx} \\
&+ \sin(\alpha)\cos(\beta)S_{xz} + \cos(\alpha)\sin(\beta)S_{zx} \\
&- \Big[(S_{yy} + \cos(\alpha)\cos(\beta)S_{xx} - \sin(\alpha)\cos(\beta)S_{zx} \\
&\quad - \cos(\alpha)\sin(\beta)S_{xz} + \sin(\alpha)\sin(\beta)S_{zz})^2 \\
&\quad + (\cos(\alpha)S_{z0} + \sin(\alpha)S_{x0} + \cos(\beta)S_{0z} + \sin(\beta)S_{0x})^2\Big]^{\frac{1}{2}}].
\end{aligned} \tag{5.3.22}$$

In terms of density matrix elements, calculated via Mathematica, Eq. 5.3.22 reads

$$\begin{aligned}
\langle W_3^5\rangle < 0 &\iff \mathrm{Re}(\rho_{23}^{y,z})^2 > \rho_{11}^{y,z}\rho_{44}^{y,z}, \\
\langle W_6^5\rangle < 0 &\iff \mathrm{Re}(\rho_{23}^{0,x})^2 > \rho_{11}^{0,x}\rho_{44}^{0,x}, \\
\langle W_9^5\rangle < 0 &\iff \mathrm{Re}(\rho_{23}^{0,y})^2 > \rho_{11}^{0,y}\rho_{44}^{0,y}.
\end{aligned} \tag{5.3.23}$$

While Mathematica was not able to symbolically match these expressions to Eq. 5.3.22, I verified in Mathematica that they agree for randomly chosen α, β over 10,000 randomly generated trace 1 Hermitian complex matrices. Of note, each expression can be made to look like $\mathrm{Re}(\rho_{14})^2 > \rho_{22}\rho_{33}$ if $\alpha \to \alpha + \pi$. I verified this numerically in Mathematica as well. The reasoning here is that by using an extra π angle, we are applying a Pauli operator that acts as a bit flip. An interesting note for further exploration is whether the α, β rotations allow you to rotate $\mathrm{Re}(\rho_{ij})$ into $\mathrm{Im}(\rho_{ij})$, or if not, whether new witnesses can be defined that follow that pattern.

To summarize the protocol thus far:

1. The $\{W^3\}$ detects entanglement for states which have

$$\mathrm{Re}(\rho_{14}^{i,0}) > \rho_{22}^{i,0}\rho_{33}^{i,0} \text{ or } \mathrm{Re}(\rho_{23}^{i,0}) > \rho_{11}^{i,0}\rho_{44}^{i,0},$$

   with $i \in \{\hat{x}, \hat{y}, \hat{z}\}$.

2. The first two witnesses in each $\{W^5\}$ triplet generalize by extending the real coherence piece to instead be its overall magnitude. They detect entanglement for states in which

$$|\rho_{23}^{(i,0)}|^2 > \rho_{11}^{i,0}\rho_{44}^{i,0} \text{ and/or } |\rho_{14}^{i,0}|^2 > \rho_{22}^{i,0}\rho_{33}^{i,0},$$

   with $i \in \{\hat{x}, \hat{y}, \hat{z}\}$.

3. The third witness in each $\{W^5\}$ triplet generalizes the $\{W^3\}$ by allowing an arbitrary rotation of the density matrix by angles α, β about a shared axis. The three used in the protocol detect states that follow the inequalities in Eq 5.3.23, as well as when you swap the bases that you compare coherence versus population for. These follow the general pattern

$$\mathrm{Re}(\rho_{23}^{i,j})^2 > \rho_{11}^{i,j}\rho_{44}^{i,j},$$

   with $i, j \in \{\hat{x}, \hat{y}, \hat{z}\}$.

5.4 Connections to PPT and the most general witness

In Sec. 5.1, I state that a witness in which the density matrix can be freely rotated about separate axes can detect any entangled state by testing the PPT criterion in every basis. I derive this here.

