The Role of Entanglement in Quantum Communication, and Analysis of the Detection Loophole

Thomas Cope

Doctor of Philosophy

University of York

Computer Science

September 2018

Dedication

To Ellen, Judith and Andrew

Abstract

Entanglement is a feature at the heart of quantum information. Its enablement of unusual correlations between particles drives a new wave of communication and computation. This thesis explores some of the ways in which the tools for studying entanglement can be used to quantify the transmission of quantum information, and compares the use of different techniques.

We begin this thesis by expanding the technique of teleportation simulation, which adds noise to the entangled resource state to mimic channel effects. By introducing classical noise in the communication step, we show it is possible to simulate more than just Pauli channels using teleportation. This new class is characterised, and studied in detail for a particular resource state, leading to a family of simulable channels named "Pauli-Damping channels" whose properties are analysed.

Also introduced are a new family of quantum states, "phase Werner" states, whose entanglement properties relate to the interesting conjecture of bound entangled states with a negative partial transpose. Holevo-Werner channels, to which these states are connected, are shown to be teleportation covariant. We exploit this to present several interesting results, including the optimal estimation of the channel-defining parameter. The minimal binary-discrimination error for Holevo-Werner channels is bounded for the first time with the analytical form of the quantum Chernoff bound. We also consider the secret key capacity of these channels, showing how different entanglement measures provide a better upper bound for different regions of these channels.

Finally, a method for generating new Bell inequalities is presented, exploiting non-physical probability distributions to obtain new inequalities. Tens of thousands of new inequivalent inequalities are generated, and their usefulness in closing the detection loophole for imperfect detectors is examined, with comparison to the current optimal construction. Two candidate Bell inequalities which may equal or beat the best construction are presented.

Contents

List of Figures

List of Tables

Acknowledgements

I am very grateful to my supervisors, Stefano Pirandola and Roger Colbeck - without them this thesis would not be possible, and their insights have been more than invaluable. My thanks also go to my peers - Riccardo, Panos, Vicky, Peter and Oliver, whose discussions and support have been helpful at every step of the process. This gratitude is extended to Carlo, Cosmo, Gaetana, Sammy and Mirjam, whose experience has been wholly beneficial. This too is accurate of my co-authors Leon, Leonardo and Kenneth; thank you for working with me. Finally, the support of my friends and especially my family has been vital, and never taken for granted.

Declaration

I declare that the research described in this thesis is original work, which I undertook at the University of York during 2015 - 2018. This work has not previously been presented for an award at this, or any other, university. Except where stated, all of the work contained within this thesis represents the original contribution of the author.

Some of the material in this thesis has been published in journals. The author of this thesis acknowledges the input of his collaborators, and has credited them appropriately throughout. A list of papers which overlap with this thesis are presented here.

- *Simulation of non-Pauli Channels*, Thomas Cope, Leon Hetzel, Leonardo Banchi and Stefano Pirandola. Published in Physical Review A **96** (2017). [30]

- *Adaptive estimation and discrimination of Holevo-Werner channels*, Thomas Cope and Stefano Pirandola. Published in Quantum Measurements and Quantum Metrology **4** (2017). [31]

- *Converse bounds for quantum and private communication over Holevo-Werner channels*, Thomas Cope, Kenneth Goodenough and Stefano Pirandola. Published in Journal of Physics A **51** (2018). [29]

Also relevant is the publication,

- *Theory of channel simulation and bounds for private communication*, Stefano Pirandola, Samuel Braunstein, Riccardo Laurenza, Carlo Ottaviani, Thomas Cope, Gaetana Spedalieri and Leonardo Banchi. Published in Quantum Science and Technology **3** (2018). [82]

Chapter 1

Foreword and Preliminaries

1.1 Foreword

The field of quantum mechanics emerged around a century ago, a reaction to new experiments which defied accepted physical theories. The work of Maxwell and Hertz had shown how light behaved as a wave, building on Young's double slit experiment. Yet observed phenomena such as black body radiation and the photoelectric effect contradicted this stance. Furthermore, discoveries were made at the atomic level suggesting a Newtonian planetary-like structure, but with strictly discrete orbital energies. From these observations emerged the idea that energy came in discrete "quanta" and that these quanta could show wave-like behaviour; an idea then extended to matter particles. From these concepts emerged mathematical frameworks, such as Schrödinger's wave functions and Heisenberg's uncertainty principle, to form the basis of quantum theory.

The theory predicted many unusual and counter-intuitive effects, with several concepts controversial even amongst the pioneers of quantum mechanics. Einstein, a Nobel prize winner for his resolution of the photoelectric effect, along with Podolsky and Rosen presented the "EPR paradox" [39], showing the predictions of quantum theory appeared to violate the principles of relativity; measurement of the position of one particle could seemingly disturb another particle distantly separated, such that a measurement of the second particle's momentum became uncertain. They proposed underlying "hidden" variables, with the uncertainty emerging only due to our lack of knowledge of the true physical description. This argument was investigated by John Bell [8], who proved the physical assumptions made by Einstein could not match all quantum behaviours. Ex-

perimental results since have all ruled against these hidden variables [3, 45, 54, 93, 99]. Schrödinger also took issue with the *interpretation* of the theory - illustrated with the famous "Schrödinger's cat" thought experiment, an intended absurdity which has become a legitimate philosophical argument, as well as a common reference point. It was here that the term "entanglement" was coined, to describe systems of particles which cannot be fully described independently of each other.

From these origins has arisen a field known as "quantum information". Abstracted away from particles and waves, instead the counter-intuitive properties are embedded within a language of Hilbert spaces, vectors, and linear operators, developed by pioneers such as von Neumann, Dirac and Weyl. Inspired by the work of Claude Shannon, who developed information theory to understand and quantify information storage, transmission and extraction, researchers in quantum information seek to study these same problems utilising quantum states. Whilst Shannon's work used the bit, a binary digit taking value either 0 or 1, as its fundamental unit of information, quantum information instead relies on the qubit, a quantum state allowing for a superposition of both 0 and 1.

Areas of research in quantum information include quantum computation, quantum cryptography and quantum communication, and in all three of those fields, many important developments are promised. These promises have been taken seriously - quantum random number generators are commerically available, whilst quantum computer chips are in development by Google, IBM and Microsoft, and governments have invested huge sums in order to be the first to develop quantum technologies for economic and security purposes. Alongside this, hundreds of academics worldwide are looking to understand the possibilities and limitations, trying to enact them in laboratories worldwide - and even beyond, to satellites orbiting Earth.

In order to achieve these amazing new developments, entangled states are often required as a resource. In the field of quantum communication, much work has been done on how to establish entanglement between remote parties, and the rate at which this can be done. Recently though, entanglement has obtained a new role as a *quantifier* of communication. Connections between quantum states and channels have allowed the communication abilities of certain channels to be determined via the entanglement properties

of related quantum states. It is this new role of entanglement theory with which the majority of this thesis concerns itself.

Chapter 1 is a primer for the rest of the thesis. It introduces the mathematical language used to understand the field of quantum information, explaining how we describe quantum states, measurements and channels. It also introduces the measures used to quantify entanglement and the properties desirable for such a measure. Also discussed are capacities of quantum channels, measures which describe the rate at which useful resources such as qubit states or entanglement may be sent or established through use of the channel.

Chapter 2 looks at the process of quantum teleportation. First introduced in [11], it allows one remote party to perfectly transmit a quantum state to another, by consuming a maximally entangled state pre-shared between them and transmitting some classical information. It was observed that by replacing the resource state shared between the parties by a less entangled state, the effect on the transmitted state was to distort it, as if it had physically been transmitted through a noisy quantum channel. In effect, the channel has been "simulated", and a connection can be made between the capacity of the channel and the entanglement of the resource state. Whilst this protocol may only simulate Pauli channels, we show that by adding classical noise into the communication step, one may expand the range of channels one can simulate, and provide a necessary criterion for non-Pauli properties. We then focus on a specific resource state, characterising the set of possible simulable channels.

Chapter 3 introduces the reader to the Werner states, a highly symmetrical family of quantum states which has been used multiple times to highlight surprising properties of quantum information theory. After explaining some unusual entanglement properties of these states, we introduce a generalisation called "phase Werner" states, and present a potential connection to the important NPT bound entanglement conjecture, including a class of phase Werner states which may exhibit this property.

Chapter 4 looks at the properties of Holevo-Werner channels, quantum channels related to Werner states by channel-state duality. By showing these channels satisfy a property known as "teleportation covariance", we are able to bound how accurately one is able

3

to estimate the channel-defining parameter via quantum measurement, either when the parameter is completely unknown or known to be one of two possible values. To do this, we establish the analytical form of the quantum Chernoff bound for Werner and isotropic states. We also look at the secret key capacity of Holevo-Werner channels, which quantifies the rate at which secure classical bits may be established between parties via this channel. By exploiting the unusual entanglement of Werner states, we are able to show the novel feature that different entanglement measures provide a better bound on this capacity, depending on the channel parameter.

Chapter 5 focuses instead on the topic of *non-locality* - the property shown by Bell to separate quantum mechanics from Einstein's hidden variable theories. Whilst entanglement is necessary for non-locality, it is not sufficient, and proving the existence of non-locality is paramount for the most secure quantum cryptography protocols. We show how one can exploit non-physical extremal no-signalling distributions in order to generate new Bell inequalities, which certify the existence of non-locality, having done so to create tens of thousands of new inequivalent inequalities. These new inequalities are then analysed to determine their usefulness in tackling the "detection loophole", a cryptographically-relevant problem in which maliciously preprogrammed distributions may appear non-local, and therefore cryptographically secure, by measurement failure. Given are two Bell inequalities which appear to match or better the detection efficiency of the current optimal construction.

Finally, Chapter 6 summarises the results and implications of the work presented, and discusses the limitations and potential further directions for future research.

1.2   Structure of this Chapter

The first section of this chapter will introduce the very basics of discrete variable quantum information, defining states, measurements and channels, along with concepts such as purity and entanglement. It will finish with two famous protocols; teleportation and key distribution, in order to show what is possible with such tools. The second section introduces the idea of entanglement measures; what such a measure requires, what properties one would like it to have, and the common measures we use. Following this, a discussion of the connections between different measures is given. This is then followed by the same

treatment for the capacities of quantum channels.

1.3 The Mathematics of Quantum Information

In order to understand an area of scientific research, we must first have a firm grasp of the underlying mathematical structure that we work with. The field of quantum information is split into two camps, *discrete* variable and *continous* variable. Discrete models have a finite number of basis states, of which the system can exists in any superposition; however, measurement of the state in any given basis can only produce a finite numbers of outcomes. By far the most studied discrete model is the qubit, in which the basis is made up of two states, explained in subsection 1.3.1. By contrast, continuous variable quantum information allows measurements to have a continuous spectrum of possible outcomes. Normally continuous variable systems are characterised by the two *quadratures*, $\hat{q}$ and $\hat{p}$, for each particle, which can be understood as position and momentum respectively. This thesis shall focus on discrete variable quantum information.

1.3.1 The Qubit

In classical computational theory, the fundamental unit of information is the *bit*, a single binary value 0 or 1. In quantum theory the basic unit of information is the *qubit* - the qubit can also be in the state $|0\rangle$ or $|1\rangle$, but also in any *superposition* of the two: a linear combination

$$|\phi\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle , \quad |\alpha_0|^2 + |\alpha_1|^2 = 1 \tag{1.1}$$

with α_0, α_1 complex.

The state $|\phi\rangle$ is not in the state $|0\rangle$ nor $|1\rangle$ - when measured it will collapse into one of the two, with probability $|\alpha_0|^2$ and $|\alpha_1|^2$ respectively. These qubit states are complex vectors $|\phi\rangle = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix}$ living in the Hilbert space $\mathcal{H}_2$, with the inner product between two states $|\phi_1\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle , |\phi_2\rangle = \beta_0 |0\rangle + \beta_1 |1\rangle$ given by

$$\langle\phi_1|\phi_2\rangle = \alpha_0^* \beta_0 + \alpha_1^* \beta_1 \tag{1.2}$$

where $\langle\phi_1| = (\alpha_0^*, \alpha_1^*)$ is the *conjugate transpose* $|\phi_1\rangle^\dagger$. Note that that this is a *complex* inner product, $\langle\phi_2|\phi_1\rangle = \langle\phi_1|\phi_2\rangle^*$. However, we always have that states $|\phi\rangle$ satisfy $\langle\phi|\phi\rangle = 1$. This idea is generalised to *qudits* - states with basis states $|0\rangle \ldots |d-1\rangle$, and

complex coefficients $\alpha_0 \ldots \alpha_{d-1}$ with $\langle \phi | \phi \rangle = \sum_i |\alpha_i|^2 = 1$.

Since these states represent physical systems, a natural question is how to combine them. Suppose I have two non-interacting particles, one in state $|\phi_A\rangle = \sum_i \alpha_i |i\rangle_A$ and one in state $|\phi_B\rangle = \sum_j \beta_j |j\rangle_B$ - is there a way I can describe the two together? This is indeed possible by making use of the *tensor product* $\otimes$:

$$|\phi_A \phi_B\rangle = |\phi_A\rangle \otimes |\phi_B\rangle = \sum_{i,j} \alpha_i \beta_j |i\rangle_A \otimes |j\rangle_B . \tag{1.3}$$

Normally we omit the $\otimes$ and group them together within one ket[1] bracket. Given $|\phi_A\rangle \in \mathcal{H_A}$, $|\phi_B\rangle \in \mathcal{H_B}$, we say the new state $|\phi_A\phi_B\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, or $\mathcal{H}_{AB}$ for short. One of the most important features of quantum theory arises from the fact that this product does *not* describe all allowable states on $\mathcal{H}_{AB}$ - clarified by the following definition.

Definition 1.3.1 A bipartite state $|\phi\rangle \in \mathcal{H}_{AB}$ is called *separable* if there exists two states $|\phi_A\rangle \in \mathcal{H}_A$, $|\phi_B\rangle \in \mathcal{H}_B$ such that $|\phi\rangle = |\phi_A\rangle \otimes |\phi_B\rangle$. If this is not possible, then $|\phi\rangle$ is said to be *entangled.*

Here it is useful to note the following distinction: in this thesis $|\phi\rangle_A$ and $|\phi\rangle_B$ represent the same state, on two different Hilbert spaces; whereas $|\phi_A\rangle$ and $|\phi_B\rangle$ generally represent two different states (which may or may not belong to different Hilbert spaces).

The most important entangled state is the *Bell pair*, or *maximally entangled state*,

$$|\Phi^+\rangle := \frac{|00\rangle + |11\rangle}{\sqrt{2}}. \tag{1.4}$$

Measuring one subsystem of this state in the $\{|0\rangle, |1\rangle\}$ basis leads to $|0\rangle$ half the time, after which the state collapses into $|00\rangle$, and $|1\rangle$ the other half, giving state $|11\rangle$. The significance of this is that if the second party also measures the $\{|0\rangle, |1\rangle\}$ basis, they will obtain a perfectly correlated outcome to the first party's measurement, despite having no prior knowledge of its outcome! This unusual feature forms the basis of many quantum cryptographic protocols, as we shall see later on.

Theorem 1.3.2 (Schmidt Decomposition) For any pure state $|\phi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, there

---

[1]The term bra refers to $\langle \cdot |$, and ket $| \cdot \rangle$ - together making a "bra-ket".

exists a set of orthonormal states $\{|i_A\rangle\}$ on $\mathcal{H}_A$ and $\{|i_B\rangle\}$ on $\mathcal{H}_B$ such that

$$|\phi\rangle = \sum_i \lambda_i |i_A\rangle |i_B\rangle \qquad (1.5)$$

where λ_i are non-negative real numbers which satisfy

$$\sum_i \lambda_i^2 = 1$$

and are known as *Schmidt* coefficients. Due to the orthonormality condition, $|\{\lambda_i\}| \leq \min\{\dim[\mathcal{H}_A], \dim[\mathcal{H}_B]\}$ with $|\{\lambda_i\}| = 1$ iff $|\phi\rangle$ is separable.

The number of non-zero λ_i is known as the *Schmidt rank*.

1.3.2 Measuring States

We have touched upon measurement and "collapse" of a quantum state, but now we define it more rigorously.

Definition 1.3.3 A quantum measurement is a collection of measurement operators $\{M_m\}$ which are linear operators acting on a Hilbert space. The probability of obtaining outcome m is given by $\langle\phi| M_m^\dagger M_m |\phi\rangle$, where $\dagger$ is the conjugate transpose. The state after outcome m is given by

$$|\phi_m\rangle = \frac{M_m |\phi\rangle}{\sqrt{\langle\phi| M_m^\dagger M_m |\phi\rangle}}. \qquad (1.6)$$

These $\{M_m\}$ must satisfy $\sum_m M_m^\dagger M_m = \mathrm{I}$.

The set of linear operators on $\mathcal{H}$ is denoted $\mathcal{L}(\mathcal{H})$, and operators can always be represented as matrices. In fact for any given orthonormal basis $\{|e_k\rangle\}$ a linear operator can be represented as $A_{kl} = \langle e_k| A |e_l\rangle$.

Definition 1.3.4 A *Projective Valued Measure* (PVM) is a measurement $\{P_m\}$ such that $P_k P_l = \delta_{kl} P_k$.

The state after measurement is therefore

$$|\phi_m\rangle = \frac{P_m |\phi\rangle}{\||P_m |\phi\rangle\|}, \qquad (1.7)$$

where $\||\phi\rangle\| = \sqrt{\langle\phi|\phi\rangle}$.

Often we are not concerned with the post-measurement state; in which case we can use a Positive Operator-Valued Measure (POVM). This is where we take $E_m = M_m^\dagger M_m$ as the elements, so that our conditions are $\sum_m E_m = \mathrm{I}$, and our probabilities $p(m) = \langle\phi| E_m |\phi\rangle$.

1.3.3 Mixed States and Density Operators

The formalism introduced so far is useful, but does not fully describe all the possibilities of quantum theory. Suppose I flip a coin and prepare state $|\phi_H\rangle$ if I obtain a heads, and $|\phi_T\rangle$ if tails. I then give the resulting state to you, without telling you the outcome of the coin flip. How would you describe the state? One would be tempted to write it as

$$|\phi_{\text{coin}}\rangle = \frac{1}{2}|\phi_H\rangle + \frac{1}{2}|\phi_T\rangle \tag{1.8}$$

but now calculating the inner product $\langle\phi_{\text{coin}}|\phi_{\text{coin}}\rangle$ we find

$$\begin{aligned}
\langle\phi_{\text{coin}}|\phi_{\text{coin}}\rangle &= \left(\frac{1}{2}\langle\phi_H| + \frac{1}{2}\langle\phi_T|\right)\left(\frac{1}{2}|\phi_H\rangle + \frac{1}{2}|\phi_T\rangle\right) \\
&= \frac{1}{4}\left(\langle\phi_H|\phi_H\rangle + \langle\phi_T|\phi_T\rangle + \langle\phi_H|\phi_T\rangle + \langle\phi_T|\phi_H\rangle\right) \\
&= \frac{1}{2} + \frac{\langle\phi_H|\phi_T\rangle + \langle\phi_T|\phi_H\rangle}{2} \neq 1
\end{aligned} \tag{1.9}$$

when $|\phi_H\rangle \neq |\phi_T\rangle$. Thus $|\phi_{\text{coin}}\rangle$ is *not* a valid quantum state. In order to describe these states, we must instead turn to *density operators*.

Definition 1.3.5 A density operator $\rho \in \mathcal{L}(\mathcal{H}_d)$ is expressed:

$$\rho = \sum_i p_i |\phi_i\rangle\langle\phi_i|, \ p_i \geq 0, \ \sum_i p_i = 1 \tag{1.10}$$

with each $|\phi_i\rangle \in \mathcal{H}_d$. They are normally represented as matrices, with $\rho_{kl} = \langle e_k|\rho|e_l\rangle$.

In our example previously, the correct way to describe the state would be

$$\rho_{\text{coin}} = \frac{1}{2}|\phi_H\rangle\langle\phi_H| + \frac{1}{2}|\phi_T\rangle\langle\phi_T|. \tag{1.11}$$

Normally we will use as shorthand $\rho \in \mathcal{H}_d$ to refer to a d-dimensional density matrix.

Lemma 1.3.6 Density matrices are:

- Hermitian: $\rho^\dagger = \rho$,

- Trace 1: $\text{Tr}[\rho] = 1$,

- Positive Semidefinite $\langle v|\rho|v\rangle \geq 0, \forall |v\rangle$.

Proof.

- Hermitian:

$$\rho^\dagger = \left(\sum_i p_i \, |\phi_i\rangle \, \langle\phi_i| \right)^\dagger = \sum_i p_i \, \langle\phi_i|^\dagger \, |\phi_i\rangle^\dagger = \sum_i p_i \, |\phi_i\rangle \, \langle\phi_i| = \rho. \qquad (1.12)$$

- Trace 1:

$$\begin{aligned}
\mathrm{Tr}[\rho] &= \sum_k \langle e_k| \, \rho \, |e_k\rangle \\
&= \sum_k \langle e_k| \left(\sum_i p_i \, |\phi_i\rangle \, \langle\phi_i| \right) |e_k\rangle \\
&= \sum_i p_i \sum_k \langle e_k|\phi_i\rangle \, \langle\phi_i|e_k\rangle \\
&= \sum_i p_i \||\phi_i\rangle\|^2 = \sum_i p_i = 1. \qquad (1.13)
\end{aligned}$$

- Positive Semidefinite:

$$\langle v| \, \rho \, |v\rangle = \langle v| \left(\sum_i p_i \, |\phi_i\rangle \, \langle\phi_i| \right) |v\rangle = \sum_i p_i \, \langle v|\phi_i\rangle \, \langle\phi_i|v\rangle = \sum_i p_i |\langle v|\phi_i\rangle|^2 \geq 0.$$

$$(1.14)$$

$$\square$$

If a matrix ρ is positive semidefinite, we express this as $\rho \geq 0$. A matrix is positive semidefinite iff its eigenvalues $\{\lambda_i\}$ satisfy $\lambda_i \geq 0$, $\forall i$.

Definition 1.3.7 A state ρ is said to be *pure* iff it can be written $\rho = |\phi\rangle \, \langle\phi|$ for some state $|\phi\rangle$. Else the state is said to be *mixed*. If a state is pure, then it can be equivalently described as ρ or $|\phi\rangle$.

One can see how for pure states, the trace 1 condition reduces to $\||\phi\rangle\| = 1$. For mixed states, the state norm $\||\phi\rangle\|$ generalises to the *trace norm*:

Definition 1.3.8 The trace norm is defined:

$$\|\rho\|_1 := \mathrm{Tr}\left[\sqrt{\rho\rho^\dagger} \right] \qquad (1.15)$$

Definition 1.3.9 The purity of a state ρ is given by $\mathrm{Tr}[\rho^2]$, with ρ pure iff $\mathrm{Tr}[\rho^2] = 1$.

The state for which the purity is minimised is the state known as the *maximally mixed* state, I_d/d. For this state the purity is $1/d$.

We can extend the concept of entanglement to density matrices.

Definition 1.3.10 A state $\rho \in \mathcal{H}_{AB}$ is separable iff it can be written in the form

$$\rho = \sum_i q_i \rho_A^i \otimes \rho_B^i, \ q_i \geq 0, \ \sum_i q_i = 1, \tag{1.16}$$

where $\rho_A^i = |\phi_i\rangle \langle\phi_i| \in \mathcal{H}_A$, $\rho_B^i = |\psi_i\rangle \langle\psi_i| \in \mathcal{H}_B$. Else ρ is said to be entangled.

The tensor product on density matrices is the natural extension from pure states:

$$\sum_i p_i \, |\phi_i\rangle \langle\phi_i| \otimes \sum_i r_j \, |\psi_j\rangle \langle\psi_j| = \sum_{i,j} p_i r_j \, |\phi_i\psi_j\rangle \langle\phi_i\psi_j| \,. \tag{1.17}$$

Given a bipartite density matrix $\rho_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$, we are often interested in what the reduced state is for just one subsystem; referred to as $\rho_A \in \mathcal{H}_A$ or $\rho_B \in \mathcal{H}_B$ respectively. we can do this via the *partial trace*.

Definition 1.3.11 The partial trace with respect to B is defined as

$$\mathrm{Tr}_B[|a_i\rangle \langle a_j| \otimes |b_k\rangle \langle b_l|] := |a_i\rangle \langle a_j| \, \mathrm{Tr}[|b_k\rangle \langle b_l|] \equiv \langle b_l|b_k\rangle \, |a_i\rangle \langle a_j| \tag{1.18}$$

and generalised to $\mathcal{H}_A \otimes \mathcal{H}_B$ by linearity. Partial trace with respect to A is defined analogously. For a bipartite state $\rho_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$, the *reduced state* on A is given by

$$\rho_A := \mathrm{Tr}_B[\rho_{AB}]. \tag{1.19}$$

1.3.3.1 Measuring Mixed States

Now that we have expanded our definition of quantum states, we need to expand how we perform measurements. Luckily the measurement operators remain unchanged - we only need to change how to apply them. The probability of measurement m is given by $\mathrm{Tr}[M_m^\dagger M_m \rho]$, whilst the postmeasurement state is

$$\rho_m = \frac{M_m \rho M_m^\dagger}{\mathrm{Tr}[M_m^\dagger M_m \rho]}. \tag{1.20}$$

Similarly, we can apply POVMs to density matrices: the probabilities we obtain are simply $p(m) = \mathrm{Tr}[E_m \rho]$.

1.3.4 Quantum Operations

Aside from measurement, there are other processes which can occur to a quantum state. The simplest of these is a unitary operation $U : \rho \to U\rho U^\dagger$, where U is a $d \times d$ unitary matrix. These are reversible operations: if $\rho' = U\rho U^\dagger$, then we may simply apply

$$U^\dagger : \rho' \to U^\dagger \rho' U = U^\dagger U \rho U^\dagger U = \rho \tag{1.21}$$

to return to our original state.

This is not the most general class of operations however; we want to consider all trace-preserving[2] completely positive maps - that is, maps $\mathcal{E}$ which satisfy

$$\mathrm{Tr}[\mathcal{E}\left(\rho\right)] = \mathrm{Tr}[\rho] = 1 \tag{1.22}$$

and

$$\forall n \in \mathbb{N}, \rho \geq 0 \Rightarrow (\mathbb{I}_n \otimes \mathcal{E})\left(\rho\right) \geq 0. \tag{1.23}$$

These requirements are simply stating that valid quantum states should be mapped to valid quantum states. The second condition (complete positivity) may seem unusual - why not require that the map is simply positive? However, one should be able to apply such a map to a subsystem of any quantum state and still obtain a valid *overall* state - it is this property that complete positivity embodies.

Maps satisfying the above criteria are known as *quantum channels*. They can be expressed as :

$$\mathcal{E}\left(\rho\right) = \sum_i K_i \rho K_i^\dagger, \tag{1.24}$$

with $\sum_i K_i^\dagger K_i = \mathrm{I}$. If $\mathcal{E} : \mathcal{H}_{d_1} \to \mathcal{H}_{d_2}$, then the K_i are of size $d_2 \times d_1$, and $\sum_i K_i^\dagger K_i = \mathrm{I}_{d_1}$.

In general, these channels are not reversible, and often we are interested in the quantum channel which accurately describes the noise (unwanted operations) that occurs when trying to send/store a quantum state.

---

[2]Of course, one can also consider operations which do not preserve trace, but are "trace non-increasing", allowing for sub-normalised states - but we do not here.

1.3.4.1 Quantum Channel Examples

- **Bit flip channel**

  Perhaps the simplest quantum channel, which has its roots in classical information theory. With probability $1 - p$ the state $|0\rangle$ is sent to $|1\rangle$ (and vice versa) whilst with probability p it is left unchanged. In the form (1.24) , we have $K_0 = \sqrt{p}\,\mathrm{I}$, $K_1 = \sqrt{1-p}\sigma_x$, where σ_x is the Pauli matrix

  $$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{1.25}$$

  the unitary which flips $|0\rangle$ and $|1\rangle$. Unless $p = 0, 1$ this channel is not unitary, and thus not reversible.

- **Depolarising channel**

  Another common channel is the depolarising channel. With probability p, the quantum state is transmitted successfully, whilst with probability $1 - p$ the state is completely lost - replaced by the maximally mixed state I/d. Thus

  $$\mathcal{D}_p := p\rho + (1 - p)\frac{\mathrm{I}}{d}. \tag{1.26}$$

  For $d = 2$, there are four Kraus operators:

  $$K_0 = \sqrt{\frac{1-3p}{4}}\mathrm{I} = \sqrt{\frac{1-3p}{4}}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad K_1 = \sqrt{\frac{p}{4}}\sigma_x = \sqrt{\frac{p}{4}}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

  $$K_2 = \sqrt{\frac{p}{4}}\sigma_y = \sqrt{\frac{p}{4}}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad K_3 = \sqrt{\frac{p}{4}}\sigma_z = \sqrt{\frac{p}{4}}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

  The three Pauli matrices $\sigma_x, \sigma_y, \sigma_z$ play an important role in qubit quantum theory, as we shall see later.

- **Amplitude damping channel**

  Another useful channel is the amplitude damping channel - it has an important operational interpretation, of the decay of an excited atom. If the atom is ground state $|0\rangle$, then it cannot decay further. However, if it is in the excited state $|1\rangle$ then it has probability γ of losing a photon into the environment and dropping to the

ground state, and probability $1 - \gamma$ of staying in the excited state. The action on both the atom and the environment is:

$$|0\rangle_A |0\rangle_E \rightarrow |0\rangle_A |0\rangle_E \,,$$

$$|1\rangle_A |0\rangle_E \rightarrow \sqrt{\gamma} |0\rangle_A |1\rangle_E + \sqrt{1 - \gamma} |1\rangle_A |0\rangle_E \,.$$

The action on the atomic system can be described as a channel with Kraus operators:

$$K_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1 - \gamma} \end{pmatrix}, \qquad K_1 = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix}. \qquad (1.27)$$

Unlike the previous examples, the Kraus operators are not Pauli matrices - a point we shall return to in chapter 2.

1.3.4.2 Local Operations and Classical Communication

A particularly important class of quantum operations is that of local operations and classical communication (LOCC). In this scenario, two parties Alice and Bob share a quantum state $\rho_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$, with Alice owning $\mathcal{H}_A$ and Bob $\mathcal{H}_B$. The class of LOCC operations are all possible processes they can do whilst being confined to separate laboratories. They can transform their subsystem, send them through channels, prepare auxillary states and measure them jointly with their subsystem - anything so long as it is confined to local Hilbert spaces. The classical communication allows them to co-ordinate their efforts - for example, they could perform the operation

$$\sum_i p_i \left(\mathcal{E}_A^i \otimes \mathcal{E}_B^i \right) (\rho_{AB}) \,, \ p_i \geq 0, \ \sum_i p_i = 1 \qquad (1.28)$$

by Alice sampling a probability distribution, and communicating the result to Bob. We shall see much more of this general class of operations throughout the thesis.

1.3.5 Quantum Protocols

In this section I shall outline two important quantum protocols, which serve to illustrate and motivate some of the research in this thesis.

1.3.5.1 Teleportation

Perhaps one of the most interesting protocols within quantum information theory - and one that always provokes a reaction when explaining one's research to strangers to the

field - is the concept of *teleportation* [11]. In this scenario, Alice owns a qubit state $|\phi\rangle_C = \alpha_0 |0\rangle + \alpha_1 |1\rangle$ she wishes to send to Bob - but they do not share a quantum channel. Fortunately, they have a classical line of communication between the two of them[3], and they had hitherto shared between themselves a maximally entangled state, $|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$. With this, it is possible for Alice to *perfectly* transmit $|\phi\rangle$ to Bob. The protocol is as follows:

1. First consider the overall state

$$|\phi\rangle_C \otimes |\Phi^+\rangle_{AB} = (\alpha_0 |0\rangle + \alpha_1 |1\rangle)_C \otimes \frac{|00\rangle_{AB} + |11\rangle_{AB}}{\sqrt{2}}$$

$$= \frac{1}{\sqrt{2}} \left(\alpha_0 |00\rangle_{CA} |0\rangle_B + \alpha_0 |01\rangle_{CA} |1\rangle_B + \alpha_1 |10\rangle_{CA} |0\rangle_B + |11\rangle_{CA} |1\rangle_B \right).$$

$$\text{(1.29)}$$

   We will re-express this overall state using a different basis for Alice: the *Bell* basis. This is a two-qubit orthonormal basis consisting of the four maximally entangled two-qubit states:

$$|\Phi^+\rangle = \mathrm{I} \otimes \mathrm{I} |\Phi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}} \qquad |\Phi^-\rangle = \mathrm{I} \otimes \sigma_z |\Phi^+\rangle = \frac{|00\rangle - |11\rangle}{\sqrt{2}}$$

$$|\Psi^+\rangle = \mathrm{I} \otimes \sigma_x |\Phi^+\rangle = \frac{|01\rangle + |10\rangle}{\sqrt{2}} \qquad |\Psi^-\rangle = \mathrm{I} \otimes i\sigma_y |\Phi^+\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}$$

   and we can express

$$|00\rangle = \frac{|\Phi^+\rangle + |\Phi^-\rangle}{\sqrt{2}}, \qquad\qquad |01\rangle = \frac{|\Psi^+\rangle + |\Psi^-\rangle}{\sqrt{2}},$$

$$|10\rangle = \frac{|\Psi^+\rangle - |\Psi^-\rangle}{\sqrt{2}}, \qquad\qquad |11\rangle = \frac{|\Phi^+\rangle - |\Phi^-\rangle}{\sqrt{2}}.$$

   This means we can rewrite our state in Eq. (1.29) as:

$$\frac{1}{2} \Bigg(|\Phi^+\rangle_{CA} \otimes (\alpha_0 |0\rangle + \alpha_1 |1\rangle)_B + |\Psi^-\rangle_{CA} \otimes (\alpha_0 |0\rangle - \alpha_1 |1\rangle)_B$$

$$+ |\Psi^+\rangle_{CA} \otimes (\alpha_0 |1\rangle + \alpha_1 |0\rangle)_B + |\Psi^+\rangle_{CA} \otimes (-\alpha_0 |1\rangle + \alpha_1 |0\rangle)_B \Bigg). \qquad \text{(1.30)}$$

2. Alice performs a projective measurement on her subsystem $\mathcal{H}_{CA}$ in the Bell basis - we can see from our rewriting that this will collapse the superposition of the four terms in Eq. (1.30) into just one of them, each with equal probability.

3. Alice then communicates her outcome over the classical communication channel - this will take two classical bits (as there are four possible outcomes).

---

[3]Or just Alice to Bob is sufficient.

4. Dependent on Alice's outcome, Bob performs one of the following unitaries:

   (a) Alice obtained $|\Phi^+\rangle$:

   Bob's state is of the form $(\alpha_0 |0\rangle + \alpha_1 |1\rangle)_B$ - this is exactly $|\phi\rangle$ and Bob performs no unitary/ the identity unitary.

   (b) Alice obtained $|\Phi^-\rangle$:

   Bob's state is of the form $(\alpha_0 |0\rangle - \alpha_1 |1\rangle)_B$ - by performing the operator σ_z, Bob obtains

   $$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \alpha_0 \\ -\alpha_1 \end{pmatrix} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = |\phi\rangle . \tag{1.31}$$

   (c) Alice obtained $|\Psi^+\rangle$:

   Bob's state is of the form $(\alpha_1 |0\rangle + \alpha_0 |1\rangle)_B$ - by performing the operator σ_x, Bob obtains

   $$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_0 \end{pmatrix} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = |\phi\rangle . \tag{1.32}$$

   (d) Alice obtained $|\Psi^-\rangle$:

   Bob's state is of the form $(\alpha_1 |0\rangle - \alpha_0 |1\rangle)_B$ - by performing the operator $i\sigma_y$, Bob obtains

   $$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ -\alpha_0 \end{pmatrix} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = |\phi\rangle . \tag{1.33}$$

   Thus regardless of Alice's outcome, Bob is able to recover $|\phi\rangle$.

Although we have only described this for pure states in detail, the protocol works for general mixed and/or entangled states. Moreover, by using the *generalised Pauli matrices* $\left\{ X_d^i, Z_d^j \right\}_{i,j=0}^{d-1}$, where

$$[X_d]_{kl} = \delta_{k,(l-1)} \bmod d, \qquad\qquad [Z_d]_{kl} = e^{\frac{2\pi i (k-1)}{d}} \delta_{kl} \tag{1.34}$$

one may generalise this protocol for qudits. For this, the measurement performed is the projection onto the d^2 states $\left\{ X_d^i Z_d^j |\Phi\rangle_d \right\}_{i,j}$ with $|\Phi\rangle_d = \sum_{i=0}^{d} |ii\rangle / \sqrt{d}$ the d-dimensional maximally entangled state.

There are a few interesting things to note about this protocol - the first, that it is an LOCC protocol - Alice performed a local measurement, Bob a local unitary, and there was classical communication only. The second, that Alice ends up owning one of the

four maximally entangled states, each with probability 1/4 - her measurement told her no information about $|\phi\rangle$, and she no longer holds a copy of $|\phi\rangle$. Finally, in order to send one arbitrary qubit, one "ebit" (a maximally entangled Bell pair) and two bits were required.

1.3.5.2 BB84 - Key Distribution

Another important application of quantum information is key distribution for cryptography purposes. The current schemes used today are classical techniques relying on the difficulty of certain number theory problems, primarily integer factorisation. However, in 1994 Peter Shor proposed [95] an algorithm which solves this problem in polynomial time - the gold standard for algorithms. This algorithm is a quantum algorithm i.e. it requires a quantum computer to run. Whilst this means communication is secure for the time being[4], with the development of practical quantum computers we must look to an alternate solution. One such scheme, which is utilised by many experimentalists, is the BB84 protocol [10], introduced by Bennett and Brassard. The protocol works as follows:

1. Alice and Bob share a perfect quantum channel $\mathbb{I} : \mathcal{H}_A \to \mathcal{H}_B$, and an unsecure classical channel. Alice generates two uniformly random bits, and depending on the result prepares one of four states:

$$
\begin{array}{cc}
\text{bits} & \text{state} \\
00 & |0\rangle \\
01 & |1\rangle \\
10 & |+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}} \\
11 & |-\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}}
\end{array}
\tag{1.35}
$$

the first bit refers to the basis, whilst the second refers to the state prepared.

2. Alice then transmits the state to Bob.

3. Bob then randomly choose a basis, $\{|0\rangle, |1\rangle\}$ or $\{|+\rangle, |-\rangle\}$, and measures - taking $|0\rangle$ and $|+\rangle$ as outcome 0 and $|1\rangle$ and $|-\rangle$ as outcome 1.

4. Alice and Bob repeat steps 1-3 many times - until Bob has an n-bit string. Bob then communicates to Alice that he has received the states.

---

[4]Provided no-one finds a non-quantum polynomial algorithm for factorisation.

5. Alice then communicates to Bob the first bit of each random pair; i.e. which basis the state was prepared in. Bob checks each of his choices, and discards all bits where the basis does not match.

6. Of the remaining bits, Alice randomly selects 1/2 of them, telling Bob their position and value. Bob checks to see if they coincide (which they should, having being measured in the same basis as they were prepared in) - if they pass up to a certain error tolerance - then they use the remaining bits to send a message by use of a *one time pad*.

The one time pad is a classical security method in which, for a message of length m bits, m uniformly random bits are added to the message via an XOR operation. If the m bits are used for only a single message, and kept completely secret, then this message is "perfectly secure" - all possible messages have a unique mapping to the encrypted string, and are thus all equally likely - without the added bit string, no information can be learned about the original message [94].

Why can we trust this protocol? There are three possible attacks to the protocol.

- **"The man in the middle" attack.**
  Our malicious eavesdropper, Eve, poses as Alice to Bob, and Bob to Alice, performing the protocol with each. This is a fundamental attack to all cryptography schemes, and thus we require the quantum channel is "authenticated" - it really goes from Alice to Bob.

- **Eve copies the state then passes it on to Bob, waiting for Alice's basis choices.**
  In classical communication, this attack would be devastating. In quantum theory however, this is made impossible by the "no-cloning theorem" [112].

  Theorem 1.3.12 (No-Cloning Theorem) There exists no unitary such that for arbitrary $|\phi\rangle$,

  $$U |\phi\rangle |r\rangle = |\phi\rangle |\phi\rangle \tag{1.36}$$

  where $|r\rangle$ is an arbitrary resource state.

  Proof. Consider the inner product of two states, $\langle\phi|\psi\rangle$. As $|r\rangle$ is a valid quantum state, $\langle\phi|\psi\rangle = \langle\phi|\psi\rangle \times 1 = \langle\phi|\psi\rangle \langle r|r\rangle$. If such a cloning unitary exists, then we could

write

$$\langle\phi|\psi\rangle\langle r|r\rangle = \langle\phi r|\psi r\rangle = \langle\phi r|\,U^\dagger U\,|\phi r\rangle - \langle\phi|\,\langle\phi|\psi\rangle\,|\psi\rangle = \langle\phi|\psi\rangle^2 \qquad (1.37)$$

this forces either $\langle\phi|\psi\rangle = 1$ or $\langle\phi|\psi\rangle = 0$, a contradiction to our arbitrary state assumption. $\square$

This proof generalises to mixed states and general operations; the important thing is Eve cannot copy the states and pass them on.

- **Eve measures the states, copying it and then passing it on to Bob.**

  In this case, Eve *can* clone her post-measurement state, as she may simply prepare the same state. However, to do this she must choose a basis to measure in. If she chooses the same basis as Alice (which occurs with probability 1/2) then she can do this without Bob detecting her measurement. However, if she chooses the wrong basis - for example, if she measures the state $|+\rangle$ in the $\{|0\rangle, |1\rangle\}$ basis she will collapse the state into either $|0\rangle = (|+\rangle + |-\rangle)/\sqrt{2}$ or $|1\rangle = (|+\rangle - |-\rangle)/\sqrt{2}$. Thus when Bob makes his measurement, if he chooses the $\{|+\rangle, |-\rangle\}$ basis, there is a 1/2 chance he will obtain $|-\rangle$, a result which would be impossible had there been no interference! Thus if a single such result (although in practice due to practical considerations, there is usually a tolerable number of fails) is found in the check, Alice and Bob should abort the protocol. Note that the protocol only establishes a secret key, so no sensitive information has been leaked.

1.3.6 Relations between dimensions

The final part of this first section states a few important results regarding the *expansion* of lower dimensional spaces to higher ones.

Theorem 1.3.13 (Purification) Given a state ρ_A on $\mathcal{H}_A$, it is always possible to define a pure state $|\phi\rangle_{AR}$ on $\mathcal{H}_A \otimes \mathcal{H}_R$, such that

$$\rho_A = \mathrm{Tr}\left(|\phi\rangle_{AR}\langle\phi|\right). \qquad (1.38)$$

System R is known the *reference* system.

Proof. Suppose that $\rho_A = \sum_i p_i |i_A\rangle\langle i_A|$, where $\{|i_A\rangle\}$ is an orthonormal basis for $\mathcal{H}_A$ and $\sum_i p_i = 1$ (we can do this for any state). Then introduce a Hilbert space $\mathcal{H}_R$ of the

same dimension as $\mathcal{H}_A$, with an orthonormal basis $\{|i_R\rangle\}$. We can define a pure state

$$|\phi\rangle_{AR} := \sum_i \sqrt{p_i}\,|i_A\rangle\,|i_R\rangle. \tag{1.39}$$

Then

$$
\begin{aligned}
\mathrm{Tr}_R\left(|\phi\rangle_{AR}\langle\phi|\right) &= \sum_{i,j} \sqrt{p_i p_j}\,|i_A\rangle\,\langle j_A|\,\mathrm{Tr}(|i_R\rangle\,\langle j_R|) \\
&= \sum_{i,j} \sqrt{p_i p_j}\,|i_A\rangle\,\langle j_A|\,\delta_{ij} \\
&= \sum_i p_i\,|i_A\rangle\,\langle i_A| \\
&= \rho_A.
\end{aligned}
$$

$\square$

Theorem 1.3.14 (Naimark Extension) Any POVM can be realised as a projective measurement, on a higher dimensional Hilbert space.

This theorem is proved in similar manner to the purification theorem above.

Theorem 1.3.15 (Stinespring's Dilation Theorem) Every channel $\mathcal{E} : \mathcal{H}_A \otimes \mathcal{H}_B$ can be seen as a transformation of the form

$$\mathcal{E}\left(\rho\right) = \mathrm{Tr}_{E'}\left[U_{AE}\left(\rho \otimes |r\rangle_E\,\langle r|\right)U_{AE}^\dagger\right] \tag{1.40}$$

with $|r\rangle \in \mathcal{H}_E$ an ancilla state, with $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_E \equiv \mathcal{H}_B \otimes \mathcal{H}_{E'}$. That is, every channel can be seen as a restriction of a unitary transformation on a larger Hilbert space.

1.4 Measures of Entanglement

Since entanglement is such an important part of quantum theory, we would like to be able to accurately quantify it, and compare whether one state is "more entangled" than another. In this section we shall introduce some of the important entanglement measures [61], focusing on the ones that we use later on in the thesis. Before this however, we would like to lay out some key features [88] that a good measure of entanglement should have. Suppose we have an entanglement measure $E\left(\rho\right)$ - we would like the following:

- *Separable states σ have $E(\sigma) = 0$.*

  Since entangled states are defined as states which are not separable, this seems like a very reasonable requirement.