The Peres-Horodecki, or Positive Partial Transpose (PPT), criterion is a necessary condition for a bipartite state to be separable. It states that ρ is separable, or not entangled, than the eigenvalues of the partial transpose of ρ with respect to subsystem B are all non-negative. So, if ρ^Γ has any negative eigenvalues, the associated state must be entangled.

Denote $\tilde{\rho} = (U_\alpha \otimes U_\beta)\rho(U_\alpha \otimes U_\beta)^\dagger$ a density matrix rotated about arbitrary axes with respect to each subsystem, with ρ initially being have $\hat{z}$ basis states. An entanglement witness with arbitrary rotational ability detects states for which

$$|\tilde{\rho}_{23}|^2 > \tilde{\rho}_{11}\tilde{\rho}_{44}. \tag{5.4.1}$$

I show that this condition is equivalent to the PPT criterion (in other words, can be derived from it) and conversely that the PPT criterion is implied from this condition.

A principal submatrix of a square matrix is obtained by deleting some set of the same subset of rows and columns from the matrix. For example,

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & i & h \end{pmatrix} \xrightarrow[\text{delete row/column 2}]{\text{principal submatrix}} \begin{pmatrix} a & c \\ g & h \end{pmatrix}. \tag{5.4.2}$$

A principal minor is the determinant of a principal submatrix. Importantly, a matrix A is positive semidefinite if and only if all principal minors of A are non-negative [9]. Consider the partial transpose of a density matrix ρ,

$$\rho^\Gamma = \begin{pmatrix} \rho^\Gamma_{11} & \rho^\Gamma_{12} & \rho^\Gamma_{13} & \rho^\Gamma_{14} \\ \rho^\Gamma_{21} & \rho^\Gamma_{22} & \rho^\Gamma_{23} & \rho^\Gamma_{24} \\ \rho^\Gamma_{31} & \rho^\Gamma_{32} & \rho^\Gamma_{33} & \rho^\Gamma_{34} \\ \rho^\Gamma_{41} & \rho^\Gamma_{42} & \rho^\Gamma_{43} & \rho^\Gamma_{44} \end{pmatrix}. \tag{5.4.3}$$

The principal submatrix of column and row $\{1,4\}$ is

$$\rho^\Gamma\{1,4\} = \begin{pmatrix} \rho^\Gamma_{11} & \rho^\Gamma_{14} \\ \rho^\Gamma_{41} & \rho^\Gamma_{44} \end{pmatrix}, \tag{5.4.4}$$

which has principal minor

$$\begin{aligned} \det\{\rho^\Gamma\{1,4\}\} &= \rho^\Gamma_{11}\rho^\Gamma_{44} - \rho^\Gamma_{14}\rho^\Gamma_{41} && (5.4.5)\\ &= \rho_{11}\rho_{44} - \rho^\Gamma_{14}\rho^{*\Gamma}_{14} && (5.4.6)\\ &= \rho_{11}\rho_{44} - |\rho^\Gamma_{14}|^2 && (5.4.7)\\ &= \rho_{11}\rho_{44} - |\langle 00|\rho|11\rangle^\Gamma|^2 && (5.4.8)\\ &= \rho_{11}\rho_{44} - |\langle 01|\rho|10\rangle|^2 && (5.4.9)\\ &= \rho_{11}\rho_{44} - |\rho_{23}|^2. && (5.4.10) \end{aligned}$$

As the principal minor must be non-negative, we obtain the condition

$$\rho_{11}\rho_{44} \geq |\rho_{23}|^2, \tag{5.4.11}$$

which is identical to the form of $\{W_1^5\}$ save for the equals to case. The principal minor for $\{2,3\}$ gives

$$\rho_{22}\rho_{33} \geq |\rho_{14}|^2, \tag{5.4.12}$$

which is identical to $\{W_2^5\}$ save for the equals to case. Importantly, note that the eigenvalues of ρ^Γ are invariant under local unitary transformations. This implies that PPT holds for ρ if and only if it holds for $\tilde{\rho}$.