- *Entanglement should not increase under LOCC operations*: For Λ, an LOCC, $E(\Lambda(\rho)) \leq E(\rho)$.

  A more operational requirement; it is easy to show that entangled states cannot be created using LOCC operations, but this statement is slightly stronger, since one could envision an "entanglement expansion" protocol. However, imagine it is possible to create state σ from ρ by a LOCC operation. Then any protocol we could do with σ (and LOCC operations) we could also do with ρ, first by transforming it using LOCC into σ and then performing the same operation. Thus we cannot say σ is more entangled than ρ, because as a resource ρ is at least as useful as σ.

- *Entanglement is invariant under local unitaries.*

  Since local unitary operations are invertible, this follows as a consequence of the previous requirement.

- *There exist maximally entangled states $|\Phi\rangle_d$ such that $E(\rho) \leq E(|\Phi\rangle_d \langle\Phi|)$, $\forall \rho \in \mathcal{H}_d \otimes \mathcal{H}_d$.*

  In particular, for bipartite systems, it should be the state[5] $|\Phi\rangle_d = \sum_{i=0}^{d-1} |ii\rangle / \sqrt{d}$ - this is because *any* other state on $\mathcal{H}_d \otimes \mathcal{H}_d$ can be formed using this state and LOCC operations. This property is limited to the bipartite case; for example the "W state" $(|001\rangle + |010\rangle + |100\rangle) / \sqrt{3}$ and the "GHZ state" $(|000\rangle + |111\rangle) / \sqrt{2}$ are both entangled, yet it has been shown there exists a partition between entangled states one can create via LOCC from the two [37]. Thus the concept of a maximally entangled state is much more nuanced for multipartite entanglement. This thesis is restricted to the bipartite scenario.

A measure that satisfies the above properties is normally referred to as an "entanglement monotone" - since it decreases monotonically under LOCC. There are other requirements

---

[5]As well as all states local unitarily equivalent.

which whilst not necessary, it would be desirable for our measure to have. These include:

- $E(\rho)$ *is well-defined for all states ρ.*

  You may be surprised that this is not on the list above, but this is not the case; there exist entanglement measures which are monotones on pure states but do not satisfy the above conditions when extended to mixed states. Thus we restrict their use to the treatment of pure states only. An example of this is the *von Neumann entropy*, which we shall see explicitly later on in this section.

- *Convexity*: $E(\sum_i p_i \rho_i) \leq \sum_i p_i E(\rho_i)$.

  Whilst there are authors who make arguments for this property based on information about the system, the primary advantage of this property is simply mathematical, as it allows for easier testing of monotonicity and calculation for mixed states.

- *Additivity*: weak: $E(\rho^{\otimes n}) = nE(\rho)$ and strong: $E(\rho \otimes \sigma) = E(\rho) + E(\sigma)$.

  This property is also extremely useful mathematically, allowing for higher dimensional tensor product states to be analysed more easily, and initially seems to make sense operationally - given two copies of an entangled state, it seems reasonable to achieve twice as much. This is a misleading thought though, as it fails to take into account joint local operations - for example, Alice could transform jointly her two subsystems by a unitary $U_{A_1 A_2} \neq U_{A_1} \otimes U_{A_2}$. Many important measures of entanglement fail this condition, even in its weaker form - whilst the strong condition is only satisfied by a few relevant measures.

- *Continuity*: $\forall \, \epsilon > 0 \, \exists \, \delta$ s.t. $\forall \rho, \rho', \, \|\rho - \rho'\|_1 \leq \delta \Rightarrow |E(\rho) - E(\rho')| \leq \epsilon$. This property is a nice mathematical property for functions to have, and also makes sense intuitively - we would expect two quantum states which are very similar to have roughly equal amounts of entanglement. Moreover, as the distance between states gets smaller and smaller then so too should the difference in entanglement between them. In fact, is is desirable for our measure to have a *stronger* form on continuity, known as *asymptotic continuity* - this requires that, given $\|\rho - \rho'\|_1 \leq \delta$, the measure satisfies a "Fannes type" inequality

$$|E(\rho) - E(\rho)| \leq k\delta \log d + O(\delta), \tag{1.41}$$

with k some constant and $O(\delta)$ a function dependent only on δ, which vanishes as $\delta \to 0$. This trait is desirable as many entanglement measures are considered in the limit of an infinite number of copies of the state - as the dimension of the space goes to infinity, we wish that the continuity distance grows sufficiently slowly that the "regularised version" $\lim_{n\to\infty} E(\rho^{\otimes n})/n$ is also continuous.

Let us have a look at some such measures.

- **Von Neumann entropy**

  For $\rho = |\phi\rangle \langle \phi|$, we can define the von Neumann entropy measure as:

  $$E_S\left(\rho\right) := S(\text{Tr}_B\left[\rho\right]) := -\text{Tr}\left[\rho_A \log \rho_A\right]. \tag{1.42}$$

  For a separable state $|\phi\rangle = |\phi_A\rangle \otimes |\phi_B\rangle$, the subsystem $|\phi_A\rangle$ is pure - and so the von Neumann entropy is 0. For the maximally entangled state $|\Phi\rangle_d = \sum_{i=0}^{d-1} |ii\rangle /\sqrt{d}$ the reduced state is the d dimensional maximally mixed state I_d/d - and so we obtain $\log d$ as our entanglement. This property of entanglement measures (where $E(|\Phi\rangle_d) = \log d$ is known as *normalisation*, and is common amongst measures. Whilst a good measure for pure states, it is simple to see that the d^2 maximally mixed state I_{d^2}/d^2 also gives value $\log d$, yet clearly $I_{d^2}/d^2 = I_d/d \otimes I_d/d$. Thus this measure is insufficient for general purposes - although defined for all states, it only provides an entanglement measure for pure states.

- **Distillable entanglement**

  We have seen already that the maximally entangled state is the "most useful" state, in that any LOCC protocol over any state can be done instead using the maximally entangled state as a starting resource. Most protocols are qubit based, and thus one often wants as many copies of $|\Phi\rangle_2 \equiv |\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ as possible. This idea gives rise to the notion of entanglement distillation - how efficiently can we convert copies of some less entangled state ρ into fewer, but maximally entangled, copies of the state $|\Phi\rangle_2$? Formally:

  $$E_D\left(\rho\right) := \sup\left\{r : \lim_{n\to\infty}\left(\inf_{\Lambda\in\text{LOCC}} \|\Lambda\left(\rho^{\otimes n}\right) - |\Phi\rangle_2 \langle\Phi|^{\otimes rn}\|_1\right) = 0\right\}. \tag{1.43}$$

  This measure reduces to the von Neumann entropy when considering pure states, but generally is difficult to calculate - a frustrating outcome given its operational usefulness.

One could ask instead the distillable entanglement using maximally entangled states $|\Phi\rangle_d$ - however, the states $|\Phi\rangle_2^{\otimes rn}$ and $|\Phi\rangle_{2^{rn}}$ are local unitarily equivalent; thus the $|\Phi\rangle_d$ distillation rate is simply $E_D / (\log d)$.

- **Entanglement cost**

  Entanglement cost can be thought of as a counterpart of the distillable entanglement, going in the reverse direction. This operational cost instead looks at the optimum rate at which our maximally entangled Bell states $|\Phi\rangle_2$ may be converted into a given state via LOCC operations. Mathematically:

  $$E_C\left(\rho\right) := \sup\left\{r : \lim_{n\to\infty}\left(\inf_{\Lambda\in\mathrm{LOCC}}\|\Lambda\left(|\Phi\rangle_2\langle\Phi|^{\otimes rn}\right) - \rho^{\otimes n}\|_1\right) = 0\right\}. \qquad (1.44)$$

  Unsurprisingly, this too is difficult to compute, due to the generality of LOCC operations. For pure states though, this reduces to von Neumann entropy once more.

- **Entanglement of formation**

  In an effort to simplify the entanglement cost, instead we may consider the entanglement of formation,

  $$E_F\left(\rho\right) := \inf\left\{\sum_i p_i E_S(|\phi_i\rangle\langle\phi_i|) \,\middle|\, \sum_i p_i |\phi_i\rangle\langle\phi_i| = \rho\right\} \qquad (1.45)$$

  which we can see looks to construct the state ρ from pure states with the minimal total entanglement. This measure only looks at a single copy of the state however, whereas entanglement cost is considered in the limit of $n \to \infty$ copies. Due to this, we are introduced to our first *regularised* measure

  $$E_F^\infty\left(\rho\right) := \lim_{n\to\infty}\frac{E_F\left(\rho^{\otimes n}\right)}{n}. \qquad (1.46)$$

  This regularised entanglement of formation was shown to satisfy $E_F^\infty\left(\rho\right) = E_C(\rho)$, and it was hoped that E_F would be additive, since it seemed to be for many ρ where both E_F and E_C were known. Ultimately though, this was shown not to be the case [52], [61] and the difficulty of LOCC optimisation seen in calculating the entanglement cost is replaced by the difficulty in regularising E_F.

- **Relative entropy of entanglement**

  This measure is one we shall repeatedly apply throughout the thesis, due to some

of the results in the literature we take advantage of. This measure aims to look at the *distance*[6] between the state and its closest separable state counterpart, by using the quantum relative entropy $S(\rho \| \sigma) = \mathrm{Tr}\left[\rho \log \rho - \rho \log \sigma\right]$. It is defined in the following way

$$E_R(\rho) := \min_{\sigma \in \mathrm{Sep}} S(\rho \| \sigma). \tag{1.47}$$

Like the entanglement of formation, the relative entropy of entanglement (REE) is also subadditive over tensor products - we shall explore this property in detail later in the thesis by looking at *Werner states* - the first such states shown to exhibit subadditivity in this measure. Thus we also have a regularised version

$$E_R^\infty(\rho) := \lim_{n \to \infty} \frac{E_R\left(\rho^{\otimes n}\right)}{n}. \tag{1.48}$$

- **Relative entropy of entanglement with respect to PPT states**

  The concept of relative entropy can easily be generalised to other sets, rather than simply separable states - as long as the set is closed under LOCC operations, then we have a valid entanglement measure. One of the most used instances of this is the set of states with *positive partial transpose* (PPT); that is $\sigma^{T_B} := (\mathbb{I} \otimes T)(\sigma) \geq 0$. This is a necessary condition [81] for separability, but is only sufficient [59] for $\mathcal{H}_2 \otimes \mathcal{H}_2$ and $\mathcal{H}_2 \otimes \mathcal{H}_3$. The formal definition of the relative entropy of entanglement with respect to positive partial transpose states (RPPT) is

  $$E_P(\rho) := \min_{\sigma, \; \sigma^{T_B} \geq 0} S(\rho \| \sigma), \tag{1.49}$$

  and like the REE admits a regularised version

  $$E_P^\infty(\rho) := \lim_{n \to \infty} \frac{E_P\left(\rho^n\right)}{n}. \tag{1.50}$$

- **Negativity and logarithmic negativity**

  As mentioned above, a partial positive transpose is a necessary condition for separable states. Thus a simple measure of entanglement is simply the absolute sum of the negative eigenvalues of ρ^{T_B}:

  $$N(\rho) := \sum_i \frac{|\lambda_i| - \lambda_i}{2} \tag{1.51}$$

---

[6]This is not a true distance, since $S(\rho \| \sigma) \neq S(\sigma \| \rho)$.

with $\{\lambda_i\}$ the spectrum of ρ^{T_B}.

This measure is convenient due to its simplicity in calculation, but is not additive over tensor products. To achieve this, a related measure, the *logarithmic negativity* was defined:

$$E_N\left(\rho\right) := \log\left(\sum_i |\lambda_i|\right) \equiv \log\left(2N(\rho) + 1\right) \tag{1.52}$$

which retains the convenience of calculation. Unfortunately, E_N is not asymptotically continuous - we shall see the significance of this limitation in the next section.

- **Squashed entanglement**

  Another measure we shall employ later in this thesis; it was motivated by ideas in classical cryptography. It is defined as

  $$E_{\text{sq}}\left(\rho\right) := \frac{1}{2} \min_{\rho'_{ABE} \in \Omega_{AB}} S(A:B|E) \tag{1.53}$$

  where Ω_{AB} is the set of density matrices ρ'_{ABE} such that $\text{Tr}_E[\rho'_{ABE}] = \rho \in \mathcal{H}_{AB}$. The ancilla space $\mathcal{H}_E$ may be of any (even infinite) dimension. The function $S(A:B|E)$ is the quantum conditional mutual information

  $$S(A:B|E) := S(\rho'_{AE}) + S(\rho'_{BE}) - S(\rho'_E) - S(\rho'_{ABE}) \tag{1.54}$$

  where $S(\rho'_{AE})$ denote the von Neumann entropy of the reduced state $\rho'_{AE} = \text{Tr}_B[\rho'_{ABE}]$. It looks to quantify how much may be learned about the full state from only the ancilla dimensions. The minimisation over all possible extensions makes this a very difficult measure to find in general - however, it is additive over tensor products, and can always be upper bounded by choosing a specific extension.

- **Key Distillation Rate**

  Another operational measure, $K(\rho)$, it is inspired from one the most important applications of quantum information theory - quantum key distribution. It quantifies the rate at which the two parties holding the entangled state may obtain shared secret bits[7] i.e. known to both of them, but not to any external eavesdroppers. Like the distillable entanglement, this measure is taken in the asymptotic limit $\lim_{n\to\infty} \rho^{\otimes n}$.

---

[7]Classical bits, not qubits!

The formal definition for this rate is:

$$K\left(\rho\right) := \sup\left\{r : \lim_{n\to\infty}\left(\inf_{\Lambda\in\text{LOCC}}\|\Lambda\left(\rho^{\otimes n}\right) - \Psi^{\otimes rn}\|_1\right) = 0\right\}. \qquad (1.55)$$

where Ψ is a "private state" - a state of the form:

$$\Psi = U_T\left(|\Phi\rangle_2\langle\Phi| \otimes \tau_{A'B'}\right)U_T^\dagger \qquad (1.56)$$

where $|\Phi\rangle_2$ is the maximally entangled state, and $\tau_{A'B'}$ is an arbitrary state often referred to as the "shield state". The unitary U_T is known as "twisting unitary" and takes the form:

$$U_T = \sum_{i,j=0}^{1}|ij\rangle_{AB}\langle ij| \otimes U_{A'B'}^{ij} \qquad (1.57)$$

where the unitaries U^{ij} are arbitrary. These states have the property that for any purification $|\Phi\rangle_{AA'BB'E}$, if Alice and Bob measure in the computational basis and trace out subsystems A', B', the resultant state is

$$\Phi_{\text{CCQ}} = \frac{1}{2}\left(|0\rangle\langle 0| + |1\rangle\langle 1|\right) \otimes \tau_E', \qquad (1.58)$$

where we see that Φ_{CCQ} is a *correlated classical quantum* (CCQ) state where Alice and Bob share one correlated classical bit, which is completely uncorrelated with the quantum state on the *environment* state on $\mathcal{H}_E$. Thus, even if a malicious eavesdropper controls the environment, she can learn nothing about their shared bit. It should be noted that the maximally entangled state itself is an example of a private state.

Alternatively, one could consider an alternate definition of $K(\rho)$ in which $|\Phi\rangle_d$ is used to define a private state instead - however, it has been shown that the two definitions coincide [19].

1.4.1 Relations Between Entanglement Measures

In this section, we present some relations between entanglement measures. Whilst a lot of these measures are interesting in their own right, many of them were created to better understand and quantify the operational entanglement measures - since at the heart of quantum information is the desire to understand and apply physical processes. In particular, we seek to understand their relationship to the entanglement cost and distillable entanglement. Before that however, we shall look at their application on *pure states only*.

Theorem 1.4.1 ([36]) If an entanglement measure E satisfies

- $E(\sigma) = 0$ for σ separable;

- $E(|\Phi\rangle_d) = \log d$;

- $E(\Lambda(\rho)) \leq E(\rho)$ for LOCC operations Λ;

- E is asymptotically continuous;

then the regularised version $E^{\infty}(\rho) := \lim_{n \to} E(\rho)/n$ coincides with the von Neumann entropy on pure states. Moreover, if E is weakly additive, then $E = E^{\infty}$ and coincides with the von Neumann entropy on pure states.

Most of the above measures satisfy these conditions: both entanglement cost and distillable entanglement do this as mentioned before; so too does the entanglement of formation (trivially), and both the regularised REE and RPPT. The logarithmic negativity does *not* however - but squashed entanglement does. In much of the literature the term "entanglement measure" is reserved for entanglement monotones satisfying this property; we have not kept to that here, but it is an important distinction worth noting.

Theorem 1.4.2 ([36]) If an entanglement measure E satisfies

- $E(|\Phi\rangle_d) = \log d$;

- $E(\Lambda(\rho)) \leq E(\rho)$ for LOCC operations Λ;

- E is asymptotically continuous[8];

- E is convex;

if E is weakly additive, then $E_D(\rho) \leq E(\rho) \leq E_C(\rho)$, and if E is subadditive, then $E_D(\rho) \leq E^{\infty}(\rho) \leq E_C(\rho)$.

It shall be useful in later chapters to understand the proof behind this theorem, in regards to upper bounding the distillable entanglement, and therefore we shall present it here.

Proof. We begin with the definition of entanglement distillation,

$$E_D(\rho) := \sup \left\{ r : \lim_{n \to \infty} \left(\inf_{\Lambda} \| \Lambda(\rho^{\otimes n}) - |\Phi\rangle_2 \langle\Phi|^{\otimes rn} \|_1 \right) = 0 \right\}. \qquad (1.59)$$

---

[8]Actually, it is sufficient for this property to be satisfied for pure states only.

Consider a specific LOCC distillation protocol, Λ with rate r. This protocol must distill n copies of ρ into rn copies of $|\Phi\rangle_2$, with some error ϵ, which vanishes in the limit $n \to \infty$. Formally, we have that:

$$\|\Lambda\left(\rho^{\otimes n}\right) - |\Phi\rangle_2\langle\Phi|^{\otimes rn}\|_1 \leq \epsilon. \tag{1.60}$$

Suppose we have an entanglement measure E which satisfies the above conditions; we may then apply asymptotic continuity[9]:

$$\|\Lambda\left(\rho^{\otimes n}\right) - |\Phi\rangle_2\langle\Phi|^{\otimes rn}\|_1 \leq \epsilon \Rightarrow |E(\Lambda\left(\rho^{\otimes n}\right) - E(|\Phi\rangle_2\langle\Phi|)^{\otimes rn}| \leq k\epsilon\log 2^{rn} + O(\epsilon)$$

$$\Rightarrow E(|\Phi\rangle_2\langle\Phi|^{\otimes rn}) \leq E(\Lambda\left(\rho^{\otimes n}\right)) + k\epsilon rn + O(\epsilon) \tag{1.61}$$

where $O(\epsilon)$ is a function that vanishes as $\epsilon \to 0$. We can now apply the condition that $E(|\Phi\rangle_d) = \log d$ to the left hand side of our inequality to obtain:

$$E(|\Phi\rangle_2\langle\Phi|^{\otimes rn}) = E(|\Phi\rangle_{2^{rn}}\langle\Phi|) = \log 2^{rn} = rn. \tag{1.62}$$

Plugging this in we have

$$E(|\Phi\rangle_2\langle\Phi|^{\otimes rn}) \leq E(\Lambda\left(\rho^{\otimes n}\right)) + k\epsilon rn + O(\epsilon)$$

$$\Rightarrow rn \leq E(\Lambda\left(\rho^{\otimes n}\right)) + k\epsilon rn + O(\epsilon)$$

$$\Rightarrow rn \leq E(\rho^{\otimes n}) + k\epsilon rn + O(\epsilon)$$

$$\tag{1.63}$$

where we have applied the monotonicity of E under LOCC. We can therefore bound r by

$$r \leq \frac{E(\rho^{\otimes n})}{n} + k\epsilon r + \frac{O(\epsilon)}{n}$$

We can now take the limit $n \to \infty$ - by definition of $E_D(\rho)$, ϵ must vanish in this limit, and thus

$$r = \lim_{n\to\infty} r \leq \lim_{n\to\infty} \frac{E(\rho^{\otimes n})}{n} + k\epsilon r + \frac{O(\epsilon)}{n} = \lim_{n\to\infty} \frac{E(\rho^{\otimes n})}{n} = E^\infty(\rho). \tag{1.64}$$

As this holds for all valid r, we thus conclude

$$E_D(\rho) \leq E^\infty(\rho). \tag{1.65}$$

We have proved our result, and when E is weakly additive, $E^\infty(\rho) = E(\rho)$. $\qquad\square$

The relative entropy of entanglement with respect to both separable and PPT states

---

[9]For this proof, we shall assume this holds for mixed states.

satisfies the above conditions, and thus both $E_R^\infty(\rho)$ and $E_P^\infty(\rho)$ upper bound $E_D(\rho)$. As this regularised measure is generally difficult to calculate, one may instead use that $E_{R/P}^\infty \leq E_{R/P}$ to provide a single-letter bound on $E_D(\rho)$ - though in general this is a looser bound. Note that we cannot do this for E_C. Another entanglement measure satisfying the conditions is the squashed entanglement - as this is an additive measure, one can use E_{sq} directly.

Despite failing to be asymptotically continuous, nevertheless the logarithmic negativity $E_N(\rho)$ also provides an upper bound for the distillable entanglement [105], although there is no known connection to the entanglement cost. As $E_N(\rho)$ is easy to calculate and strongly additive, this provides the simplest way to provide upper bounds - although often it provides a looser bound than those measures satisfying the conditions above.

Another important comparison to make is that between secret key rate and distillable entanglement. One maximally entangled state $(|00\rangle + |11\rangle)/\sqrt{2}$ can give one bit of secret key, as both Alice and Bob can measure in the $\{|0\rangle, |1\rangle\}$ basis[10] to obtain a shared bit, which is secure by the monogamy of entanglement [24] - showing $E_D(\rho) \leq K(\rho)$. The two are nonequivalent though, as there exist states with a positive partial transpose known as *bound entangled states*. These states are undistillable - for these states $E_D(\rho) \leq E_P^\infty(\rho) = 0$ - but some of these states have been shown to have a non-zero key distillation rate [55].

One may also generalise the proof of theorem 1.4.2 in order to provide an upper bound for the secret key distillation rate - replacing $|\Phi\rangle_2 \langle\Phi|$ by Ψ, a suitable private state. In order to apply the same methodology, there is the extra condition of E that $E(\Psi^{\otimes rn}) \geq rn$, a property satisfied by the relative entropy of entanglement [56] and squashed entanglement [17][11], but *not* by the RPPT. One also requires that the dimension of the private state $\dim[\Psi^{\otimes rn}] \leq 2^{d_n n}$, with $\liminf_{n\to} d_n = d_k$, a constant. This is to ensure the term $(k\epsilon \dim[\Psi^{\otimes rn}])/n$ vanishes as $n \to \infty$. Although this restricts the set of private states one may consider, it was shown in [18] that the key distillation rate for all states may be achieved considering these restricted private states only - therefore one needs only consider distillation protocols Λ for which the dimensional constraint holds.

---

[10]Or indeed, any basis.

[11]These two papers were the first show this upper bound for E_R^∞ and E_{sq} respectively.

1.5   Capacities of Quantum Channels

Before we explore the realm of quantum channel capacities, it is worth briefly returning to the classical world and looking at what capacity means in terms of a classical channel. To do this we need to define exactly what a *classical channel* is. In mathematical terms, it is simply a map from an alphabet $\mathcal{X}$ to another alphabet $\mathcal{Y}$ defined by a conditional probability distribution $p(Y = y | X = x)$. Note this is a *digital channel* - this concept can be generalised to analogue by considering instead sets $\mathcal{X}$, $\mathcal{Y}$ allowing continuous variables, and a conditional probability density function $f_Y(y|x)$. The capacity of such a channel is the maximal amount of information that be transmitted over the channel per use, such that the probability of error can be made arbitrarily small. Claude Shannon, the father of information theory, showed that the capacity C of a channel π, was given by

$$C(\pi) = \sup_{p_X(x)} I(X:Y) := H(Y) - H(Y|X) \tag{1.66}$$

where $I(X : Y)$ is the mutual information between two random variables, given by the Shannon entropy of Y, $H(Y) = \sum_y -p_Y(y) \log p_Y(y)$, whilst $H(Y|X)$ is the conditional Shannon entropy $H(Y|X) = \sum_x p_X(x) \left(\sum_y -p_{Y|X}(y|x) \log p_{Y|X}(y|x) \right)$. The mutual information is optimised over all probability distributions for the input alphabet, to allow for biases in the channel in transmitting certain letters.

In the above scenario, one is concerned with transmitting bits across a classical channel. For quantum channels however, the situation is slightly different. Quantum channels are maps $\mathcal{E} : \mathcal{H}_A \to \mathcal{H}_B$ such that

$$\mathcal{E}(\rho) = \sum_i K_i \rho K_i^\dagger, \;\; \sum_i K_i^\dagger K_i = I_A. \tag{1.67}$$

Since we are now transmitting quantum states, some ambiguity arises regarding capacities - should we alter our definition to instead consider quantum states, rather than bits? Or should we retain the desire to transmit bits, with the additional power of a quantum channel? The solution to this conundrum is instead for each channel to have *multiple* capacities, each referring to the transmission of a specific type of desired information. We also must consider a capacity C in the context of three separate scenarios - *no-communication*, denoted C, where both sender and receiver are allowed local quantum operations, but no additional classical communication; *one-way communication*, denoted C_1, where the sender may also transmit classical communication to the receiver with each use of the

channel, to aid them in their local operations, and *two-way communication* C_2, where between each use of the channel sender and receiver are allowed to communicate freely before and after each use of the channel - the general term for such protocols is *adaptive.* This may seem to defeat the purpose of transmitting information though the quantum channel at all, but as we shall see, the nature of the capacities we wish to determine allow for a rigorous treatment of all three of these scenarios - assuming the classical channel is "unsecure" - meaning any eavesdropper may obtain the information sent. Naturally for any capacity the following holds true:

$$C \leq C_1 \leq C_2. \tag{1.68}$$

Some of the most important capacities are:

- **Entanglement distillation capacity**

  This capacity quantifies the optimal rate at which a channel may be used to establish maximally entangled states $|\Phi\rangle_2$ between sender and receiver - it can be thought of as analogous to the distillable entanglement. The two-way entanglement distillation capacity is defined:

  $$D_2(\mathcal{E}) := \sup \left\{ r : \lim_{n \to \infty} \left(\inf_\Lambda \|\rho_{AB}^n - |\Phi\rangle_2 \langle\Phi|^{\otimes rn}\|_1 \right) = 0 \right\}. \tag{1.69}$$

  Where ρ_{AB}^n is the overall joint state between sender and receiver after n uses of $\mathcal{E}$, with LOCC operation $\Lambda_i, i \in \{0 \ldots n+1\}$ performed after the ith use of the channel - we refer to these as the overall LOCC operation Λ. For the no-communication and one-way versions, the Λ_i are restricted to the relevant communication operations. Since entanglement cannot be increased by LOCC operations, we see that the two-way communication does not "break" this capacity - distillation is impossible over the classical channel alone.

- **Quantum capacity**

  This capacity seeks to quantify how many qubits may be sent through the channel per use - note that this refers to an arbitrary qubit - thus the amplitude damping channel, which adds noise to the state $|1\rangle$ but leaves $|0\rangle$ unchanged, does not have capacity 1. It is defined in a similar manner to D_2, namely:

  $$Q_2(\mathcal{E}) := \sup \left\{ r : \lim_{n \to \infty} \left(\sup_\tau \inf_\Lambda \|\rho_{AB}^n - \tau\|_1 \right) = 0 \right\}, \quad \tau \in \mathcal{H}_2^{\otimes rn} \tag{1.70}$$

where the supremum over τ ensures an arbitrary rn qubit state can be sent. Like the entanglement distillation capacity, this capacity is not made unbounded by the addition of classical capacity - since it is known that even an infinite amount of classical communication is insufficient to perfectly send an arbitrary quantum state.

- **Secret key capacity**

  This capacity is the analog of the key distillation rate - the rate at which shared, secret key bits can be established between parties. This can be formulated in the same manner as the previous two:

  $$K_2(\mathcal{E}) = \sup\left\{r : \lim_{n\to\infty}\left(\inf_{\Lambda}\|\rho_{AB}^n - \Psi^{\otimes rn}\|_1\right) = 0\right\} \tag{1.71}$$

  where Ψ is a private state, as seen in Eq. (1.56). Notice again how no secret key bits can be sent directly using the auxiliary classical channel, due to our assumption the classical channel is unsecure. As this capacity is almost always considered with the assistance of two-way communication, we shall write K to mean K_2 for the rest of this thesis.

- **Private communication capacity**

  A similar concept to the secret key capacity above, the private communication capacity (P_2) quantifies how many bits of secure classical information may be transmitted from the sender to receiver. We shall not define a supremum-based definition like the others as $P_2 = K$, so any results about secret key capacity can be directly translated.

1.5.1   Relations Between Capacities

Aside from the relations concerning the allowance of various amounts of additional classical communication, there are also some important relations between the two-way capacities. The first is that for all channels, $Q_2 = D_2$. That $D_2 \leq Q_2$ is immediately obvious - Alice (the sender) and Bob (the receiver) can distill r maximally entangled states between them, and use them to send r qubits via teleportation. The reverse is also surprisingly simple, once we remember that τ^n was an arbitrary rn qubit string - we can take this to be rn halves of maximally entangled qubit pairs to achieve the second inequality.

We also know that $D_2 \leq K$ - since maximally entangled states are specific examples of private states. It is likely there exist channels for which $D_2 < K$, akin to bound entangled states, although no such channel has yet been shown. The final relation states that $K = P_2$ - a secret key string of length rn can send securely rn bits by use of a one-time pad, whilst if Alice has been able to send rn private uniform bits securely to Bob, then they would also be able to use that string as a secure key string.

1.5.2 Connecting Capacities and Entanglement Measures - and LOCC Simulation

It is easy to see the similarities between many of the defined entanglement measures for channels and capacities of channels. It turns out they can be connected in terms of their values also.

Theorem 1.5.1 For a channel $\mathcal{E}$, we may lower bound either the entanglement distillation capacity or secret key capacity in the following way:

$$E_D(\chi_{\mathcal{E}}) \leq D_2\left(\mathcal{E}\right) \qquad\qquad K(\chi_{\mathcal{E}}) \leq K_2\left(\mathcal{E}\right) \qquad\qquad (1.72)$$

where $\chi_{\mathcal{E}}$ is the *Choi matrix* of $\mathcal{E} : \mathcal{H}_A \otimes \mathcal{H}_B$, defined as:

$$\chi_{\mathcal{E}} := \left(\mathrm{I}_A \otimes \mathcal{E}\right)\left(|\Phi\rangle_A \langle\Phi|\right). \qquad\qquad (1.73)$$

Proof. Given $\mathcal{H}_A$ is of dimension d, Alice may prepare the maximally entangled state $|\Phi\rangle_d$, and send half of the state through $\mathcal{E}$, resulting in exactly one copy of $\chi_{\mathcal{E}}$ per use of the channel. Alice and Bob may then use the optimal distillation procedure on these states, obtaining distillation/key rate r - since the number of channel uses to states is one-to-one, this is a specific example of a rate r protocol for $\mathcal{E}$ - thus it is a lower bound, since the supremum must be at least r. $\qquad\qquad\square$

In determining capacities, lower bounding it using the entanglement distillation of the Choi matrix only solves some of the difficulty - since entanglement distillation itself requires considering all LOCC operations on any possible number of copies of the state. Moreover as we have seen in section 1.4.1, most of our bounds of entanglement distillation are upper bounds - and thus we cannot state the relation between them and the capacity in question.

This problem has been tackled by a theoretical tool known as *channel simulation*, which will be discussed in more detail in the next chapter; but we shall outline the general idea here. Suppose Alice and Bob share a state $\tau \in \mathcal{H}_A \otimes \mathcal{H}_B$, and that, for any state $\rho \in \mathcal{H}_{A'}$ (another Hilbert space belonging to Alice) they are able to perform an LOCC operation Λ *independent* of ρ, such that $\mathrm{Tr}_{AA'}[\Lambda(\rho \otimes \tau)] = \mathcal{E}(\rho) \in \mathcal{H}_B$. We can see that they have in effect performed one transmission "through" $\mathcal{E}$ using one copy of τ. When this is possible, then the adaptive protocol achieving the capacity of $\mathcal{E}$ can therefore be seen as a specific distillation protocol for τ, giving [85]:

$$D_2(\mathcal{E}) \leq E_D(\tau), \qquad\qquad K_2(\mathcal{E}) \leq K(\tau), \qquad\qquad (1.74)$$

and the upper bounds of $E_D(\tau)$ and $K(\tau)$ can then be used to upper bound the capacities of $\mathcal{E}$. Moreover, in some cases $\tau = \chi_\mathcal{E}$, thus establishing $D_2(\mathcal{E}) = E_D(\chi_\mathcal{E})$ and $K_2(\mathcal{E}) = K(\chi_\mathcal{E})$. Although channel simulation was first shown for Pauli channels[12] in [12], the most general formulation of channel simulation for Pauli channels and many more was given by [85].

1.6 Summary

The aim of this chapter is to provide the reader with the basic tools to understand the processes and results that follow in the next few chapters, and to give an overview of the work done to characterise that subject-defining property that is entanglement. Try explaining the concept of entanglement to a lay person and already one struggles to pin it down; the neat analogy of the perfectly correlated coins does not accurately portray the full extent of entanglement's subtlety. Trying to characterise the "amount" of entanglement has led to all of the measures above, almost all of them requiring the consideration of infinite dimensions to determine. Even the relationship between entanglement and the other great feature of quantum information, non-locality, is ambiguous; seemingly entanglement is not sufficient for non-locality, nor is a greater quantity of entanglement indicative of a greater degree of non-locality. These foibles will be looked at further in the coming chapters.

---

[12]See chapter 2.

Chapter 2

Simulation of Non-Pauli Channels

The work presented in this chapter forms the basis of the paper "Simulation of non-Pauli Channels", whose authors are (in order) Thomas Cope, Leon Hetzel, Leonardo Banchi, and Stefano Pirandola. Any work by authors other than myself in this chapter will be credited appropriately.

2.1   Structure of this Chapter

This chapter will begin with an overview of channel simulation, defining the concept and giving a history of previous results up until this point. This will include the motivation for studying such a concept. Once this is completed, we will move to look at the specific technique of teleportation simulation, before moving into the new developments presented in this work allowing for the simulation of novel channels which do not have the Pauli property. Once the proofs are fully presented, a summary of results will be given, along with a discussion for future topics of research.

2.2   The History of Channel Simulation

The first introduction of the important tool of channel simulation was, surprisingly, not used to simulate channels at all. The 1996 paper by Bennett, DiVincenzo, Smolin and Wootters [12] was interested in two vital concepts in quantum information theory. The first was entanglement purification, in which pure entangled states are distilled from mixed entangled states shared between two parties - typically distilling many mixed states into fewer maximally entangled states. The other concept was error correction, a staple in

classical information theory to protect information from noise induced by the transmission through a channel. Their insight was to consider the standard teleportation protocol (explained in chapter 1) but replacing the maximally entangled resource state by a mixed entangled state - the authors noted that, instead of teleporting the input state exactly, it was distorted as if sent through a quantum channel. Given this information, they were able to deduce to two important equalities for *Bell-diagonal* states:

$$E_{D_1}(\sigma) = Q(\mathcal{M}(\sigma)) \tag{2.1}$$

and

$$E_D(\sigma) = Q_2(\mathcal{M}(\sigma)) \tag{2.2}$$

with $\mathcal{M}$ the mapping from states to channels achieved by performing teleportation over them. Bell-diagonal states are convex combinations of the four Bell pairs, and E_{D_1} is the distillable entanglement limited to one-way communication only, whilst $Q(Q_2)$ refers the quantum capacity of a channel with no(two-way) communication. These results were derived from their explicit construction of a distillation protocol for σ from an error correcting code of $\mathcal{M}(\sigma)$, and vice versa. They also noted that Bell-diagonal states σ always gave rise to a *Pauli channel*[1]: one whose action on a state ρ may be written[2]

$$\mathcal{E}_P(\rho) = p_0 I \rho I + p_1 \sigma_x \rho \sigma_x + p_2 \sigma_y \rho \sigma_y + p_3 \sigma_z \rho \sigma_z, \;\; p_i \geq 0, \;\; \sum_{i=0}^{3} p_i = 1. \tag{2.3}$$

and that there was a bijection between the two - the weights p_i correspond to the weights of the four Bell states. This result was expanded in the work of Bowen and Bose [13] who were able to show that teleportation over *any* arbitrary two-qubit state would result in a Pauli channel. Furthermore for any two-qudit state, using the generalised teleportation protocol will produce a channel of the form

$$\mathcal{E}(\rho) = \sum_{i=0}^{d^2-1} p_i \sigma_i \rho \sigma_i^{\dagger}, \;\; p_i \geq 0, \;\; \sum_{i=0}^{d^2-1} p_i = 1, \tag{2.4}$$

with $\{\sigma_i\}$ the set of generalised Pauli matrices introduced in section 1.3.5.1.

Another generalisation in a different direction was performed by Reinhard Werner [111], who looked at maintaining perfect teleportation fidelity, but altering the shared state,

---

[1] These are also called *generalised depolarising channels*.
[2] The Pauli matrices are Hermitian.

Alice's measurements and Bob's corrective operation. This departure from the standard protocol will be mirrored by our approach later in the chapter. He concluded that, under the condition that the protocol is "tight" - teleporting a d-dimensional state with a d^2-dimensional resource state and a d^2-outcome measurement, that perfect teleportation occurs iff the resource state σ is maximally entangled, Alice's measurement distinguishes between d^2 orthogonal maximally entangled states, and Bob's corrective operations are the unitaries generating these orthogonal states from resource state σ.

In 2015, the property of *teleportation covariance* was studied by [68], and then by [85] for systems of any dimension.

Definition 2.2.1 A channel $\mathcal{E} : \mathcal{H}_d \to \mathcal{H}_{d'}$ is teleporation covariant iff it satisfies

$$\mathcal{E}\left(U_i \rho U_i^\dagger\right) = V_i \mathcal{E}\left(\rho\right) V_i^\dagger, \forall \rho \in \mathcal{H}_d \tag{2.5}$$

with $\{U_i\} \equiv \{\sigma_i\}$ the generalised Pauli matrices for dimension d, and $\{V_i\}$ some corrective unitaries.

This property is extremely significant, since it allows a channel to be simulated by its own Choi matrix:

$$\chi_\mathcal{E} := \left(\mathbb{I} \otimes \mathcal{E}\right)\left(|\Phi\rangle_d \langle\Phi|\right). \tag{2.6}$$

This we can show in the following way: Consider the teleportation of ρ over the maximally entangled state $|\Phi\rangle_d$. Then the conditional state Bob obtains before learning of Alice's measurement is the state $U_i \rho U_i^\dagger$. Suppose he then applies the channel $\mathcal{E}$ to this state, to obtain $\mathcal{E}\left(U_i \rho U_i^\dagger\right) = V_i \mathcal{E}\left(\rho\right) V_i^\dagger$ by teleportation covariance. Thus when he obtains Alice's result i, by performing the corrective unitary $V_i^\dagger$ he obtains the state $\mathcal{E}\left(\rho\right)$. As Bob applied the channel $\mathcal{E}$ to his state before receiving Alice's result, we can instead consider him applying the channel *before* the measurement took place. As the two operations take place on locally separated subsystems, this will not affect the output. As $\left(\mathbb{I}_A \otimes \mathcal{E}_B\right)\left(|\Phi\rangle_d \langle\Phi|\right) = \chi_\mathcal{E}$, this means the same protocol can instead be thought of performing the conventional teleportation protocol over the Choi matrix, still obtaining $\mathcal{E}\left(\rho\right)$. The consequence of this is that we have enacted the channel $\mathcal{E}$ simply by performing local operations and classical communication (LOCC) operations on the state $\chi_\mathcal{E}$ - we call this *simulating* the channel, defined formally in 2.2.3.

Leung and Matthew [68] used this teleportation covariance property to upper bound the performance of coding schemes which preserved a positive partial transpose (PPT). Its use in the field of channel simulation was greatly expanded by Pirandola, Laurenza, Laurenza and Banchi [85] which introduced many new concepts and results, which we shall go over here.

First we must introduce the concept of an "adaptive protocol". Alice and Bob begin with general Hilbert spaces, $\mathcal{H}_A$, $\mathcal{H}_B$, with which they may prepare any states locally, with free classical communication between them - we label this state ρ_{AB}^0, prepared by LOCC operations Λ_0. Alice then picks a subsystem to send through the channel $\mathcal{E}$ - we label the Hilbert space of this subsystem $\mathcal{H}_{A_1}$ - and transmits it to Bob through $\mathcal{E}$ - the received state living in space $\mathcal{H}_{B_1}$. This is followed by another round of local operations, Λ_1, resulting in a new shared state $\rho_{AB}^1 \in \mathcal{H}_{A \setminus A_1} \otimes \mathcal{H}_{B \cup B_1}$ - we then update $\mathcal{H}_A = \mathcal{H}_{A \setminus A_1}$ and $\mathcal{H}_B = \mathcal{H}_{B \cup B_1}$. This process is then repeated n times, resulting in the shared state ρ_{AB}^n, dependent on the LOCC operations $\mathbf{\Lambda} = (\Lambda_0, \ldots, \Lambda_n)$. Such an adaptive protocol is said to have rate r_n if $||\rho_{AB}^n - \phi_n||_1 \leq \varepsilon$, where ϕ_n is the target state consisting of nr_n copies of some "desired resource". For example, $\phi_n = |\Phi\rangle_2^{\otimes nr_n}$, if the aim to distill maximally entangled Bell pairs (ebits). The error ε must be "sufficiently small"[3] - it vanishes as $n \to \infty$.

This concept of an adaptive protocol allows the following definition of a two-way capacity.

Definition 2.2.2 A two-way assisted capacity of quantum channel $\mathcal{E}$ is given by the optimisation

$$\mathcal{C}(\mathcal{E}) := \sup_{\mathbf{\Lambda} \in \text{LOCC}} \lim_{n \to \infty} r_n. \tag{2.7}$$

This may be interpreted as the optimal asymptotic rate of a desired resource generation over all adaptive protocols[4]. Generally, this optimisation is almost impossible to do, due to the extremely general nature of $\mathbf{\Lambda}$. However, we can simplify the process using an LOCC simulation of the channel.

---

[3]In the definitions presented in chapter 1, this condition is presented more explicitly.

[4]See chapter 1 for some specific capacities.
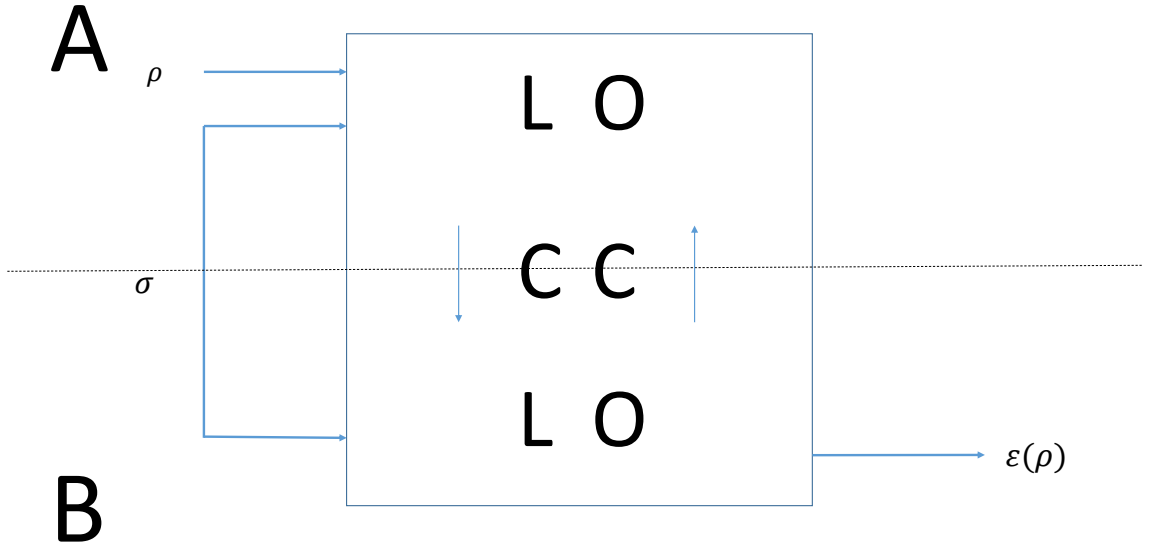
Figure 2.1: The general idea behind LOCC simulation: Alice begins with an arbitrary input state ρ, and the two share some resource state σ. Via an LOCC protocol, they aim for Bob to end up with the state $\mathcal{E}(\rho)$.

Definition 2.2.3 ([85]) A channel $\mathcal{E} : \mathcal{H}_A \to \mathcal{H}_B$ is σ-*simulable* if there exists a state $\sigma \in \mathcal{H}_{A'} \otimes (\mathcal{H}_B \otimes \mathcal{H}_{B'})$ such that, for any state $\rho \in \mathcal{H}_A$, Alice and Bob can transform using trace-preserving LOCCs $\rho \otimes \sigma \to \tau_{\mathcal{E},\rho}$ satisfying $\mathrm{Tr}_{AA'B'}[\tau_{\mathcal{E},\rho}] = \mathcal{E}(\rho) \in \mathcal{H}_B$.

This definition requires that the LOCCs are trace-preserving; this imposes that the channel is simulated with certainty, rather than a probabilistic simulation where many attempts are made to simulated the channel, and the successful ones are post-selected.

Now consider the final state of an adaptive protocol ρ^n_{AB}. This state was constructed

$$\rho^n_{AB} = \Lambda_n \left(\mathbb{I}_{AB \backslash A_N} \otimes \mathcal{E}_{A_N \to B_N} \right) \left(\rho^{(n-1)}_{AB} \right) \tag{2.8}$$

from $\rho^{(n-1)}_{AB}$, the state after $n-1$ channel transmissions and LOCCs. If $\mathcal{E}$ is σ-simulable, we may replace the nth channel use $\mathcal{E}_{A_N \to B_N}$ by it's LOCC simulation, to obtain

$$\rho^n_{AB} = \Lambda_n \Lambda_s \left(\rho^{(n-1)}_{AB} \otimes \sigma \right) = \Lambda'_n \left(\rho^{(n-1)}_{AB} \otimes \sigma \right) \tag{2.9}$$

with Λ_s the LOCC simulation protocol. Furthermore, we may iterate this process repeatedly to obtain [85] $\rho^n_{AB} = \Lambda'_n \Lambda'_{(n-1)} \dots \Lambda'_0 (\sigma^{\otimes n}) = \bar{\Lambda}(\sigma^{\otimes n})$. Note that the input state ρ^0_{AB}

does not feature in this formula, since this state is locally separable and was also prepared by LOCC. This technique is known as "stretching of the protocol".

The consequence of this is that any adaptive protocol for a σ-simulable channel $\mathcal{E}$ may be simulated using n copies of the resource state σ. We can take advantage of this in order to simplify the calculation of the two-way assisted capacity. We explain how this can be done, following the methodology in [82].

Suppose we have an entanglement measure E which is:

- *Asymptotically continuous*: $\|\rho - \sigma\|_1 \leq \delta \Rightarrow |E(\rho) - E(\sigma)| \leq k\delta\log d + O(\delta)$,

- *Monotonic under LOCC*: $E(\Lambda(\rho)) \leq E(\rho)$,

- *Weakly subadditive*: $E(\rho^{\otimes n}) \leq nE(\rho)$.

Consider an adaptive protocol with rate r_n; by definition this means that we have the output state ρ_{AB}^n satisfies $\|\rho_{AB}^n - \phi_n\|_1 \leq \varepsilon$, with $\lim_{n\to\infty} \varepsilon = 0$. By asymptotic continuity, this implies

$$|E(\rho_{AB}^n) - E(\phi_n)| \leq k\varepsilon\log d + O(\varepsilon) \Rightarrow E(\phi_n) \leq E(\rho_{AB}^n) + k\varepsilon \log d + O(\varepsilon) \qquad (2.10)$$

where $d = \dim[\phi_n]$. If the adaptive protocol is over a σ-simulable channel, we may write:

$$E(\phi_n) \leq E(\rho_{AB}^n) + k\varepsilon\log d + O(\varepsilon)$$
$$= E(\bar{\Lambda}\left(\sigma^{\otimes n}\right)) + k\varepsilon\log d + O(\varepsilon)$$
$$\leq E(\sigma^{\otimes n}) + k\varepsilon\log d + O(\varepsilon) \qquad (2.11)$$

where we have applied the monotonicity of E. In order to progress further, we require two more conditions - this time on ϕ_n, the target state;

- *Normalisation*: $E(\phi_n) \geq nr_n$.

- *Exponential Growth*: $\dim[\phi_n] \leq 2^{d_n n}$, where $\liminf_{n\to\infty} d_n = d_k$, a constant value.

40

Applying the normalisation condition, we can further simplify:

$$E(\phi_n) \leq E(\sigma^{\otimes n}) + k\varepsilon \log d + O(\varepsilon)$$

$$\Rightarrow nr_n \leq E(\sigma^{\otimes n}) + k\varepsilon \log d + O(\varepsilon)$$

$$\leq E(\sigma^{\otimes n}) + k\varepsilon \log\left(2^{d_n n}\right) + O(\varepsilon)$$

$$\leq E(\sigma^{\otimes n}) + k\varepsilon d_n n + O(\varepsilon)$$

$$\Rightarrow r_n \leq \frac{E(\sigma^{\otimes n})}{n} + k\varepsilon d_n + \frac{O(\varepsilon)}{n}. \tag{2.12}$$

Finally we can substitute this into our definition of a two-way capacity [82, 85],

$$\mathcal{C}(\mathcal{E}) = \sup_{\Lambda \in \mathrm{LOCC}} \lim_{n \to \infty} r_n \leq \sup_{\Lambda \in \mathrm{LOCC}} \lim_{n \to \infty} \frac{E(\sigma^{\otimes n})}{n} + k\varepsilon d_n + \frac{O(\varepsilon)}{n} = E^{\infty}(\sigma) \tag{2.13}$$

with $E^{\infty}(\sigma)$ the regularised entanglement measure $\lim_{n \to \infty} E(\sigma^{\otimes n})/n \leq E(\sigma)$. The other terms disappear because we imposed that $\lim_{n \to \infty} \varepsilon = 0$. Thus $\lim_{n \to \infty} k\varepsilon d_n = 0$, as does $\lim_{n \to \infty} O(\varepsilon)/n = 0$, as asymptotic continuity requires $\lim_{\varepsilon \to 0} O(\varepsilon) = 0$.

We can see how this method adapts the techniques for bounding the distillable entanglement of states, for use in determining channel capacities. For the choice of entanglement measure E, the two we shall use primarily in this thesis are the relative entropy of entanglement (REE) E_R, and the squashed entanglement E_{sq}. Both satisfy the requirements presented above [17, 20, 97, 103]. One of the most commonly investigated capacities is the quantum capacity Q_2. It has been shown that Q_2 is equal to the entanglement distillation capacity D_2, where the target state is $\phi_n = |\Phi\rangle_2^{\otimes nr_n}$. We have that $E_R\left(|\Phi\rangle_2^{\otimes nr_n}\right) = E_{\mathrm{sq}}\left(|\Phi\rangle_2^{\otimes nr_n}\right) = nr_n$, and that $\dim\left[|\Phi\rangle_2^{\otimes nr_n}\right] = 2^{nr_n}$ implying $d_n = d_k = 1$.

Another commonly investigated capacity is K - the secret key capacity. The target state here is $\Psi^{\otimes nr_n}$, with $\Psi = U_T\left(|\Phi\rangle_2 \langle\Phi| \otimes \tau_{A'B'}\right) U_T^{\dagger}$ a "private state" - $\tau_{A'B'}$ is an arbitrary "shield state", and U_T a "twisting unitary". It has been shown $E_R\left(\Psi^{\otimes nr_n}\right) \geq nr_n$ [56] and $E_{\mathrm{sq}}\left(\Psi^{\otimes nr_n}\right) \geq nr_n$ [17]. There has been some controversy over the exponential growth requirement for private states; as $\tau_{A'B'}$ is allowed to be any size. However, in [18] it was shown that the key distillation rate for states, $K(\rho)$, could be achieved by only considering private states with at most exponential growth in the number of copies of the state; this proof was adapted in [85] to show secret key capacity need only consider adaptive

protocols with exponential growth private states; and so the above method may be applied.

Finally, another entanglement measure worth considering is the relative entropy with respect to positive partial transpose states (RPPT), E_P. This too is asymptotically continous, monotonic and subadditive [97, 103]; however, whilst $E_P \left(|\Phi\rangle_2^{\otimes nr_n} \right) = nr_n$, $E_P \left(\Psi_2^{\otimes nr_n} \right) \not\gtrsim nr_n$. This means E_P can be used as an upper bound for Q_2, but not K.

These bounds are *weak converse* bounds - they bound the rate of *error*-free communication ($\varepsilon \to 0$).

It is interesting to note that the concept of LOCC simulation can be applied to *any* channel - however, the most general method only gives us a trivial upper bound on two-way capacities.

Lemma 2.2.4 Every channel is trivially simulable.

Proof. Suppose we have a channel $\mathcal{E} : \mathcal{H}_{d_A} \to \mathcal{H}_{d_B}$ between Alice and Bob. If Alice and Bob instead shared the maximally entangled state $|\Phi\rangle_{d_A}$, they can perform the teleportation protocol to send the input state ρ from Alice to Bob with perfect fidelity. Bob can then apply the channel $\mathcal{E}$ locally. $\qquad\square$

This is a valid LOCC protocol, which clearly $|\Phi\rangle_{d_A}$-simulates $\mathcal{E}$. However, this is not useful to us, since it implies $\mathcal{Q}_2 \left(\mathcal{E} \right) \leq K \left(\mathcal{E} \right) \leq E_R^\infty \left(|\Phi\rangle_{d_A} \right) = \log d_A$. However, this is a trivial upper bound on the capacity of any channel with input Hilbert space $\mathcal{H}_{d_A}$, as $\mathcal{Q}_2(\mathbb{I}_{d_A}) = K(\mathbb{I}_{d_A}) = \log d_A$. In the next section though, we shall see we can use the teleportation protocol in order to obtain non-trivial upper bounds.

2.3 Channel Simulation via Teleportation

As previously discussed, it was shown in [13] that teleportation over *any* arbitrary two-qubit state would result in a Pauli channel. We shall present that result here:

Theorem 2.3.1 ([13]) The standard teleportation protocol over an arbitrary two-qubit state σ will produce a channel of the form

$$\mathcal{E}_P \left(\rho \right) = \sum_{i=0}^{3} p_i \sigma_i \rho \sigma_i \qquad (2.14)$$
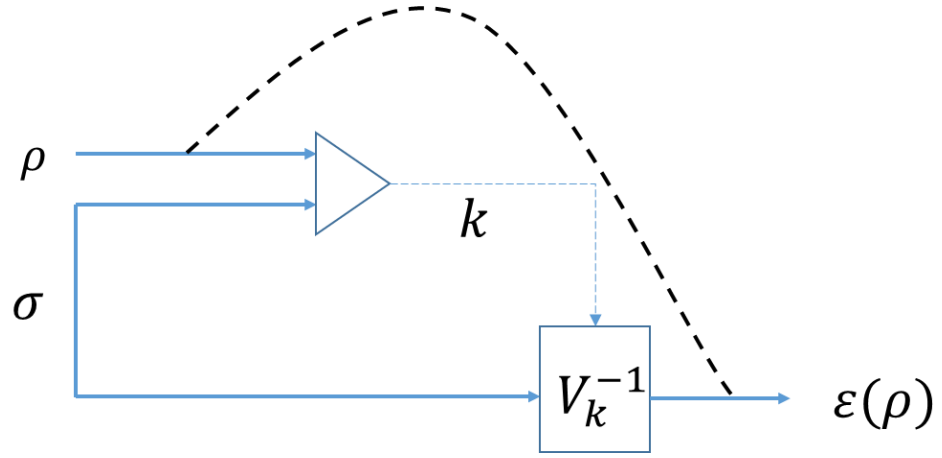
Figure 2.2: Teleportation simulation - by replacing the maximally entangled state with a less entangled state σ, the teleportation protocol has the overall effect of enacting the channel $\mathcal{E}$.

with $p_i = \mathrm{Tr}\,[E_i \sigma]$ and

$$E_0 := |\Phi^+\rangle\langle\Phi^+|, \quad |\Phi^+\rangle = \frac{1}{\sqrt{2}}|00\rangle + |11\rangle,$$

$$E_1 := |\Psi^+\rangle\langle\Psi^+|, \quad |\Psi^+\rangle = \frac{1}{\sqrt{2}}|01\rangle + |10\rangle,$$

$$E_2 := |\Psi^-\rangle\langle\Psi^-|, \quad |\Psi^-\rangle = \frac{1}{\sqrt{2}}|01\rangle - |10\rangle,$$

$$E_3 := |\Phi^-\rangle\langle\Phi^-|, \quad |\Phi^-\rangle = \frac{1}{\sqrt{2}}|00\rangle - |11\rangle.$$

For the rest of the chapter, it shall be useful to think of qubits in their *Bloch* representation.

Definition 2.3.2 An arbitrary qubit state can be written in the *Bloch representation*:

$$\rho = \frac{1}{2}\begin{pmatrix} 1+z & x-iy \\ x+iy & 1-z \end{pmatrix} \tag{2.15}$$

with $\mathbf{x} = (x, y, z)$ the *Bloch vector* of ρ, and $\rho = (\mathrm{I} + \mathbf{x} \cdot \boldsymbol{\sigma})/2$, $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$.

Definition 2.3.3 The *Bloch sphere* $\|\mathbf{x}\| \leq 1$ defines all valid qubit states. The state is pure iff $\|\mathbf{x}\| = 1$.

43
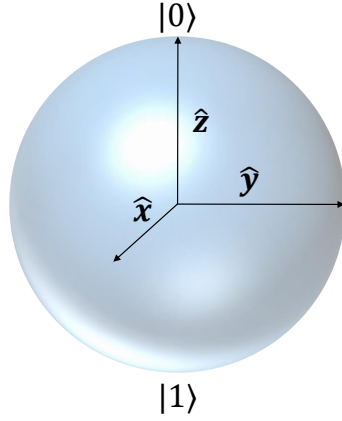
Figure 2.3: The Bloch Sphere - all qubit states can be represented as a point within the sphere.

We can also express an arbitrary two-qubit state in a "Bloch-style" representation:

$$\sigma = \frac{1}{4}\left(\mathrm{I} \otimes \mathrm{I} + \sum_{i=1}^{3} a_i \sigma_i \otimes \mathrm{I} + \sum_{j=1}^{3} \mathrm{I} \otimes b_j \sigma_j + \sum_{i,j=1}^{3} t_{ij} \sigma_i \otimes \sigma_j \right). \tag{2.16}$$

Corollary 2.3.4 In Bloch form, the channel simulated over teleportation by σ may be written:

$$\mathcal{E} : (x, y, z) \rightarrow (t_{11}x, -t_{22}y, t_{33}z) \tag{2.17}$$

with conversion between expressions given by:

$$t_{11} = p_0 + p_1 - p_2 - p_3 = 1 - 2p_2 - 2p_3, \tag{2.18}$$

$$t_{22} = -p_0 + p_1 - p_2 + p_3 = -1 + 2p_1 + 2p_3, \tag{2.19}$$

$$t_{33} = p_0 - p_1 - p_2 + p_3 = 1 - 2p_1 - 2p_2. \tag{2.20}$$

These points satisfy

$$t_{11} + t_{22} + t_{33} \leq 1, \tag{2.21}$$

$$t_{11} - t_{22} - t_{33} \leq 1, \tag{2.22}$$

$$-t_{11} + t_{22} - t_{33} \leq 1, \tag{2.23}$$

$$-t_{11} - t_{22} + t_{33} \leq 1, \tag{2.24}$$
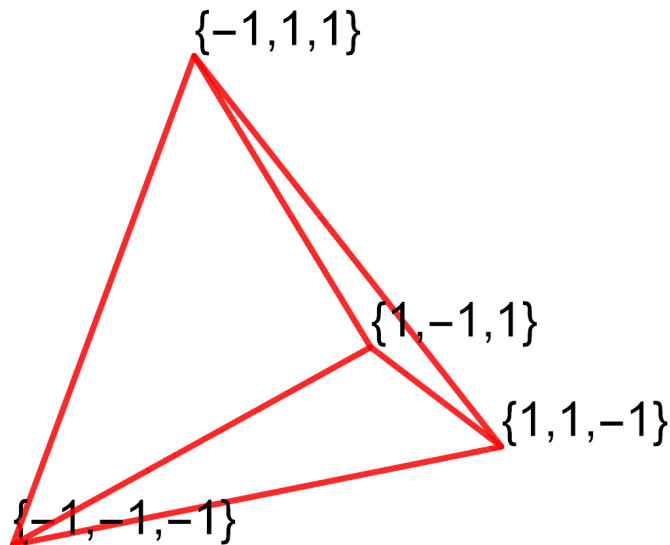
Figure 2.4: The Pauli tetrahedron, $\mathcal{T}$. The four extremal points correspond to the four Bell states.

which implies that the vector (t_{11}, t_{22}, t_{33}) characterising the Pauli channel belongs to the tetrahedron $\mathcal{T}$ defined by the convex combination of the four points

$$\mathbf{e}_0 = (1, -1, 1), \qquad\qquad \mathbf{e}_1 = (1, 1, -1), \qquad\qquad (2.25)$$
$$\mathbf{e}_2 = (-1, -1, -1), \qquad\qquad \mathbf{e}_3 = (-1, 1, 1).$$

If these points are substituted into Eq. (2.16), and the remaining parameters set[5] to 0 then we obtain the four density matrices of the Bell pairs, $\{E_i\}$.

According to corollary 2.3.4, there is a simple way to simulate a Pauli channel with arbitrary probability distribution $\{p_i\}$. One may just take the resource state

$$\sigma = \frac{1}{4}\left(\mathrm{I} \otimes \mathrm{I} + \sum_{i=1}^{3} t_{ii}\sigma_i \otimes \sigma_i\right), \qquad\qquad (2.26)$$

with t_{ii} being connected to $\{p_i\}$ by Eqs. (2.18)-(2.20). Note that this resource state is *Bell diagonal*, i.e., a mixture of the four Bell states. Thus this description presents the result of [12] neatly.

---

[5]This is necessary in order to be a valid state.

2.4 Introducing a Noisy Teleportation Protocol

Pauli channels are important for accurately modelling noise, but are restrictive in terms of the transformations we may apply; corollary 2.3.4 shows that they only allow for rescaling of individual elements of the Bloch vector. It was our goal to expand the possible channels simulable to include more general transformations. Some notable non-Pauli channels have been shown simulable, including the erasure and amplitude damping channels [85], but none using a two-qubit resource state.

To achieve this, it will be necessary to alter the teleportation protocol, whilst retaining its LOCC character. A natural way to do this is to change the *classical channel* used by Alice and Bob - formerly always assumed to be just the identity channel. Since we are retaining the original measurement and correction unitaries, we can without loss of generality consider an arbitrary classical channel as a conditional probability distribution:

$$\Pi = \left\{ p_{l|k} \right\}, \quad p_{l|k} \geq 0, \quad \sum_{k=0}^{3} p_{l|k} = 1. \tag{2.27}$$

This means given Alice receives output k, rather than Bob performing deterministically corrective unitary[6] σ_k, he instead performs σ_l with probability $p_{l|k}$. For the original protocol we have $p_{l|k} = \delta_{lk}$.

Theorem 2.4.1 Consider a teleportation protocol based on a Bell detection and Pauli correction unitaries, but where the resource state is a generic two-qubit state σ and the CCs from Alice to Bob are subject to a classical channel Π ("noisy teleportation"). In this way, we simulate a quantum channel $\mathcal{E}_f$ whose action on the Bloch sphere is described by

$$\mathcal{E}_f : (x, y, z) \to (f_{10} + f_{11}x + f_{12}y + f_{13}z,$$
$$f_{20} + f_{21}x + f_{22}y + f_{23}z,$$
$$f_{30} + f_{31}x + f_{32}y + f_{33}z) \tag{2.28}$$

where f_{ij} is given by the formula $f_{ij} = t'_{ji}S_{ij}$, with

$$S_{ij} := \frac{1}{4} \sum_{k,l=0}^{3} -1^{\delta_{k,0}+\delta_{j,2}+\delta_{j,0}+\delta_{k,j}+\delta_{i,l}+\delta_{0,l}} p_{l|k}, \tag{2.29}$$

---

[6]We use the convention that $\sigma_0 = I$.
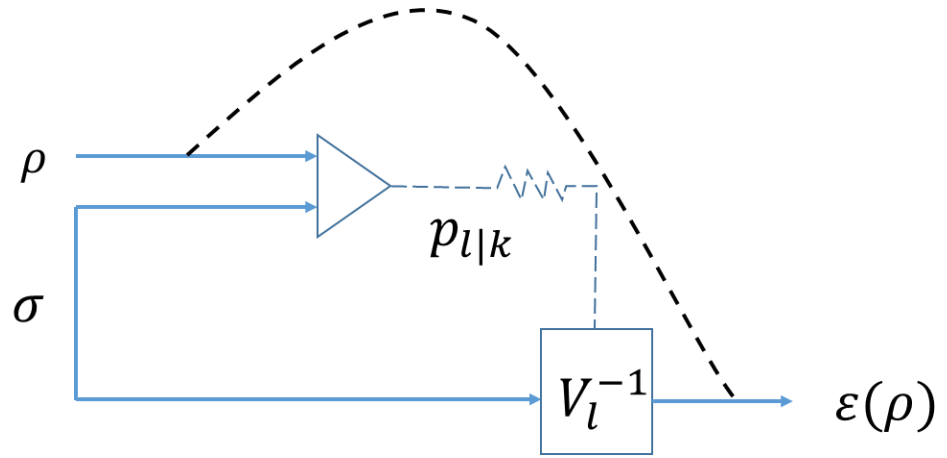
Figure 2.5: By replacing both the resource state and the classical channel between parties, we hope to increase the possible $\mathcal{E}$ the protocol can simulate.

and T' is defined as the *augmented* T matrix,

$$t'_{ji} = \begin{cases} b_i & j = 0 \\ t_{ji} & j \in \{1,2,3\} \end{cases} \qquad i \in \{1,2,3\}, \tag{2.30}$$

taking t_{ji} from the T matrix of Eq. (2.16).

$$\left(\begin{array}{c}\left(\begin{array}{cccc}1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1\end{array}\right) \left(\begin{array}{cccc}1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & -1\end{array}\right) \left(\begin{array}{cccc}-1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1\end{array}\right) \left(\begin{array}{cccc}1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1\end{array}\right)\end{array}\right)$$

Figure 2.6: The matrix $\mathbf{S}$, a succinct representation of the 12 possible forms for S_{ij}. The rows of $\mathbf{S}$ corresponds to $i = 1, 2, 3$ respectively, whilst the columns give $j = 0 \ldots 3$. Given i, j^{th} element $\boldsymbol{S}_{ij}$, S_{ij} can be obtained by the sum $1/4 \sum_{k=0,l=0}^{3,3}[\boldsymbol{S}_{ij}]_{k,l}\; p_{l|k}$, starting the row/column count at 0.

Definition 2.4.2 A qubit quantum channel $\mathcal{E} : (x, y, z) \to (x', y', z')$ can be described by its F matrix $F_{\mathcal{E}}$, where

$$\begin{pmatrix} 1 \\ x' \\ y' \\ z' \end{pmatrix} = F_{\mathcal{E}} \begin{pmatrix} 1 \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ f_{10} & f_{11} & f_{12} & f_{13} \\ f_{20} & f_{21} & f_{22} & f_{23} \\ f_{30} & f_{31} & f_{32} & f_{33} \end{pmatrix} \begin{pmatrix} 1 \\ x \\ y \\ z \end{pmatrix} \tag{2.31}$$

describes its action on the Bloch vector of a qubit state.

As an example, a Pauli channel $\mathcal{E}_P : (x, y, z) \to (t_{11}x, -t_{22}y, t_{33}z)$ has F matrix

$$F_P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & t_{11} & 0 & 0 \\ 0 & 0 & -t_{22} & 0 \\ 0 & 0 & 0 & t_{33} \end{pmatrix}. \tag{2.32}$$

We see immediately that this addition of classical noise greatly expands the range of simulable channel actions; It is now possible to include constant terms in the actions,

as well as interdependence between components of the Bloch vector; much more than previously possible in the Pauli framework. Before we go on to discuss the possibilities for two arbitrary qubits, it is worth considering the restriction to Bell-diagonal resource states - for which we can prove the following no-go theorem.

Theorem 2.4.3 Using a Bell-diagonal resource state, i.e. of the form in Eq. (2.26), it is only possible to simulate Pauli channels regardless of the classical channel in place between the two parties.

Proof. We know from the structure of Bell-diagonal states means that only t_{ii} are non-zero; this immediately limits the action of any simulated channel to

$$\mathcal{E} : (x, y, z) \to (t_{11}S_{11}x, t_{22}S_{22}y, t_{33}S_{33}z) . \tag{2.33}$$

Looking at the structure of the sums S_{ii} for $i \in \{1, 2, 3\}$, it is possible to verify[7] that for any given $p_{l|k}$, the induced action is one of the four transformations

$$\mathcal{E}_{p_{l|k}} : (x, y, z) \to (\quad t_{11}x, \quad -t_{22}y, \quad t_{33}z) \tag{2.34}$$

$$\to (\quad t_{11}x, \quad t_{22}y, \quad -t_{33}z) \tag{2.35}$$

$$\to (-t_{11}x, \quad -t_{22}y, \quad -t_{33}z) \tag{2.36}$$

$$\to (-t_{11}x, \quad t_{22}y, \quad t_{33}z) , \tag{2.37}$$

which are the four Pauli transformations induced by simulation over the respective states defined by

$$(\quad t_{11}, \quad t_{22}, \quad t_{33}), \qquad\qquad (\quad t_{11}, -t_{22}, -t_{33}),$$
$$(-t_{11}, \quad t_{22}, -t_{33}), \qquad\qquad (-t_{11}, -t_{22}, \quad t_{33}),$$

with perfect classical communication. Since $\{t_{ii}\}$ define a Bell state, they may be given by a convex weighting of the four Bell pairs, using the probabilities given in Eqs. (2.18)-(2.20). By permuting the probabilities in the following way, it is possible to generate the four states above.

State	$\lvert\Phi^+\rangle$	$\lvert\Psi^+\rangle$	$\lvert\Psi^-\rangle$	$\lvert\Phi^-\rangle$
$(\quad t_{11}, \quad t_{22}, \quad t_{33})$	p_0	p_1	p_2	p_3
$(\quad t_{11}, -t_{22}, -t_{33})$	p_1	p_0	p_3	p_2
$(-t_{11}, \quad t_{22}, -t_{33})$	p_2	p_3	p_0	p_1
$(-t_{11}, -t_{22}, \quad t_{33})$	p_3	p_2	p_1	p_0

---

[7] By looking at the sign of each $p_{l|k}$ in S_{11}, S_{22}, S_{33}.

This means that, for any given $p_{l|k}$, a valid Pauli channel is simulated. As $1/4 \sum p_{l|k} = 1$ and by linearity of the action, we can consider the overall channel a convex combination of these channels, and thus a Pauli channel also. $\qquad\square$

It is important to understand the difference between theorem 2.3.1 and theorem 2.4.3. Theorem 2.3.1 tells us that an *arbitrary* two-qubit resource state with *perfect* CC from Alice to Bob may only simulate Pauli channels, whereas theorem 2.4.3 states that a *Bell-diagonal* resource with an *arbitrary* classical channel for the CC from Alice to Bob may only simulate Pauli channels. As a result, we have the following corollary which will drive us in the choice of the resource state σ in the next section.

Corollary 2.4.4 In order to simulate a non-Pauli channel via noisy teleportation, the resource state σ of Eq. (2.16) must have $\mathbf{b} \neq 0$ or T non-diagonal. This means σ cannot be Bell-diagonal (and thus cannot be the Choi matrix of a Pauli channel).

2.5   Pauli-Damping Channels

Corollary 2.4.4 tells us that a non-Bell-diagonal state is required to simulate non-Pauli channels. We chose the Choi matrix of the amplitude damping channel, motivated by three reasons. It is the most studied (dimension preserving) non-Pauli channel; the Choi matrix has a relatively high number of zero parameters in the Bloch description, and the fact that it was (and still remains) an open question as to whether the amplitude damping channel is Choi-simulable i.e. LOCC simulable by its own Choi matrix - indeed, by any discrete variable resource non-trivially.

The action of the amplitude damping channel on the Bloch sphere is given by:

$$\mathcal{E}_\gamma : (x, y, z) \to \left(\sqrt{1-\gamma}x, \sqrt{1-\gamma}y, \gamma + (1-\gamma)z \right). \qquad (2.38)$$

where $\gamma \in [0, 1]$ is the probability of damping. The Choi matrix of this channel is

$$\chi_\gamma = \begin{pmatrix} \frac{1}{2} & 0 & 0 & \frac{\sqrt{1-\gamma}}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{\gamma}{2} & 0 \\ \frac{\sqrt{1-\gamma}}{2} & 0 & 0 & \frac{1-\gamma}{2} \end{pmatrix}, \qquad (2.39)$$

which is a resource state of the form (2.16), where the non-zero entries are only

$$b_3 = \gamma, \; t_{11} = \sqrt{1 - \gamma}, \; t_{22} = -\sqrt{1 - \gamma}, \; t_{33} = 1 - \gamma. \tag{2.40}$$

We may also describe this channel using its F matrix:

$$F_\gamma = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{1-\gamma} & 0 & 0 \\ 0 & 0 & \sqrt{1-\gamma} & 0 \\ \gamma & 0 & 0 & 1-\gamma \end{pmatrix}. \tag{2.41}$$

Using this resource, we state one of our two main results:

Theorem 2.5.1 All channels that are simulable by noisy teleportation over the amplitude damping Choi matrix χ_γ can be uniquely[8] decomposed in the following way:

$$\mathcal{E}_{\text{sim}} = \sigma_x^u \circ \mathcal{E}_\eta \circ \mathcal{E}_P \tag{2.42}$$

where $u = 0$ or 1, σ_x is the Pauli unitary map $\sigma_x(\rho) = \sigma_x \rho \sigma_x^\dagger$, $\mathcal{E}_\eta$ is an amplitude damping channel with parameter η, and $\mathcal{E}_P$ is a Pauli channel with suitable parameters $\mathbf{q} = (q_1, q_2, q_3)$ belonging to the tetrahedron $\mathcal{T}$.

Proof. Making use the formula in Eq. (2.28) we know that any channel $\mathcal{E}_{\text{sim}}$ simulated with χ_γ will have an F matrix of the form

$$F_{\text{sim}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{1-\gamma}S_{11} & 0 & 0 \\ 0 & 0 & -\sqrt{1-\gamma}S_{22} & 0 \\ \gamma S_{30} & 0 & 0 & (1-\gamma)S_{33} \end{pmatrix}. \tag{2.43}$$

Before we continue this proof, it is important to state that if two channels have equal F matrices then they are equivalent. This is because they both enact the same action on an arbitrary qubit state (and thus act identically on the entire input space). Thus we aim to prove the theorem by equating the above F matrix of a simulated channel with that of our decomposition defined in Eq. (2.42). From the F matrices of $\mathcal{E}_\eta$ and $\mathcal{E}_P$, we derive

---

[8]Except in the special cases $\eta = 0, 1$.

that $\mathcal{E}_+ := \mathcal{E}_\eta \circ \mathcal{E}_P$ and $\mathcal{E}_- := \sigma_x \circ \mathcal{E}_\eta \circ \mathcal{E}_P$ have F matrices:

$$
F_+ = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{1-\eta}q_1 & 0 & 0 \\ 0 & 0 & -\sqrt{1-\eta}q_2 & 0 \\ \eta & 0 & 0 & (1-\eta)q_3 \end{pmatrix}, \tag{2.44}
$$

$$
F_- = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{1-\eta}q_1 & 0 & 0 \\ 0 & 0 & \sqrt{1-\eta}q_2 & 0 \\ -\eta & 0 & 0 & -(1-\eta)q_3 \end{pmatrix}, \tag{2.45}
$$

where $(q_1, q_2, q_3) \in \mathcal{T}$. As $\gamma \geq 0$, but $S_{30} \in [-1, 1]$ (a fact easy to see from its structure), this means we can reduce theorem 2.5.1 to this equivalent proposition.

Proposition 2.5.2 For any channel $\mathcal{E}_{\text{sim}}$ simulated by resource state χ_γ using noisy teleportation, we may write the following equality:

$$
F_{\text{sim}} = \begin{cases} F_+ & \text{if } S_{30} \geq 0, \\ F_- & \text{if } S_{30} \leq 0. \end{cases} \tag{2.46}
$$

with F_+/F_- generated by valid Pauli and amplitude damping channels. One can see that this equality defines a unique value for u and η, and a unique vector $\mathbf{q}$ - except when $\eta = 1$, in which case $\mathbf{q}$ is arbitrary, and $\eta = 0$, where all decompositions have degeneracy $u = 0, \mathbf{q} = (q_1, q_2, q_3)$ and $u = 1, \mathbf{q} = (q_1, -q_2, -q_3)$.

We thus split this proof into two cases.

Case 1: $S_{30} \geq 0$.

First we equate the f_{30} components: $\eta = \gamma S_{30}$. Since both $\gamma, S_{30} \in [0, 1]$ this is a valid η value. We now equate the diagonal components of the two sides to obtain

$$
\left(\sqrt{1-\eta}q_1, -\sqrt{1-\eta}q_2, (1-\eta)q_3 \right) = \left(\sqrt{1-\gamma}S_{11}, -\sqrt{1-\gamma}S_{22}, (1-\gamma)S_{33} \right) \tag{2.47}
$$

$$
\Rightarrow (q_1, q_2, q_3) = \left(\sqrt{\frac{1-\gamma}{1-\eta}}S_{11}, \sqrt{\frac{1-\gamma}{1-\eta}}S_{22}, \frac{1-\gamma}{1-\eta}S_{33} \right)
$$

$$
\Rightarrow (q_1, q_2, q_3) = \left(\sqrt{\frac{1-\gamma}{1-\gamma S_{30}}}S_{11}, \sqrt{\frac{1-\gamma}{1-\gamma S_{30}}}S_{22}, \frac{1-\gamma}{1-\gamma S_{30}}S_{33} \right). \tag{2.48}
$$

We can always do this rearrangement unless $\eta = 1$ - this forces both $\gamma, S_{30} = 1$ also, which trivially satisfy Eq. (2.47). For the rest of the proof, we shall assume $\eta \in [0, 1)$. It remains to prove that this vector describes a valid Pauli channel, and from corollary 2.3.4 this is equivalent to proving that the vector $\mathbf{q} \in \mathcal{T}$.

Lemma 2.5.3 For all classical channels Π, as defined in our noisy teleportation protocol, we have that (S_{11}, S_{22}, S_{33}) belongs to the tetrahedron $\mathcal{T}$.

Proof. An alternative way to define $\mathcal{T}$ is by four inequalities which are satisfied by all points within the tetrahedron, namely

$$x + y + z \leq 1,$$
$$x - y - z \leq 1,$$
$$-x + y - z \leq 1,$$
$$-x - y + z \leq 1.$$

We have already seen these used in corollary 2.3.4. We may substitute into these S_{11}, S_{22} and S_{33}, obtaining

$$S_{11} + S_{22} + S_{33} = 1 - (p_{02} + p_{13} + p_{20} + p_{31}) \leq 1,$$
$$S_{11} - S_{22} - S_{33} = 1 - (p_{03} + p_{12} + p_{21} + p_{30}) \leq 1,$$
$$-S_{11} + S_{22} - S_{33} = 1 - (p_{00} + p_{11} + p_{22} + p_{33}) \leq 1,$$
$$-S_{11} - S_{22} + S_{33} = 1 - (p_{01} + p_{10} + p_{23} + p_{32}) \leq 1.$$

From this, we can conclude that all (S_{11}, S_{22}, S_{33}) possible belong to the tetrahedron. $\quad\square$

Lemma 2.5.4 If a point (x, y, z) belongs to the tetrahedron defined by $\mathcal{T}$, then so too does the point $(\sqrt{\alpha}x, \sqrt{\alpha}y, \alpha z)$, where $\alpha \in [0, 1]$.

Proof. Since any point in the tetrahedron can be expressed as a convex combination of the four extremal points in $\mathcal{T}$, it is sufficient to show that the four points,

$$
\begin{array}{ll}
(\quad\sqrt{\alpha}, -\sqrt{\alpha}, \quad\alpha), & (\quad\sqrt{\alpha}, \quad\sqrt{\alpha}, -\alpha), \\
(-\sqrt{\alpha}, -\sqrt{\alpha}, -\alpha), & (-\sqrt{\alpha}, \quad\sqrt{\alpha}, \quad\alpha),
\end{array}
\qquad (2.49)
$$

belong to the tetrahedron (i.e. are themselves a convex combination of the four extremal points), and thus any rescaled tetrahedron point also still remains with the full tetrahedron,

since it may be written as a convex combination of the four points above.

Expressing an arbitrary point in $\mathcal{T}$ as

$$(x, y, z) = \sum_{i=0}^{3} p_i \mathbf{e}_i, \quad \sum_{i=0}^{3} p_i = 1, \quad p_i \geq 0, \tag{2.50}$$

then we can achieve the points in Eq. (2.49)

$$
\begin{array}{ccccc}
\text{Point} & p_0 & p_1 & p_2 & p_3 \\
(\sqrt{\alpha}, -\sqrt{\alpha}, \alpha) & \frac{(1+\sqrt{\alpha})^2}{4} & \frac{1-\alpha}{4} & \frac{1-\alpha}{4} & \frac{(1-\sqrt{\alpha})^2}{4} \\
(\sqrt{\alpha}, \sqrt{\alpha}, -\alpha) & \frac{1-\alpha}{4} & \frac{(1-\sqrt{\alpha})^2}{4} & \frac{(1+\sqrt{\alpha})^2}{4} & \frac{1-\alpha}{4} \\
(-\sqrt{\alpha}, -\sqrt{\alpha}, -\alpha) & \frac{1-\alpha}{4} & \frac{(1+\sqrt{\alpha})^2}{4} & \frac{(1-\sqrt{\alpha})^2}{4} & \frac{1-\alpha}{4} \\
(-\sqrt{\alpha}, \sqrt{\alpha}, \alpha) & \frac{(1-\sqrt{\alpha})^2}{4} & \frac{1-\alpha}{4} & \frac{1-\alpha}{4} & \frac{(1+\sqrt{\alpha})^2}{4}
\end{array}
\tag{2.51}
$$

The normalisation condition is easy to verify:

$$\frac{(1+\sqrt{\alpha})^2}{4} + \frac{1-\alpha}{4} + \frac{1-\alpha}{4} + \frac{(1-\sqrt{\alpha})^2}{4} = 1. \tag{2.52}$$

$\square$

Lemma 2.5.5 Given $S_{30} \geq 0$, $(1 - \gamma)/(1 - \gamma S_{30}) \in [0, 1]$.

Proof. As $\gamma \in [0, 1]$, we have $(1 - \gamma) \in [0, 1]$ also. Moreover, $S_{30} \in [0, 1]$, and we can therefore write

$$1 \geq \gamma \geq \gamma S_{30}$$

$$\Rightarrow -1 \leq -\gamma \leq -\gamma S_{30}$$

$$\Rightarrow 0 \leq 1 - \gamma \leq 1 - \gamma S_{30}$$

$$\Rightarrow 0 \leq \frac{1 - \gamma}{1 - \gamma S_{30}} \leq 1$$

as we took $\eta = \gamma S_{30} \in [0, 1)$. $\square$

Combining the results of lemmas 2.5.4 and 2.5.5 by setting $\alpha = (1 - \gamma)/(1 - \gamma S_{30})$, we can conclude that if (S_{11}, S_{22}, S_{33}) is in $\mathcal{T}$, so too is (q_1, q_2, q_3). Since (S_{11}, S_{22}, S_{33}) is in $\mathcal{T}$ by lemma 2.5.3, we may conclude the defined $\mathcal{E}_P$ is a valid Pauli channel, verifying our decomposition.

Case 2: $S_{30} \leq 0$.

Equating F_{sim} and F_-, we now find $\eta = -\gamma S_{30}$, though due to the sign of S_{30} we still find

$\eta \in [0,1]$. The diagonal elements now give us that

$$(q_1, q_2, q_3) = \left(\frac{\sqrt{1-\gamma}}{\sqrt{1+\gamma S_{30}}} S_{11}, -\frac{\sqrt{1-\gamma}}{\sqrt{1+\gamma S_{30}}} S_{22} - \frac{1-\gamma}{1+\gamma S_{30}} S_{33} \right). \qquad (2.53)$$

Again, this rearrangement is acceptable unless $\eta = 1$, forcing $\gamma = 1$ and $S_{30} = -1$ - which trivially satisfy $F_{\text{sim}} = F_-$ as in case 1. Thus we assume $\eta < 1$ for the rest of this proof.

Lemma 2.5.6 For all classical channels Π, as defined in our noisy teleportation protocol, we have that $(S_{11}, -S_{22}, -S_{33})$ belongs to the tetrahedron $\mathcal{T}$.

Proof. Consider the four equations in corollary 2.3.4, with the vector $(S_{11}, -S_{22}, -S_{33})$:

$$\begin{aligned}
S_{11} + (-S_{22}) + (-S_{33}) &= S_{11} - S_{22} - S_{33} \leq 1, \\
S_{11} - (-S_{22}) - (-S_{33}) &= S_{11} + S_{22} + S_{33} \leq 1, \\
-S_{11} + (-S_{22}) - (-S_{33}) &= -S_{11} - S_{22} + S_{33} \leq 1, \\
-S_{11} - (-S_{22}) + (-S_{33}) &= -S_{11} + S_{22} - S_{33} \leq 1.
\end{aligned}$$

where the inequalities come from the proof of lemma 2.5.3. $\qquad \square$

Lemma 2.5.7 Given $S_{30} \leq 0$, $(1-\gamma)/(1+\gamma S_{30}) \in [0,1]$.

Proof. As $-S_{30} \in [0,1]$:

$$1 \geq \gamma \geq -\gamma S_{30}$$
$$\Rightarrow -1 \leq -\gamma \leq \gamma S_{30}$$
$$\Rightarrow 0 \leq 1 - \gamma \leq 1 + \gamma S_{30}$$
$$\Rightarrow 0 \leq \frac{1-\gamma}{1+\gamma S_{30}} \leq 1$$

$\qquad \square$

We may then repeat the same logic used in the proof of case 1, applying lemma 2.5.4 with $\alpha = (1-\gamma)(1+\gamma S_{30})$ and $(S_{11}, -S_{22}, -S_{33}) \in \mathcal{T}$.

These two cases are sufficient to cover all χ_γ-simulated channels, and thus our result is proved. $\qquad \square$

Whilst we have shown that all the channels χ_γ simulable by noisy teleportation are necessarily of the form (2.42), we must also consider the converse - which channels of the

form (2.42) are χ_γ-simulable? We answer that question in the following theorem.

Theorem 2.5.8 Using noisy teleportation over the amplitude damping Choi matrix χ_γ with $\gamma \in (0,1)$, it is only possible to simulate channels of the form in Eq. (2.42) where $\eta \in [0,\gamma]$ and $\mathbf{q} = (q_1, q_2, q_3)$ belonging to the convex space bounded by the points

$$
\begin{aligned}
&\left(\sqrt{\frac{1-\gamma}{1-\eta}}, \qquad \pm\sqrt{\frac{1-\gamma}{1-\eta}}\left(1-\frac{\eta}{\gamma}\right), \mp\frac{1-\gamma}{1-\eta}\left(1-\frac{\eta}{\gamma}\right)\right), \\
&\left(\pm\sqrt{\frac{1-\gamma}{1-\eta}}\left(1-\frac{\eta}{\gamma}\right), \quad \sqrt{\frac{1-\gamma}{1-\eta}}, \qquad \mp\frac{1-\gamma}{1-\eta}\left(1-\frac{\eta}{\gamma}\right)\right), \\
&\left(-\sqrt{\frac{1-\gamma}{1-\eta}}, \qquad \pm\sqrt{\frac{1-\gamma}{1-\eta}}\left(1-\frac{\eta}{\gamma}\right), \pm\frac{1-\gamma}{1-\eta}\left(1-\frac{\eta}{\gamma}\right)\right), \\
&\left(\pm\sqrt{\frac{1-\gamma}{1-\eta}}\left(1-\frac{\eta}{\gamma}\right), -\sqrt{\frac{1-\gamma}{1-\eta}}, \qquad \pm\frac{1-\gamma}{1-\eta}\left(1-\frac{\eta}{\gamma}\right)\right). \quad (2.54)
\end{aligned}
$$

These points correspond to the extremal points of the tetrahedron $\mathcal{T}$ truncated by the two planes $z = \pm(1-\eta/\gamma)$, and shrunk by the transformation

$$
(x, y, z) \rightarrow \left(\sqrt{\frac{1-\gamma}{1-\eta}}x, \sqrt{\frac{1-\gamma}{1-\eta}}y, \frac{1-\gamma}{1-\eta}z \right). \quad (2.55)
$$

For $\gamma = 0$, $\eta = 0$ and $\mathcal{E}_P$ is any valid Pauli channel $\mathbf{q} \in \mathcal{T}$, whilst for $\gamma = 1$, $\eta \in [0,1]$ and $\mathcal{E}_P$ is in the convex space (2.54) unless $\eta = 1$ also - in which case $\mathcal{E}_P$ is irrelevant and we set $\mathbf{q} = \mathbf{0}$.

Before we prove this result, we take from it the following definition:

Definition 2.5.9 We define the Pauli-damping channels as the class of qubit channels that are simulable by teleporting over the amplitude damping Choi matrix χ_γ and using a classical channel Π for the CCs. They have a unique decomposition of the form in theorem 2.5.1, and must satisfy the criteria in theorem 2.5.8.