Let ρ be an entangled state. I show that the principal minor discussed above detects entanglement for arbitrary ρ. As ρ is entangled, ρ^Γ has a negative eigenvalue $\lambda < 0$. By theorem 1 of [21], λ is the only negative eigenvalue of ρ^Γ. The associated eigenvector, $|\phi\rangle$ is a vector in $\mathbb{C}^2 \otimes \mathbb{C}^2$ and so admits a Schmidt decomposition in terms of some orthonormal bases $\{a_0, a_1\}, \{b_0, b_1\}$, $|\phi\rangle = \cos\theta\,|a_0\rangle\,|b_0\rangle +$

$\sin\theta\,|a_1\rangle\,|b_2\rangle$. Apply a unitary that rotates the eigenvector into

$$|\tilde{\phi}\rangle = (U_A \otimes U_B^*)\,|\phi\rangle = \cos\theta\,|00\rangle + \sin\theta\,|11\rangle\,, \tag{5.4.13}$$

and notice that $|\tilde{\phi}\rangle$ lives entirely in the principal submatrix of $\{1,4\}$, as given in Eq. 5.4.4—call this matrix $M_{\{1,4\}}$. This implies that

$$\langle\tilde{\phi}|\tilde{\rho}^{\Gamma}|\tilde{\phi}\rangle = \langle\tilde{\phi}|M|\tilde{\phi}\rangle = \lambda < 0. \tag{5.4.14}$$

We know that this submatrix's diagonal values are real and greater than zero for ρ to be a valid density matrix, implying $\mathrm{Tr}(M_{\{1,4\}}) = \lambda + \lambda_+ \geq 0$, where the trace being non-negative implies that the second eigenvalue is positive; call it λ_+. The determinant of $M_{\{1,4\}}$ is the product of its eigenvalues and thus looks like

$$\det\{M_{\{1,4\}}\} = \lambda\lambda_+ < 0, \tag{5.4.15}$$

$$= \tilde{\rho}_{11}\tilde{\rho}_{44} - |\tilde{\rho}_{23}|^2 < 0, \tag{5.4.16}$$

from Eq. 5.4.5. This implies that

$$\tilde{\rho}_{11}\tilde{\rho}_{44} < |\tilde{\rho}_{23}|^2, \tag{5.4.17}$$

which is precisely the entanglement witnessing condition from above. This proves that there exists unitaries U_A, U_B such that the rotated density matrix of any entangled ρ can be found to satisfy Eq. 5.4.17. In the converse direction, the $\{W\}$-type witnesses satisfying Eq. 5.4.17 implies entanglement because, as we have shown, this requires that $\rho^{\Gamma} < 0$. We have thus shown that the condition on an arbitrary rotation $\{W^5\}$-type witness detecting entanglement is the same as the PPT condition for any entangled state.

Chapter 6

Conclusion

In this thesis, I explore the optimal performance of an adaptive entanglement witnessing protocol meant to detect the entanglement of unknown bipartite quantum states of two qubits from few measurements. I define a lower bound on the success rate of a Bayes Classifier model for the optimal adaptive choice and estimate its value using a k-dimensional tree to bin a large data set. I analytically minimize the majority of two sets of optimal decomposable entanglement witnesses to improve protocol run time and resolve inaccurate witness value evaluation due to numerical minimization. I use quantities derived from these analytically minimized expressions to bin the k-dimensional tree, finding that the corresponding estimated Bayes classifier success rate is equal to our current best choice method to 0.02% across all computationally generated random states, and that it exceeds our best choice model on states with a concurrence greater than 0.05.

I also show that the population model—an analytic adaptive choice model that was heuristically derived by previous group members [30]—works by maximizing the magnitude of certain quantities in the analytic minimization of these entanglement witnesses. Furthermore, I use these minimizations to derive a pair of entangled states that have identical inputs to the adaptive choice model but require different outputs, thus proving the existence of an information-theoretic bound on the adaptive choice model's success rate. I then use this logic to place a bound on how well an experimental system could implement our adaptive witnessing protocol given it is only able to measure a state's Stokes parameters to some accuracy ϵ.

The analytic minimization also opens the doors to a new understanding of the structure of these entanglement witnesses. When expressed as density matrix elements, I prove that the analytically minimized witness expressions are equivalent to the PPT criterion when restricted to only representing the density matrix in a constrained set of basis states and in some cases, by weakening the PPT criterion. I interpret this result geometrically, showing that our protocol is

unable to witness any entangled state for which the eigenvector associated with the negative eigenvalue of the state's partially transposed density matrix has basis vectors of a Schmidt decomposition that do not lie on or near great circles in the xy, yz, or, xz planes on the Bloch sphere of all possible Schmidt decomposition basis vectors. I give reasoning for why is the only mechanism by which our protocol is unable to detect an entangled state, thus fully characterizing what states our protocol can and cannot detect.

There are multiple avenues that warrant deeper exploration. For one, the question remains of why the protocol's limited set of witnesses is sufficient to cover all states detected by extremal decomposable witnesses of the form extrapolated from Riccardi et al. [22]. Furthermore, it would be interesting to confirm whether the current protocol's constrained set of witnesses is the best possible set for an adaptive protocol of this kind, especially considering if we drop the current Stokes parameter exchange symmetry present in our entanglement witnesses. It would also be interesting to apply the gap-based bounding model to an experimental system to see the effects of noise on our protocol's success rate. Separately, there are also interesting mathematical reformulations of the Bayes classifier that could be attempted, such as restating the Bayes success as a volume ratio of fiber bundles from the space of all 15 Stokes parameters to the 9 measured for the $\{W^3\}$ witness set.

Bibliography

[1] Charles H. Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K. Wootters, *Teleporting an unknown quantum state via dual classical and einstein-podolsky-rosen channels*, Phys. Rev. Lett. **70** (1993), 1895–1899.

[2] Charles H. Bennett, Gilles Brassard, and N. David Mermin, *Quantum cryptography without bell's theorem*, Phys. Rev. Lett. **68** (1992), 557–559.

[3] Charles H Bennett and Stephen J Wiesner, *Communication via one-and two-particle operators on einstein-podolsky-rosen states*, Physical review letters **69** (1992), no. 20, 2881.

[4] Lukas Biewald, *Experiment tracking with weights and biases*, 2020, Software available from wandb.com.

[5] François Chollet et al., *Keras*, `https://keras.io`, 2015.

[6] Dariusz Chruściński and Gniewomir Sarbicki, *Entanglement witnesses: construction, analysis and classification*, Journal of Physics A: Mathematical and Theoretical **47** (2014), no. 48, 483001.

[7] Luc Devroye, László Györfi, and Gábor Lugosi, *A probabilistic theory of pattern recognition*, Springer New York, NY, 1996.

[8] Artur K Ekert, *Quantum cryptography based on bell's theorem*, Physical review letters **67** (1991), no. 6, 661.

[9] F.R. Gantmacher, *The theory of matrices. vol i.*, Chelsea Publishing Company, 1977.

[10] Otfried Gühne and Géza Tóth, *Entanglement detection*, Physics Reports **474** (2009), no. 1-6, 1–75.

[11] Sam A. Hill and William K. Wootters, *Entanglement of a pair of quantum bits*, Phys. Rev. Lett. **78** (1997), 5022–5025.

[12] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki, *Separability of mixed states: necessary and sufficient conditions*, Physics Letters A **223** (1996), no. 1, 1–8.

[13] Ryszard Horodecki and Michal/ Horodecki, *Information-theoretic aspects of inseparability of mixed states*, Phys. Rev. A **54** (1996), 1838–1843.