Proof. First we deal with the two extremal cases, $\gamma = 0, 1$. When $\gamma = 0$, we must have that $\eta = |\gamma S_{30}| = 0$ also. This means S_{30} is not fixed and $(S_{11}, S_{22}, S_{33}) = \mathbf{q}$. By setting:

$$
\begin{aligned}
p_{0|0} = p_{1|1} = p_{2|2} = p_{3|3} = 1 &\rightarrow \mathbf{q} = (1, -1, 1), \\
p_{1|0} = p_{0|1} = p_{3|2} = p_{2|3} = 1 &\rightarrow \mathbf{q} = (1, 1, -1), \\
p_{2|0} = p_{3|1} = p_{0|2} = p_{1|3} = 1 &\rightarrow \mathbf{q} = (-1, -1, -1), \\
p_{3|0} = p_{2|1} = p_{1|2} = p_{0|3} = 1 &\rightarrow \mathbf{q} = (-1, 1, 1),
\end{aligned}
$$

we can produce the extremal points of $\mathcal{T}$ - consequently we can set $\mathbf{q}$ to any point in $\mathcal{T}$, by taking convex combinations of these distributions.

For $\gamma = 1$, the proof below for $\gamma \in (0, 1)$ holds, unless $\eta = 1$ also. However for this case the channel $\mathcal{E}_\eta$ sends every state to $|0\rangle$ - making our Pauli channel in the decomposition irrelevant. We can therefore choose it to be the channel defined by $\mathbf{q} = \mathbf{0}$.

Fixing $\gamma \in (0, 1)$, we know the sum S_{30} may take any value in $[-1, 1]$, and that $\eta = |\gamma S_{30}|$ - so we may conclude we are free to choose $\eta \in [0, \gamma]$. For the Pauli channel $\mathcal{E}_P$ in the decomposition, we know its defining vector $\mathbf{q}$ must satisfy (using the same case 1: $S_{30} \geq 0$, case 2: $S_{30} \leq 0$ split):

Case 1:

$$\mathbf{q} = \left(\frac{\sqrt{1-\gamma}}{\sqrt{1-\gamma S_{30}}} S_{11}, \quad \frac{\sqrt{1-\gamma}}{\sqrt{1-\gamma S_{30}}} S_{22}, \quad \frac{1-\gamma}{1-\gamma S_{30}} S_{33} \right) \tag{2.56}$$

$$= \left(\frac{\sqrt{1-\gamma}}{\sqrt{1-|\gamma S_{30}|}} S_{11}, \quad \frac{\sqrt{1-\gamma}}{\sqrt{1-|\gamma S_{30}|}} S_{22}, \quad \frac{1-\gamma}{1-|\gamma S_{30}|} S_{33} \right). \tag{2.57}$$

Case 2:

$$\mathbf{q} = \left(\frac{\sqrt{1-\gamma}}{\sqrt{1+\gamma S_{30}}} S_{11}, \quad -\frac{\sqrt{1-\gamma}}{\sqrt{1+\gamma S_{30}}} S_{22}, \quad -\frac{1-\gamma}{1+\gamma S_{30}} S_{33} \right) \tag{2.58}$$

$$= \left(\frac{\sqrt{1-\gamma}}{\sqrt{1-|\gamma S_{30}|}} S_{11}, \quad -\frac{\sqrt{1-\gamma}}{\sqrt{1-|\gamma S_{30}|}} S_{22}, \quad -\frac{1-\gamma}{1-|\gamma S_{30}|} S_{33} \right). \tag{2.59}$$

We have already shown that the two vectors (S_{11}, S_{22}, S_{33}), $(S_{11}, -S_{22}, -S_{33})$ lie within $\mathcal{T}$, so we know with certainty that allowable Pauli channels will certainly lie within the shrunk tetrahedron with extremal points:

$$\left(\frac{\sqrt{1-\gamma}}{\sqrt{1-\eta}}, -\frac{\sqrt{1-\gamma}}{\sqrt{1-\eta}}, \frac{1-\gamma}{1-\eta} \right)$$

$$\left(\frac{\sqrt{1-\gamma}}{\sqrt{1-\eta}}, \frac{\sqrt{1-\gamma}}{\sqrt{1-\eta}}, -\frac{1-\gamma}{1-\eta} \right)$$

$$\left(-\frac{\sqrt{1-\gamma}}{\sqrt{1-\eta}}, -\frac{\sqrt{1-\gamma}}{\sqrt{1-\eta}}, -\frac{1-\gamma}{1-\eta} \right)$$

$$\left(-\frac{\sqrt{1-\gamma}}{\sqrt{1-\eta}}, \frac{\sqrt{1-\gamma}}{\sqrt{1-\eta}}, \frac{1-\gamma}{1-\eta} \right) \tag{2.60}$$

where we have used that $\eta = |\gamma S_{30}|$. However, we have now fixed our η value, forcing $S_{30} = \pm \eta / \gamma$. As the summations S_{11}, S_{22}, S_{33} are over the same parameters as S_{30} - the conditional probabilities $p_{l|k}$ - this imposes some structure on them.

S_{30}	$\#(\{Q_m\})$	$\#(\{\mathcal{S}_\pm(Q_m^\pm)\})$	$\#(\mathcal{S})$.
$\lvert S_{30}\rvert = 0$	96	19	4
$0 < \lvert S_{30}\rvert < \frac{1}{2}$	384	80	8
$\lvert S_{30}\rvert = \frac{1}{2}$	64	16	8
$\frac{1}{2} < \lvert S_{30}\rvert < 1$	128	48	8
$\lvert S_{30}\rvert = 1$	16	9	4

Table 2.1: The cardinality of sets constrained by the condition $S_{30} = k$.

To analyse the possible values for S_{11}, S_{22}, S_{33}, we use a technique known as *vertex enumeration* (described in detail in section 5.6). This obtains the extremal points of a set X given that $\forall \mathbf{x} \in X$ satisfy a set of inequality constraints $\mathbf{c}_i^T \mathbf{x} \leq b_i$, $\forall i$. In this case we are interested in the vectors $\boldsymbol{\pi} = (p_{0|0} \ldots p_{3|3})$ in the sets

$$\mathcal{P}_\eta^\pm = \left\{ \boldsymbol{\pi} \,\middle|\, p_{l|k} \geq 0, \ \sum_{k=0}^{3} p_{l|k} = 1, \ S_{30} = \pm\frac{\eta}{\gamma} \right\}. \tag{2.61}$$

We denote the sets of extremal points $\{Q_m^\pm\}_m$. Now we may consider (S_{11}, S_{22}, S_{33}), $(S_{11}, -S_{22}, -S_{33})$ as two linear functions, $\mathcal{S}_+$ and $\mathcal{S}_-$, which map

$$\mathcal{S}_\pm : \mathcal{P}_\eta^\pm \to \mathcal{T}.$$

Since $\{Q_m^\pm\}$ are extremal points of $\mathcal{P}_\eta^\pm$ we may express any $\boldsymbol{\pi} \in \mathcal{P}_\eta^\pm$ as $\boldsymbol{\pi} = \sum_m \pi_m Q_m^\pm$, $\pi_m \geq 0$, $\sum_m \pi_m = 1$. This means the transformations $\mathcal{S}_\pm$ act as the following

$$\mathcal{S}_\pm(\boldsymbol{\pi}) = \mathcal{S}_\pm\left(\sum_m \pi_m Q_m^\pm\right) = \sum_m \pi_m \mathcal{S}_\pm\left(Q_m^\pm\right) \tag{2.62}$$

due to the linearity of $S_\pm$. From this, we may conclude that the extremal points of the allowable (S_{11}, S_{22}, S_{33}), $(S_{11}, -S_{22}, -S_{33})$ are simply a subset $\mathcal{S} \subseteq \{\mathcal{S}_\pm(Q_m^\pm)\}_m$.

Using the vertex enumeration program PANDA [71] to find the extremal points $Q_m^\pm$, the set $\{\mathcal{S}_\pm(Q_m^\pm)\}$ was then calculated; points $\mathbf{s} \in \{\mathcal{S}_\pm(Q_m^\pm)\}$ were tested for extremality by testing whether they could be written as a convex combination of $\{\mathcal{S}_\pm(Q_m^\pm)\} \setminus \{\mathbf{s}\}$ - if it cannot, it must be extremal. Thus $\mathcal{S}$ was obtained. A summary of the process is presented in table 2.1.

The 8 points referenced in table 2.1 (which reduce to four points in the special $S_{30} = 0, \pm 1$ cases) are the points

$$\left(\begin{array}{ccc} 1 & , \pm\left(1 - \dfrac{\eta}{\gamma}\right) & , \mp\left(1 - \dfrac{\eta}{\gamma}\right) \end{array} \right), \left(\begin{array}{ccc} \pm\left(1 - \dfrac{\eta}{\gamma}\right), & 1 & , \mp\left(1 - \dfrac{\eta}{\gamma}\right) \end{array} \right),$$
$$\left(\begin{array}{ccc} -1 & , \pm\left(1 - \dfrac{\eta}{\gamma}\right) & , \pm\left(1 - \dfrac{\eta}{\gamma}\right) \end{array} \right), \left(\begin{array}{ccc} \pm\left(1 - \dfrac{\eta}{\gamma}\right), & -1 & , \pm\left(1 - \dfrac{\eta}{\gamma}\right) \end{array} \right), \quad (2.63)$$

regardless of the case (S_{30} positive or negative). These points correspond to $\mathcal{T}$, truncated by two planes at $z = \pm(1 - \eta/\gamma)$. Combining this result with the shrinking factors of $\left(\sqrt{(1-\gamma)/(1-\eta)}, \sqrt{(1-\gamma)/(1-\eta)}, (1-\gamma)/(1-\eta)\right)$ gets our final result. $\qquad\square$

Corollary 2.5.10 Any "non-trivial" amplitude damping channel $\mathcal{E}_\gamma$ (i.e. $\gamma \neq 0, 1$) cannot be χ_γ-simulated using the noisy teleportation protocol.

Proof. In order for this to occur, we require $\eta = \gamma$, and $\mathcal{E}_P = \mathbb{I}$, which have seen already equates to $\mathbf{q} = (1, -1, 1)$. However, $\eta = \gamma$ means every extremal point $\mathbf{s} \in S$ has z coefficient $\pm(1 - \gamma/\gamma) = 0$, and thus every allowable $\mathbf{q}$ must have z component 0 also. $\qquad\square$

Special consideration should also be given to the two extremal cases, $\gamma = 0, 1$. For $\gamma = 0$, we have $\mathcal{E}_\gamma = \mathbb{I}$, and $\chi_\gamma = |\Phi^+\rangle\langle\Phi^+|$. We can Choi-simulate $\mathcal{E}_\gamma$ simply by following the standard teleportation protocol. When $\gamma = 1$, $\mathcal{E}_\gamma$ maps every state to $|0\rangle\langle0|$, and so the Choi matrix is $I/2 \otimes |0\rangle\langle0|$. Therefore Bob may simply apply the identity matrix to simulate the channel[9], regardless of Alice's measurement result. For this resource state, only convex combinations of $|0\rangle\langle0|$ and $|1\rangle\langle1|$ are possible from the application of the corrective Pauli unitaries. This follows from our decomposition, as the constraint on $\mathbf{q}$ forces $\mathcal{E}_P = \mathcal{D}_1$ - the depolarising channel sending all states to $I/2$, with $\mathcal{E}_\eta$ and σ_x^u then determining the exact convex combination.

2.5.1 Distinguishability of Pauli-Damping Channels

We have shown how we can simulate non-Pauli channels using noisy teleportation. In this section we quantify the distance between the Pauli-damping channels and the closest Pauli channel, known as the *distinguishability.*

---

[9]Although this channel clearly has 0 capacity.
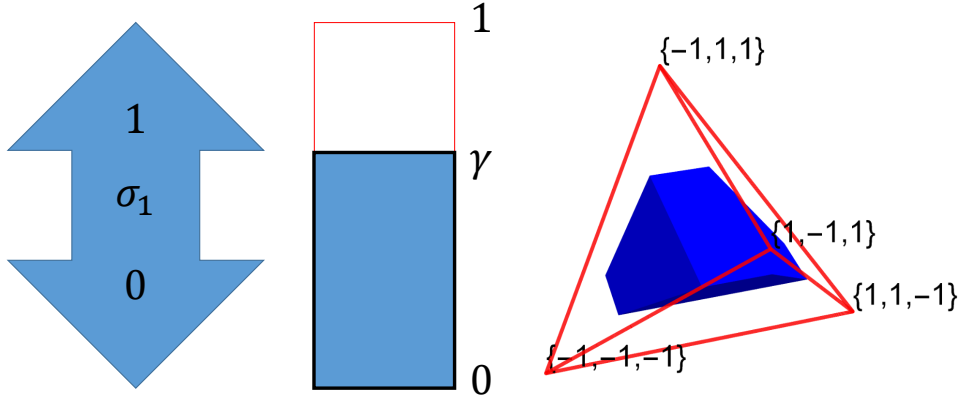
Figure 2.7: A representation of the three components of the Pauli-damping channels.

2.5.1.1 Distance in Trace Norm

Definition 2.5.11 The trace norm distance between two quantum channels $\mathcal{E}_1, \mathcal{E}_2$ is defined as

$$\|\mathcal{E}_1 - \mathcal{E}_2\|_1 := \sup_{\rho} \|\mathcal{E}_1(\rho) - \mathcal{E}_2(\rho)\|_1, \tag{2.64}$$

where $\|\sigma\|_1 = \mathrm{Tr}\left[\sqrt{\sigma\sigma^\dagger}\right]$.

For Hermitian σ, this is equivalent to the sum of the absolute values of the eigenvalues of σ. The trace norm between two channels also has the following operational meaning.

Lemma 2.5.12 ([53]) Given a single copy of a channel $\mathcal{E} : \mathcal{H}_A \to \mathcal{H}_B$ chosen uniformly at random from $\{\mathcal{E}_1, \mathcal{E}_2\}$, the minimal probability of incorrectly identifying the channel using an input state $\rho \in \mathcal{H}_A$ is given by:

$$p_{\min} = \frac{1}{2} - \frac{\|\mathcal{E}_1 - \mathcal{E}_2\|_1}{4}. \tag{2.65}$$

Proposition 2.5.13 Given a decomposition $\mathcal{E}_{\mathrm{sim}} = \sigma_x^u \circ \mathcal{E}_\eta \circ \mathcal{E}_P$ characterised by η and (q_1, q_2, q_3) respectively, then the trace norm between $\mathcal{E}_{\mathrm{sim}}$ and the closest Pauli channel $\mathcal{E}_{\mathrm{cl}}$ is simply η. Moreover, the closest Pauli channel has F matrix

$$(f_{11}, f_{22}, f_{33}) = \begin{cases} \left(\sqrt{1-\eta}q_1, -\sqrt{1-\eta}q_2, \quad (1-\eta)q_3\right) & \text{for } u = 0, \\ \left(\sqrt{1-\eta}q_1, \quad \sqrt{1-\eta}q_2, -(1-\eta)q_3\right) & \text{for } u = 1. \end{cases} \tag{2.66}$$

with $f_{ij} = 0$, $i \neq j$.

Proof. For qubit states, the trace norm between two states is equivalent to the Euclidean distance between their respective Bloch vectors. We shall split the proof into two cases:

Case 1: $u = 0$.

When $u = 0$, the action of $\mathcal{E}_{\text{sim}}$ is

$$\mathcal{E}_{\text{sim}} : (x, y, z) \rightarrow \left(\sqrt{1 - \eta} q_1 x, -\sqrt{1 - \eta} q_2 y, (1 - \eta) q_3 z + \eta \right) \tag{2.67}$$

whilst an arbitrary Pauli channel[10] $\mathcal{E}_C$ with parameters $(c_1, c_2, c_3) \in \mathcal{T}$ has action:

$$\mathcal{E}_C : (x, y, z) \rightarrow (c_1 x, -c_2 y, c_3 z). \tag{2.68}$$

This means we can reformulate our problem to:

$$\min_{(c_1, c_2, c_3) \in \mathcal{T}} \quad \max_{x, y, z: \, x^2 + y^2 + z^2 \leq 1}$$
$$\left((\sqrt{1 - \eta} q_1 - c_1) x \right)^2 + \left(-(\sqrt{1 - \eta} q_2 - c_2) y \right)^2 + \left(((1 - \eta) q_3 - c_3) z + \eta \right)^2, \tag{2.69}$$

minimising the square of the trace norm between the two output states. Looking at the final term $(((1 - \eta) q_3 - c_3) z + \eta)^2$, given that the maximum occurs for a given $|z|$ value, this term can be rewritten as

$$\max \left\{ \left(((1 - \eta) q_3 - c_3) |z| + \eta \right)^2, \left(-((1 - \eta) q_3 - c_3) |z| + \eta \right)^2 \right\} = (|(1 - \eta) q_3 - c_3||z| + \eta)^2. \tag{2.70}$$

This term is minimised at $c_3 = (1 - \eta) q_3$ regardless of the value of $|z|$, and has value η^2. For the term

$$\left((\sqrt{1 - \eta} q_1 - c_1) x \right)^2 \tag{2.71}$$

we can see this will be minimised when $c_1 = \sqrt{1 - \eta} q_1$ and will have value 0, regardless of $|x|$. Similarly,

$$\left(-(\sqrt{1 - \eta} q_2 - c_2) y \right)^2 \tag{2.72}$$

is minimised when $c_2 = \sqrt{1 - \eta} q_2$.

Summarising this, we have found our trace norm-minimising channel provided

$$\mathcal{E}_C : (x, y, z) \rightarrow \left(\sqrt{1 - \eta} q_1 x, -\sqrt{1 - \eta} q_2 y, (1 - \eta) q_3 z \right) \tag{2.73}$$

---

[10]We use subscript C, to distinguish it from $\mathcal{E}_{\text{cl}}$ and $\mathcal{E}_P$.

is a valid Pauli channel, equivalent to the condition $\left(\sqrt{1-\eta}q_1, \sqrt{1-\eta}q_2, (1-\eta)q_3\right) \in \mathcal{T}$. Since $(q_1, q_2, q_3) \in \mathcal{T}$ and $(1-\eta) \in [0,1]$, we may apply lemma 2.5.4 to conclude this is indeed the case.

Case 2: $u = 1$.

Using a similar method, our problem can be written in this case as

$$\min_{(c_1,c_2,c_3)\in\mathcal{T}} \quad \max_{x,y,z:x^2+y^2+z^2\leq 1}$$

$$\left((\sqrt{1-\eta}q_1 - c_1)x\right)^2 + \left((-\sqrt{1-\eta}q_2 - c_2)y\right)^2 + \left((-(1-\eta)q_3 - c_3)z - \eta\right)^2. \quad (2.74)$$

Again, we begin by looking at the final part of the sum. For a fixed value of $|z|$, this term will be

$$\max\left\{\left((-(1-\eta)q_3 - c_3)\,|z| - \eta\right)^2, \left(((1-\eta)q_3 + c_3)\,|z| - \eta\right)^2\right\}$$

$$= \max\left\{\left(((1-\eta)q_3 + c_3)\,|z| + \eta\right)^2, \left(-((1-\eta)q_3 + c_3)\,|z| + \eta\right)^2\right\}$$

$$= \left(|(1-\eta)q_3 + c_3|\,|z| + \eta\right)^2. \quad (2.75)$$

This is clearly minimised when $c_3 = -(1-\eta)q_3$. For the x and y terms, they are minimised for

$$c_1 = \sqrt{1-\eta}q_1, \ c_2 = -\sqrt{1-\eta}q_2.$$

To see that this vector defines a valid Pauli channel, one can notice that $\mathcal{T}$ is invariant under the Bloch action of σ_x as it simply permutes the extremal points. Therefore we can conclude that if (q_1, q_2, q_3) belongs to the tetrahedron so too does $(q_1, -q_2, -q_3)$. As this is exactly the relation between the terms (c_1, c_2, c_3) for cases 1 and 2, we may infer that (c_1, c_2, c_3) again defines a valid Pauli channel, and the second part of our proposition is proved. □

In general the trace norm is *not* the most accurate way of measuring channel distinguishability. This is due to the detail that, by sending part of an entangled state through the channel, some channels may be distinguished with better probability than that provided by the trace norm. A better distance measure is given by the *diamond norm*, defined in the following way:

Definition 2.5.14 The diamond norm distance between two quantum channels $\mathcal{E}_1, \mathcal{E}_2$ is defined as

$$\|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond := \sup_{\rho \in \kappa \otimes \mathcal{H}} \|(\mathbb{I}_\kappa \otimes \mathcal{E}_1)(\rho) - (\mathbb{I}_\kappa \otimes \mathcal{E}_2)(\rho)\|_1. \tag{2.76}$$

with κ an ancillary Hilbert space of any dimension.

It is immediately obvious that $\|\mathcal{E}_1 - \mathcal{E}_2\|_1 \leq \|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond$ - and like the trace norm, this measure too has an operational motivation.

Lemma 2.5.15 ([109]) Given a single copy of a channel $\mathcal{E} : \mathcal{H}_A \to \mathcal{H}_B$ chosen uniformly at random from $\{\mathcal{E}_1, \mathcal{E}_2\}$, the minimal probability of incorrectly identifying the channel is given by:

$$p_{\min} = \frac{1}{2} - \frac{\|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond}{4}. \tag{2.77}$$

Proposition 2.5.16 Given a decomposition $\mathcal{E}_{\text{sim}} = \sigma_x^u \circ \mathcal{E}_\eta \circ \mathcal{E}_P$ characterised by η and (q_1, q_2, q_3) respectively, then the diamond norm between $\mathcal{E}_{\text{sim}}$ and the closest Pauli channel $\mathcal{E}_{\text{cl}}$ is η, equivalent to the trace norm, with the closest channel under $\|\cdot\|_\diamond$ the same as $\|\cdot\|_1$ given in proposition 2.5.13.

Although the diamond norm appears to be a difficult quantity to obtain, due to the unbounded dimension of the ancillary Hilbert space, we may use the following lemma to greatly simplify the problem.

Lemma 2.5.17 ([9]) There exists a $\rho \in \mathcal{H} \otimes \mathcal{H}$ which achieves the diamond norm between two channels $\mathcal{E}_1, \mathcal{E}_2 : \mathcal{H} \to \mathcal{H}'$.

Corollary 2.5.18 ([9]) The diamond norm between two qubit channels $\mathcal{E}_1, \mathcal{E}_2$ is achieved by a two-qubit state.

Proof. Once more we split the proof of proposition 2.5.16 into two cases, depending on the value of u.

Case 1: $u = 0$.

We start with the knowledge that

$$\min_{\mathcal{E}' \in \text{Pauli}} \|\mathcal{E}_{\text{sim}} - \mathcal{E}'\|_\diamond \leq \|\mathcal{E}_{\text{sim}} - \mathcal{E}_{\text{cl}}\|_\diamond. \tag{2.78}$$

since $\mathcal{E}_{\text{cl}}$ is itself Pauli. To find $\|\mathcal{E}_{\text{sim}} - \mathcal{E}_{\text{cl}}\|_\diamond$, we shall look at the quantity

$$\|(\mathbb{I}_2 \otimes \mathcal{E}_{\text{sim}})(\rho) - (\mathbb{I}_2 \otimes \mathcal{E}_{\text{cl}})(\rho)\|_1 \tag{2.79}$$

for an arbitrary two-qubit state ρ, using the decomposition given in Eq. (2.16). We find that $M_D = (\mathbb{I}_2 \otimes \mathcal{E}_{\text{sim}})(\rho) - (\mathbb{I}_2 \otimes \mathcal{E}_{\text{cl}})(\rho)$ is given by

$$
M_D = \frac{1}{4} \begin{pmatrix}
(1+a_3)\eta & 0 & (a_1 - ia_2)\eta & 0 \\
0 & (1+a_3)\eta & 0 & -(a_1 - ia_2)\eta \\
(a_1 + ia_2)\eta & 0 & (1-a_3)\eta & 0 \\
0 & -(a_1 + ia_2)\eta & 0 & (-1+a_3)\eta
\end{pmatrix}. \tag{2.80}
$$

As M_D is Hermitian, the trace norm is equal to the sum of the absolute values of eigenvalues. The eigenvalues of M_D are:

$$
\begin{aligned}
&\frac{1}{4}\left(-1-\sqrt{a_1^2 + a_2^2 + a_3^2}\right)\eta, && \frac{1}{4}\left(1-\sqrt{a_1^2 + a_2^2 + a_3^2}\right)\eta, \\
&\frac{1}{4}\left(-1+\sqrt{a_1^2 + a_2^2 + a_3^2}\right)\eta, && \frac{1}{4}\left(1+\sqrt{a_1^2 + a_2^2 + a_3^2}\right)\eta.
\end{aligned} \tag{2.81}
$$

Since $a_1^2 + a_2^2 + a_3^2 \leq 1$, the absolute values are:

$$
\begin{aligned}
&\frac{1}{4}\left(1+\sqrt{a_1^2 + a_2^2 + a_3^2}\right)\eta, && \frac{1}{4}\left(1-\sqrt{a_1^2 + a_2^2 + a_3^2}\right)\eta, \\
&\frac{1}{4}\left(1-\sqrt{a_1^2 + a_2^2 + a_3^2}\right)\eta, && \frac{1}{4}\left(1+\sqrt{a_1^2 + a_2^2 + a_3^2}\right)\eta.
\end{aligned} \tag{2.82}
$$

We can see that the sum of these values is η, irrespective of the input state ρ. We can thus conclude that $\|\mathcal{E}_{\text{sim}} - \mathcal{E}_{\text{cl}}\|_\diamond = \eta = \|\mathcal{E}_{\text{sim}} - \mathcal{E}_{\text{cl}}\|_1$.

Now let us suppose that there exists a channel $\mathcal{E}'$ which has a strictly smaller diamond norm from $\mathcal{E}_{\text{sim}}$ than our posited closest channel $\mathcal{E}_{\text{cl}}$. This means we may write the following chain of inequalities:

$$
\|\mathcal{E}_{\text{sim}} - \mathcal{E}'\|_1 \leq \|\mathcal{E}_{\text{sim}} - \mathcal{E}'\|_\diamond < \|\mathcal{E}_{\text{sim}} - \mathcal{E}_{\text{cl}}\|_\diamond = \eta = \|\mathcal{E}_{\text{sim}} - \mathcal{E}_{\text{cl}}\|_1 \tag{2.83}
$$

however, we proved in proposition 2.5.13 that the closest channel under trace norm was $\mathcal{E}_{\text{cl}}$, thus leading to a contradiction. We can therefore conclude that the diamond norm between $\mathcal{E}_{\text{sim}}$ and $\mathcal{E}_{\text{cl}}$ is minimal.

Case 2: $u = 1$.

In this case, we have our simulated channel is defined by $\mathcal{E}_{\text{sim}} = \sigma_x \circ \mathcal{E}_\eta \circ \mathcal{E}_P$ - we can define a related channel, $\mathcal{E}_{\text{pos}} = \mathcal{E}_\eta \circ \mathcal{E}_P$. From case 1, we know that the closest channel to $\mathcal{E}_{\text{pos}}$ is the channel $\mathcal{E}_{\text{poscl}}$ with action $(f_{11}, f_{22}, f_{33}) = \left(\sqrt{1-\eta}q_1, -\sqrt{1-\eta}q_2, (1-\eta)q_3\right)$,

with diamond norm distance η. This means we can write:

$$\eta = \|\mathcal{E}_{\text{pos}} - \mathcal{E}_{\text{poscl}}\|_\diamond$$
$$= \sup_\rho \|(\mathbb{I}_2 \otimes \mathcal{E}_{\text{pos}})(\rho) - (\mathbb{I}_2 \otimes \mathcal{E}_{\text{poscl}})(\rho)\|_1$$
$$= \sup_\rho \|(\mathbb{I}_2 \otimes \sigma_x)(\mathbb{I}_2 \otimes \mathcal{E}_{\text{pos}})(\rho) - (\mathbb{I}_2 \otimes \sigma_x)(\mathbb{I}_2 \otimes \mathcal{E}_{\text{poscl}})(\rho)\|_1$$
$$= \sup_\rho \|(\mathbb{I}_2 \otimes \mathcal{E}_{\text{sim}})(\rho) - (\mathbb{I}_2 \otimes \mathcal{E}_{\text{cl}})(\rho)\|_1$$
$$= \|\mathcal{E}_{\text{sim}} - \mathcal{E}_{\text{cl}}\|_\diamond, \tag{2.84}$$

where we have used the fact that the trace norm is invariant under unitary operations, and that $\mathcal{E}_{\text{cl}}$ for the case $u = 1$ is equivalent to $\sigma_x \circ \mathcal{E}_{\text{poscl}}$. We can now use the same argument used in Eq. (2.83) to conclude that this norm is indeed minimal. $\square$

From this proposition we can conclude that, given resource state χ_γ, we can simulate a channel at most $\|\cdot\|_\diamond = \gamma$ distinguishable from the set of Pauli channels (since that is the largest allowable η). We can also state that, given a Pauli-damping channel $\mathcal{E}_{\text{sim}} = \sigma_x^u \circ \mathcal{E}_\eta \circ \mathcal{E}_P$, we can one-copy distinguish it from any Pauli channel with probability $p \geq 1/2 + \eta/4$, and give the Pauli channel for which equality holds.[11]

2.5.2 Capacities of Pauli-Damping Channels

As mentioned in the first section of this chapter, if a channel can be shown to be LOCC simulated over a pre-shared resource, we may upper bound the two-way entanglement distillation, quantum, private and secret key capacities by the REE of the resource state. Since the Pauli-damping channels were explicitly constructed by an LOCC protocol over resource state χ_γ, we may apply this result for a χ_γ-constructed Pauli-damping channel $\mathcal{E}$ to obtain

$$D_2(\mathcal{E}) = Q_2(\mathcal{E}) \leq P_2(\mathcal{E}) = K(\mathcal{E}),$$
$$K(\mathcal{E}) \leq E_R(\chi_\gamma) \leq \Phi(\mathcal{E}_\gamma) := \frac{1}{2} - \frac{1-\gamma}{2}\log_2\left(\frac{1-\gamma}{2}\right) + \frac{2-\gamma}{2}\log_2\left(\frac{2-\gamma}{2}\right). \tag{2.85}$$

This specific bound is obtained by calculating the relative entropy to a specific separable state, $\sigma_\gamma = (\mathbb{I} \otimes \mathcal{E}_\gamma)((|00\rangle\langle00| + |11\rangle\langle11|)/2)$. It is possible to see that this bound may

---

[11]In fact, for this particular channel, we may obtain the optimal probability by sending the maximally mixed state through the channel, measuring σ_z, and choosing $\mathcal{E}_{\text{sim}}$ if the output is 0 and $\mathcal{E}_{\text{cl}}$ otherwise.

not always be optimal; for example, if $\gamma = 0$ we have the maximally entangled state as our resource state, and $E_R(\chi_0) = E_R(|\Phi^+\rangle\langle\Phi^+|) = 1$ is our upper bound. However, if the Pauli-damping channel we are considering is the completely depolarising channel $\mathcal{D}_1 : \rho \to I/2$, then the capacity of the channel is 0. This is an extreme example, but it illustrates how the bound can fail to portray the true nature of the channel.

To counter this, we can consider an alternative upper bound, obtained by a *different* LOCC simulation protocol. This simulation takes advantage of the fact Pauli-damping channels are composite.

- Take a specific Pauli-damping channel $\mathcal{E}_{\text{sim}} = \sigma_x^u \circ \mathcal{E}_\eta \circ \mathcal{E}_P$.

- Alice and Bob have the pre-shared resource $\chi_{\mathcal{E}_P}$ - the Choi matrix of channel $\mathcal{E}_P$ in the decomposition of $\mathcal{E}_{\text{sim}}$.

- Alice and Bob perform the standard teleportation protocol over $\chi_{\mathcal{E}_P}$ - by corollary 2.3.4 this simulates the channel $\mathcal{E}_P$ from Alice to Bob.

- Bob then applies locally the channel $\sigma_x^u \circ \mathcal{E}_\eta$ to his received state; his final state is thus $\sigma_x^u \circ \mathcal{E}_\eta \circ \mathcal{E}_P(\rho) \equiv \mathcal{E}_{\text{sim}}(\rho)$.

- The channel $\mathcal{E}_{\text{sim}}$ has been $\chi_{\mathcal{E}_P}$-simulated, and therefore

$$D_2(\mathcal{E}) = Q_2(\mathcal{E}) \leq P_2(\mathcal{E}) = K(\mathcal{E}) \leq E_R(\chi_{\mathcal{E}_P}) \leq \Phi(\mathcal{E}_P) := \begin{cases} 1 - H_2(p_{\max}) & p_{\max} \geq \frac{1}{2}, \\ 0 & p_{\max} \leq \frac{1}{2}, \end{cases}$$
$$(2.86)$$

  with the final inequality coming from [85], and $p_{\max} = \max_i\{p_i\}$ in the decomposition (2.4) of $\mathcal{E}_P$. $\Phi(\mathcal{E}_P)$ is obtained by choosing the specific separable state $\sigma_P = (\mathbb{I} \otimes \mathcal{E}_P)((|00\rangle\langle00| + |11\rangle\langle11|)/2)$.

This upper bound is effectively the same principle that for composite channels, $\mathcal{C}(\mathcal{E}_2 \circ \mathcal{E}_1) \leq \mathcal{C}(\mathcal{E}_1)$. Although the upper bound is decomposition specific, we can compare the efficacy with the resource-based bound by noting that $\Phi(\mathcal{E}_P)$ is a convex function over the tetrahedron $\mathcal{T}$ - the four values p_i corresponding to the weights of each corner in the convex decomposition. Therefore, for a given γ, we can compare the upper bound given by χ_γ to that given by the extremal Pauli channels of the restricted tetrahedron in theorem 2.5.8. The Pauli-based bound will be greatest when the restricted tetrahedron is largest; occuring when $\eta = 0$. Our convex space is then bounded by the points:

$$\left(\sqrt{1-\gamma}, \pm\sqrt{1-\gamma}, \mp(1-\gamma) \right),$$

$$\left(\pm\sqrt{1-\gamma}, \sqrt{1-\gamma}, \mp(1-\gamma) \right),$$

$$\left(-\sqrt{1-\gamma}, \pm\sqrt{1-\gamma}, \pm(1-\gamma) \right),$$

$$\left(\pm\sqrt{1-\gamma}, -\sqrt{1-\gamma}, \pm(1-\gamma) \right).$$

Each of these is generated by a permutation of the four probabilities:

$$\left\{ \frac{\gamma}{4}, \frac{\left(1-\sqrt{1-\gamma}\right)^2}{4}, \frac{\left(1+\sqrt{1-\gamma}\right)^2}{4}, \frac{\gamma}{4} \right\}. \tag{2.87}$$

As $\gamma \in [0,1]$, it must be the case that $p_{\max} = \left(1+\sqrt{1-\gamma}\right)^2/4$, and $p_{\max} \geq 1/2$ when $\gamma \leq 2(\sqrt{2}-1)$. Therefore, applying our convexity argument, we have that

$$\Phi\left(\mathcal{E}_P\right) \leq \begin{cases} 1 - H_2\left(\frac{\left(1+\sqrt{1-\gamma}\right)^2}{4} \right) & \gamma \leq 2(\sqrt{2}-1), \\ 0 & \gamma \geq 2(\sqrt{2}-1). \end{cases} \tag{2.88}$$

In figure 2.8, we compare the the two bounds for varying γ. We see that the bound obtained by the Pauli channel in the decomposition is always tighter than that obtained by using the resource state! This means that, although we were able to simulate the Pauli-damping channel in a non-trivial way using a novel LOCC protocol, it is actually more beneficial to simulate part of the channel trivially (by local application of the $\sigma_x^u \circ \mathcal{E}_\eta$) allowing us to use an alternate resource state.

It should be noted that for both of our analytical bounds, we did not find the exact REE of our resource state, instead choosing a particular separable state to calculate the relative entropy - providing an upper bound for the REE. With this in mind, it is possible that the $E_R(\chi_\gamma)$ provides a better upper bound than $E_R(\chi_{\mathcal{E}_P})$ when the two quantities are exactly calculated. To see if this is the case, we note that for $\mathcal{H}_2 \otimes \mathcal{H}_2$ the REE and the RPPT coincide [58]. We use a semidefinite programming-based[12] numerical approximation of the RPPT, "CVXQUAD" [40, 41]. The numerical analysis of the two resource states is given in figure 2.9, showing that $E_R(\chi_{\mathcal{E}_P})$ still provides a better bound than $E_R(\chi_\gamma)$.

---

[12]These problems are explained in detail in chapter 5.
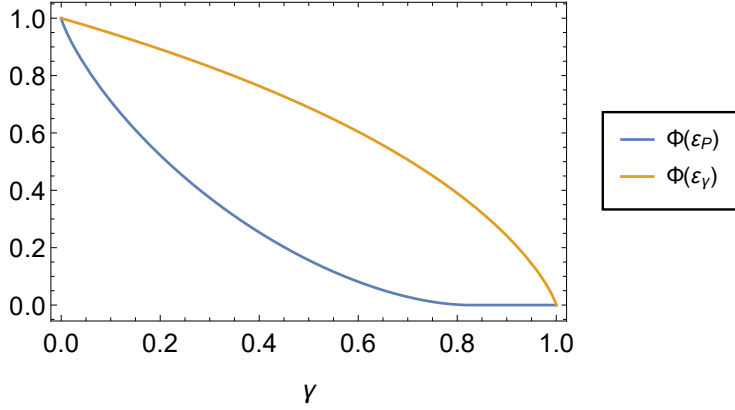
Figure 2.8: Here we compare our two analytical bounds for Pauli-damping channels, one based on the non-Bell-diagonal resource state χ_γ, the other from the Pauli channel in the decomposition. The Pauli-based bound is considerably tighter.

2.5.3 The "Squared" Pauli-Damping Channel

Corollary 2.5.10 proved that it was impossible to χ_γ-simulate $\mathcal{E}_\gamma$ (for $\gamma \neq 0, 1$); but we were still interested in how close a channel it *was* possible to simulate. Motivated by this, my summer project student Leon Hetzel performed the least-squares minimisation

$$\min_{\mathcal{E}_{\text{sim}}} \sum_{i,j} |[F_\gamma]_{ij} - [F_{\text{sim}}]_{ij}|^2 \tag{2.89}$$

with F_{sim} the F-matrix of a Pauli-damping channel,

$$F_{\text{sim}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{1-\eta}q_1 & 0 & 0 \\ 0 & 0 & -\sqrt{1-\eta}q_2 & 0 \\ \eta & 0 & 0 & (1-\eta)q_3 \end{pmatrix}. \tag{2.90}$$

Note that generally this type of minimisation does *not* coincide with the minimisation of the trace/diamond norm. The reason why we chose this measure was because we wanted to find the channel with the closest action *regardless* of the input state, rather than the trace/diamond norms, which only considers the worst case scenario input state.

The result of the optimisation (2.89) was the channel

$$\mathcal{E}_{\text{sq}} : (x, y, z) \to \left(\sqrt{1-\gamma}\left(1 - \frac{\gamma}{2}\right)x, \sqrt{1-\gamma}\left(1 - \frac{\gamma}{2}\right)y, \gamma^2 + (1-\gamma^2)z \right) \tag{2.91}$$
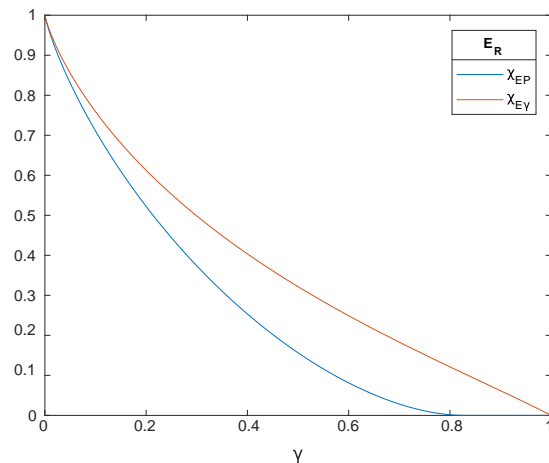
Figure 2.9: This figure compares the numerical REE bounds for secret key capacity of Pauli-damping channels. Although we see an improvement on $\Phi(\mathcal{E}_\gamma)$, the upper bound provided by $E_R(\chi_\gamma)$ is still looser than $E_R(\chi_{\mathcal{E}_P})$.

with F matrix

$$
F_{\mathrm{sq}} = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & \sqrt{1-\gamma}\left(1-\frac{\gamma}{2}\right) & 0 & 0 \\
0 & 0 & \sqrt{1-\gamma}\left(1-\frac{\gamma}{2}\right) & 0 \\
\gamma^2 & 0 & 0 & (1-\gamma^2)
\end{pmatrix}. \tag{2.92}
$$

We name this channel the *squared* channel due to its similarity of action on the z component of the Bloch vector to the original channel $\mathcal{E}_\gamma$, except with the square exponent.

This channel is a Pauli-damping channel (being χ_γ-simulable by noisy teleportation) and thus may be decomposed into the catenation of channels given in theorem 2.5.1. This particular channel has decomposition $u = 0$, $\eta = \gamma^2$, and

$$
\mathbf{q} = \left(\frac{1-\frac{\gamma}{2}}{\sqrt{1+\gamma}}, -\frac{1-\frac{\gamma}{2}}{\sqrt{1+\gamma}}, \frac{1-\gamma}{1+\gamma} \right). \tag{2.93}
$$

2.6 Discussion and Further Directions

The tool of LOCC-simulation in order to provide weak converse bounds of channel capacities is one with lots of potential; but comes with the caveat that a suitable resource state and LOCC protocol need to be found in order to apply the result. The most commonly studied LOCC protocol for this task has been the standard teleportation protocol, over

an arbitrary resource state. We extended this protocol by introducing classical noise into the communication step between Alice and Bob, allowing teleportation simulation of more than just the Pauli channels. One could generalise this further by allowing Bob different operations conditional on Alice's measurement outcome, or changing Alice's measurement. Not all non-Pauli channels are simulable using this new protocol though; although we have been able to introduce constant terms in the Bloch vector, as well as interdependent terms between elements, the interdependence of the sums S_{ij} constrains these terms. We were able to prove a no-go theorem which states that a non-Bell-diagonal resource state is necessary in order to simulate non-Pauli channels with this method, and were able to characterise the channels possible to simulate using the specific resource states χ_γ. Upon further inspection however, the composite nature of these channels meant that the standard teleportation protocol, followed by a local application of the remainder of the channel, provided a better upper bound on these channels' capacities. This is perhaps because the noise in the classical channel Π leads to a lower-capacity simulated channel, yet our upper bound is independent of Π. Whether the noisy teleportation protocol can provide better bounds for other channels remains uncertain; an important step towards discerning this would be to see if all channels simulated this way are composite, potentially with the first component channel always a Pauli channel.

A natural generalisation of this research is the extension to $d > 2$ - akin to the generalisation of Bowen and Bose. Though the Pauli matrices and noisy classical channel have natural generalisations to higher dimensions, there exists no concise Bloch representation for either the input nor the resource state, making the F-matrix representation redundant. Generally, the simulated channel would depend on $d^4 - 1$ parameters in the resource state, subject to positivity constraints, and $d^4 - d^2$ free parameters for the classical communication channel. Most interest in higher dimensional channels is limited to channels which are already teleportation covariant, and therefore this generalisation would be hard to present and likely of little relevance to the field. A more reasonable application of theorem 2.4.1 could be to numerically optimise over resource states, in order to find the tightest bound for general qubit channels. I have written a Matlab program using a least-squares approach to determine whether there exists σ, Π which simulate a given channel - this approach could be combined with E_R-approximating semidefinite programs to find an upper bound.

Another area of investigation is *port based teleportation* [63], as another LOCC protocol candidate [84]. In this process, many copies of the maximally entangled states are employed, to allow for teleportation of a state *without* the requirement of a corrective unitary - at the cost of fidelity. This has been shown to LOCC simulate a depolarising channel, with the probability of depolarisation decreasing as the number of maximally entangled states used increases. By replacing these maximally entangled states with more general qudit states, we will be able to simulate much more general channels - moreover, it appears likely that methods such as "parallel teleportation" (in which one sends over many copies of the state) and entanglement recycling (in which the resource state is reused for another round of port based teleportation) can provide tight, asymptotic upper bounds.

Overall, this work takes its place in the literature as a proof-of-concept that the teleportation protocol may be used as a template for LOCC simulation for more general channels than previously thought - and provides an explicit example, the Pauli-damping channels which are simulated this way. They are one of the few non-Pauli channels to be shown non-trivially simulable, and the first using a two-qubit resource state.

Chapter 3

Werner States and Phase Werner States

The work presented in this chapter stems from a discussion of work between myself, Stefano Pirandola, and Kenneth Goodenough, a PhD student at TU Delft. The original concept to introduce a phase into the extremal Werner state was Kenneth's idea, which I then expanded to the full phase Werner and isotropic families, with input from the two parties mentioned. I hope to develop these ideas sufficiently into a paper in the future.

3.1   Structure of this Chapter

The first section of this chapter will begin by introducing the family of Werner states, along with the related isotropic states, and explain their importance among the development and understanding of quantum information. The next section will then introduce the concept of the *phase* Werner state, and its generalisation to the whole family, extending the concept to phase isotropic states. The final section will then prove some properties of these states, and discuss some conjectured (and disproved) results.

3.2   Werner States and their Role in Quantum Theory

Werner states, as they are now called, were first introduced by Reinhard Werner in [110], in order to prove a remarkable result - there exist entangled states for which a local hidden variable model can describe the measurement statistics of any local (projective) measurements upon the state. This means that despite being entangled, these states

cannot violate a Bell inequality. This result was extended in [7] to POVM measurements. The consequence of this is that entanglement is a *necessary* condition for non-locality, but *not* a sufficient one. We shall look at non-locality and Bell inequalities in more detail in chapter 5, but non-locality is often thought of as what makes measurement statistics "truly quantum", and forms the basis of many quantum cryptographic protocols.

Definition 3.2.1 ([110]) d-dimensional Werner states $W_{\eta,d}$ are exactly the set of states such that, for all d-dimensional unitaries U_d,

$$(U_d \otimes U_d) \, W_{\eta,d} \, (U_d \otimes U_d)^\dagger = W_{\eta,d}. \tag{3.1}$$

Definition 3.2.2 The d-dimensional flip operator $\mathbb{F}_d$, is defined as

$$\mathbb{F}_d := \sum_{i,j=0}^{d-1} |ij\rangle \langle ji| \tag{3.2}$$

and has the effect of "swapping" two subsystems, with $\mathbb{F}_d \, |\phi\rangle \otimes |\psi\rangle = |\psi\rangle \otimes |\phi\rangle$.

Definition 3.2.3 ([110]) The d-dimensional Werner states $W_{\eta,d}$ can be parametrised by one parameter $\eta = \text{Tr} \, [\mathbb{F}_d W_{\eta,d}]$, and explicitly expressed as

$$W_{\eta,d} := \frac{(d - \eta)\text{I}_{d^2} + (d\eta - 1)\mathbb{F}_d}{d^3 - d} \tag{3.3}$$

with I_{d^2} the d^2 identity operator. The parameter η satisfies $\eta \in [-1, 1]$.

These states have continued to have a remarkable influence on the development of quantum information theory - often providing a counter example to a seemingly logical claim. One of the most striking examples of this is the following theorem:

Theorem 3.2.4 ([106]) The relative entropy of entanglement, $E_R(\rho)$, is *not* additive in general. In particular, we have

$$\frac{E_R(W_{-1,d} \otimes W_{-1,d})}{2} < E_R(W_{-1,d}), \ \ d \geq 3. \tag{3.4}$$

It is results such as this which has motivated the study of the regularised entanglement measures we have seen in previous chapters. This property of subadditivity motivates the work in the next chapter.

3.2.1 Isotropic States

Sitting alongside their close cousins, isotropic states have also become an integral part of quantum theory. They too exhibit the unusual property of admitting no non-local probability distributions for some levels of entanglement [1]. Moreover, they have been shown to exhibit "super-activation of non-locality" - joint measurements on multiple copies of entangled isotropic states can produce non-local distributions, despite being unable to do so on a single copy.

Like the Werner states, isotropic states also exhibit an invariance under symmetry.

Definition 3.2.5 ([57]) d-dimensional isotropic states $I_{\eta,d}$ are exactly the set of states such that, for all d-dimensional unitaries U_d,

$$(U_d \otimes U_d^*) I_{\eta,d} (U_d \otimes U_d^*)^\dagger = I_{\eta,d}. \tag{3.5}$$

Definition 3.2.6 The d-dimensional maximally entangled operator, $\mathbb{M}_d$ is defined as:

$$\mathbb{M}_d := \sum_{i,j=0}^{d-1} |ii\rangle \langle jj|. \tag{3.6}$$

This is the (unnormalised) density matrix of the d-dimensional maximally entangled state, $|\Phi\rangle_d$. It relates to $\mathbb{F}_d$ in that $\mathbb{F}_d^{T_B} = \mathbb{M}_d$.

Using this operator, we can give the exact expression of the isotropic states.

Definition 3.2.7 ([57]) The d-dimensional isotropic states $I_{\eta,d}$ can be parametrised by one parameter $\eta = \text{Tr} [\mathbb{M}_d I_{\eta,d}]$, and explicitly expressed as

$$I_{\eta,d} := \frac{(d-\eta)\text{I}_{d^2} + (d\eta - 1)\mathbb{M}_d}{d^3 - d} \tag{3.7}$$

and are valid states for $\eta \in [0, d]$.

From here on, we shall drop d, d^2 subscripts from the operators $\text{I}, \mathbb{F}$ and $\mathbb{M}$, as they shall be clear from context.

Lemma 3.2.8 Werner states are related to isotropic states by partial transpose $W_{\eta,d}^{T_B} = I_{\eta,d}$ when $\eta \in [0, 1]$. This is the separable region for both families.

Proof. The proof of this lemma is simple:

$$\begin{aligned}
W_{\eta,d}^{T_B} &= \left(\frac{(d-\eta)\mathrm{I} + (d\eta - 1)\mathbb{F}}{d^3 - d}\right)^{T_B} \\
&= \frac{(d-\eta)\mathrm{I}^{T_B} + (d\eta - 1)\mathbb{F}^{T_B}}{d^3 - d} \\
&= \frac{(d-\eta)\mathrm{I} + (d\eta - 1)\mathbb{M}}{d^3 - d} = I_{\eta,d}.
\end{aligned}$$

$\square$

3.2.2 Basic Properties of Werner and Isotropic states

In this part, we outline some of the basic properties of these families of states, so that we may better understand them. The focus here will mainly be *entanglement*-based properties.

Lemma 3.2.9 Werner states are separable for $\eta \in [0,1]$ and entangled for $\eta \in [-1,0)$. Isotropic states are separable for $\eta \in [0,1]$, and entangled for $\eta \in (1,d]$.

Lemma 3.2.10 The relative entropy of entanglement of $W_{\eta,d}$ is strictly subadditive for the region $\eta \in [-1, -2/d)$ and additive elsewhere, whilst the relative entropy of entanglement is additive for isotropic states.

3.2.2.1 Eigensystems of Werner and Isotropic States

Werner states have the following eigensystem:

- $d(d-1)/2$ eigenvectors of the form $(|ij\rangle + |ji\rangle)/\sqrt{2}$, with eigenvalue $(1+\eta)/(d(d+1))$,

- d eigenvectors of the form $|ii\rangle$, each with eigenvalue $(1+\eta)/(d(d+1))$.

- $d(d-1)/2$ eigenvectors of the form $(|ij\rangle - |ji\rangle)/\sqrt{2}$, $(i > j)$ with eigenvalue $(1-\eta)/(d(d-1))$.

These are easy to verify since $(|ij\rangle + |ji\rangle)/\sqrt{2}, |ii\rangle$ are eigenvectors of the flip operator with value 1, whilst $(|ij\rangle - |ji\rangle)/\sqrt{2}$ are eigenvectors with value -1; as Werner states are linear combinations of $\mathbb{F}$ and I, it is easy to generate the above system.

One noticeable property is that the eigenvectors of $W_{\eta,d}$ are independent of η, meaning the family of Werner states (for a given dimension) are *simultaneously diagonalisable* - a

convenient property for calculating things such as relative entropy, as we shall see in the next chapter.

Isotropic states have a similarly well structured eigensystem. They have:

- 1 eigenvector $\sum_{i=0}^{d-1} |ii\rangle / \sqrt{d}$, with eigenvalue η/d.

- $d(d-1)$ eigenvectors $|ij\rangle$, $i \neq j$, each with eigenvalue $(d-\eta) / (d^3 - d)$.

- $d-1$ eigenvectors $|v_k\rangle = \sqrt{k/(k+1)} |kk\rangle - \sum_{j=0}^{k-1} |jj\rangle / \sqrt{k(k+1)}$, $k \in \{1, \ldots d-1\}$. These also have eigenvalue $(d-\eta) / (d^3 - d)$.

As before, these can be relatively easily verified - the eigenvector $\sum_{i=0}^{d-1} |ii\rangle / \sqrt{d}$ is the sole eigenvector of the rank 1 matrix $\mathbb{M}$, with eigenvalue d - the other eigenvectors provide an orthonormal basis for the nullspace of $\mathbb{M}$. We can then use the same linearity argument as for Werner states.

Given that we have these decompositions, we may easily give the negativity of both states:

$$N(W_{\eta,d}) = \frac{|\eta| + |d - \eta| - d}{2d}, \tag{3.8}$$

$$N(I_{\eta,d}) = \frac{|1 - \eta| + |1 + \eta| - 2}{4}. \tag{3.9}$$

3.2.3 Separability Criterion and Bound Entanglement

The isotropic states were introduced by Michal and Pawel Horodecki in order to provide a sufficiency criterion for distilling states: the *reduction criterion*. By transforming states into the form (3.7), and then providing an explicit protocol for the distillation of isotropic states, this gives a method for distilling quantum states into maximally entangled Bell pairs. The transformation used to obtain these states is known as twirling, which we define in section 3.2.4.
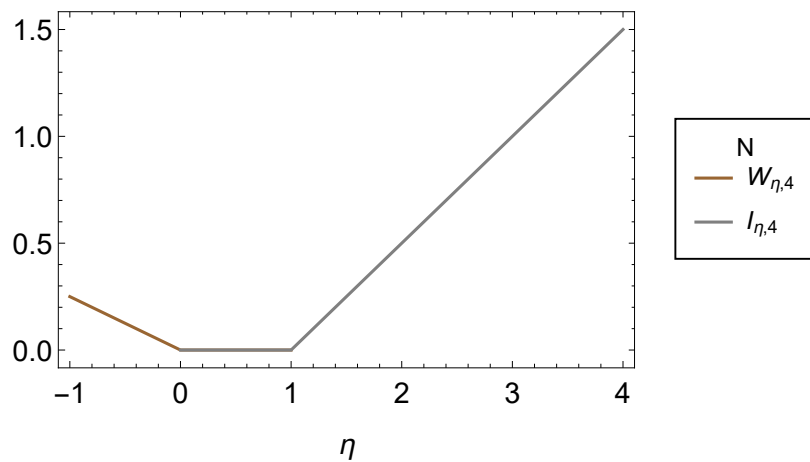
Figure 3.1: The negativity of the states $W_{\eta,4}$ and $I_{\eta,4}$. Both are 0 in the separable range $[0,1]$.
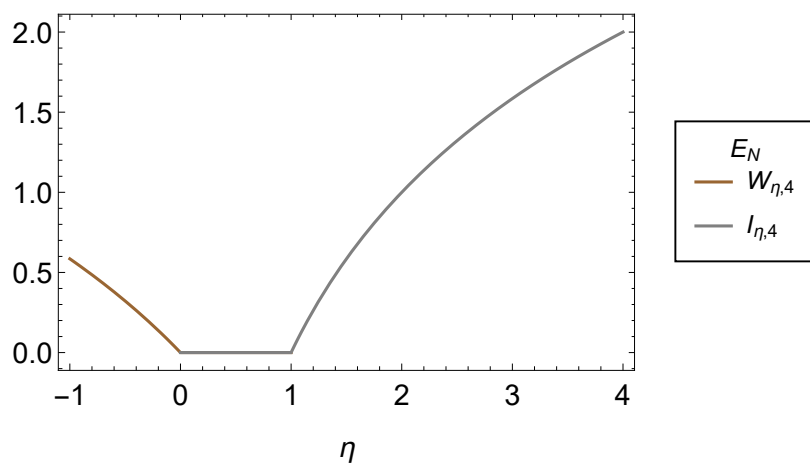
Figure 3.2: We also plot the log negativity of $W_{\eta,4}$, $I_{\eta,4}$, which provides an upper bound to the entanglement of distillation.

Using these states, the Horodeckis were able to prove the following:

Theorem 3.2.11 (Reduction Criterion) For any separable state ρ_{AB} the following must hold:

$$\rho_A \otimes I_B - \rho_{AB} \geq 0, \tag{3.10}$$

$$I_A \otimes \rho_B - \rho_{AB} \geq 0 \tag{3.11}$$

with $\rho_A = \text{Tr}_B[\rho_{AB}]$, $\rho_B = \text{Tr}_A[\rho_{AB}]$. If a state violates either of these conditions, then it is necessarily *distillable*. This means that we can distill n copies of the state into m Bell pairs using LOCC operations only, achieving a conversion rate $m/n > 0$ - the optimal rate is the distillable entanglement seen in chapter 1.

The reduction criterion is an example of a *separability criterion*, the violation of which proves entanglement of the state. Another common separability criterion is the positive partial transpose (PPT) criterion, which states:

Theorem 3.2.12 ([81]) For any separable state ρ_{AB}, the partial transpose $\rho^{T_B} = (\mathbb{I} \otimes T_B)(\rho)$ must be positive semidefinite. Thus if $\rho^{T_B} \not\geq 0$ then ρ is entangled.

For identifying entangled states, the PPT criterion is stronger than the reduction criterion; for example, entangled Werner states $d \geq 3$ all satisfy the reduction criterion, but violate the PPT criterion. Unlike the reduction criterion, violation of the PPT criterion does not guarantee distillability. Instead, the following is true.

Theorem 3.2.13 If an entangled state ρ satisfies the PPT criterion, it is undistillable. We call such states *bound entangled* states.

For $\mathcal{H}_2 \otimes \mathcal{H}_2$ and $\mathcal{H}_2 \otimes \mathcal{H}_3$, the PPT criterion is sufficient for separability; but for higher dimensions explicit examples of bound entangled states have been found [60]. A natural question that arises is - is PPT *necessary* for undistillability? Or do there exist negative partial transpose (NPT) states $\rho^{T_B} \not\geq 0$ which are undistillable also? This has led to the following conjecture:

Conjecture 3.2.14 ([57]) Entangled Werner states $W_{\eta,d}$ with $d \geq 3$ and $\eta \in [(2-d)/(2d-1), 0)$ are NPT bound entangled.

Much like channel capacities, determining the distillability of states is in general difficult, since it requires optimisation over all LOCC protocols, over any number of copies of the state. Thus this conjecture has remained unresolved. Generally the conjecture is believed true though, due to the following result.

Theorem 3.2.15 ([35]) There are Werner states which are n-copy pseudo-undistillable for all finite n.

Definition 3.2.16 ([35]) a state ρ is n-copy pseudo-undistillable if, for all $|\phi\rangle$ with Schmidt rank 2, $\langle\phi| (\rho^{\otimes n})^{T_B} |\phi\rangle \geq 0$.

This is called pseudo-undistillability as it implies there are no possible local projections onto an entangled $\mathcal{H}_2 \otimes \mathcal{H}_2$ state (which are always distillable [59]). If such a state $|\phi\rangle = \sqrt{\lambda_0} |a_0\rangle |b_0\rangle + \sqrt{\lambda_1} |a_1\rangle |b_1\rangle$ with $\langle\phi| (\rho^{\otimes n})^{T_B} |\phi\rangle < 0$ exists, the local projection is $P_A \otimes P_B = (|a_0\rangle \langle a_0| + |a_1\rangle \langle a_1|) \otimes (|b_0\rangle \langle b_0| + |b_1\rangle \langle b_1|)$.

Lemma 3.2.17 Werner states $W_{\eta,d}$, $\eta \in [-1, (2 - d) / (2d - 1))$ are 1-distillable.

Proof. In order to show this, we require there is a state $|\phi_0\rangle$ such that $\langle\phi_0| (W_{\eta,d})^{T_B} |\phi_0\rangle = \langle\phi_0| I_{\eta,d} |\phi_0\rangle < 0$. Choosing $|\phi_0\rangle = |00\rangle + |11\rangle /\sqrt{2}$ gives

$$\langle\phi_0| I_{\eta,d} |\phi_0\rangle = \frac{(d - 2) + (2d - 1)\eta}{d^3 - d} \tag{3.12}$$

which is negative for $\eta < (2 - d) / (2d - 1)$. $\qquad\square$

For the region $[(2 - d) / (2d - 1), 0)$, Werner states are 1-copy pseudo-undistillable, and conjectured to be NPT bound entangled.

The phase Werner states we introduce in section 3.3 have some properties which imply they may be relevant to this problem, and we shall discuss some of the attempts we made to strengthen this connection.

3.2.4   Twirling

When Werner states were first introduced, so too was an operation to generate such states, based on their property that $(U \otimes U) W_{\eta,d}(U \otimes U)^\dagger = W_{\eta,d}$.

Theorem 3.2.18 ([110]) For any state ρ, we have that

$$\int (U \otimes U) \rho(U \otimes U)^\dagger dU = W_{\eta,d} \tag{3.13}$$

with $\eta = \text{Tr}\,[\rho\mathbb{F}]$.

The integral is done via the *Haar* measure of the unitary group - conceptually a volume allowing "densities" of unitaries within the infinite group to be defined. Operationally, this integration may be achieved by the application of random unitaries, or of certain well-constructed finite sets [48]. The twirl operation is an LOCC operation, and one may prove that the states $W_{\eta,d}$, $\eta \in [0,1]$ are separable by twirling the separable states $|u\rangle\langle u| \otimes (\alpha\,|u\rangle\langle u| + (1-\alpha)\,|v\rangle\langle v|)$ to produce them [109]. As the partial transpose of a separable state is itself separable, this also gives the separability of $I_{\eta,d}$, $\eta \in [0,1]$.

Isotropic states satisfy a similar property:

Theorem 3.2.19 ([57]) For any state ρ, we have that

$$\int (U \otimes U^*)\,\rho(U \otimes U^*)^\dagger dU = I_{\eta,d} \tag{3.14}$$

with $\eta = \text{Tr}\,[\rho\mathbb{M}]$.

It is easy to think that there must exist many more similar twirling operations of the form $U \otimes \tilde{U}$, but this turns out not to be the case. Twirls such as $(U \otimes VUV^\dagger)$ are equivalent to Eq. (3.13) followed by a change in basis on the second subsystem. The twirling $U \otimes U^\dagger$ is not valid, since the map $h : U \to \tilde{U}$ needs to be a group homomorphism i.e. $h(U_1)h(U_2) = h(U_1U_2)$; for the conjugate transpose operation this is not the case, since $(U_1)^\dagger(U_2)^\dagger = U_1^\dagger U_2^\dagger = (U_2U_1)^\dagger \neq (U_1U_2)^\dagger$.

It is possible to use subsets of unitaries to define other, interesting twirls. An example of this is the twirl $O \otimes O$ with O representative of an arbitrary real orthogonal unitary (the unitaries exactly satisfying $O = O^*$). This maps to a set of states which are a linear combination of both Werner and isotropic states.

3.2.5 Representations of Werner States

As mentioned before, Werner states are oft-studied and as a consequence are presented (equivalently) in a variety of ways. I outline some of the most common here.

- **Expectation representation**

  The original representation introduced by Werner, expressed by the formula

  $$W_{\eta,d} = \frac{(d - \eta)\mathrm{I} + (d\eta - 1)\mathbb{F}}{d^3 - d}.$$ (3.15)

  With the nice property that $\eta = \mathrm{Tr}\,[\mathbb{F}W_{\eta,d}]$, this is the representation that will be primarily used in the following chapters. The defining parameter η ranges between $[-1, 1]$, with $\eta \in [-1, 0)$ entangled and $\eta \in [0, 1]$ separable. The two other important regions are $\eta \in [-1, -2/d)$ for which the REE is subadditive, and $\eta \in [(2 - d)/(2d - 1), 0)$, the region conjectured to be NPT bound entangled.

- **α-representation**

  Another common representation, the α-representation has explicit construction

  $$W_{\alpha,d} := \frac{\mathrm{I} - \alpha\mathbb{F}}{d^2 - \alpha d},$$ (3.16)

  with $\alpha \in [-1, 1]$. For this representation though, it is when $\alpha \in (1/d, 1]$ that the Werner state is entangled, and for $\alpha \in [-1, 1/d]$ they are separable. The subadditive region is $\alpha \in (3d/(d^2 + 2), 1]$, but the conjectured NPT region has the nice range $\alpha \in (1/d, 1/2]$, and it is in this context that you normally see this representation used.

- **Symmetric representation**

  In order to understand this representation, we need to introduce two operators:

  $$\mathbb{P}_{sym} := \frac{1}{2}\left(\mathrm{I} + \mathbb{F}\right) \qquad\qquad \mathbb{P}_{asym} := \frac{1}{2}\left(\mathrm{I} - \mathbb{F}\right).$$ (3.17)

  Acting on the Hilbert space, $\mathbb{P}_{\text{sym}}$ and $\mathbb{P}_{\text{asym}}$ project onto the symmetric and antisymmetric subspaces respectively. The symmetric subspace is spanned by the states $|ii\rangle, (|ij\rangle + |ji\rangle)/\sqrt{2}$, which are left invariant by $\mathbb{F}$, whilst the antisymmetric subspace is spanned by $(|ij\rangle - |ji\rangle)/\sqrt{2}$, whose sign is flipped by $\mathbb{F}$. We may then write the Werner state as:

  $$W_{p,d} := (1 - p)\frac{2P_{\text{sym}}}{d^2 + d} + p\frac{2P_{\text{asym}}}{d^2 - d}$$ (3.18)

  The fractions serve just to normalise the projectors to trace 1 - and these normalised versions are the extremal Werner states. The parameter p ranges between 0 and

Representation	Sep. Extremal	Ent. Extremal	Sep. Boundary	Subadditive Boundary	Conj. NPT Boundary
Expectation	1	-1	0	$\frac{-2}{d}$	$\frac{2-d}{2d-1}$
α	-1	1	$\frac{1}{d}$	$\frac{3d}{d^2+2}$	$\frac{1}{2}$
Symmetric	0	1	$\frac{1}{2}$	$\frac{1}{2}+\frac{1}{d}$	$\frac{3(d-1)}{4d-2}$
Antisymmetric	$\frac{-1}{d-1}$	1	$\frac{1}{d+1}$	$\frac{3}{d+1}$	$\frac{d-1}{2d-1}$

Table 3.1: A comparison of various representations of Werner states, showing the important parameter values for each.

1, and with entanglement for $p \in (1/2, 1]$ (and thus separable for $p \in [0, 1/2]$). Here the subadditive region is $p \in (1/2 + 1/d, 1]$, and the conjectured NPT region is $p \in (1/2, 3(d-1)/(4d-2)]$. This representation can be useful studying convex combinations of Werner channels, and is related to the expectation representation by the simple relation $p = (1 - \eta)/2$.

- **Antisymmetric representation**

  This slightly unusual representation portrays the Werner state as a mixture of a pseudo-antisymmetric operator, and the identity matrix:

  $$W_{t,d} := t\frac{\mathrm{I} - d\mathbb{F}}{d^2(d-1)} + \frac{\mathrm{I}}{d^2}. \tag{3.19}$$

  For this description entangled states live in the region $t \in (1/(d+1), 1]$, and separable $t \in [-1/(d-1), 1/(d+1)]$. The subadditive region is $t \in (3/(d+1), 1]$, and the NPT conjectured region $t \in (1/(d+1), (d-1)/(2d-1)]$. This representation is relatively rare, and we shall not make use of it here.

A brief summary of this information is provided in table 3.1.

3.3 Introducing a Phase to Werner States

Now that we feel comfortable understanding Werner states, we now look to generalise them. The first step in the process was done by Kenneth Goodenough, a PhD student at Delft. He started with the initial fact that the extremal entangled Werner state is the (normalised) antisymmetric projector,

$$W_{-1,d} = \frac{2P_{\mathrm{asym}}}{d^2 - d} = \frac{2}{d^2 - d}\sum_{i>j}\frac{|ij\rangle - |ji\rangle}{\sqrt{2}}\frac{\langle ij| - \langle ji|}{\sqrt{2}}. \tag{3.20}$$

From this, we can define the extremal phase Werner state

$$W_{-1,d}^{\theta} := \frac{2}{d^2 - d} \sum_{i>j} \frac{|ij\rangle - e^{i\theta}|ji\rangle}{\sqrt{2}} \frac{\langle ij| - e^{-i\theta}\langle ji|}{\sqrt{2}}. \tag{3.21}$$

We can see that the eigenvalues of this state remain unchanged, but the eigenvectors have been changed by definition.

Learning about this, I thought a natural next step would be to generalise this state for η, like the Werner states themselves. Since Werner states are characterised by $\mathbb{F}$, the logical step is to define $\mathbb{F}^{\theta}$.

Definition 3.3.1 The phase flip operator $\mathbb{F}^{\theta}$ is defined:

$$\mathbb{F}^{\theta} := \mathrm{I} - 2P_{\mathrm{asym}}^{\theta} = \mathrm{I} - 2\sum_{i>j} \frac{|ij\rangle - e^{i\theta}|ji\rangle}{\sqrt{2}} \frac{\langle ij| - e^{-i\theta}\langle ji|}{\sqrt{2}} \tag{3.22}$$

motivated by rearranging $P_{\mathrm{asym}} = (\mathrm{I} - \mathbb{F})/2$.

This operator has eigensystem:

$$\mathbb{F}^{\theta}|ii\rangle = |ii\rangle \tag{3.23}$$

$$\mathbb{F}^{\theta}\frac{|ij\rangle + e^{i\theta}|ji\rangle}{\sqrt{2}} = \frac{|ij\rangle + e^{i\theta}|ji\rangle}{\sqrt{2}}, i > j \tag{3.24}$$

$$\mathbb{F}^{\theta}\frac{|ij\rangle - e^{i\theta}|ji\rangle}{\sqrt{2}} = -\frac{|ij\rangle - e^{i\theta}|ji\rangle}{\sqrt{2}}, i > j. \tag{3.25}$$

Moreover, it can be written as

$$\mathbb{F}^{\theta} = \mathbb{F} \odot \begin{pmatrix} 1 & e^{i\theta} & \dots & \dots & e^{i\theta} \\ e^{-i\theta} & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & e^{i\theta} \\ e^{-i\theta} & \dots & \dots & e^{-i\theta} & 1 \end{pmatrix} \tag{3.26}$$

where $\odot$ denotes elementwise multiplication $[A\odot B]_{ij} = A_{ij}B_{ij}$. We now have the operator required to define the family of phase Werner states.

Definition 3.3.2 The phase Werner state $W_{\eta,d}^{\theta}$ is defined as

$$W_{\eta,d}^{\theta} := \frac{(d-\eta)\mathrm{I} + (d\eta - 1)\mathbb{F}^{\theta}}{d^3 - d}. \tag{3.27}$$

The eigenvalues of $W_{\eta,d}$ and $W_{\eta,d}^{\theta}$ coincide, and thus $\eta \in [-1,1]$ define valid phase Werner states.

Lemma 3.3.3 For $d = 2$, phase Werner states are equivalent under local unitaries.

Proof. For $d = 2$, the extremal Werner state is $W_{-1,2} = (|10\rangle - |01\rangle)(\langle 10| - \langle 01|)/2$, just the singlet state. Thus the unitary

$$\mathrm{I} \otimes Z_\theta := \mathrm{I} \otimes \begin{pmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{pmatrix} \text{ gives } (\mathrm{I} \otimes Z_\theta) W_{-1,2} (\mathrm{I} \otimes Z_\theta)^\dagger = W_{-1,2}^\theta. \tag{3.28}$$

Applying Z_θ to $W_{\eta,2}$, we find that $(\mathrm{I} \otimes Z_\theta) W_{\eta,2} (\mathrm{I} \otimes Z_\theta)^\dagger = W_{\eta,2}^\theta$ for all η. $\qquad \square$

Lemma 3.3.4 The states $W_{\eta,d}^\theta$ and $W_{\eta,d}^{-\theta}$ are equivalent under local unitary.

Proof. We begin by defining the d-dimensional unitary $\tilde{U} : |i\rangle \to |(d-1) - i\rangle$ - this is indeed a unitary since it just permutes the computational basis. Applying this to P_{asym}^θ

$$\left(\tilde{U} \otimes \tilde{U}\right) P_{\mathrm{asym}}^\theta \left(\tilde{U} \otimes \tilde{U}\right)^\dagger = \left(\tilde{U} \otimes \tilde{U}\right) \frac{2}{d^2 - d} \sum_{i>j} \frac{|ij\rangle - e^{i\theta} |ji\rangle}{\sqrt{2}} \frac{\langle ij| - e^{-i\theta} \langle ji|}{\sqrt{2}} \left(\tilde{U} \otimes \tilde{U}\right)^\dagger.$$

By defining $i' = (d-1-j)$, $j' = (d-1-i)$, and noting $i > j \Rightarrow i' > j'$ this becomes

$$\left(\tilde{U} \otimes \tilde{U}\right) P_{\mathrm{asym}}^\theta \left(\tilde{U} \otimes \tilde{U}\right)^\dagger = \frac{2}{d^2 - d} \sum_{i'>j'} \frac{|j'i'\rangle - e^{i\theta} |i'j'\rangle}{\sqrt{2}} \frac{\langle j'i'| - e^{-i\theta} \langle i'j'|}{\sqrt{2}}$$

$$= \frac{2}{d^2 - d} \sum_{i'>j'} (-e^{i\theta}) \frac{|i'j'\rangle - e^{-i\theta} |j'i'\rangle}{\sqrt{2}} \frac{\langle i'j'| - e^{i\theta} \langle j'i'|}{\sqrt{2}} (-e^{-i\theta})$$

$$= \frac{2}{d^2 - d} \sum_{i'>j'} \frac{|i'j'\rangle - e^{-i\theta} |j'i'\rangle}{\sqrt{2}} \frac{\langle i'j'| - e^{i\theta} \langle j'i'|}{\sqrt{2}}$$

$$= P_{\mathrm{asym}}^{-\theta}.$$

We can use this result to show:

$$\left(\tilde{U} \otimes \tilde{U}\right) F^\theta \left(\tilde{U} \otimes \tilde{U}\right)^\dagger = \left(\tilde{U} \otimes \tilde{U}\right) \left(\mathrm{I} - 2P_{\mathrm{asym}}^\theta\right) \left(\tilde{U} \otimes \tilde{U}\right)^\dagger = \left(\mathrm{I} - 2P_{\mathrm{asym}}^{-\theta}\right) = \mathbb{F}^{-\theta}$$

and

$$\left(\tilde{U} \otimes \tilde{U}\right) W_{\eta,d}^\theta \left(\tilde{U}^\dagger \otimes \tilde{U}^\dagger\right) = \left(\tilde{U} \otimes \tilde{U}\right) \left(\frac{(d-\eta)\mathrm{I} + (d\eta - 1)\mathbb{F}^\theta}{d^3 - d}\right) \left(\tilde{U} \otimes \tilde{U}\right)^\dagger$$

$$= \frac{(d-\eta)\mathrm{I} + (d\eta - 1)\mathbb{F}^{-\theta}}{d^3 - d}$$

$$= W_{\eta,d}^{-\theta}.$$

$\square$

Given that we have defined these new phase Werner states, the next natural step to ask is whether we can can define *phase isotropic states*? It looks like we are faced with a choice - should we extend our definition of phase Werner states using partial transpose, or alter the state-defining operator $\mathbb{M}$ as we altered $\mathbb{F}$? It turns out that it is of no consequence - both define the *same* family of states.

Definition 3.3.5 The phase isotropic state $I^{\theta}_{\eta,d}$ is defined as

$$I^{\theta}_{\eta,d} := \frac{(d-\eta)\mathrm{I} + (d\eta-1)\mathbb{M}^{\theta}}{d^3 - d},\tag{3.29}$$

where

$$\mathbb{M}^{\theta} = \mathbb{M} \odot \begin{pmatrix} 1 & e^{i\theta} & \dots & \dots & e^{i\theta} \\ e^{-i\theta} & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & e^{i\theta} \\ e^{-i\theta} & \dots & \dots & e^{-i\theta} & 1 \end{pmatrix}.\tag{3.30}$$

$I^{\theta}_{\eta,d} = \left(W^{\theta}_{\eta,d}\right)^{T_B}$ and in general its eigenvalues are dependent on θ.

In order to see that this is indeed the case, let us apply the partial transpose to $W^{\theta}_{\eta,d}$. Since partial transpose respects addition, and the identity is invariant under partial transposition, we need just consider $\left(\mathbb{F}^{\theta}\right)^{T_B}$. Using Eq. (3.26) we see that there are three types of nonzero element in $\mathbb{F}^{\theta}$:

- $|ii\rangle\langle ii|$ - diagonal elements. These are left unaffected by the partial transpose.

- $e^{i\theta}|ij\rangle\langle ji|$, $i > j$ - these are in the upper diagonal due to our ordering: $00, 01\dots$ These are mapped to $e^{i\theta}|ii\rangle\langle jj|$, so remain in the upper diagonal, as $i > j$.

- $e^{-i\theta}|ij\rangle\langle ji|$, $i < j$ - these are in the lower diagonal and are mapped to $e^{-i\theta}|ii\rangle\langle jj|$; since $i < j$ these remain in the lower diagonal.

From this we can conclude that indeed $\mathbb{M}^{\theta} = \left(\mathbb{F}^{\theta}\right)^{T_B}$ is given by Eq. (3.30).

Corollary 3.3.6 Phase isotropic states are equivalent under local unitaries for $d = 2$.

Proof. We may write the following chain of equalities.

$$
\begin{aligned}
I_{\eta,d}^{\theta} = \left(W_{\eta,d}^{\theta} \right)^{T_B} &= \left((\mathrm{I} \otimes Z_{\theta}) W_{\eta,2} (\mathrm{I} \otimes Z_{\theta})^{\dagger} \right)^{T_B} \\
&= (\mathrm{I} \otimes Z_{\theta}^*) W_{\eta,2}^{T_B} (\mathrm{I} \otimes Z_{\theta}^*)^{\dagger} \\
&= (\mathrm{I} \otimes Z_{\theta}^*) I_{\eta,2} (\mathrm{I} \otimes Z_{\theta}^*)^{\dagger}.
\end{aligned}
$$

$\square$

Corollary 3.3.7 The states $I_{\eta,d}^{\theta}$ and $I_{\eta,d}^{-\theta}$ are equivalent under local unitary.

Proof. Taking the same unitary $\tilde{U}$ as defined in lemma 3.3.4, we can see that

$$
\left(\tilde{U} \otimes \tilde{U}^* \right) I_{\eta,d}^{\theta} \left(\tilde{U} \otimes \tilde{U}^* \right)^{\dagger} = I_{\eta,d}^{-\theta}
$$

following the same logic as corollary 3.3.6. $\square$

The eigenvalues of the phase isotropic states are generally *not* independent of θ - in fact, their dependence of θ is far from simple. Kenneth found them to be:

- $(d - \eta) / \left(d^3 - d \right)$ with multiplicity $d(d-1)$, and

- $((d - \eta) + (d\eta - 1)(1 + \lambda_k)) / \left(d^3 - d \right)$, $k \in \{0 \ldots d-1\}$ where

$$
\lambda_k := \frac{\sin \left(\frac{(d-1)\theta - k\pi}{d} \right)}{\sin \left(\frac{\theta + k\pi}{d} \right)}. \tag{3.31}
$$

These eigenvalues restrict the region for which $I_{\eta,d}^{\theta}$ is a valid quantum state[1]. We plot this region for $d = 3, 5$ in figure 3.3. We see it is widest when $\theta = 0$, ranging from 0 to d as we expect, and narrowest when $\theta = \pi$, where the phase isotropic state is valid in the range $\eta \in [(2 - d) / (2d - 1), 2/(d - 1)]$. The bound of $\eta_c := (2 - d) / (2d - 1)$ may look familiar - it is the *conjectured* bound for NPT undistillable entanglement of Werner states! As $I_{\eta_c,d}^{\pi}$ is a valid quantum state, that means that $W_{\eta_c,d}^{\pi}$ is a PPT state, and hence undistillable. This seemed to me a big coincidence and, along with the fact that for $d = 2$ the phase Werner states are locally equivalent, led me to make the following conjecture.

---

[1]Since we require the state to be positive semidefinite.
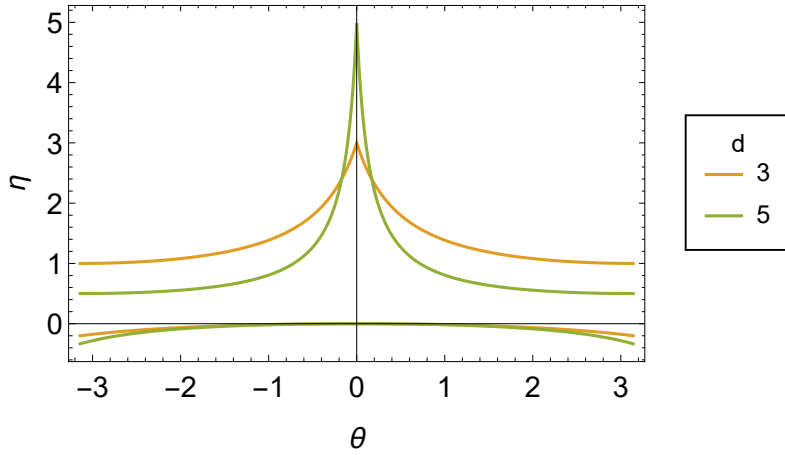
Figure 3.3: Comparison of the allowable range of η for valid $I_{\eta,d}^{\theta}$ states. Note the steep incline around $\theta = 0$ and the dip at $\theta = \pi$.

Conjecture 3.3.8 The entanglement distillation of $W_{\eta,d}^{\theta}$ is equal for all θ.

This conjecture is strictly stronger than conjecture 3.2.14 as not only does it conjecture that for Werner states $W_{\eta,d}$ the region $\eta \in [(2-d)/(2d-1), 0)$ is NPT bound entangled, but also that *any* phase Werner state $W_{\eta,d}^{\theta}$ with a negative partial transpose in the region $\eta \in [(2-d)/(2d-1), 1]$ would *also* be bound entangled. This is possible in two regions: $[(2-d)/(2d-1), 0)$ and $(2/(d-1), 1]$.

Unfortunately, this stronger conjecture is *false*; the key to showing this is to limit our consideration to the two extremal angles, $\theta = 0, \pi$. This is because for these particular values the eigenstates coincide, with different associated eigenvalues.

3.3.1 Entanglement Properties of π-Werner States

The eigensystem of $I_{\pi,d}$ is markedly simpler than that for general θ;

- $d(d-1)$ eigenvectors $|ij\rangle$, $i \neq j$, each with eigenvalue $(d-\eta)/(d^3-d)$.

- 1 eigenvector $\sum_{i=0}^{d-1} |ii\rangle / \sqrt{d}$, with eigenvalue $(2-(d-1)\eta)/(d(d+1))$.

- $d-1$ eigenvectors $|v_k\rangle$, $k \in \{1, \ldots d-1\}$. These have eigenvalue $((d-2)+(2d-1)\eta)/(d^3-d)$.
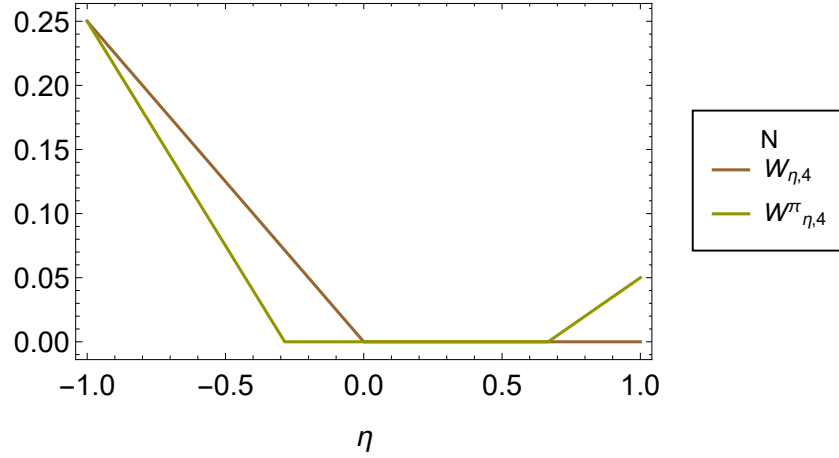
Figure 3.4: The negativity of Werner and π-Werner states; unlike Werner states, there exists NPT π-Werner states with $\eta > 0$. Note the coincidence at $\eta = -1$; this holds for all d.

The states $|v_k\rangle$ are the same defined in section 3.2.2.1. From this, we see that $I_{\pi,d}$ is a valid state for $\eta \in [(2-d)/(2d-1), 2/(d-1)]$. Therefore, the negativity of $W_{\eta,d}^{\pi}$ is non-zero for $\eta \in [-1, (2-d)/(2d-1))$ and $\eta \in (2/(d-1), 1]$, and so for these regions $W_{\eta,d}^{\pi}$ is NPT entangled - as shown in figure 3.4.

We can however, learn much more about these states by exploiting the results of [35]. In this paper, the authors study the states:

$$\rho_{b,c,d} := a \sum_{i=0}^{d-1} |ii\rangle\langle ii| + b \sum_{i,j=0,\,i<j}^{d-1} \frac{|ij\rangle - |ji\rangle}{\sqrt{2}} \frac{\langle ij| - \langle ji|}{\sqrt{2}} + c \sum_{i,j=0,\,i<j}^{d-1} \frac{|ij\rangle + |ji\rangle}{\sqrt{2}} \frac{\langle ij| + \langle ji|}{\sqrt{2}}$$

(3.32)

with $a = 1/d - (b+c)(d-1)/2$, by the normalisation condition.

Whilst generally non-equal to the phase-Werner states, they coincide for

$$\theta = 0 \qquad\qquad b = \frac{1-\eta}{d(d-1)} \qquad\qquad c = \frac{1+\eta}{d(d+1)}, \qquad (3.33)$$

$$\theta = \pi \qquad\qquad b = \frac{1+\eta}{d(d+1)} \qquad\qquad c = \frac{1-\eta}{d(d-1)}. \qquad (3.34)$$

Using this, we can interpret their results for variables b, c to learn more about π-Werner states. This gives us three important results.

Theorem 3.3.9 States $W_{\eta,d}^{\pi}$, $\eta \in [(2-d)/(2d-1), 2/(d-1)]$ are separable.

These states are exactly the π-Werner states with a PPT. As a consequence, this tells us *all* π-isotropic states are separable. This means that for $\theta = 0$, PPT is equivalent to separability for both Werner and isotropic states, and so too for $\theta = \pi$. We can therefore speculate that this equivalence holds for all values of θ.

Theorem 3.3.10 States $W_{\eta,d}^{\pi}$, $\eta \in [-1, (2-d)/(2d-1))$ are 1-distillable.

We show this is the case using the condition for 1-distillability - that there exists a state $|\phi_{\pi}\rangle$ satisfying $\langle \phi_{\pi} | (W_{\eta,d}^{\pi})^{T_B} |\phi_{\pi}\rangle = \langle \phi_{\pi} | I_{\eta,d}^{\pi} |\phi_{\pi}\rangle < 0$. The state $|\phi_{\pi}\rangle = (|00\rangle - |11\rangle)/\sqrt{2}$ is an eigenvector of $I_{\eta,d}^{\pi}$ with eigenvalue $((d-2) + (2d-1)\eta)/(d^3 - d)$. Clearly, this has Schmidt rank 2, and shows the 1-distillability of $W_{\eta,d}^{\pi}$ when $\eta < (2-d)/(2d-1)$.

Theorem 3.3.11 States $W_{\eta,d}^{\pi}$, $\eta \in ((d^2 + 2d - 4)/(d(2d-3)), 1]$ are 1-distillable.

In the same manner as theorem 3.3.10, we choose an explicit $|\psi_{\pi}\rangle$ with Schmidt rank 2 satisfying $\langle \psi_{\pi} | I_{\eta,d}^{\pi} |\psi_{\pi}\rangle < 0$. The state $|\psi_{\pi}\rangle$ is defined[2]

$$|\psi_{\pi}\rangle := \frac{1}{d\sqrt{2}} \left(\left(\sum_{j=0}^{d-1} |j\rangle \right) \otimes \left(\sum_{k=0}^{d-1} |k\rangle \right) + \left(\sum_{j=0}^{d-1} e^{2\pi ij} |j\rangle \right) \otimes \left(\sum_{k=0}^{d-1} e^{-2\pi ik} |k\rangle \right) \right). \quad (3.35)$$

The expectation for this state gives:

$$\langle \psi_{\pi} | I_{\eta,d}^{\pi} |\psi_{\pi}\rangle = \frac{(d^2 + 2d - 4) - (d(2d-3))\eta}{d^2(d^2 - 1)} \quad (3.36)$$

which is negative when $\eta > (d^2 + 2d - 4)/(d(2d-3))$.

For $d > 4$, the region $((d^2 + 2d - 4)/(d(2d-3)), 1]$ is non-trivial. This means we have a set of states $W_{\eta,d}^{\pi}$, $\eta \in ((d^2 + 2d - 4)/(d(2d-3)), 1]$ which are 1-distillable, while their $W_{\eta,d}$ counterparts are separable. This provides a counterexample to conjecture 3.3.8. An interesting point is that

$$\frac{2}{d-1} \leq \frac{d^2 + 2d - 4}{d(2d-3)} \quad (3.37)$$

for all $d > 2$, and thus for $d \geq 4$ the distillability of π-Werner states with $\eta \in (2/(d-1), (d^2 + 2d - 4)/(d(2d-3))]$ is still an open question - though [35] gives they are pseudo 1-copy undistillable.

---

[2]In [35] this state is incorrectly defined with no negative in the exponent of the final sum.