[14] M. Lewenstein, B. Kraus, J. I. Cirac, and P. Horodecki, *Optimization of entanglement witnesses*, Phys. Rev. A **62** (2000), 052310.

[15] Sheng-Kai Liao, Wen-Qi Cai, Wei-Yue Liu, Liang Zhang, Yang Li, Ji-Gang Ren, Juan Yin, Qi Shen, Yuan Cao, Zheng-Ping Li, et al., *Satellite-to-ground quantum key distribution*, Nature **549** (2017), no. 7670, 43–47.

[16] Xiao Liu, Jun Hu, Zong-Feng Li, Xue Li, Pei-Yun Li, Peng-Jun Liang, Zong-Quan Zhou, Chuan-Feng Li, and Guang-Can Guo, *Heralded entanglement distribution between two absorptive quantum memories*, Nature **594** (2021), no. 7861, 41–45.

[17] Jonas Maziero, *Random sampling of quantum states: a sur- vey of methods*, Brazilian Journal of Physics **11** (2015), 575–583.

[18] Paco Navarro, *A geometric approach to adaptive entanglement witnesses*, (2025).

[19] Asher Peres, *Separability criterion for density matrices*, Phys. Rev. Lett. **77** (1996), 1413–1415.

[20] John Preskill, *Lecture notes for ph219/cs219: Quantum information and computation*, 2015, Chapter 2, Section 2.4.

[21] Swapan Rana, *Negative eigenvalues of partial transposition of arbitrary bipartite states*, Phys. Rev. A **87** (2013), 054301.

[22] Alberto Riccardi, Dariusz Chruściński, and Chiara Macchiavello, *Optimal entanglement witnesses from limited local measurements*, Physical Review A **101** (2020), no. 6, 062319.

[23] Jan Roik, Karol Bartkiewicz, Antonín Černoch, and Karel Lemr, *Accuracy of Entanglement Detection via Artificial Neural Networks and Human-Designed Entanglement Witnesses*, Physical Review Applied **15** (2021), no. 5, 054006, Publisher: American Physical Society.

[24] Tobias Schmale, Moritz Reh, and Martin Gärttner, *Efficient quantum state tomography with convolutional neural networks*, npj Quantum Information **8** (2022), no. 1, 115.

[25] Oscar Scholin, Alec Roberson, and Richard Zheng, *Summer 2023 writeup*, internal document (2023).

[26] David S Simon, Gregg Jaeger, and Alexander V Sergienko, *Entangled-coherent-state quantum key distribution with entanglement witnessing*, Physical Review A **89** (2014), no. 1, 012315.

[27] Barbara M. Terhal, *Detecting quantum entanglement*, Theoretical Computer Science **287** (2002), no. 1, 313–335, Natural Computing.

[28] Giuseppe Viola, Nikolai Miklin, Mariami Gachechiladze, and Marcin Pawłowski, *Entanglement witnessing with untrusted detectors*, Journal of Physics A: Mathematical and Theoretical **56** (2023), no. 42, 425301.

[29] Wikimedia Commons, *Bloch sphere representations*, https://commons.wikimedia.org/w/index.php?curid=179712998, 2025, Accessed: 2026-04-23.

[30] Eritas Yang, Becca Verghese, and Ben Hartley, *Entanglement witness write-up*, internal document (2022).

[31] Juan Yin, Yu-Huai Li, Sheng-Kai Liao, Meng Yang, Yuan Cao, Liang Zhang, Ji-Gang Ren, Wen-Qi Cai, Wei-Yue Liu, Shuang-Lin Li, et al., *Entanglement-based secure quantum cryptography over 1,120 kilometres*, Nature **582** (2020), no. 7813, 501–505.

[32] Karol Zyczkowski and Hans-Jurgen Sommers, *Average fidelity between random quantum states*, Physical Review A **71** (2003).