3.3.2 Distillation of Phase Werner States

3.3.2.1 Separability of Phase Werner States

We have seen that all Werner and π-Werner states with a positive partial transpose are separable. This result follows from [35], in which all such states are written as convex combinations of separable states of the form (3.32). Thus, we cannot apply this result for general phase Werner states. Instead, we exploit a *sufficient* condition for separability, given in [51]. The idea behind it is intuitive; if a state is "sufficiently close" to the maximally mixed state, it is unentangled. This is formalised mathematically in the following theorem.

Theorem 3.3.12 ([51]) The state $\rho \in \mathcal{H}_{d_\rho}$ is separable if

$$\text{Tr}[\rho^2] \leq \frac{1}{d_\rho - 1} \tag{3.38}$$

where $\text{Tr}[\rho^2]$ is the purity of ρ, with $\text{Tr}[\rho^2] = 1$ iff ρ is pure.

We can calculate

$$\text{Tr}\left[\left(W_{\eta,d}^\theta\right)^2\right] = \frac{d - 2\eta + d\eta^2}{d^3 - d}. \tag{3.39}$$

For phase Werner states $d_\rho = \dim\left[W_{\eta,d}^\theta\right] = d^2$, and thus we require

$$\frac{d - 2\eta + d\eta^2}{d^3 - d} \leq \frac{1}{d^2 - 1} \Rightarrow \eta \in \left[0, \frac{2}{d}\right]. \tag{3.40}$$

This criterion is not tight for $W_{\eta,d}$ nor $W_{\eta,d}^\pi$, whilst for general phase Werner states we have two regions of uncertainty for the PPT states with $\eta < 0$ and $\eta > 2/d$. We conjecture these are also separable, since this occurs for the extremal angles $0, \pi$.

Although it is interesting for completeness whether these PPT phase Werner states are separable, we know that these states are undistillable by theorem 3.2.13. Of more interest to us is the distillation of NPT phase Werner states. Whilst we cannot use directly the results of [35], we can generalise the result of theorem 3.3.10.

Theorem 3.3.13 States $W_{\eta,d}^\theta$, $\eta \in [-1, (2-d)/(2d-1))$ are distillable.

Proof. To prove this result, we again provide an explicit $|\phi_\theta\rangle$ such that $\langle\phi_\theta| (W_{\eta,d}^\theta)^{T_B} |\phi_\theta\rangle < 0$, implying 1-distillability. Inspired by the result that for $\theta = 0$, $|\phi_0\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$,

and for $\theta = \pi$, $|\phi_\pi\rangle = (|00\rangle - |11\rangle)/\sqrt{2}$, we make the ansantz that $|\phi_\theta\rangle = (|00\rangle + e^{-i\theta}|11\rangle)/\sqrt{2}$. We find that

$$\langle\phi_\theta|\, I_{\eta,d}^\theta\, |\phi_\theta\rangle = \frac{(d-2) + (2d-1)\eta}{d^3 - d} \tag{3.41}$$

implying that phase Werner states with $\eta < (2-d)/(2d-1)$ are 1-distillable. Except for $\theta = \pi$, the state $|\phi_\theta\rangle$ is *not* an eigenstate of $I_{\eta,d}^\theta$. $\qquad\square$

For the NPT region $2/(d-1) < \eta$, we have no ansantz state akin to the one used to prove theorem 3.3.11. Instead we use another distillation technique, relying on *twirling*.

3.3.2.2 Twirling Phase Werner States

Whilst phase Werner states are only equivalent under local unitaries for the special case $d = 2$, we can still create Werner states from them in higher dimensions via twirling. Labelling the Werner twirl (Eq. (3.13)) as T_W, we necessarily have $E_D(T_W(\rho)) \leq E_D(\rho)$, since this twirl is an LOCC operation.

Performing a Werner twirl of $W_{\eta,d}^\theta$ we find:

$$T_W(W_{\eta,d}^\theta) = W_{\zeta_\theta,d}, \ \ \zeta_\theta = \frac{(1+\eta) + (d\eta - 1)\cos\theta}{d+1}. \tag{3.42}$$

We know Werner states $W_{\zeta,d}$ with $\zeta < (2-d)/(2d-1)$ are distillable $(E_D(W_{\zeta,d}) > 0)$. Thus, if

$$\frac{(1+\eta) + (d\eta - 1)\cos\theta}{d+1} < \frac{2-d}{2d-1} \tag{3.43}$$

then $W_{\eta,d}^\theta$ is distillable by first twirling to a Werner state, then the projection onto the $\{|0\rangle, |1\rangle\} \otimes \{|0\rangle, |1\rangle\}$ subspace. Rearranging Eq. (3.43) we require

$$\eta\,(1 + d\cos\theta) < \frac{(2-d)(d+1)}{2d-1} - (1 - \cos\theta). \tag{3.44}$$

For $\cos\theta < -1/d$, this defines a range of *positive* values of η for which $W_{\eta,d}^\theta$ are distillable via this method. Figure 3.5 gives a plot of distillable η for $d = 8$, varying over θ - we see that there indeed exist valid $W_{\eta,d}^\theta$, $\eta > 0$ distillable via this method - more counterexamples to conjecture 3.3.8.

For $\theta = \pi$, the condition (3.44) simplifies to:

$$\eta > \frac{4+d}{2d-1}. \tag{3.45}$$
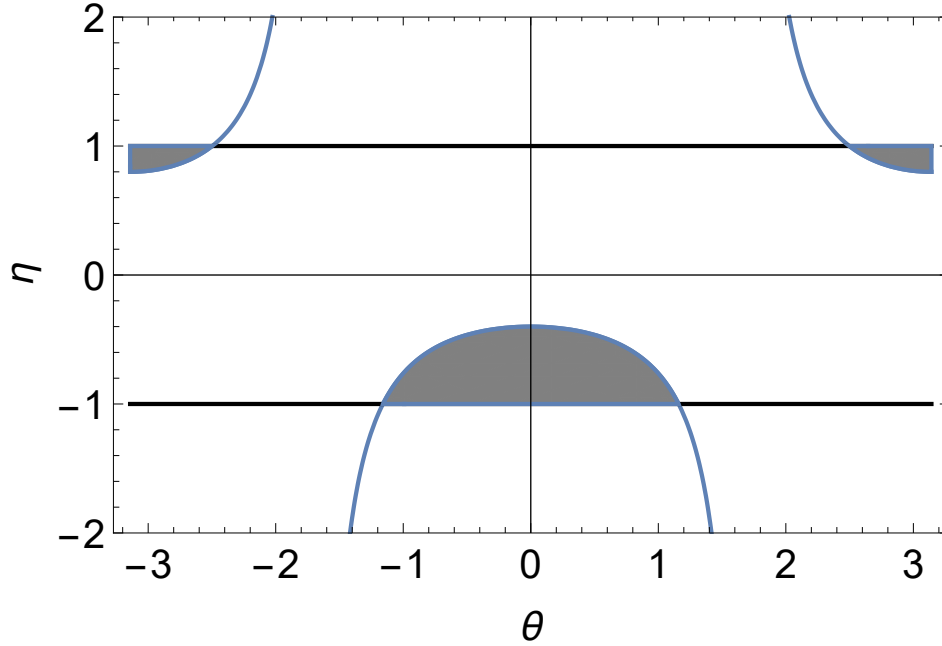
Figure 3.5: The shaded region denotes states $W_{\eta,8}^{\theta}$ which are distillable after the Werner twirling T_W. The blue line is the criticial boundary given by Eq. (3.44).

This gives a non-trivial region $\eta \in ((4+d)/(2d-1), 1]$ of states $W_{\eta,d}^{\pi}$ which are distillable via Werner twirl when $d > 5$. This is illustrated in figure 3.6. We can compare this condition to that of theorem 3.3.11 to find:

$$\frac{d^2 + 2d - 4}{d(2d-3)} < \frac{4+d}{2d-1} \tag{3.46}$$

when $d > 2$. Thus our twirling method is not an optimal distillation method, as there exists 1-distillable π-Werner which are twirled to 1-copy pseudo-undistillable Werner states. It is therefore likely that condition (3.44) only defines a subset of the distillable phase Werner states.

One more note should be made regarding twirling, especially as a lower bound for distillability. We can also consider the isotropic twirl, T_I, whose definition is given in Eq. (3.14). We can calculate $E_D(T_I(\rho))$ exactly, since the twirled state is an isotropic state. For θ-Werner states we obtain

$$T_I(W_{\eta,d}^{\theta}) = I_{\mu,d}, \ \mu = \frac{1+\eta}{d+1}. \tag{3.47}$$

$I_{\mu,d}$ is separable for all η, so we learn nothing about $E_D(W_{\eta,d}^{\theta})$. A more interesting twirl is $T_I(I_{2,2}^{\pi})$, where we obtain the state $I_{0,2}$ - which is completely separable. From corollary
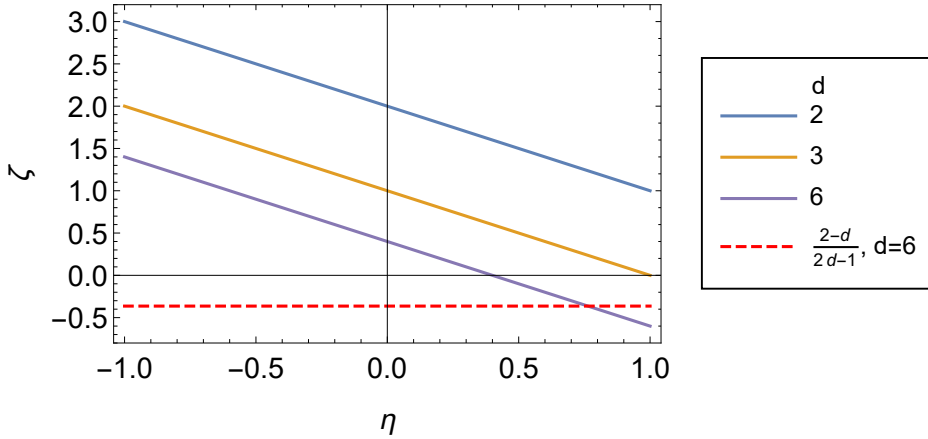
Figure 3.6: The transformation $T_W \left(W^\pi_{\eta,d} \right) = W_{\zeta,d}$, for $d = 2, 3, 6$. The dashed line shows when the resultant states becomes guaranteably distillable.

3.3.6 we know that $I^\pi_{2,2}$ is locally equivalent to the maximally entangled state $I_{2,2} = |\Phi\rangle_2 \langle\Phi|$; by twirling the state we have lost all of its entanglement! This example highlights how such a twirl can completely miss the true entanglement of the state. A potential solution (which would avoid the scenario above, but may not in general) would be to define the "optimal twirl"

$$T_{I,\text{opt}} = \sup_{V \in U(d)} \int (U \otimes U^*) (\mathbb{I} \otimes V) \rho (\mathbb{I} \otimes V)^\dagger (U \otimes U^*)^\dagger dU \tag{3.48}$$

and could potentially lead to better bounds.

3.4 Another Generalisation of Werner States

To end this chapter we introduce another set of states inspired by Werner states. They are inspired by the observation that Werner states can be expressed as a convex combination of all possible Bell pairs:

$$|\Phi^+_{ij}\rangle = \frac{|ii\rangle + |jj\rangle}{\sqrt{2}} \qquad\qquad |\Phi^-_{ij}\rangle = \frac{|ii\rangle - |jj\rangle}{\sqrt{2}}$$

$$|\Psi^+_{ij}\rangle = \frac{|ij\rangle + |ji\rangle}{\sqrt{2}} \qquad\qquad |\Psi^-_{ij}\rangle = \frac{|ij\rangle - |ji\rangle}{\sqrt{2}}$$

taking $i < j$ and $i, j \in \{0, \ldots d-1\}$. Let us consider convex combinations for which the weights are independent of i, j; these are of the form:

$$\rho = \sum_{i,j=0,\ i<j}^{d-1} w_s^+ \, |\Phi_{ij}^+\rangle \langle\Phi_{ij}^+| + w_s^- \, |\Phi_{ij}^-\rangle \langle\Phi_{ij}^-| + w_a^+ \, |\Psi_{ij}^+\rangle \langle\Psi_{ij}^+| + w_a^- \, |\Psi_{ij}^-\rangle \langle\Psi_{ij}^-|. \qquad (3.49)$$

For Werner states, we have that

$$w_s^+ = \frac{1+\eta}{d(d^2-1)} \quad w_s^- = \frac{1+\eta}{d(d^2-1)} \quad w_a^+ = \frac{(d-1)(1+\eta)}{d(d^2-1)} \quad w_a^- = \frac{(d+1)(1-\eta)}{d(d^2-1)}. \qquad (3.50)$$

We may also express the π-Werner states in this manner:

$$w_s^+ = \frac{1+\eta}{d(d^2-1)} \quad w_s^- = \frac{1+\eta}{d(d^2-1)} \quad w_a^+ = \frac{(d+1)(1-\eta)}{d(d^2-1)} \quad w_a^- = \frac{(d-1)(1+\eta)}{d(d^2-1)}. \qquad (3.51)$$

We see that the weights of $|\Psi_{ij}^+\rangle$ and $|\Psi_{ij}^-\rangle$ have been swapped[3]. We name this the z-permutation of the weights. This is because, for $d = 2$, the unitary operation $I \otimes \sigma_z$ swaps $|\Psi^+\rangle$ and $|\Psi^-\rangle$ and $|\Phi^+\rangle$ and $|\Phi^-\rangle$. It is important to note however, that this permutation is *not* a local operation (except for $d = 2$). We define $W_{\eta,d}^z \equiv W_{\eta,d}^\pi$ as the set of states with the weightings in Eq. (3.51).

From this observation, we can see this naturally implies two more permutations to consider: the x-permutation, performing $|\Phi_{ij}^+\rangle \leftrightarrow |\Psi_{ij}^+\rangle$ and $|\Phi_{ij}^-\rangle \leftrightarrow |\Psi_{ij}^-\rangle$; and the y-permutation $|\Phi_{ij}^+\rangle \leftrightarrow |\Psi_{ij}^-\rangle$ and $|\Phi_{ij}^-\rangle \leftrightarrow |\Psi_{ij}^+\rangle$. These create new families of states $W_{\eta,d}^x$ and $W_{\eta,d}^y$ respectively.

Explicitly, the weights for $W_{\eta,d}^x$ are:

$$w_s^+ = \frac{(d-1)(1+\eta)}{d(d^2-1)} \quad w_s^- = \frac{(d+1)(1-\eta)}{d(d^2-1)} \quad w_a^+ = \frac{1+\eta}{d(d^2-1)} \quad w_a^- = \frac{1+\eta}{d(d^2-1)}. \qquad (3.52)$$

The weights for $W_{\eta,d}^y$ are:

$$w_s^+ = \frac{(d+1)(1-\eta)}{d(d^2-1)} \quad w_s^- = \frac{(d-1)(1+\eta)}{d(d^2-1)} \quad w_a^+ = \frac{1+\eta}{d(d^2-1)} \quad w_a^- = \frac{1+\eta}{d(d^2-1)}. \qquad (3.53)$$

The states $W_{\eta,d}^x$ are valid for the range $\eta \in [-1, (d^2-d+1)/(2d-1)]$ - the only one of the three permutations with a range larger[4] than $[-1, 1]$. These states are NPT entangled

---

[3]This matches the definition of the phase Werner states.

[4]For $\eta > 1$, the weight w_s^- is negative, and therefore no longer expressible as a convex combination.

for two regions: $\eta \in [-1, 0)$ and $\eta \in (2/(d-1), (d^2 - d + 1)/(2d - 1)]$. Interestingly, one can calculate

$$\langle \Psi_{ij}^+ | (W_{\eta,d}^x)^{T_B} | \Psi_{ij}^+ \rangle = \frac{\eta}{d(d-1)}, \qquad \Rightarrow \langle \Psi_{ij}^+ | (W_{\eta,d}^x)^{T_B} | \Psi_{ij}^+ \rangle < 0 \text{ when } \eta < 0, \quad (3.54)$$

$$\langle \Psi_{ij}^- | (W_{\eta,d}^x)^{T_B} | \Psi_{ij}^- \rangle = \frac{2 - (d-1)\eta}{d(d^2 - 1)}, \quad \Rightarrow \langle \Psi_{ij}^- | (W_{\eta,d}^x)^{T_B} | \Psi_{ij}^- \rangle < 0 \text{ when } \eta > \frac{2}{d-1}.$$
$$(3.55)$$

The states $W_{\eta,d}^y$ are valid for the range $\eta \in [-1, 1]$, and also NPT entangled for the regions: $\eta \in [-1, 0)$ and $\eta \in (2/(d-1), 1]$. For these states one obtains:

$$\langle \Psi_{ij}^- | (W_{\eta,d}^y)^{T_B} | \Psi_{ij}^- \rangle = \frac{\eta}{d(d-1)}, \qquad \Rightarrow \langle \Psi_{ij}^- | (W_{\eta,d}^y)^{T_B} | \Psi_{ij}^- \rangle < 0 \text{ when } \eta < 0, \quad (3.56)$$

$$\langle \Psi_{ij}^+ | (W_{\eta,d}^y)^{T_B} | \Psi_{ij}^+ \rangle = \frac{2 - (d-1)\eta}{d(d^2 - 1)}, \quad \Rightarrow \langle \Psi_{ij}^+ | (W_{\eta,d}^y)^{T_B} | \Psi_{ij}^+ \rangle < 0 \text{ when } \eta > \frac{2}{d-1}.$$
$$(3.57)$$

This means that NPT entanglement is equivalent to 1-distillability for the states $W_{\eta,d}^x$, $W_{\eta,d}^y$, whilst the same cannot be said for $W_{\eta,d}$, $W_{\eta,d}^z$.

States of form in Eq. (3.49) reduce to those of the form Eq. (3.32) when

$$w_s^+ = \frac{a}{d-1} \qquad w_s^- = \frac{a}{d-1} \qquad w_a^+ = c \qquad w_a^- = b. \qquad (3.58)$$

We speculate that these more general states (and the specific permutations applied to the Werner states) could help in the process to understand NPT entanglement.

3.5 Discussion and Further Directions

The main aim of this chapter is to familiarise the reader with Werner and isotropic states, knowledge of which is required for the results in the next chapter. Of particular use are the *eigensystems* of the two classes of states - the independence of their eigenvectors from the defining parameter η simplifies many of the calculations.

Also introduced was a new class of quantum states, the *phase Werner* states, along with their counterpart, *phase isotropic* states. We conjectured a stronger version of the conjecture made by Michal Horodecki and Pawel Horodecki in [57], which has remained unsolved for over 20 years. Our stronger conjecture can be shown to be false from results in the literature, notably [35] from which we obtained many properties of the specific phase

Werner state when $\theta = \pi$.

For the region $\eta \in [-1, (2-d)/(2d-1))$, in which the Werner states are 1-distillable, we were able to show that general phase-Werner states are also 1-distillable. Phase Werner states with $\eta \in [(2-d)/(2d-1), 0)$ and a negative partial transpose provide new candidates for NPT bound entanglement, and looking at these in more detail would be a valid line of future research - the addition of a phase adding an extra layer of complexity. Another interesting line of investigation would be try to extend the result for π-Werner states regarding 1-distillability in the region $\eta \in ((d^2 + 2d - 4)/(d(2d-3)), 1]$ for general phase Werner states. This region is unlike $[-1, (2-d)/(2d-1))$ in that there exist both NPT and PPT phase Werner states, and so the lower bound for the 1-distillable region will be dependent on θ. We were able to provide an (implicit) bound via the twirling process, although comparison for $\theta = \pi$ implies that this is not tight, as there are π-Werner states which are 1-distillable which do not result in a 1-distillable state after twirling. Something we observed during calculations is that negativity of the state before and after the twirl is preserved - an illustration of how negativity does not fully capture the notion of distillable entanglement.

We also introduced an alternative characterisation of the Werner states in terms of convex combinations of Bell pairs with symmetric weightings, and that these convex combinations generalise the states studied in [35]. Here we only proved some simple properties for specific states of this form, and I believe a rigorous investigation of these states would be worthwhile. In the case $d = 2$, the states simply reduce to Bell diagonal states[5], which are well understood, but for $d > 2$ many interesting results could be found.

The observation that the conjectured boundary for NPT bound entangled states at $(2-d)/(2d-1)$ coincides with the boundary for PPT π-Werner states provides strong motivation for studying these states further. We provide two generalisations of Werner states which we hope may prove illuminating to researchers in the field.

---

[5]Which play an important role in chapter 2.

Chapter 4

Holevo-Werner Channels

The work in this chapter will be based on two papers: "Adaptive estimation and discrimination of Holevo-Werner channels" and "Converse bounds for quantum and private communication over Holevo-Werner channels". The authors of the former are Thomas Cope and Stefano Pirandola, and the latter Thomas Cope, Kenneth Goodenough and Stefano Pirandola.

4.1 Structure of this Chapter

This chapter begins by introducing the Holevo-Werner (HW) channels, through channel-state duality with the Werner states discussed in the previous chapter. We are then introduced to the field of quantum metrology, and prove some results bounding the optimal discrimination of Holevo-Werner channels. Finally, we use the channel simulation results introduced in chapter 2 in order to bound the secret key and quantum capacities of these channels, exploiting the unusual entanglement properties of Werner states in order to show the necessity of considering several entanglement measures.

4.2 Holevo-Werner Channels

We have already been introduced to Choi matrices in chapter 2, but we revisit them here. The Choi matrix of a channel $\mathcal{E} : \mathcal{H}_{d_1} \to \mathcal{H}_{d_2}$ is the state

$$(\mathbb{I}_{d_1} \otimes \mathcal{E}) \left(|\Phi\rangle_{d_1} \langle \Phi| \right) \tag{4.1}$$

with $|\Phi\rangle_{d_1} = \sum_{i=0}^{d_1-1} |ii\rangle / \sqrt{d_1}$. This map is an isomorphism between completely positive trace-preserving maps $\mathcal{E} : \mathcal{H}_{d_1} \to \mathcal{H}_{d_2}$ and density matrices on $\mathcal{H}_{d_1} \otimes \mathcal{H}_{d_2}$.

We can use this map to consider the channel whose Choi matrix is the Werner state $W_{\eta,d}$. The corresponding channel is known as the *Holevo-Werner* channel, defined as:

$$\mathcal{W}_{\eta,d}\left(\rho\right) := \frac{1}{d^2-1}\left((d-\eta)\operatorname{Tr}[\rho]\mathrm{I} + (d\eta-1)\,\rho^T\right) \tag{4.2}$$

with $\eta \in [-1,1]$. When ρ is a density matrix this reduces to

$$\mathcal{W}_{\eta,d}\left(\rho\right) := \frac{1}{d^2-1}\left((d-\eta)\,\mathrm{I} + (d\eta-1)\,\rho^T\right) \tag{4.3}$$

and hitherto we shall omit the $\operatorname{Tr}[\rho]$ term, as we are primarily concerned with sending valid quantum states. It is interesting to note that, whilst transposition is *not* a completely positive map (and thus not a valid quantum channel), it appears as the input-dependent term of this channel. A particular instance of this channel which has been shown much interest is the "extremal" Holevo-Werner channel,

$$\mathcal{W}_{-1,d}\left(\rho\right) = \frac{1}{d-1}\left(\mathrm{I} - \rho^T\right). \tag{4.4}$$

Before we explore Holevo-Werner channels further, it is also worth looking at the channel whose Choi matrix is the isotropic state $I_{\eta,d}$. We find this to be

$$\mathcal{D}_{\eta,d}\left(\rho\right) := \frac{1}{d^2-1}\left((d-\eta)\,\mathrm{I} + (d\eta-1)\,\rho\right), \tag{4.5}$$

with $\eta \in [0,d]$. This has a similar form to Eq. (4.3), but without transposition. This channel is the well known *depolarising channel*, most commonly parametrised in the following way:

$$D_p\left(\rho\right) = p\rho + (1-p)\frac{\mathrm{I}}{d}, \quad p \in \left[-\frac{1}{d^2-1}, 1\right] \tag{4.6}$$

where with some probability[1] p the state is transmitted, else it is instead mapped to the maximally mixed state. The two forms are related by $p = (d\eta-1)/(d^2-1)$.

Lemma 4.2.1 Depolarising channels and Holevo-Werner channels are teleportation covariant.

Proof. We saw in chapter 2 that for a channel $\mathcal{E}$ to be teleportation covariant, there must exists a set of unitaries $\{V_k\}$ satisfying

$$\mathcal{E}\left(U_k\rho U_k^\dagger\right) = V_k\mathcal{E}\left(\rho\right)V_k^\dagger, \quad \forall\rho \tag{4.7}$$

---

[1]Technically pseudo-probability, as p can take negative values.

where $\{U_k\}$ are the generalised Pauli matrices used in d-dimensional teleportation[2]. For depolarising channels we can see that for an arbitrary unitary U:

$$\mathcal{D}_{\eta,d}\left(U\rho U^\dagger\right) = \frac{1}{d^2-1}\left((d-\eta)\,\mathrm{I} + (d\eta-1)\,U\rho U^\dagger\right)$$

$$= \frac{1}{d^2-1}\left((d-\eta)\,U\mathrm{I}U^\dagger + (d\eta-1)\,U\rho U^\dagger\right)$$

$$= U\mathcal{D}_{\eta,d}\left(\rho\right)U^\dagger \tag{4.8}$$

whilst for the Holevo-Werner channels:

$$\mathcal{W}_{\eta,d}\left(U\rho U^\dagger\right) = \frac{1}{d^2-1}\left((d-\eta)\,\mathrm{I} + (d\eta-1)\,(U\rho U^\dagger)^T\right)$$

$$= \frac{1}{d^2-1}\left((d-\eta)\,U^*\mathrm{I}U^T + (d\eta-1)\,U^*\rho^T U^T\right)$$

$$= U^*\mathcal{W}_{\eta,d}\left(\rho\right)(U^*)^\dagger. \tag{4.9}$$

Thus we set $V_k = U_k$ for depolarising channels, and $V_k = U_k^*$ for Holevo-Werner channels. $\square$

4.3 Metrology of Quantum Channels

Metrology, both is the classical and quantum regimes, is a huge area of study [101], and an overview of the field could easily fill a thesis by itself. Most generally, the term "metrology" refers to the study of measurement; this encompasses such concepts as the definition and implementation of SI units, calibration of industrial devices, and the gravitational wave detection experiments, among many others. Quantum metrology combines this field with that of quantum information, looking to perform highly precise measurements by exploiting features of quantum theory such as entanglement.

In this chapter, we shall focus on a specific scenario in which one is given n copies of an unknown quantum state, which is known to be parametrised by a single variable θ. Alternatively, one is allowed to utilise an unknown quantum channel (also parametrised by a single variable θ) n times. This value n is referred to as the the number of *probings* one is allowed to make. We want to know the precision, $\Delta\theta$, with which one can optimally estimate θ, and how this precision scales with n. We are looking at *adaptive* estimation

---

[2]See chapter 1.

of θ, where the choice of probe may change conditionally on the outcome of the previous probe. For states, one can think of a probe as a measurement, while for channels it is sending a known quantum state (or part of an entangled state) through the channel, followed by measurement.

In classical systems the optimal parameter estimation precision necessarily scales as $(\Delta\theta)^2 \approx 1/N$, known as the *shot noise limit*. In quantum systems this scaling is also observed when performing probing without entanglement. By incorporating entanglement into the probing - for example sending part of a maximally entangled state through the channel - then in some scenarios optimal probing may achieve $(\Delta\theta)^2 \approx 1/N^2$ - this is known as the *Heisenberg limit*. It is this improvement which motivates much of the study of quantum metrology.

We shall also in this chapter consider the related problem of *binary discrimination* - in this scenario, one is again given an unknown state or channel belonging to a family of states/channels which is parametrised by a single variable θ. This time however, one is asked to determine between two possible values of θ. This is also known as *quantum hypothesis testing*. We wish to know the minimal probability of an incorrect choice, p_{err}. In chapter 2 we saw the one-shot[3] version of this problem for channels is characterised by the diamond-norm.

This chapter takes advantage of the results in [86], which allow for the simplification of adaptive metrological protocols for teleportation covariant channels - we shall present and explain later in the chapter. First though we introduce some metrological concepts, in order to understand the results presented here.

Consider sampling a (classical) probability distribution X, which is determined by some unknown parameter θ. This gives a conditional probability distribution function $f(x|\theta)$, defining the probability of outcome x given a specific θ. We could also consider this a function of θ, in which we are given the "likelihood"[4] of the underlying value determining X being a certain θ value, given that we observed value x when X was sampled. From

---

[3]In which only a single probe is used.

[4]This function does not need to integrate to 1 over all θ - hence the term likelihood over probability.

this, the *Fisher Information* $I(\theta)$ is defined, as the variance of the natural logarithm of this likelihood function. The Fisher information is a measure of how much we can learn about θ by sampling X, and we have the following important result:

Theorem 4.3.1 The Cramér-Rao bound states that any unbiased estimator $\hat{\theta}$ of a fixed unknown θ is bounded

$$(\Delta\hat{\theta})^2 = \mathrm{Var}\left(\hat{\theta}\right) \geq \frac{1}{I(\theta)} \tag{4.10}$$

where an unbiased estimator is a function $\hat{\theta}(X)$ such that $E\left(\hat{\theta}\right) = \theta$.

In the quantum scenario, we have a similar task, except instead of sampling a probability distribution we are measuring a quantum state, with the outcomes being determined using quantum theory. The Fisher information was generalised for quantum scenarios in [15]. In the quantum setting, we are trying to determine parameter θ using n probings of the channel[5]. We are considering the most general adaptive case - this may involve global quantum operations on output and to-be-inputted states, along with ancilla dimensions and outcome dependent operations - however we are limited to n probings of the channel. Once we have performed n probes subject to the adaptive procedure Λ, we process the resulting state space is into an unbiased estimator $\hat{\theta}$ for θ. As the quantum behaviour of the channel is determined by the unknown parameter θ, for a given n, Λ estimation there is associated a quantum Fisher information[6] $I_\theta^{n,\Lambda}$. In [15] it was shown that, given an adaptive estimation method Λ, the associated variance of the estimator $\hat{\theta}$ satisfies the *quantum Cramér-Rao bound*,

$$\mathrm{Var}\left(\hat{\theta}\right) \geq \frac{1}{I_\theta^{n,\Lambda}} \tag{4.11}$$

and thus the *optimal* estimator $\tilde{\theta}$ (in which the variance is minimised) satisfies

$$\mathrm{Var}\left(\tilde{\theta}\right) = \inf_\Lambda \mathrm{Var}\left(\hat{\theta}\right) \geq \frac{1}{\sup_\Lambda I_\theta^{n,\Lambda}} =: \frac{1}{I_\theta^n}, \tag{4.12}$$

with I_θ^n the quantum Fisher information optimised over all adaptive protocols.

We have discussed in previous chapters the difficulty of optimisation over adaptive protocols in the context of entanglement measures, and the same is true in the metrological

---

[5]We focus on channels in this chapter, though the same argument applies equally to states.

[6]The superscript n refers to the n probings of the channel - this could equally be applied classically by constraining n samplings of X.

context. This is where the results of [86] come in. They prove that, due to channel simulation of teleportation covariant channels, the following holds true.

Theorem 4.3.2 ([86]) For a set of jointly[7] teleportation covariant channels $\{\mathcal{E}_\theta\}$, dependent on a single parameter θ, the optimal quantum Fisher information is given by

$$I_\theta^n = \lim_{\delta\theta \to 0} 8n \frac{1 - F(\chi_{\mathcal{E}_\theta}, \chi_{\mathcal{E}_{\theta+\delta\theta}})}{d\theta^2} \tag{4.13}$$

with $F(\rho, \sigma) = \text{Tr}\left[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right]$ the quantum fidelity.

Corollary 4.3.3 Parameter estimation for θ defining $\{\mathcal{E}_\theta\}$, a set of jointly teleportation covariant channels, cannot beat the shot noise precision.

Proof. The optimal quantum Fisher information for such channels is given by

$$I_\theta^n = \lim_{\delta\theta \to 0} 8n \frac{1 - F(\chi_{\mathcal{E}_\theta}, \chi_{\mathcal{E}_{\theta+\delta\theta}})}{d\theta^2} = n \lim_{\delta\theta \to 0} 8 \frac{1 - F(\chi_{\mathcal{E}_\theta}, \chi_{\mathcal{E}_{\theta+\delta\theta}})}{d\theta^2} := nL(\theta). \tag{4.14}$$

Thus, for the optimal estimator of θ, $\tilde{\theta}$:

$$\text{Var}(\tilde{\theta}) \geq \frac{1}{I_\theta^n} = \frac{1}{nL(\theta)}. \tag{4.15}$$

As $L(\theta)$ is independent of n , at best $\text{Var}(\tilde{\theta})$ can scale with with $1/n$ - this is the shot noise limit. □

4.4 Metrology of Holevo-Werner Channels

The Choi matrices of Holevo-Werner channels are the Werner states. We have seen in lemma 4.2.1 that Holevo-Werner channels of a given dimension are jointly teleportation covariant; we may therefore apply theorem 4.3.2 to these channels. In order to do this, we require the fidelity between two Werner states. In chapter 3 it was shown that the eigenvectors of these states are dependent only on their dimension. We can therefore do all calculations in this basis, where all Werner states are diagonal - a property we refer to as "simultaneously diagonalisable". This means that the fidelity simplifies:

$$F(W_{\eta,d}, W_{\zeta,d}) = \text{Tr}\left[\sqrt{\sqrt{W_{\eta,d}}W_{\zeta,d}\sqrt{W_{\eta,d}}}\right] = \sum_i \sqrt{\sqrt{p_i}q_i\sqrt{p_i}} = \sum_i \sqrt{p_i q_i} \tag{4.16}$$

with $\{p_i\}, \{q_i\}$ the eigenvalues of $W_{\eta,d}, W_{\zeta,d}$ respectively. The eigenvalues of $W_{\eta,d}$ are[8]:

---

[7]This jointly means the covariance unitaries V_k in Eq. (4.7) are θ-independent.
[8]See chapter 3.

- $(d(d+1))/2$ eigenvalues with value $(1+\eta)/(d(d+1))$,

- $(d(d-1))/2$ eigenvalues with value $(1-\eta)/(d(d-1))$.

This means the fidelity between the two Werner states is:

$$F(W_{\eta,d}, W_{\zeta,d}) = \frac{\sqrt{1+\eta}\sqrt{1+\zeta}}{2} + \frac{\sqrt{1-\eta}\sqrt{1-\zeta}}{2}. \tag{4.17}$$

Therefore we have that:

$$F(W_{\eta,d}, W_{\eta+\delta\eta,d}) = \frac{\sqrt{1+\eta}\sqrt{1+(\eta+\delta\eta)}}{2} + \frac{\sqrt{1-\eta}\sqrt{1-(\eta+\delta\eta)}}{2} := F(\delta\eta). \tag{4.18}$$

In order to use this in Eq. (4.13), we shall Taylor expand the above function around $\delta\eta \approx 0$. This is justified, since we take the limit $\delta\eta \to 0$ in the final formula.

First we consider $F(0)$. It is easy to verify that

$$F(0) = \frac{1+\eta}{2} + \frac{1-\eta}{2} = 1. \tag{4.19}$$

We now differentiate $F(\delta\eta)$ (with respect to $\delta\eta$), to obtain:

$$F'(\delta\eta) = \frac{(1+\eta)^{\frac{1}{2}}}{4\left(1+(\eta+\delta\eta)\right)^{\frac{1}{2}}} - \frac{(1-\eta)^{\frac{1}{2}}}{4\left(1-(\eta+\delta\eta)\right)^{\frac{1}{2}}} \tag{4.20}$$

and so

$$F'(0) = \frac{(1+\eta)^{\frac{1}{2}}}{4\left(1+\eta\right)^{\frac{1}{2}}} - \frac{(1-\eta)^{\frac{1}{2}}}{4\left(1-\eta\right)^{\frac{1}{2}}} = 0. \tag{4.21}$$

Differentiating again, we find:

$$F''(\delta\eta) = -\frac{(1+\eta)^{\frac{1}{2}}}{8\left(1+(\eta+\delta\eta)\right)^{\frac{3}{2}}} - \frac{(1-\eta)^{\frac{1}{2}}}{8\left(1-(\eta+\delta\eta)\right)^{\frac{3}{2}}} \tag{4.22}$$

which we then evaluate at $\delta\eta = 0$;

$$
\begin{aligned}
F''(0) &= -\frac{(1+\eta)^{\frac{1}{2}}}{8\left(1+\eta\right)^{\frac{3}{2}}} - \frac{(1-\eta)^{\frac{1}{2}}}{8\left(1-\eta\right)^{\frac{3}{2}}} \\
&= -\frac{1}{8(1+\eta)} - \frac{1}{8(1-\eta)} \\
&= \frac{-1}{4(1-\eta^2)}.
\end{aligned}
\tag{4.23}
$$

Substituting these into the Taylor expansion

$$F(\delta\eta) = 1 + F'(\delta\eta)\delta\eta + F''(\delta\eta)\frac{\delta\eta^2}{2} + O(\delta\eta^3), \tag{4.24}$$
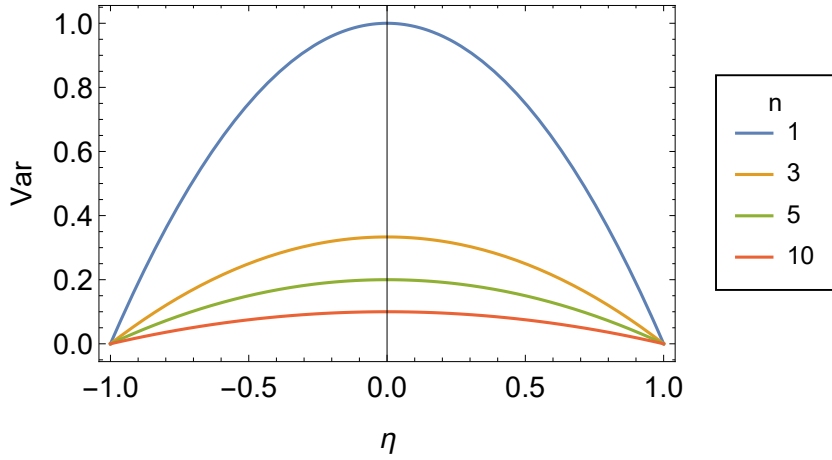
Figure 4.1: This figure shows how the variance bound varies with η, as the number of probes increases. This holds for all dimensions d.

we find

$$F(\delta\eta) = 1 - \frac{\delta\eta^2}{8(1-\eta^2)} + O(\delta\eta^3)\ldots \tag{4.25}$$

and so

$$I_\eta^n \leq \lim_{\delta\eta\to 0} 8n\frac{1 - F(\chi_{\mathcal{E}_\eta}, \chi_{\mathcal{E}_{\eta+\delta\eta}})}{\delta\eta^2} = \frac{n}{1-\eta^2}. \tag{4.26}$$

This gives us our first main result.

Theorem 4.4.1 For a fixed-dimension Holevo-Werner channel with unknown parameter η, the optimal unbiased estimator $\hat{\eta}$ must satisfy:

$$\text{Var}(\hat{\eta}) \geq \frac{1-\eta^2}{n} \tag{4.27}$$

and this limit is achievable in large n [86]. This is the standard shot noise limit.

This result has two interesting consequences: the first is that this bound is dimension-independent; a surprising result which implies η is equally difficult to estimate regardless of the channel's dimension. Another interesting note is that the two channels $\mathcal{W}_{\zeta,d}, \mathcal{W}_{-\zeta,d}, \zeta > 0$ are identically bounded, even though $\mathcal{W}_{\zeta,d}$ is an entanglement-breaking channel whereas $\mathcal{W}_{-\zeta,d}$ is not. It is the trace norm $\|W_{\zeta,d} - W_{0,d}\|_1 = |\zeta|$ which determines optimal estimation.
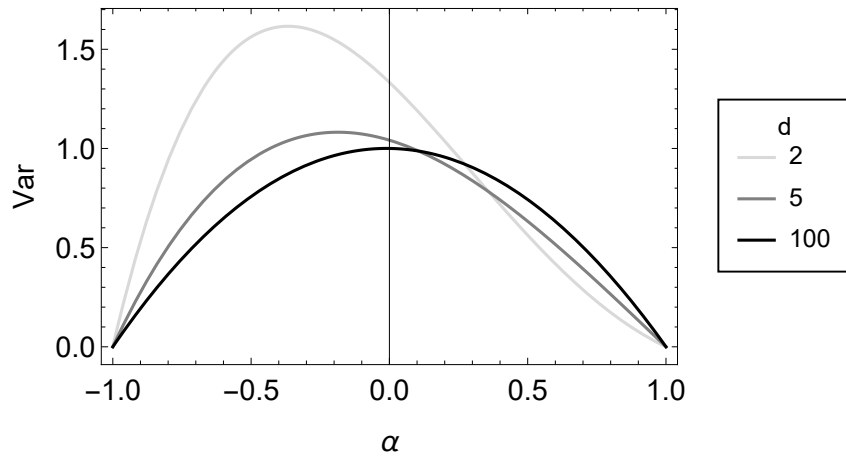
Figure 4.2: This graph for a single probe shows how the bound varies with the channel dimension. Surprisingly, estimation of certain α values improves with higher dimension, but deteriorates for others.

It is worth noting the dimensional independence is particular to the expectation representation of the Holevo-Werner channels. If instead we were to use α-representation, we would find:

$$\mathrm{Var}(\hat{\alpha}) \geq \frac{\left(1 - \alpha^2\right)(d - \alpha)^2}{n(d^2 - 1)}. \tag{4.28}$$

Curiously, we find that this bound is greatest when $\alpha = \left(d - \sqrt{d^2 + 8}\right)/4$ - corresponding to $\eta = \left(4 - d^2 + d\sqrt{d^2 + 8}\right)/\left(3d + \sqrt{d^2 + 8}\right)$. Since this does *not* correspond to the maximal bound for $\mathrm{Var}(\hat{\eta})$ - simply $\eta = 0$ - we find that, depending on the parametrisation used, the ease with which we can estimate a particular channel relative to the other channels in the family varies. Despite this, we cannot exploit this property - for example by estimating α with $\hat{\alpha}$, then converting[9] to $\hat{\eta} = -\left(1 - \hat{\alpha}d\right)/\left(\hat{\alpha} - d\right)$. This is because of the generalised Cramér-Rao bound, which states that an unbiased estimator $T(X)$ of a function $f(\theta)$ must satisfy

$$\mathrm{Var}(T) \geq \frac{|f'(\theta)|^2}{I(\theta)}, \tag{4.29}$$

---

[9]This is a valid unbiased estimator for η.

and thus

$$\text{Var}(\hat{\eta}) \geq \frac{\left(1 - \alpha^2\right)\left(d^2 - 1\right)}{n(d - \alpha)^2} := v(\eta(\alpha)) \tag{4.30}$$

where $\eta(\alpha) = -\left(1 - \alpha d\right) / \left(\alpha - d\right)$. One finds that $v(\eta(\alpha)) = \left(1 - \eta(\alpha)^2\right)/n$, giving us the same bound as Eq. (4.27).

4.4.1 Binary Discrimination of Holevo-Werner Channels

As stated before, we are also interested in the binary discrimination problem for Holevo-Werner channels. We will consider the situation where we have one of two possible Holevo-Werner channels of the same dimension, and we try to discern which using n channel probes. As the two channels are teleportation covariant, we may employ the following theorem.

Theorem 4.4.2 ([86]) For two jointly teleportation covariant channels $\mathcal{E}_0, \mathcal{E}_1$, the optimum error probability in discerning the channels after n probes is bounded by

$$\frac{1 - \sqrt{\min\{1 - F^{2n}, nS\}}}{2} \leq p_{\min} \leq \frac{Q^n}{2} \leq \frac{F^n}{2} \tag{4.31}$$

where $F = F(\chi_{\mathcal{E}_0}, \chi_{\mathcal{E}_1})$ is the fidelity, Q is the *quantum Chernoff bound* (QCB) defined below, and S is a function of the quantum relative entropy given by

$$S := \left(\ln \sqrt{2}\right) \min\{S(\chi_{\mathcal{E}_0} \| \chi_{\mathcal{E}_1}), S(\chi_{\mathcal{E}_1} \| \chi_{\mathcal{E}_0})\}. \tag{4.32}$$

The quantum Chernoff bound is defined as:

$$Q := \inf_{s \in [0,1]} \text{Tr} \left[\chi_{\mathcal{E}_0}^s \chi_{\mathcal{E}_1}^{1-s} \right] \tag{4.33}$$

and has the following operational definition.

Definition 4.4.3 ([4]) Given n copies of a state known to be either ρ_0 or ρ_1, the minimal probability of incorrectly identifying the state is bounded by:

$$p_{\min} \leq \frac{Q^n}{2} = \frac{1}{2} \left(\inf_{s \in [0,1]} \text{Tr} \left[\rho_0^s \rho_1^{1-s} \right] \right)^n. \tag{4.34}$$

We can see that this operational definition extends to teleportation covariant channels.

We have already seen that for Werner states the fidelity is

$$F = \frac{\sqrt{1 + \eta}\sqrt{1 + \zeta} + \sqrt{1 - \eta}\sqrt{1 - \zeta}}{2}. \tag{4.35}$$
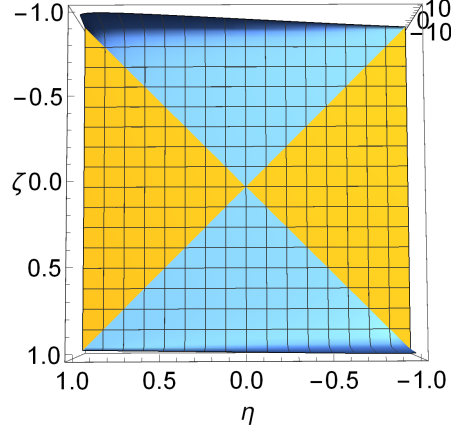
Figure 4.3: A plot of ΔS. The yellow region denotes $\Delta S < 0$, attained when $|\eta| > |\zeta|$.

In order to calculate S we first calculate the relative entropy between two Werner states:

$$S(W_{\eta,d}\|W_{\zeta,d}) = \mathrm{Tr}[W_{\eta,d}\log(W_{\eta,d}) - W_{\eta,d}\log(W_{\zeta,d})] = \sum_i p_i \log \frac{p_i}{q_i} \qquad (4.36)$$

where we have again exploited the shared eigenbasis between the two states. Substituting in the eigenvalues of $W_{\eta,d}, W_{\zeta,d}$ this gives:

$$
\begin{aligned}
S(W_{\eta,d}\|W_{\zeta,d}) &= \frac{d(d+1)}{2}\frac{(1+\eta)}{d(d+1)}\log\left(\frac{1+\eta}{1+\zeta}\right) + \frac{d(d-1)}{2}\frac{(1-\eta)}{d(d-1)}\log\left(\frac{1-\eta}{1-\zeta}\right) \\
&= \frac{1+\eta}{2}\log\left(\frac{1+\eta}{1+\zeta}\right) + \frac{1-\eta}{2}\log\left(\frac{1-\eta}{1-\zeta}\right). \qquad (4.37)
\end{aligned}
$$

In order to determine which relative entropy to use for S, we may look at the function

$$
\begin{aligned}
\Delta S := {}& S(W_{\eta,d}\|W_{\zeta,d}) - S(W_{\zeta,d}\|W_{\eta,d}) \\
={}& \frac{1+\eta}{2}\log\left(\frac{1+\eta}{1+\zeta}\right) + \frac{1-\eta}{2}\log\left(\frac{1-\eta}{1-\zeta}\right) - \frac{1+\zeta}{2}\log\left(\frac{1+\zeta}{1+\eta}\right) - \frac{1-\zeta}{2}\log\left(\frac{1-\zeta}{1-\eta}\right) \\
={}& \left(1 + \frac{\eta+\zeta}{2}\right)\log\left(\frac{1+\eta}{1+\zeta}\right) + \left(1 - \frac{\eta+\zeta}{2}\right)\log\left(\frac{1-\eta}{1-\zeta}\right). \qquad (4.38)
\end{aligned}
$$

Clearly we can see that $\Delta S = 0$ when $|\eta| = |\zeta|$. By plotting ΔS, we can show (numerically) that $\Delta S < 0$ when $|\eta| > |\zeta|$, giving us that:

$$
S = \begin{cases}
\ln\left(\sqrt{2}\right)\frac{1+\eta}{2}\log\left(\frac{1+\eta}{1+\zeta}\right) + \frac{1-\eta}{2}\log\left(\frac{1-\eta}{1-\zeta}\right) & |\eta| \geq |\zeta|, \\[2mm]
\ln\left(\sqrt{2}\right)\frac{1+\zeta}{2}\log\left(\frac{1+\zeta}{1+\eta}\right) + \frac{1-\zeta}{2}\log\left(\frac{1-\zeta}{1-\eta}\right) & |\eta| \leq |\zeta|.
\end{cases} \qquad (4.39)
$$

For the QCB we begin by noting again that our ability to simultaneously diagonalise Werner states of equal dimension allows us to write:

$$Q_s := \text{Tr}\left[W_{\eta,d}^s W_{\eta,d}^{1-s}\right] = \sum_i p_i^s q_i^{1-s}$$
$$= \frac{d(d+1)}{2}\left(\frac{1+\eta}{d(d+1)}\right)^s\left(\frac{1+\zeta}{d(d+1)}\right)^{1-s} + \frac{d(d-1)}{2}\left(\frac{1-\eta}{d(d-1)}\right)^s\left(\frac{1-\zeta}{d(d-1)}\right)^{1-s}$$
$$= \frac{1+\zeta}{2}\left(\frac{1+\eta}{1+\zeta}\right)^s + \frac{1-\zeta}{2}\left(\frac{1-\eta}{1-\zeta}\right)^s. \tag{4.40}$$

It is clear that the infimum $Q = \inf_{s\in[0,1]} Q_s$ will not occur[10] at $s = 0, 1$ as $Q_0 = Q_1 = 1$. Thus we can instead take the infimum as $s \in (0,1)$. First, we must consider some singular cases of Q_s:

- $\eta = \zeta$: Clearly for this scenario $Q_s = 1$ for all values of s. We shall choose to take $s = 1/2$ (the fidelity) here.

- $\zeta = 1$: This simplifies $Q_s = ((1+\eta)/2)^s$. Since $(1+\eta)/2 \in [0,1]$, we achieve $\inf_s Q_s = (1+\eta)/2$ as $s \to 1^-$.

- $\zeta = -1$: This time we obtain the simplification $Q_s = ((1-\eta)/2)^s$. Since $(1-\eta)/2 \in [0,1]$ also, we achieve $\inf_s Q_s = (1-\eta)/2$ as $s \to 1^-$.

- $\eta = 1$: In this case we find $Q_s = ((1+\zeta)/2)^{1-s}$. $(1+\zeta)/2 \in [0,1]$ also, giving $\inf_s Q_s = (1+\zeta)/2$ as $s \to 0^+$.

- $\eta = -1$: For the final singular case, $Q_s = ((1-\zeta)/2)^{1-s}$. Again $(1-\zeta)/2 \in [0,1]$, meaning that $\inf_s Q_s = (1-\zeta)/2$ as $s \to 0^+$.

For other values of η, ζ, our function (4.40) behaves well and we can find the infimum by more conventional means. Let us define the following:

$$k_\pm := \frac{1\pm\zeta}{2}, \ a := \frac{1+\eta}{1+\zeta}, \ m := \frac{1-\eta}{1-\zeta}. \tag{4.41}$$

so that

$$Q_s = k_+ a^s + k_- m^s = k_+ e^{s\ln(a)} + k_- e^{s\ln(m)}. \tag{4.42}$$

Let us now compute the derivative in s

$$\frac{dQ_s}{ds} = k_+\ln(a)e^{s\ln(a)} + k_-\ln(m)e^{s\ln(m)}. \tag{4.43}$$

---

[10]Unless trivially the QCB is 1.

By setting $dQ_s/ds = 0$, we derive

$$0 = k_+ \ln(a)e^{s \ln(a)} + k_- \ln(m)e^{s \ln(m)}$$

$$k_+ \ln(a)e^{s \ln(a)} = -k_- \ln(m)e^{s \ln(m)}$$

$$\frac{e^{s \ln(a)}}{e^{s \ln(m)}} = \frac{-k_- \ln(m)}{k_+ \ln(a)}$$

$$e^{s(\ln(a)-\ln(m))} = \frac{-k_- \ln(m)}{k_+ \ln(a)}$$

$$s\left(\ln(a) - \ln(m)\right) = \ln\left(\frac{-k_- \ln(m)}{k_+ \ln(a)}\right)$$

$$s = \frac{\ln\left(\frac{-k_- \ln(m)}{k_+ \ln(a)}\right)}{\ln\left(\frac{a}{m}\right)}. \tag{4.44}$$

Substituting back in our definitions of $k_\pm$, a and m, we obtain

$$s = \frac{\ln\left(\frac{\zeta-1}{\zeta+1}\frac{\ln\left(\frac{1-\eta}{1-\zeta}\right)}{\ln\left(\frac{1+\eta}{1+\zeta}\right)}\right)}{\ln\frac{(1+\eta)(1-\zeta)}{(1+\zeta)(1-\eta)}} =: s_{\eta,\zeta}. \tag{4.45}$$

To show this stationary point in a suitable candidate for the infimum, we must first show $s_{\eta,\zeta} \in [0,1]$, $\forall\ \eta, \zeta \in (-1,1)$, $\eta \neq \zeta$. We first show that $s_{\eta,\zeta}$ is positive - we do this by considering both the denominator and numerator separately, and in two different cases:

Denominator, Case 1: $-1 < \zeta < \eta < 1$.

In this scenario, both fractions $(1+\eta)/(1+\zeta)$ and $(1-\zeta)/(1-\eta)$ must necessarily be greater than 1; thus the overall denominator is the logarithm of a value greater than 1, and therefore positive.

Denominator, Case 2: $-1 < \eta < \zeta < 1$.

Conversely, in this case both fractions $(1+\eta)/(1+\zeta)$ and $(1-\zeta)/(1-\eta)$ are less than 1, but positive, and so too is their product; forcing the overall denominator to be negative when the logarithm is taken.

Numerator, Case 1: $-1 < \zeta < \eta < 1$.

Since the denominator is positive, we thus require the numerator to be non-negative. Equivalently, we require the following inequality holds:

$$\frac{\zeta-1}{\zeta+1}\frac{\ln\left(\frac{1-\eta}{1-\zeta}\right)}{\ln\left(\frac{1+\eta}{1+\zeta}\right)} \geq 1. \tag{4.46}$$

Since $\zeta + 1 > 0$, and $(1 + \eta)/(1 + \zeta) > 1$, we have the denominator of Eq. (4.46) is positive, and thus we may rearrange the inequality to:

$$(\zeta - 1)\ln\left(\frac{1-\eta}{1-\zeta}\right) - (\zeta + 1)\ln\left(\frac{1+\eta}{1+\zeta}\right) \geq 0. \tag{4.47}$$

Numerator, Case 2: $-1 < \eta < \zeta < 1$.

Due to the negative denominator, we require a non-positive numerator - this requires the inequality

$$\frac{\zeta - 1}{\zeta + 1} \frac{\ln\left(\frac{1-\eta}{1-\zeta}\right)}{\ln\left(\frac{1+\eta}{1+\zeta}\right)} \leq 1. \tag{4.48}$$

Again $\zeta + 1 > 0$, but in this case $(1 + \eta)/(1 + \zeta) < 1$, and thus when multiplying out by $\ln((1 + \eta)/(1 + \zeta))$ we must flip the sign, to obtain the equation:

$$(\zeta - 1)\ln\left(\frac{1-\eta}{1-\zeta}\right) - (\zeta + 1)\ln\left(\frac{1+\eta}{1+\zeta}\right) \geq 0. \tag{4.49}$$

Comparing Eq. (4.47) and Eq. (4.49), we see the condition for $s_{\eta,\zeta}$ to be positive is the same for both cases. Moreover, we can see this inequality holds iff

$$\left(\frac{1-\zeta}{2}\right)\ln\left(\frac{\frac{1-\zeta}{2}}{\frac{1-\eta}{2}}\right) + \left(\frac{\zeta+1}{2}\right)\ln\left(\frac{\frac{1+\zeta}{2}}{\frac{1+\eta}{2}}\right) \geq 0 \tag{4.50}$$

holds. Substituting $p_\eta = (1 - \eta)/2$, $p_\zeta = (1 - \zeta)/2$, we can rewrite the left hand side as

$$p_\zeta \ln\left(\frac{p_\zeta}{p_\eta}\right) + (1 - p_\zeta)\ln\left(\frac{1 - p_\zeta}{1 - p_\eta}\right) \tag{4.51}$$

with $p_\eta, p_\zeta \in (0, 1)$. This formula is the *classical relative entropy*, also known as the Kullback-Leibler (KL) divergence, of two biased coin flips (in a different logarithmic basis). Therefore we may use Gibbs' inequality, which states that this quantity is always non-negative (regardless of logarithmic basis). This proves the first statement, that $s_{\eta,\zeta} \geq 0$.

In order to show $s_{\eta,\zeta} \leq 1$, consider the sum of two non-negative values $s_{\eta,\zeta} + s_{\zeta,\eta}$. Using formula (4.45) we obtain:

$$s_{\eta,\zeta} + s_{\zeta,\eta} = \frac{\ln\left(\frac{\zeta-1}{\zeta+1}\frac{\ln\left(\frac{1-\eta}{1-\zeta}\right)}{\ln\left(\frac{1+\eta}{1+\zeta}\right)}\right)}{\ln\frac{(1+\eta)(1-\zeta)}{(1+\zeta)(1-\eta)}} + \frac{\ln\left(\frac{\eta+1}{\eta-1}\frac{\ln\left(\frac{1+\zeta}{1+\eta}\right)}{\ln\left(\frac{1-\zeta}{1-\eta}\right)}\right)}{\ln\frac{(1+\eta)(1-\zeta)}{(1+\zeta)(1-\eta)}} \tag{4.52}$$

where we have used $-\log(x) = \log(x^{-1})$ on both numerator and denominator of $s_{\zeta,\eta}$. We
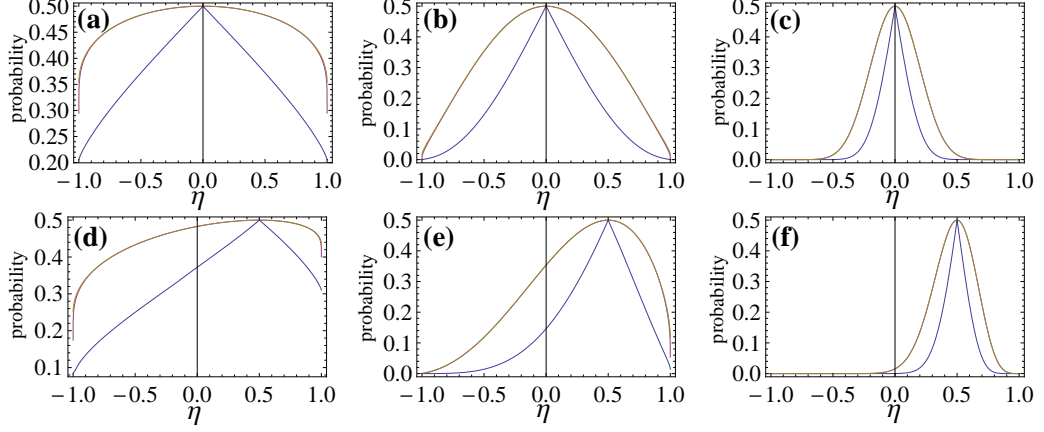
Figure 4.4: We plot the fidelity-based lower bound and the QCB (upper bound) to the optimal error probability $p_{\min}$ in Eq. (4.31) for two HW channels $\mathcal{W}_{\eta,d}$ and $\mathcal{W}_{\zeta,d}$ in an arbitrary finite dimension $d \geq 2$. In panels (a)-(c), we set $\zeta = 0$ and we plot the bounds as a function of η, considering (a) $n = 1$, (b) $n = 10$, and (c) $n = 100$. In panels (d)-(f), we repeat the study with the same parameters as before but setting $\zeta = 1/2$.

see they share a denominator, so we shall focus on the sum of the numerators:

$$
\ln\left(\frac{\zeta-1}{\zeta+1}\frac{\ln\left(\frac{1-\eta}{1-\zeta}\right)}{\ln\left(\frac{1+\eta}{1+\zeta}\right)}\right) + \ln\left(\frac{\eta+1}{\eta-1}\frac{\ln\left(\frac{1+\zeta}{1+\eta}\right)}{\ln\left(\frac{1-\zeta}{1-\eta}\right)}\right)
$$
$$
= \ln\frac{(1+\eta)\,(1-\zeta)}{(1+\zeta)\,(1-\eta)}, \tag{4.53}
$$

where we have used the following basic properties: $\log(x) + \log(y) = \log(xy)$, $-\log(x) = \log(x^{-1})$, and $-1^2 = 1$. We can see that the numerator and denominator coincide, and so $s_{\eta,\zeta} + s_{\zeta,\eta} = 1$ - since both these values are non-negative, we can conclude $0 \leq s_{\eta,\zeta} \leq 1$, as required. Remembering that $\frac{dQ_s}{ds}|_{s=s_{\eta,\zeta}} = 0$, it remains to show this value is a minima. Looking at the second order differential, we find

$$
\frac{d^2 Q_s}{ds^2} = k_+[\ln(a)]^2 e^{s\ln(a)} + k_-[\ln(m)]^2 e^{s\ln(m)}. \tag{4.54}
$$

For all $\zeta \in (-1, 1)$, both k_+, k_- are strictly positive, and when $\eta \neq \zeta$ we also have $\ln(a), \ln(m) \neq 0$ - thus their squares are positive too. Finally, e to the power of any real value is strictly positive, and we so we may conclude $\frac{d^2 Q_s}{ds^2} > 0$ for all values of s, including $s_{\eta,\zeta}$. Thus our value $s_{\eta,\zeta}$ is a true minima. Moreover, the limits $s \to 0^+, s \to 1^-$ for the non-singular η, ζ are 1, and therefore we may be confident our minimum achieves the infimum. Incorporating this result alongside our singular cases, we have provided the analytical QCB for Holevo-Werner channels.

4.4.2 Quantum Chernoff Bound for Isotropic States

In addition to an analytic result for Holevo-Werner channels, we also present a closed form for the QCB of two isotropic states - this could then be used to bound the binary discrimination of two depolarising channels, as they are also jointly teleportation covariant (from lemma 4.2.1). Like Werner states, isotropic states of the same dimension are also simultaneously diagonalisable, and thus we have for two isotropic states $I_{\alpha,d}, I_{\beta,d}$ that

$$Q = \inf_{s \in (0,1)} Q_s, \ Q_s := \text{Tr}[I_{\alpha,d}^s I_{\beta,d}^{1-s}] \tag{4.55}$$

reduces to

$$
\begin{aligned}
Q_s &= \sum_i r_i^s t_i^{1-s} \\
&= \left(\frac{\alpha}{d}\right)^s \left(\frac{\beta}{d}\right)^{1-s} + (d^2 - 1)\left(\frac{d-\alpha}{d(d^2-1)}\right)^s \left(\frac{d-\beta}{d(d^2-1)}\right)^{1-s} \\
&= \frac{\beta}{d}\left(\frac{\alpha}{\beta}\right)^s + \frac{d-\beta}{d}\left(\frac{d-\alpha}{d-\beta}\right)^s
\end{aligned}
\tag{4.56}
$$

where we have taken $\{r_i\}$, $\{t_i\}$ as the eigenvalues of $I_{\alpha,d}, I_{\beta,d}$ respectively (which we saw in chapter 3). Once again, we must first look at the singular scenarios:

- $\alpha = \beta$: As with Werner states, we have $Q_s = 1$, for all s. We again choose $s = 1/2$.

- $\beta = d$: This simplifies $Q_s = (\alpha/d)^s$. Here we achieve infimum α/d at $s \to 1^-$, since $\alpha/d \in [0,1]$.

- $\beta = 0$: As $Q_s = ((d-\alpha)/d)^s$ in this scenario, we obtain infimum $(d-\alpha)/d$ at $s \to 1^-$ using that $(d-\alpha)/d \in [0,1]$ also.

- $\alpha = d$: This simplifies $Q_s = (\beta/d)^{1-s}$, and thus the infimum is β/d at $s \to 0^+$.

- $\alpha = 0$: As this gives $Q_s = ((d-\beta)/d)^{1-s}$, the infimum is $(d-\beta)/d$ at $s \to 0^+$.

  For all other values, we may differentiate Q_s in the same manner as Eq. (4.44) to obtain

  $$\frac{dQ_s}{ds} = l_+ \ln(a_\Omega)e^{s\ln(a_\Omega)} + l_- \ln(m_\Omega)e^{s\ln(m_\Omega)}, \tag{4.57}$$

  where we have set

  $$l_+ = \frac{\beta}{d}, \ l_- = \frac{d-\beta}{d}, \ a_\Omega = \frac{\alpha}{\beta}, \ m_\Omega = \frac{d-\alpha}{d-\beta}. \tag{4.58}$$

114

Setting this to 0, rearranging, and substituting back in our definitions, we obtain the critical point:

$$s = \frac{\ln\left(\frac{\beta-d}{\beta}\frac{\ln\left(\frac{d-\alpha}{d-\beta}\right)}{\ln\left(\frac{\alpha}{\beta}\right)}\right)}{\ln\left(\frac{\alpha(d-\beta)}{\beta(d-\alpha)}\right)} =: s^\Omega_{\alpha,\beta}. \tag{4.59}$$

Although this point is dimension-dependent, we may transform α, β to $\eta = (2\alpha - d)/d \in [-1, 1]$, $\zeta = (2\beta - d)/d \in [-1, 1]$ to obtain the dimension-independent

$$s^\Omega_{\eta,\zeta} = \frac{\ln\left(\frac{\zeta-1}{\zeta+1}\frac{\ln\left(\frac{1-\eta}{1-\zeta}\right)}{\ln\left(\frac{1+\eta}{1+\zeta}\right)}\right)}{\ln\frac{(1+\eta)(1-\zeta)}{(1+\zeta)(1-\eta)}} \tag{4.60}$$

and we see that actually $s^\Omega_{\eta,\zeta} = s_{\eta,\zeta}$. We have already shown $s_{\eta,\zeta}$ to be a valid minimum in the open interval $(-1, 1)$, and therefore we may conclude $s^\Omega_{\alpha,\beta}$ achieves the isotropic QCB, for the non-singular cases.

4.5 Capacities of Holevo-Werner Channels

In this section, we apply results from [82] and [85], which allow for the bounding of teleportation covariant channel capacities using entanglement measures. We have already seen the application of some of these results in chapter 2, in the case of Pauli-damping channels; however these bounds take on a specific character for Holevo-Werner channels in that they may be *subadditive* - thus motivating the use of *regularised* entanglement measures where possible.

To begin with, we recap some results that prove to be useful.

Definition 4.5.1 The relative entropy of entanglement (REE) of a state ρ is given by

$$E_R(\rho) := \min_{\sigma\in\text{sep}} S(\rho\|\sigma) \tag{4.61}$$

and is a valid entanglement monotone. This can be generalised to the n-copy version

$$E^n_R(\rho) := \frac{1}{n}\min_{\sigma\in\text{sep}} S(\rho^{\otimes n}\|\sigma) \tag{4.62}$$

and the *regularised* version,

$$E^\infty_R(\rho) := \lim_{n\to\infty}\frac{1}{n}\min_{\sigma\in\text{sep}} S(\rho^{\otimes n}\|\sigma). \tag{4.63}$$

This measure is (weakly) subadditive, namely $E^\infty_R(\rho) \leq \dots E^{(n+1)}_R(\rho) \leq E^n_R(\rho) \leq E_R(\rho)$.

Theorem 4.5.2 ([85]) For any teleportation covariant channel $\mathcal{E}$, we can bound the secret key capacity[11] of the channel $K(\mathcal{E})$ by

$$K(\mathcal{E}) \leq E_R^\infty(\chi_\mathcal{E}), \tag{4.64}$$

with $\chi_\mathcal{E}$ the Choi matrix of the channel.

We saw the proof of this in chapter 2 - in which it was shown that a channel $\mathcal{E}$ which can be LOCC simulated over resource state σ satisfies the bound $K(\mathcal{E}) \leq E_R^\infty(\sigma)$. We also saw that for teleportation covariant channels we can take the resource state to be the Choi matrix of the channel, $\chi_\mathcal{E}$. This gives the above result.

Corollary 4.5.3 For any $n \in \mathbb{N}$, $E_R^n(\chi_\mathcal{E})$ is a valid upper bound on $K(\mathcal{E})$ for a $\mathcal{E}$ teleportation covariant.

Proof. This simply follows from theorem 4.5.2, and the subadditivity of the REE. $\qquad\square$

For many teleportation covariant channels, it occurs that $E_R^\infty(\chi_\mathcal{E}) = E_R(\chi_\mathcal{E})$, and thus the so-called *single-letter* bound $E_R(\chi_\mathcal{E})$ is sufficient to provide the tightest bound (via this method) - this occurs for Pauli channels, including the depolarising channel[12] $D_{\eta,d}$. However, the extremal Werner state $W_{-1,d}$, $d \geq 3$ was the counterexample used to disprove additivity of the REE, and thus we see that for the extremal Holevo-Werner channel $\mathcal{W}_{-1,d}$ when $d \geq 3$ the n-copy and the regularised REE will provide tighter secret key bounds for this channel than the single-letter version. Unfortunately $E_R^\infty(W_{-1,d})$ is not known; but we may exploit some other results in order to gain a more accurate picture.

For $W_{\eta,d}$ the relative entropy of entanglement *with respect to partial positive transpose states* (RPPT) is known, and so too is the *regularised* RPPT. We were introduced to this measure in chapter 1, which is defined:

$$E_P(\rho) := \min_{\sigma \text{ s.t. } \sigma^{T_B} \geq 0} S(\rho \| \sigma). \tag{4.65}$$

---

[11]Also the two-way entanglement distillation and quantum capacities, since $D_2 = Q_2 \leq K$.

[12]whose Choi matrix is the isotropic state.

Theorem 4.5.4 ([5]) The regularised RPPT for $W_{\eta,d}$, is given by

$$E_P^\infty (W_{\eta,d}) = \begin{cases} 0 = E_P (W_{\eta,d}) & \text{if } \eta \geq 0, \\ \frac{1+\eta}{2} \log (1 + \eta) + \frac{1-\eta}{2} \log (1 - \eta) = E_P (W_{\eta,d}) & \text{if } -\frac{2}{d} \leq \eta \leq 0, \\ \log \left(\frac{d+2}{d}\right) + \frac{1+\eta}{2} \log \left(\frac{d-2}{d+2}\right) \neq E_P (W_{\eta,d}) & \text{if } \eta \leq -\frac{2}{d}. \end{cases} \quad (4.66)$$

You can see that for $\eta \in [-2/d, 1]$, the RPPT is additive. Moreover, since $E_R(W_{\eta,d}) = E_P(W_{\eta,d})$ - that is, the one-copy values coincide - we can conclude that $E_R(W_{\eta,d})$ must also be additive in this region, using the following chain of inequalities.

$$E_R(W_{\eta,d}) = E_P(W_{\eta,d}) = E_P^\infty(W_{\eta,d}) \leq E_R^\infty(W_{\eta,d}) \leq E_R(W_{\eta,d}) \quad (4.67)$$

remembering that for all separable states σ, $\sigma^{\mathrm{T_B}} \geq 0$, and therefore minimising over all PPT states will always give a lower/equal value than over separable states.

Whilst we are unable to provide the tightest bound for the region $\eta < -2/d$, we are able to improve on the single letter bound - to do this, we exploit a result in [106], in which it was proved that the state σ minimising $E_R^n(W_{\eta,d})$ necessarily satisfies:

$$\left(U_d^1 \otimes U_d^1 \otimes \ldots U_d^n \otimes U_d^n\right) \sigma(U_d^1 \otimes U_d^1 \otimes \ldots U_d^n \otimes U_d^n)^\dagger, \quad (4.68)$$

for all unitaries on $\mathcal{H}_d$, where each $U_d^i \otimes U_d^i$ acts on the $d \times d$ Hilbert space occupied by the ith copy of $W_{\eta,d}$. States which are invariant under this action may be expressed as the convex combination [5]:

$$\sigma_{\mathbf{x}}^n = x_0 W_{-1,d}^{\otimes n} + \frac{x_1}{n} \left(W_{-1,d}^{\otimes(n-1)} \otimes W_{1,d} + \ldots W_{1,d} \otimes W_{-1,d}^{\otimes(n-1)}\right)$$
$$+ \ldots \frac{x_k}{\binom{n}{k}} \left(W_{-1,d}^{\otimes(n-k)} \otimes W_{1,d}^{\otimes k} \ldots W_{1,d}^{\otimes k} \otimes W_{-1,d}^{\otimes(n-k)}\right) + \ldots + x_n W_{1,d}^{\otimes n}, \quad (4.69)$$

where $\mathbf{x} = (x_0, x_1, \ldots, x_n)^T$ satisfies $x_i \geq 0$ and $\sum_{i=0}^n x_i = 1$. We also have an explicit condition on $\mathbf{x}$ to ensure that $\sigma_{\mathbf{x}}^n$ is PPT. This is [5]

$$\begin{pmatrix} -1 & 1 \\ 1 & \frac{d-1}{d+1} \end{pmatrix}^{\otimes n} \mathbf{x}' \geq 0, \quad (4.70)$$

where

$$\mathbf{x}' = \left(x_0, \frac{x_1}{n}^{\times n} \ldots \frac{x_k}{\binom{n}{k}}^{\times \binom{n}{k}} \ldots x_n\right)^T. \quad (4.71)$$

It was via this result that theorem 4.5.4 was obtained, by taking the limit as $n \to \infty$. In the paper [106], it was proved that for $n = 2$ the set of states of the form (4.69) which are

separable and those which are PPT exactly coincide. Thus to minimise $E_R^2(W_{\eta,d})$, we just need to consider the states

$$\sigma_{\mathbf{x}}^2 = x_0 W_{-1,d}^{\otimes 2} + \frac{x_1}{2}\left(W_{-1,d} \otimes W_{1,d} + W_{1,d} \otimes W_{-1,d}\right) + (1 - x_1 - x_2) W_{1,d}^{\otimes 2} \tag{4.72}$$

which satisfy

$$1 - 2x_1 \geq 0, \tag{4.73}$$

$$(d-1) - 2dx_0 + (2-d)x_1 \geq 0, \tag{4.74}$$

$$(d-1)^2 + 4dx_0 + 2(d-1)x_1 \geq 0, \tag{4.75}$$

where we have eliminated the dependent variable x_2.

As Werner states of a given dimension are simultaneously diagonalisable, so too are states of the form $\sigma_{\mathbf{x}}^n$ - and moreover, the relative entropy between $\sigma_{\mathbf{y}}^n$ and $\sigma_{\mathbf{x}}^n$ may be expressed simply as

$$S(\sigma_{\mathbf{y}}^n \| \sigma_{\mathbf{x}}^n) = \sum_{i=0}^n y_i \log\left(\frac{y_i}{x_i}\right). \tag{4.76}$$

Note that the state $W_{\eta,d}^{\otimes n}$ is invariant under operation given in Eq. (4.68), and so may also be expressed in the form (4.69). Setting $W_{\eta,d}^{\otimes n} = \sigma_{\mathbf{y}}^n$ gives

$$y_i = \frac{\binom{n}{i}(1-\eta)^{(n-i)}(1+\eta)^i}{2^n}. \tag{4.77}$$

Plugging these values into Eq. (4.76) for $n = 2$ we find that

$$E_R^2(W_{\eta,d}) = \min_{x_0, x_1} \frac{(1-\eta)^2}{8} \log \frac{(1-\eta)^2}{4x_0} + \frac{(1+\eta)(1-\eta)}{4} \log \frac{(1+\eta)(1-\eta)}{2x_1}$$
$$+ \frac{(1+\eta)^2}{8} \log \frac{(1+\eta)^2}{4(1 - x_0 - x_1)} \tag{4.78}$$

subject to conditions (4.73)-(4.75). For convenience, we shall substitute variables with $p = (1 - \eta)/2$ - the *symmetric* representation - making our equation

$$E_R^2(W_{p,d}) = \min_{x_0, x_1} \frac{p^2}{2} \log \frac{p^2}{x_0} + p(1-p) \log \frac{2p(1-p)}{x_1}$$
$$+ \frac{(1-p)^2}{2} \log \frac{(1-p)^2}{(1 - x_0 - x_1)}. \tag{4.79}$$

It is worth remembering in this representation that $W_{p,d}$ are entangled for $p \geq 1/2$, and $E_R(W_{p,d})$ is subadditive for $p > 1/2 + 1/d$.

In order to perform this minimisation, we shall take advantage of the Karush–Kuhn–Tucker (KKT) conditions [14] - necessary conditions for a solution of a non-linear optimisation to be optimal, which generalise the concept of Lagrangian multipliers.

Theorem 4.5.5 Consider an optimisation problem of the following form:

$$\text{Minimise } f(\mathbf{x}), \text{ subject to } g_i(\mathbf{x}) \leq 0.$$

If $\mathbf{x}^*$ is a local minima and $f(\mathbf{x})$ is continuously differentiable at $\mathbf{x}^*$, then the following must be true: there exist λ_i, called *KKT multipliers* such that

$$-\nabla f(\mathbf{x}^*) = \sum_i \lambda_i \nabla g_i(\mathbf{x}^*), \tag{4.80}$$

holds, along with the following conditions:

$$g_i(\mathbf{x}^*) \leq 0 \,\forall i, \tag{4.81}$$

$$\lambda_i \geq 0 \,\forall i, \tag{4.82}$$

$$\lambda_i g_i(\mathbf{x}^*) = 0 \,\forall i. \tag{4.83}$$

Moreover, if $f(\mathbf{x}), g_i(\mathbf{x})$ are convex, then such a point $\mathbf{x}^*$ satisfying these conditions is the global minima.

In our case,

$$f(\mathbf{x}) = \frac{p^2}{2} \log \frac{p^2}{x_0} + p(1-p) \log \frac{2p(1-p)}{x_1} + \frac{(1-p)^2}{2} \log \frac{(1-p)^2}{(1-x_0-x_1)}, \tag{4.84}$$

and

$$g_1(\mathbf{x}) = 2x_1 - 1, \tag{4.85}$$

$$g_2(\mathbf{x}) = 2dx_0 + (d-2)x_1 - (d-1). \tag{4.86}$$

$f(\mathbf{x})$ is not well defined outside of the region $X = \{\mathbf{x} \,|\, x_0 \geq 0, x_1 \geq 0, x_0 + x_1 \leq 1\}$, and thus the minimisation is limited to this region implicitly, without the need for explicit $g_i(\mathbf{x})$ - it has also enabled us to remove condition (4.75), since it is satisfied by all $x_0, x_1 \geq 0$. $f(\mathbf{x})$ is continuously differentiable everywhere in X except $x_0 = 0$ when $p \neq 0$, $x_1 = 0$ when $p \notin \{0, 1\}$, or $1 - x_0 - x_1 = 0$ when $p \neq 1$ - in all cases it takes value ∞. Thus the minimum value cannot occur at such points, and we can be confident any $\mathbf{x}^*$ found is the true global minima by the sufficiency condition. We may employ the sufficiency condition

since $g_1(\mathbf{x})$ and $g_2(\mathbf{x})$ are linear, and thus convex, whilst the quantum relative entropy is jointly convex in both terms - since we have fixed the first input (dependent on p), the function $f(\mathbf{x})$ is convex in $\mathbf{x}$, since there is a one-to-one mapping between states of the form (4.69) and $\mathbf{x}$.

Substituting these functions into Eq. (4.80) gives us two equations

$$\frac{p^2}{2x_0 \ln(2)} - \frac{(1-p)^2}{2\ln(2)(1-x_0-x_1)} = 2d\lambda_2, \tag{4.87}$$

$$\frac{(1-p)p}{x_1 \ln(2)} - \frac{(1-p)^2}{2\ln(2)(1-x_0-x_1)} = 2\lambda_1 + (d-2)\lambda_2. \tag{4.88}$$

whilst the conditions

$$\lambda_1 (2x_1 - 1) = 0 \qquad\qquad \lambda_2 (2dx_0 + (d-2)x_1 - (d-1)) = 0 \tag{4.89}$$

give us four possible cases.

Case 1: $\lambda_1 = \lambda_2 = 0$.

Substituting these into Eq. (4.87)-(4.88), we find that

$$x_0 = p^2, \qquad\qquad x_1 = 2p(1-p). \tag{4.90}$$

Using these values we obtain the condition that $g_2(\mathbf{x}) = (2p-1)(2p+d-1) \leq 0$; this only holds for $p \in [(1-d)/2, 1/2]$, which provides a valid x_1 value when $p \in [0, 1/2]$. This solution gives $x_0 = y_0, \ x_1 = y_1$, and thus corresponds to the minimal solution $f(\mathbf{x}) = 0$ for when $W_{p,d}$ is separable - exactly the region $p \in [0, 1/2]$.

Case 2: $\lambda_1 = 0, x_1 = (d - 1 - 2dx_0)/(d-2)$.

In this case, the solutions for our two unknown variables are much more complicated, and are twofold:

$$\lambda_{2,\alpha} = \frac{a - \sqrt{b}}{4d(d+2)} \qquad\qquad x_{0,\alpha} = \frac{a + 2(d-1) + \sqrt{b}}{4d(d+2)}, \tag{4.91}$$

$$\lambda_{2,\beta} = \frac{a + \sqrt{b}}{4d(d+2)} \qquad\qquad x_{0,\beta} = \frac{a + 2(d-1) - \sqrt{b}}{4d(d+2)}. \tag{4.92}$$

with

$$a = 2d^2 p^2 - 2d^2 p + d^2 + 2dp - d - 4p^2 + 4p, \tag{4.93}$$

$$b = \left(d^2 \left(2p^2 - 2p + 1\right) + 2dp + d - 4p^2 + 4p - 2\right)^2 - 8(d-1)d(d+2)p^2. \tag{4.94}$$

Although complicated, we may still analyse whether they satisfy the KKT conditions. $\lambda_{2,\alpha} \geq 0$ iff $p \geq 1/2$, whilst $\lambda_{2,\beta} \geq 0$ for all values of p. However $g_2(\mathbf{x}_\beta) > 0$ (and thus invalid) for all values of p; but $g_2(\mathbf{x}_\alpha) \leq 0$ in the regions $p \in [0, 1/2]$ and $p \in [1/2 + 1/d, 1]$ - from this, we may conclude that the α solution is the global minima for the region $p \in [1/2 + 1/d, 1]$.

Case 3: $x_1 = 1/2, \lambda_2 = 0$.

Solving Eq. (4.87) and Eq. (4.88) we find that

$$x_0 = \frac{p^2}{4p^2 - 4p + 2} \qquad \lambda_1 = \frac{-4p^2 + 4p - 1}{2\ln(2)}. \qquad (4.95)$$

However, $\left(-4p^2 + 4p - 1\right) / \left(2\ln(2)\right) < 0$ for all values of p except $p = 1/2$, and thus is not valid as a minima except for this value. Moreover, for $p = 1/2$ we have $x_1 = 1/2, x_2 = 1/4$ - so this solution is a special instance of the case below.

Case 4: $x_1 = 1/2, x_2 = 1/4$.

This time we obtain solutions for the two KKT multipliers:

$$\lambda_1 = -\frac{4dp - d + 4p - 2 - 4dp^2}{2d\ln(2)} \qquad \lambda_2 = \frac{2p - 1}{d\ln(2)}. \qquad (4.96)$$

Clearly our choices of x_i are feasible, and satisfy the condition (4.83) - thus it remains only to check when the values of λ_i are non-negative. For λ_2 this occurs when $p \geq 1/2$, whilst for λ_1 this occurs only for $p \in [1/2, 1/2 + 1/d]$. This should not surprise us, as the choice of x_0, x_1 correspond to the separable state $W_{p=1/2,d} \otimes W_{p=1/2,d}$, giving the $E_R^2 = E_R$ for the additive region.

For the subadditive region $p \in (1/2 + 1/d, 1]$ our closest separable state is that given by $\mathbf{x}_\alpha$ - which coincides with the separable state $W_{p=1/2,d} \otimes W_{p=1/2,d}$ at the subadditivity boundary $p = 1/2 + 1/d$. We can convert this state back into the expectation representation, in order to gain our result for $\eta < -2/d$.

$$\begin{aligned}
E_R^2(W_{\eta,d}) = &\frac{(1-\eta)^2}{8} \log \frac{(1-\eta)^2}{4x_0} + \frac{(1+\eta)(1-\eta)}{4} \log \frac{(1+\eta)(1-\eta)}{2x_1} \\
&+ \frac{(1+\eta)^2}{8} \log \frac{(1+\eta)^2}{4(1-x_0-x_1)}
\end{aligned} \qquad (4.97)$$

with

$$x_0 = \frac{c + \sqrt{d}}{8d(d+2)} \qquad x_1 = \frac{4(d+2)(d-1) - c - \sqrt{d}}{4(d^2 - 4)}, \qquad (4.98)$$

121
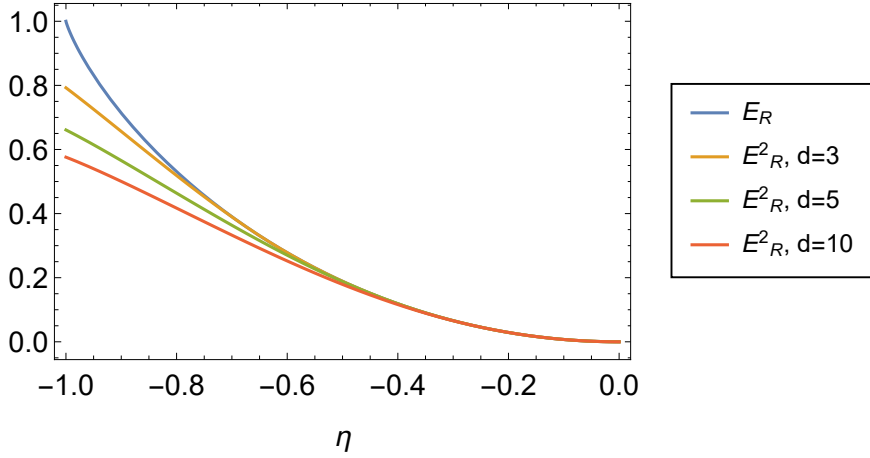
Figure 4.5: A comparison of the one-copy and two-copy Werner states' REE, for varying dimensions $d > 2$. These provide upper bounds for the secret key capacity of $\mathcal{W}_{\eta,d}$.

where

$$c = d^2 \left(\eta^2 + 1\right) - 2d(\eta - 2) - 2\eta^2 - 2, \tag{4.99}$$

$$d = d^4 \left(\eta^2 + 1\right)^2 - 4d^3\eta \left(\eta^2 - 3\right) - 4d^2 \left(\eta^4 + 3\eta^2 - 1\right) + 8d\eta \left(\eta^2 - 3\right) + 4 \left(\eta^2 + 1\right)^2. \tag{4.100}$$

We plot this result for varying dimension in figure 4.5; by consequence of theorem 4.5.2, it provides a tighter bound of the secret key rate of $\mathcal{W}_{\eta,d}$ than the single-copy REE.

4.5.1 Squashed Entanglement as a Secret Key Bound

In chapter 2, we referenced how there were *two* entanglement measures for which we may upper bound the secret key capacity of a σ-simulable channel; The first is the one just discussed, the (regularised) REE. As the REE is subadditive, it does not provide the best bound for Holevo-Werner channels - the one-copy bound is not tight, whilst for $n > 2$ the n-copy value is not known. Thus we look to the other entanglement measure to provide a better bound, *squashed entanglement* [20]. We recap its definition here.

Definition 4.5.6 The squashed entanglement of a state $\rho_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ is given by:

$$E_{\text{sq}}\left(\rho_{AB}\right) = \frac{1}{2} \min_{\rho'_{ABE} \in \Omega_{AB}} S(A : B|E) \tag{4.101}$$

where Ω_{AB} is the set of density matrices ρ'_{ABE} such that $\mathrm{Tr}_E[\rho'_{ABE}] = \rho_{AB}$ - the size of the ancilla space $\mathcal{H}_E$ is generally unbounded. The value $S(A : B|E)$ is the *quantum conditional mutual information*, given by

$$S(A : B|E) := S(\rho'_{AE}) + S(\rho'_{BE}) - S(\rho'_E) - S(\rho'_{ABE}) \tag{4.102}$$

where $S(\rho'_{AE})$ denotes the von Neumann entropy of the reduced state $\rho'_{AE} = \mathrm{Tr}_B[\rho'_{ABE}]$.

This entanglement measure is strictly *additive* over tensor products - thus we avoid the need for regularisation, which is our difficulty with the REE. Unfortunately, there is a trade-off for this simplification - the optimisation over Ω_{AB} is extremely difficult[13] - this is because of the unbounded ancilla dimensions $\mathcal{H}_E$. For Holevo-Werner states, we may upper bound the secret key capacity by the squashed entanglement of its Choi matrix $W_{\eta,d}$ - which we then upper bound by choosing a particular ρ'_{ABE} - thus we obtain an analytical bound at the cost of having potentially chosen the non-minimising extension.

In order to upper bound $E_{\mathrm{sq}}(W_{\eta,d})$, we choose the extension where ρ'_{ABE} is the *purification* of $W_{\eta,d}$ - the pure $\mathcal{H}_{d^4}$ state $\tilde{W}_{\eta,d} \in \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_{\tilde{A}} \otimes \mathcal{H}_{\tilde{B}}$ which satisfies $\mathrm{Tr}_{\tilde{A}\tilde{B}}[\tilde{W}_{\eta,d}] = W_{\eta,d}$.

As $\tilde{W}_{\eta,d}$ is pure, we have that $S(\rho'_{ABE}) = S(\tilde{W}_{\eta,d}) = 0$. Moreover, $\rho'_E = W_{\eta,d}$, and so

$$S(\rho'_E) = S(W_{\eta,d}) = -\sum_i p_i \log p_i = -\left(\frac{1-\eta}{2} \log\left(\frac{1-\eta}{d(d-1)} \right) + \frac{1+\eta}{2} \log\left(\frac{1+\eta}{d(d+1)} \right) \right). \tag{4.103}$$

Finally, we compute $S(\rho'_{AE}) = S(\rho'_{BE}) = \log d$. This means we have

$$E_{\mathrm{sq}}(W_{\eta,d}) \leq \tilde{E}_{\mathrm{sq}}(\eta) := \log d + \frac{1-\eta}{4} \log\left(\frac{1-\eta}{d(d-1)} \right) + \frac{1+\eta}{4} \log\left(\frac{1+\eta}{d(d+1)} \right). \tag{4.104}$$

We present a comparison of this bound in various dimensions in figure 4.6. We can see from this formula we have certainly not chosen the optimal ρ'_{ABE}, as for $\eta = 0$ we have $\tilde{E}_{\mathrm{sq}}(\eta) = \log\left(d^2 / \left(d^2 - 1 \right) \right) / 4$, whereas the optimal choice of ρ'_{ABE} would give $E_{\mathrm{sq}}\left(W_{0,d} \right) = 0$, since this state is separable. Nevertheless , $\tilde{E}_{\mathrm{sq}}(\eta)$ provides a valid, easy to calculate upper bound for the squashed entanglement.

---

[13]NP hard in fact [62].
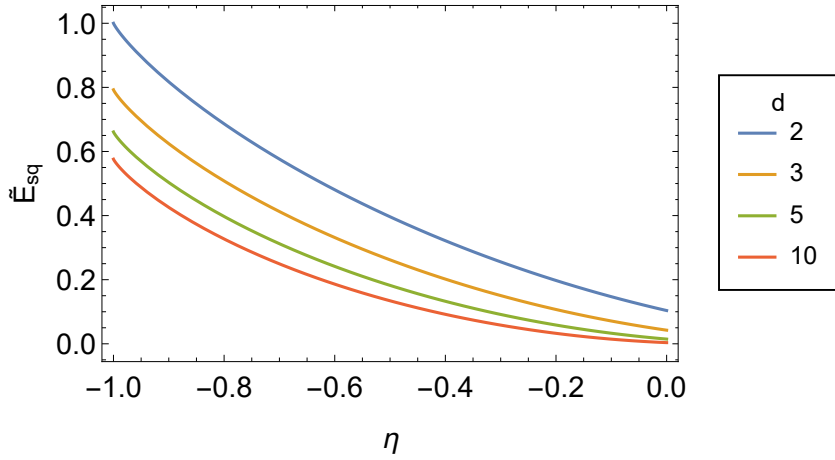
Figure 4.6: A comparison of $\tilde{E}_{\mathrm{sq}}(W_{\eta,d})$ for entangled Werner states of varying dimension - note the failure to reach 0 at $\eta = 0$.

There is one final bound to secret key capacity we may exploit for Werner states, one which also relies on the squashed entanglement. This exploits the *convexity* of the squashed entanglement [17]. We first note that

$$W_{\eta,d} = \frac{(d-\eta)\mathrm{I} + (d\eta-1)\mathbb{F}}{d^3-d} = (1+\eta)\frac{d\mathrm{I}-\mathbb{F}}{d^3-d} + (-\eta)\frac{(d+1)\mathrm{I}+(-d-1)\mathbb{F}}{d^3-d}$$

$$= (1+\eta)W_{0,d} + (-\eta)W_{-1,d}. \tag{4.105}$$

Therefore for entangled states $\eta \in [-1,0)$ we may write $W_{\eta,d}$ as a convex combination of the separable state $W_{0,d}$ and the extremal entangled state $W_{-1,d}$. This means we may write

$$E_{\mathrm{sq}}(W_{\eta,d}) \leq (1+\eta)E_{\mathrm{sq}}(W_{0,d}) - \eta E_{\mathrm{sq}}(W_{-1,d}) = -\eta E_{\mathrm{sq}}(W_{-1,d}) \tag{4.106}$$

since $E_{\mathrm{sq}}(W_{0,d}) = 0$. In the paper [21], an analysis of the squashed entanglement for the extremal state $W_{-1,d}$ was performed, and the following bound obtained:

$$E_{\mathrm{sq}}(W_{-1,d}) \leq \begin{cases} \log\left(\frac{d+2}{d}\right) & \text{if } d \text{ even,} \\ \frac{1}{2}\log\left(\frac{d+3}{d-1}\right) & \text{if } d \text{ odd} \end{cases} \tag{4.107}$$

124

(a) $d = 4$ (b) $d = 5$
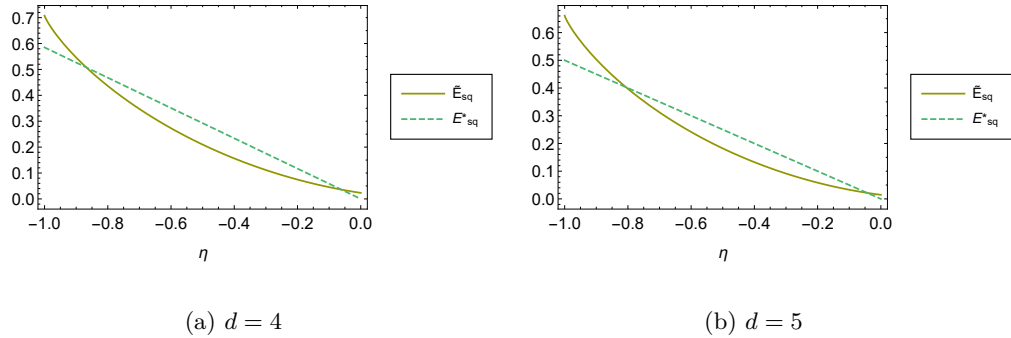
Figure 4.7: A comparison of our two upper bounds for $E_{\mathrm{sq}}(W_{\eta,d})$, for dimensions $d = 4, 5$. Which bound is tighter is dependent on the value of η.

This gives us the following useful bound on the squashed entanglement

$$E_{\mathrm{sq}}(W_{\eta,d}) \leq E_{\mathrm{sq}}^*(W_{\eta,d}) := \begin{cases} -\eta \log\left(\frac{d+2}{d}\right) & \text{if } d \text{ even,} \\ -\frac{\eta}{2} \log\left(\frac{d+3}{d-1}\right) & \text{if } d \text{ odd.} \end{cases} \tag{4.108}$$

4.5.2 Comparing bounds, and use of the RPPT

For Werner states the regularised RPPT is exactly known; and although we cannot use $E_P^\infty(W_{\eta,d})$ as an upper bound to the secret key capacity, we can use it as an upper bound[14] for Q_2. As for teleportation covariant channels we have that $Q_2(\mathcal{E}) = E_D(\chi_{\mathcal{E}})$, we can compare this bound to the logarithmic negativity[15] of $W_{\eta,d}$, which we do in figure 4.8. In figure 4.9, we present a comparison of the four considered secret key capacity bounds for $\mathcal{W}_{\eta,d}$, along with the regularised RPPT which bounds the two-way quantum capacity. In figure 4.10, we show only the tightest bound across the whole entangled range $\eta < 0$. Each of the four considered bounds is the tightest for some region of η, and therefore in order to obtain the most accurate picture of a channel's capacity we should not limit ourselves to one entanglement measure.

---

[14]See chapter 2.

[15]Which is an upper bound on distillable entanglement for states.
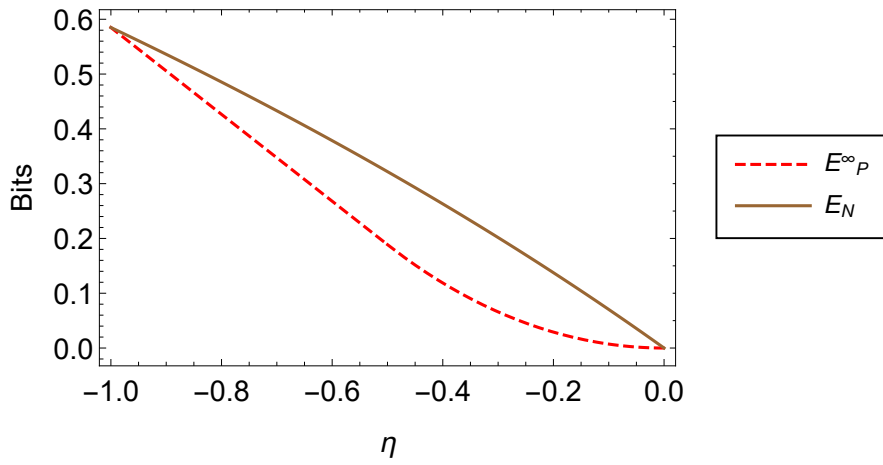
Figure 4.8: We can see that the regularised RPPT provides a tighter bound for $Q_2(\mathcal{W}_{\eta,4})$ for all entangled η than the logarithmic negativity.
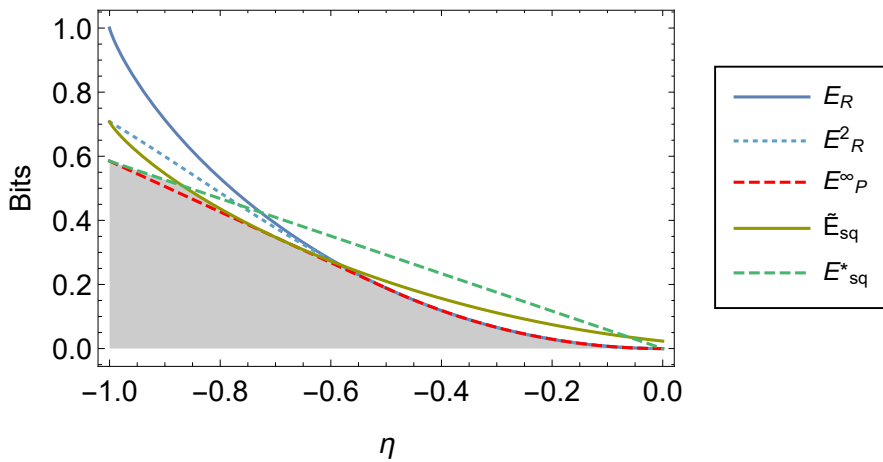
Figure 4.9: A comparison of the various upper bounds for the secret key capacity $K(\mathcal{W}_{\eta,4})$ - the allowable range is given by the grey shading.

126
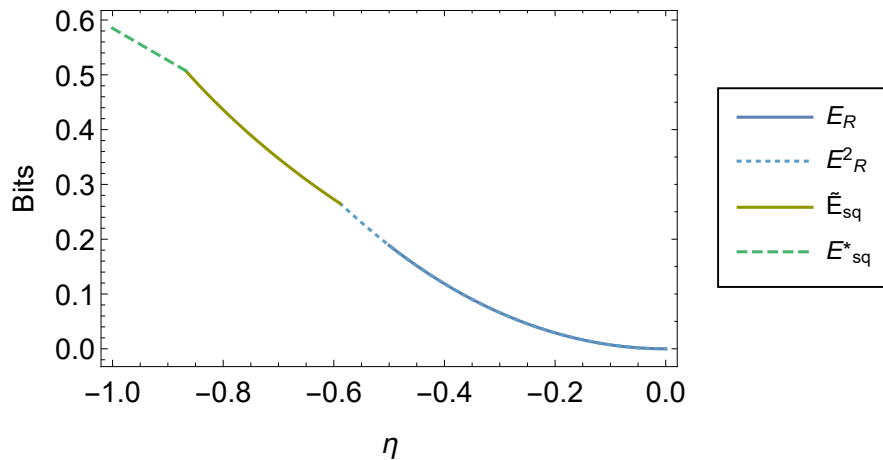
Figure 4.10: The tightest bound for $K(\mathcal{W}_{\eta,4})$ - note how all four entanglement measures provide the best bound for certain regions of η.

4.6 Discussion and Further Directions

The first part of this chapter applied some cutting edge results in metrology to the unusual Holevo-Werner channels, which we were able to exploit by proving these channels are teleportation covariant - one of the few non-Pauli channels[16] shown to have this property. As a consequence of this, the optimal parameter estimation may only achieve the shot noise limit. We also obtained the surprising result that entanglement-breaking is not the determining factor in optimal estimation for these channels - although for teleportation covariant states the optimal precision may be achieved [86] by sending half of a maximally entangled state through the channel, the lack of dependence on entanglement-breaking implies the same precision may be achievable with unentangled states also. We also found that the choice of parametrisation is important for determining the role of the channel's dimension - for the expectation representation it plays no role, whilst for the α-representation different values change their difficulty as the channel's dimension increases.

We then progressed to the problem of binary discrimination of Holevo-Werner channels, providing for the first time analytic forms for the true quantum Chernoff bound for

---

[16]When $d > 2$.

both Holevo-Werner and depolarising channels, allowing for more accurate bounding of their binary discrimination.

After the discussion on metrology, we looked at bounding the capacities of Holevo-Werner channels. The ability to bound the secret key distillation rate of a state by various entanglement measures is a useful tool, and mirrors the use of entanglement measures to bound other operational measures such as the distillable entanglement and entanglement cost. The extension of this to *channels* in [85] by use of LOCC simulation was a huge step; in the cases where the channel is teleportation covariant, it seemed to work particularly well, as for many channels the upper bound provided by the single-copy REE coincided with the lower bound provided by coherent [69, 91] and reverse coherent [43, 83] information, achievable lower bounds on quantum and secret key capacities, allowing for the first time the capacities of several channels to be exactly determined; for example the dephasing and erasure channels, along with important continuous variable channels such as the pure loss channel. The aim of the work presented in this chapter was to look at a channel for which, by construction, this upper bound would *not* work as well, since the REE was known to be subadditive for Werner states. By calculating the analytical form for E_R^2, the two-copy REE, we saw that indeed there was a substantial difference between the two, at the extremal value $\eta = -1$ already reaching a gap of 0.2 bits for the smallest subadditive dimension $d = 3$. In the limit as $d \to \infty$ this gap reaches 0.5 bits, so the one-copy bound becomes quite loose indeed. This suggests that for these states, an alternative entanglement measure is preferable - from figure 4.10 we see that, for much of the subadditive region, the best upper bound is provided by one of the two bounds motivated by the squashed entanglement, which is strictly additive over tensor products.

In order to try and tighten the upper bound for $K(\mathcal{W}_{\eta,d})$, one could try to obtain $E_R^3(W_{\eta,d})$. One possible way to do this would be to try and prove that all states of the form shown in Eq. (4.69) which are PPT are also separable, for $n = 3$. This could be done by using vertex enumeration[17] to find the extremal states of this set subject to the linear constraints in Eq. (4.70) with $n = 3$, and then show each of these extremal states is separable. One could then apply the KKT conditions as with $n = 2$. There is no guar-

---

[17]Explained in detail in chapter 5.

antee that separable and PPT states coincide for these states though, and furthermore we cannot say how well E_R^3 will compare to our squashed entanglement-based bounds, for the region where they are tightest. With this in mind, perhaps a more fruitful approach could be try to obtain the true value of $E_{\text{sq}}(W_{\eta,d})$ - as noted in [21], the coincidence of $E_{\text{sq}}^*\left(W_{-1,d}\right)$ for even d, $E_P^\infty\left(W_{-1,d}\right)$ and $E_N\left(W_{-1,d}\right)$ implies this may be the *true* entanglement of the extremal state, suggesting the best way to approach this problem would be to generalise an extension of $W_{-1,d}$ achieving this value. An interesting question would be to investigate whether the optimal extension ρ'_{ABE} could change dependent on the range of η, akin to the way either $E_{\text{sq}}^* < \tilde{E}_{\text{sq}}$ or $\tilde{E}_{\text{sq}} < E_{\text{sq}}^*$ depending on the region.

Another consideration to take into account is the conjecture that Werner states with $\eta \in \left[(2-d)/(2d-1),0\right)$ are NPT bound entangled. If this conjecture is true, then $Q_2(\mathcal{W}_{\eta,d}) = 0$ for this region, as for teleportation covariant channels states $Q_2(\mathcal{E}) = D_2(\mathcal{E}) = E_D(\chi_{\mathcal{E}})$. This would illustrate that our tightest bound for Q_2, the RPPT, would still be very loose. As there exist PPT bound entangled states from which a secure secret key can be distilled, it is likely that we still have $K(\mathcal{W}_{\eta,d}) > 0$ in this region and therefore we have strong motivation to study these entanglement-based upper bounds.

Chapter 5

Bell Polytopes and the Detection Loophole

This chapter details work done in collaboration with Roger Colbeck, in which we have used a linear programming algorithm to generate new Bell inequalities, and tested their efficacy with regards to the detection loophole. We hope to turn this into a paper in the near future.

5.1 Structure of this Chapter

This chapter begins with a review of Bell inequalities and their relevance to quantum information, as well as an overview of some useful mathematical tools for studying them, including polytopes and both linear and semidefinite programming. We then move on to explain the linear programming algorithm used to obtain new Bell inequalities, before presenting evidence for how they perform in regards to determining non-locality with detection failures.

5.2 Bell Correlations and Non-locality: An Introduction

In 1935 Einstein, Podolsky and Rosen published a paper [39] containing a thought experiment aimed to challenge the Heisenberg uncertainty principle. It started with two particles (which we shall refer to as A and B) maximally entangled in position and momentum - one can measure the one particle's position or momentum, and the other particle will have exactly the opposite value. Quantum mechanics allows for particles to have this property,

and to retain it when separated such that they cannot interact. The paradox proposed was as follows: the uncertainty principle asserts that both the position and momentum of particle B cannot be exactly known. However, by measuring particle A, one can determine exactly the position or momentum of particle B - without particle B having been disturbed in any way. Thus one would expect to be able to measure the other property exactly on B, since it has not previously been disturbed by measurement. In actuality when measuring B this turns out not to be the case, and the value of the quantity not measured on particle A turns out to be uncertain. Assuming that information cannot be transferred from A to B about the measurement - due to the limitations of the speed of light - and that every particle has an intrinsic position/momentum determining the outcome of measurements, this gives a paradox - why does quantum theory not allow us to measure exactly the position and momentum of particle B, when they must have exact values - seeing as how we could measure either the position or momentum of A, in order to determine B's corresponding counterpart?

The solution favoured by the authors was that of *hidden variables* - that when the two particles interacted to form the entanglement, there existed more fundamental variables which determined the true properties of the particles - the quantum mechanical description only describes the behaviour because we do not have a more complex theory, in which both position and momentum can be described exactly. Perhaps even we are unable to ever measure these variables - yet they still determine the particles' behaviours exactly.

Einstein was always a vocal critic of quantum uncertainty, disliking the idea that probability is fundamental to physical behaviour. The argument over probability versus hidden determinism continued until 1964, when John Bell [8] approached the problem. He began with the assumptions made by Einstein et al. - so called "local realism" - that there exists real properties intrinsic to the particles determining the outcomes of measurement, and that measurements in one location cannot remotely affect the outcomes of distant measurements. From these assumptions he was able to show that any model determined by local hidden variables must satisfy a certain correlation relation for measurements - known as *Bell's inequality*. Furthermore, this inequality is *violated* by quantum mechanics; and thus the assumptions made by Einstein et al. (and therefore the local hidden variable model) must be false. Quantum mechanics is intrinsically random - there cannot be an

underlying deterministic behaviour.

5.2.1 CHSH

Bell's original inequality concerned the spin of maximally entangled particles - in particular requiring that the two measured particles are perfectly anti-correlated; measuring in opposite directions will result in an identical outcome, whilst measuring in the same direction giving exactly opposite outcomes - the same correlations analysed by the authors of [39]. Whilst such perfect correlation is predicted by quantum mechanics, experimentally this is impossible to achieve. Thus a big step was acheived in [23], in which the CHSH inequality (named after the four authors) was introduced. They stated that, for any set of two party measurements $\{A_1, A_2\}$, $\{B_1, B_2\}$ with outcome values ± 1, then the following holds

$$\langle A_1 B_1 \rangle + \langle A_1 B_2 \rangle + \langle A_2 B_1 \rangle - \langle A_2 B_2 \rangle \leq 2 \tag{5.1}$$

for any local hidden variable theory. In quantum mechanics, it is possible to achieve $2\sqrt{2}$, known as Tsirelson's bound, using maximally entangled states. The advantage of this inequality is that *any* value greater than 2 provides a counterexample to the hidden variable theory, allowing it to be tested with noisy states and measurements.

5.3 Preliminaries

5.3.1 Polytope Theory

Before we look in more detail at the world of Bell inequalities, it is worthwhile to take a detour and familiarise ourselves with some concepts from geometry. In particular, it is useful to understand at least the basics of a branch of geometry looking at *polytopes*. There are two ways of defining a polytope.

Definition 5.3.1 The vertex representation (V-representation) of a polytope $\mathcal{P}$ is the set of points $\{\mathbf{x}_i\}$ such that

$$\mathcal{P} = \left\{ \mathbf{x} \,\middle|\, \mathbf{x} = \sum_i \lambda_i \mathbf{x}_i, \ \lambda_i \geq 0, \ \sum_i \lambda_i = 1 \right\} \tag{5.2}$$

i.e. all convex combinations of the set of points $\mathbf{x}_i$.

Obviously this definition is not unique - we could always take the set $\{\mathbf{x}_i\} \cup \{(\mathbf{x_1} + \mathbf{x_2})/2\}$ and it would define the same polytope. Thus when referring to the V-representation of a

polytope, we are implicitly referring to the *minimal* V-representation, which is unique [49]. A polytope with this definition is also known as a *convex hull*, as $\mathcal{P}$ is the convex hull[1] of the points $\{\mathbf{x}_i\}$.

Definition 5.3.2 The half-space representation (H-representation) of a polytope $\mathcal{P}$ is a finite set of linear inequalities, such that:

$$\mathcal{P} = \{\mathbf{x} \,|\, A\mathbf{x} \leq \mathbf{b}\}. \tag{5.3}$$

Once again this representation is not unique - there may be redundant inequalities. Therefore we refer to the minimal H-representation [49], which is again unique[2]. The inequalities in the H-representation are known as "facet-defining inequalities", as each equation $A_i\mathbf{x} = b_i$ defines a *supporting hyperplane* of dimension strictly one less that that of the polytope; and the intersection of the hyperplane with the polytope $\mathcal{P}$ defines a face of the polytope (again of dimension $\dim[\mathcal{P}] - 1$) - these faces are known as *facets*. Any point in the polytope satisfying $A_i\mathbf{x} = b_i$ is said to *saturate* the inequality.

The concept of a face is closely linked to that of *interior points* of a polytope. An interior point $\mathbf{x}_{\text{int}}$ of the polytope is a point such that there exists an $\epsilon > 0$ where

$$B_\epsilon(\mathbf{x}_{\text{int}}) := \left\{\mathbf{x}' \,\middle|\, \|\mathbf{x}' - \mathbf{x}_{\text{int}}\| \leq \epsilon\right\} \subseteq \mathcal{P}. \tag{5.4}$$

Faces are then defined as non-empty intersections of a half-plane with the polytope, such that no interior points of the polytope lie within the intersection. Given a d-dimensional polytope, faces can range in dimension from the $(d-1)$-dimensional facets down to the 0-dimensional vertices.

5.3.1.1 A Simple Example: The Cube

Consider a cube whose sides are of length 2, centred around the origin. This is a 3-dimensional polytope, and its V-representation is:

$$V_{\text{cube}} = \{\mathbf{x}_1 = (-1, -1, -1), \quad \mathbf{x}_2 = (-1, -1, 1), \quad \mathbf{x}_3 = (-1, 1, -1), \quad \mathbf{x}_4 = (-1, 1, 1),$$
$$\mathbf{x}_5 = (1, -1, -1), \quad \mathbf{x}_6 = (1, -1, 1), \quad \mathbf{x}_7 = (1, 1, -1), \quad \mathbf{x}_8 = (1, 1, 1)\}$$
$$\tag{5.5}$$

---

[1]The convex hull of a set X is the smallest convex set $C(X)$ such that $X \subseteq C(X)$.
[2]Up to the trivial multiplication of inequalities by a constant.

which are the corners/vertices of the cube. We also give the H-representation

$$H_{\text{cube}} = \left(A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \right) \tag{5.6}$$

where we have just provided the set-defining matrix A and vector $\mathbf{b}$. As you see, each row $A_i \mathbf{x} = b_i$ gives a 2-dimensional face (and thus a facet) of the cube. Moreover, we may obtain the 1-dimensional faces (edges) of the cube by taking two compatible equalities: the first two rows define the edge $\lambda(1, 1, -1) + (1 - \lambda)(1, 1, 1),\ \lambda \in [0, 1]$, and the first three rows taken together define the 0-dimensional face (vertex) $(1, 1, 1)$. Generally, a $(d - k)$-dimensional face of a d-polytope[3] will satisfy exactly k equalities in the H-representation, and can thus be seen as the intersection of exactly k supporting hyperplanes.

Note that we could express the cube polytope in *both* a vertex representation and a half-plane representation. It turns out that this is always the case, and this forms the result often called the the *main theorem of polytope theory.*

Theorem 5.3.3 (Minkowski-Weyl Theorem) Every bounded[4] polytope admits a V-representation and an H-representation of the form described above. If instead the polytope is unbounded, both representations still exist with the V-representation instead given by two sets: the vertex set $\{\mathbf{x}_i\}$ and the ray set $\{\mathbf{r}_j\}$. The polytope then consists of all points expressible as

$$\mathcal{P} = \left\{ \mathbf{x} \,\middle|\, \mathbf{x} = \sum_i \lambda_i \mathbf{x}_i + \sum_j \alpha_j \mathbf{r}_j,\ \lambda_i \geq 0,\ \sum_i \lambda_i = 1,\ \alpha_i \geq 0, \right\}. \tag{5.7}$$

5.3.1.2 Facet Enumeration

The problem of finding the set of facets (the H-representation) given the extremal points (the V-representation) is known as *facet enumeration*, or the *convex hull problem.* This problem is generally difficult; one of the most popular algorithms [6] runs in $O(ndk)$, with

---

[3]A polytope of dimension d.

[4]There exists N such that $\|\mathbf{x}\| < N,\ \forall \mathbf{x} \in \mathcal{P}$.

n the number of extremal points, d the polytope's dimension, and k the number of facets. In the worst case this can be of $O(n^{\lfloor \frac{d}{2} \rfloor})$ [75]. The converse problem, known as *vertex enumeration*, is equivalently hard - taking n as the number of facets and k as the number of extremal points instead. This problem is known to be NP-hard in general [67]. As we shall see later in the chapter, the problem of facet enumeration is particularly relevant in quantum theory.

5.3.1.3 Dual Polytopes

Given a polytope $\mathcal{P}$, it is possible to construct a graph known as a "face lattice" - an ordered graph detailing subset inclusion of all faces of the polytope (from vertices up to facets), the empty set, and the polytope itself. By taking this graph and reversing the ordering one may create a new "hypothetical" face lattice. Any polytope $\mathcal{D}$ whose own face lattice structure matches this hypothetical lattice is said to be a *combinatorial dual* of $\mathcal{P}$. Every polytope admits a dual and there exists an explicit construction we can use. First we require that the origin of the polytope belongs to the interior of $\mathcal{P}$; if this not the case, a coordinate transform is required. Given this, the polar dual of $\mathcal{P}$ is defined:

$$\mathcal{P}^{\Delta} := \left\{ \mathbf{y} \,\middle|\, \mathbf{x}^T \mathbf{y} \leq 1, \ \forall \mathbf{x} \in \mathcal{P} \right\}. \tag{5.8}$$

Importantly these polar duals introduce the idea of duality between vertices and facets; they have the easily verifiable property that:

$$\mathcal{P}^{\Delta} = \left\{ \mathbf{y} \,\middle|\, \mathbf{x}_i^T \mathbf{y} \leq 1 \ \forall \mathbf{x}_i \in V_{\mathcal{P}} \right\} \tag{5.9}$$

and furthermore this turns out to be exactly the H-representation of $\mathcal{P}^{\Delta}$. Thus the vertices of the primal polytope $\mathcal{P}$ are exactly the facets of the polar dual polytope $\mathcal{P}^{\Delta}$. Moreover, $\left(\mathcal{P}^{\Delta}\right)^{\Delta} = \mathcal{P}$ - meaning that the vertices of $\mathcal{P}^{\Delta}$ are the rows of $A \in H_{\mathcal{P}}$, provided the inequalities have been normalised so that $\mathbf{b} = \mathbf{1}$. This concept of the polar dual is one that we shall return to later.

For our simple example of the cube, we can see that we have already cleverly chosen our inequalities to be of the form $\mathbf{b} = \mathbf{1}$. Thus we can immediately form the polar dual.

$$\begin{aligned} V_{\text{cube}^{\Delta}} = \{ \mathbf{y}_1 = (1,0,0), \qquad &\mathbf{y}_2 = (0,1,0), \qquad \mathbf{y}_3 = (0,0,1), \\ \mathbf{y}_4 = (-1,0,0), \qquad &\mathbf{y}_5 = (0,-1,0), \qquad \mathbf{y}_6 = (0,0,-1) \}, \end{aligned} \tag{5.10}$$

$$H_{\text{cube}^\triangle} = \left(A = \begin{pmatrix} -1 & -1 & -1 \\ -1 & -1 & 1 \\ -1 & 1 & -1 \\ -1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \right). \tag{5.11}$$

This eight-sided shape with six vertices is a regular octahedron centred around $\mathbf{0}$.

5.3.1.4 Dimensions, Affine and Vector Spaces

So far we have been discussing the dimensions of polytopes without explicitly defining it. To understand exactly what we refer to, we shall briefly outline affine and vector spaces.

For a vector space $\mathcal{V}$ (or rather subspace - we are taking subspaces of $\mathbb{R}^d$, for some dimension d) we may always find a *basis*, a set of vectors $\{\mathbf{z}_i\}_{i=1}^k$ which are linearly independent and any $\mathbf{v} \in \mathcal{V}$ can be expressed: $\mathbf{v} = \sum_i^k \beta_i \mathbf{z}_i, \ \beta_i \in \mathbb{R}$. The dimension of this vector space is k, the number of basis vectors. Given a set of vectors $X = \{\mathbf{x}_i\}$ we can construct a vector space $\mathcal{V}(X)$ by taking all combinations $\sum_i \beta_i \mathbf{x}_i, \ \beta_i \in \mathbb{R}$, and find the dimension of this space by taking the size of the largest linearly independent subset of X.

Given that same set X, we could also construct an *affine* subspace, $\mathcal{A}(X)$. This is the set of points $\mathbf{x} = \mathbf{x}_1 + \beta_i (\mathbf{x}_i - \mathbf{x}_1), \ \beta_i \in \mathbb{R}$. The dimension of this affine space is given by the vector space obtained when $\mathcal{A}(X)$ is translated such that $\mathbf{x}_1$ is taken to the origin, and thus $\dim [\mathcal{A}(X)] = \dim [\mathcal{V}(\{\mathbf{x}_i - \mathbf{x}_1\}]$. The choice of $\mathbf{x}_1$ is not unique, and one may choose any $\mathbf{x} \in \mathcal{A}(X)$.

Finally, the dimension of a polytope[5] is taken to be the dimension of the affine subspace of its vertex set. Although the polytope (being restricted to convex combinations) is a subset of the affine space, it is clear its dimensionality must be equal to that of the affine

---

[5]Or a face of a polytope - which is itself a polytope.

space, since its extremal points are the vectors $\mathbf{x}_1$, $\mathbf{x}_1 + (\mathbf{x}_2 - \mathbf{x}_1) \ldots$ and therefore the number of basis vectors required to fully describe the space must exactly coincide with that of the full affine space.

5.3.2 An Introduction to Linear Programming

Linear programming refers to the methods used to solve a class of problems which require the optimisation of a linear objective function over a set of variables constrained by a series of linear equalities and/or inequalities. They have a strong practical application, with much development of their study being motivated by economic theory and military strategy; often these problems are formulated as the optimal use of limited resources. In this section we shall introduce the canonical form of a linear programming problem, and the most common methods for solving them. We shall also explain their relation to polytope theory [107].

The canonical form of a linear programming problem is as follows: given a fixed $\mathbf{c}, \mathbf{q}$ and A,

$$\text{maximise } \mathbf{c}^T \mathbf{x} \text{ subject to: } A\mathbf{x} \leq \mathbf{q}, \ \mathbf{x} \geq \mathbf{0}. \tag{5.12}$$

We also refer to this as the (canonical) *primal* form. It should be noted that all linear programming problems can be converted into this form; if the problem instead asks to minimise $\mathbf{c}^T \mathbf{x}$ then we may set $\mathbf{c}' = -\mathbf{c}$; clearly our original problem is minimised when $\mathbf{c}'^T \mathbf{x}$ is maximised. If one of our constraints has that $A_k \mathbf{x} \geq q_k$, we may instead use the equivalent constraint that $-A_k \mathbf{x} \leq -q_k$. Another common issue is that one of our variables x_k is bounded by some value other than zero. We may either add this constraint to the matrix A, or redefine the variable x_k so that it is correctly bounded.

Given a linear programming problem, there are three possibilities: if there are no $\mathbf{x}$ which satisfy the constraints, then the problem is *infeasible*. If the domain of the variables is unbounded in such a way that we may increase the objective function's[6] value indefinitely, then we say the problem is *unbounded*. If neither of these are true, then the region satisfying the constraints is known as the *feasible region*, and there exists with certainty a finite optimal solution on this space.

---

[6]Objective function refers to $\mathbf{c}^T \mathbf{x}$.

Note that the constraints are of the form $A\mathbf{x} \leq \mathbf{q}$ and $\mathbf{x} \geq 0$ - these are half-spaces, and form a H-representation of a polytope, which is our solution space. Thus we have a strong connection between linear programming and polytope theory.

Lemma 5.3.4 (The Maximum Principle [90]) The optimum value of the objective function is achieved at an extremal point of the solution space polytope.

5.3.2.1 The Dual Problem

Every primal linear programming problem has a corresponding linear problem which we call the *dual problem*; given a primal problem in the form (5.12), the dual problem is given as

$$\text{minimise } \mathbf{q}^T\mathbf{y} \text{ subject to: } A^T\mathbf{y} \geq \mathbf{c}, \ \mathbf{y} \geq \mathbf{0}. \tag{5.13}$$

Note that this problem too takes place over a polytope - one should be careful to not to automatically connect this to the *dual* polytope, due to the manner in which it has been constructed; note that the origin is never in the interior of the solution space polytope. However we shall see later on there are some connections we may make.

Lemma 5.3.5 Linear programming problems are *strongly dual*; given an optimum solution $\mathbf{x}^*$ for (5.12), and $\mathbf{y}^*$ for (5.13), then $\mathbf{c}^T\mathbf{x}^* = \mathbf{q}^T\mathbf{y}^*$. If either the primal or dual is unbounded, then the other is infeasible.

We can see that changing from the primal to the dual form switches the number of constraints and variables; the most efficient algorithms for linear programs utilise both the primal and dual problem to find the solution. There is also another extremely important result within linear programming, which we will take advantage of later on in the paper.

Theorem 5.3.6 (Complimentary Slackness.) The solutions $\mathbf{x}^*, \mathbf{y}^*$ to the primal/dual of a linear program are optimal iff

$$(\mathbf{q} - A\mathbf{x}^*)^T \mathbf{y}^* = 0, \text{ and} \tag{5.14}$$

$$\left(A^T\mathbf{y}^* - \mathbf{c}\right)^T \mathbf{x}^* = 0. \tag{5.15}$$

5.3.2.2 Methods for Solving Linear Programming Problems

There are two main approaches for solving this class of problems. We shall provide a qualitative explanation of them both here.

- **The simplex algorithm**.

  Formulated by George Dantzig [32], this algorithm begins by calculating the objective function's value at one vertex of the feasible polytope, known as the *initial feasible solution*. It then moves along the polytope's edges vertex to vertex in such a way that the objective function is always non-decreasing. Provided the choice of edge to follow in the scenario where there are multiple options is well chosen[7], then this method is guaranteed to find the global optimum in finite time. This method can be done using exact values, and is practically efficient, although there exist non-polynomial worst case scenarios. There exist other similar edge-following algorithms; an example is the criss-cross method ([98], [108]); however the simplex is the most widely implemented, and the one we use here.

- **Interior point methods** [77].

  Unlike the previous approach, interior point methods stay strictly inside the interior of the feasible region, and takes successive steps closer to the optimal solution until the sequence converges to within some allowed tolerance. Though different algorithms choose their step determination differently, the general idea is to move along the direction of the greatest change in the objective function with each successive step. There exist particular algorithms which are polynomial for all problems, and modern methods are efficient for numerical precision problems.

5.3.3   Semidefinite Programming

Related to linear programming is an area of convex optimisation known as *semidefinite programming*. The canonical form of a semidefinite programming problem is:

$$\text{minimise Tr}\left[C^T X\right] \text{ subject to Tr}\left[A_k^T X\right] = b_k, \ k = 1, \dots m, \ X \geq 0. \tag{5.16}$$

Similar to linear programming, we have a linear objective function we wish to maximise: $\text{Tr}\left[C^T X\right] = \sum_{i,j} C_{ij} X_{ij}$. This time however, our variable space is that of the positive semidefinite matrices - subject to a set of linear equalities, given by the matrices $\{A_k\}$ and vector $\mathbf{b}$.

---

[7]Many suitable rules for this choice exist - Bland's rule for example.

Lemma 5.3.7 The problem:

$$\text{minimise Tr}\left[C^T X\right] \text{ subject to}$$

$$\text{Tr}\left[A_k^T X\right] = b_k, \ k = 1, \ldots m_1,$$

$$\text{Tr}\left[N_j^T X\right] \leq q_j, \ j = 1, \ldots m_2, \ X \geq 0 \tag{5.17}$$

can be written as a semidefinite programming problem.

Proof. For each inequality $j = 1 \ldots m_2$ define a slack variable z_j. We can rewrite our variable matrix as the block diagonal matrix

$$X' = \begin{pmatrix} X & & & & \\ & z_1 & & & \\ & & \ddots & \\ & & & z_{m_2} \end{pmatrix}. \tag{5.18}$$

Clearly $X' \geq 0$ iff $X \geq 0$ and $z_j \geq 0, \ \forall j$. We can now write the problem as

$$\text{minimise Tr}\left[C'^T X'\right] \text{ subject to}$$

$$\text{Tr}\left[A_k'^T X'\right] = b_k, \ k = 1, \ldots m_1,$$

$$\text{Tr}\left[N_j'^T X'\right] = q_j, \ j = 1, \ldots m_2, \ X' \geq 0 \tag{5.19}$$

where

$$C' = \begin{pmatrix} C & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix} \quad A_k' = \begin{pmatrix} A_k & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix} \quad N_j' = \begin{pmatrix} N_j & & & \\ & \delta_{j1} & & \\ & & \ddots & \\ & & & \delta_{jm_2} \end{pmatrix}.$$

$$\tag{5.20}$$

This is the canonical form of a semidefinite programming problem. $\qquad\square$

Corollary 5.3.8 Every linear program can be written as a semidefinite programming problem.

Proof. One may convert the objective function

$$\text{maximise } \mathbf{c}^T \mathbf{x} = \sum_i c_i x_i = \sum_i C_{ii} X_{ii} = \text{Tr}\left[C^T X\right] \equiv \text{minimise Tr}\left[(-C)^T X\right] \tag{5.21}$$

where C and X are diagonal matrices with $C_{ii} = c_i$, $X_{ii} = x_i$. Each linear programming constraint thus becomes

$$[A]_j \mathbf{x} \leq q_j \rightarrow \text{Tr}[N_j^T X] \leq q_j \tag{5.22}$$

with N_j a diagonal matrix $[N_j]_{ii} = [A]_{ji}$. We than then turn these inequalities into equalities using the slack variables described in lemma 5.17. $\square$

Similar to linear programming, every canonical semidefinite program has a dual form. For a problem of the form (5.16), the dual is given by:

$$\text{maximise } \mathbf{b}^T \mathbf{y} \text{ subject to } \sum_k y_k A_k + S = C, \ S \geq 0. \tag{5.23}$$

You can see that for the dual there are two variables; a semidefinite matrix S, and a real vector $\mathbf{y}$ (sometimes called the multiplier vector).

Lemma 5.3.9 Semidefinite programming problems are *weakly* dual; the solution to the primal X^* and the solution to the dual $\{\mathbf{y}^*, S^*\}$ satisfy

$$\text{Tr}\left[C^T X^*\right] \geq \mathbf{b}^T \mathbf{y}^*. \tag{5.24}$$

Lemma 5.3.10 (Slater's Condition) If both the primal and dual are feasible, and there exists a strictly feasible solution $\tilde{X} > 0$ to the primal or $(\tilde{\mathbf{y}}, \tilde{S})$ with $\tilde{S} > 0$ to the dual, then the optimal values of the primal and dual are achievable and coincide.

Thus in many cases, Slater's condition gives us strong duality - though unlike linear programming it is not guaranteed.

5.3.3.1 Methods for Solving Semidefinite Programming Problems

There does not exists a direct simplex method analog for semidefinite programs; this is because the boundary consists of an infinite number of extremal points (in general). Thus the majority of solvers utilise interior point methods; once again relying on objective-reducing steps converging to the optimal solution. Like linear programming this can be done in polynomial time, often utilising both the primal and dual form of the problem.

5.3.4 Conditional Probability Distributions and Polytopes.

Returning to quantum theory, we have seen how the idea of non-locality is key to what is possible with quantum theory. In particular, we have seen how non-locality can be

verified by use of a Bell inequality. This subsection will explain exactly how this is done in more detail. The first step is to expand on Bell's set-up. Consider a situation in which n parties are spatially separated[8], and each has m measurement choices (often referred to as *inputs*), each of which has k outcomes (referred to as *outputs*). This is commonly called a n-party, m, k Bell scenario. Since we will almost exclusively deal with two-party scenarios in this chapter, we shall use the notation (m, k) to represent a two-party, m measurement k outcome scenario. We will also use the notation (m_A, m_B, k_A, k_B) for situations where the two parties have differing numbers of measurements/outcomes.

In the two party case, we can characterise the situation by the *conditional probabilities* $p(ab|xy)$ - that is, the joint probability of Alice obtaining outcome a given measurement x whilst Bob obtains outcome b from y. Naturally we require that $p(ab|xy) \geq 0$, and that $\sum_{a,b} p(ab|xy) = 1$. We shall express this either as a vector $\boldsymbol{\pi} \in \mathbb{R}^{m^2k^2}$, or as a matrix $\Pi \in \mathbb{R}^{mk \times mk}$ in the following way:

$$
\Pi = \left(
\begin{array}{ccc|cccc|ccc}
p(11|11) & \dots & p(1k_B|11) & \dots & \dots & \dots & \dots & p(11|1m_B) & \dots & p(1k_B|1m_B) \\
\vdots & \ddots & \vdots & \dots & \dots & \dots & \dots & \vdots & \ddots & \vdots \\
p(k_A1|11) & \dots & p(k_Ak_B|11) & \dots & \dots & \dots & \dots & p(k_A1|1m_B) & \dots & p(k_Ak_B|1m_B) \\
\hline
\vdots & \vdots & \vdots & \ddots & & & & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & & \ddots & & & \vdots & \vdots & \vdots \\
\hline
p(11|m_A1) & \dots & p(1k_B|m_A1) & \dots & \dots & \dots & \dots & p(11|m_Am_B) & \dots & p(1k_B|m_Am_B) \\
\vdots & \ddots & \vdots & \dots & \dots & \dots & \dots & \vdots & \ddots & \vdots \\
p(k_A1|m_A1) & \dots & p(k_Ak_B|m_A1) & \dots & \dots & \dots & \dots & p(k_A1|m_Am_B) & \dots & p(k_Ak_B|m_Am_B)
\end{array}
\right).
$$

$$(5.25)$$

These two representations should be thought of as equivalent, and we shall move from one to the other when convenient. We refer to the submatrix of Π corresponding to measurement choices x, y as $\Pi_{|xy}$, and $\Pi(ab|xy)$ as shorthand for $p(ab|xy)$ for a given distribution Π.

These are not the only constraints on our distribution though - it is widely accepted[9] that, although non-local, spatially separated quantum probability distributions are *no-signalling* - it is impossible to transmit information between parties by local operations (including

---

[8]Sufficiently separated to prevent communication.

[9]If one assumes the Born rule, this follows as a consequence.

measurements) - mathematically this forces:

$$\sum_b p(ab|xy) = \sum_b p(ab|xy'), \quad \forall a, x, y, y' \tag{5.26}$$

$$\sum_a p(ab|xy) = \sum_a p(ab|x'y), \quad \forall b, y, x, x' \tag{5.27}$$

You can see this condition means that Alice and Bob cannot learn the other's measurement choice from their local choices and outcomes alone. Although the vector $\boldsymbol{\pi}$ is in the space $\mathbb{R}^{m^2 k^2}$, the no-signalling and normalisation constraint means that the true dimension of the no-signalling space is

$$t = m^2(k-1)^2 + 2m(k-1) \tag{5.28}$$

or

$$t = m_A m_B (k_A - 1)(k_B - 1) + m_A(k_A - 1) + m_B(k_B - 1) \tag{5.29}$$

when $m_A \neq m_B$, $k_A \neq k_B$.

An interesting result is that there exists no-signalling correlations which *cannot* be achieved by quantum mechanics; the most famous example of these is the "PR box" [89] which gives the following correlations, for outcomes $a, b \in \{0, 1\}$ and measurement choices $x, y \in \{0, 1\}$:

$$p(ab|xy) = \begin{cases} \frac{1}{2} & \text{if } a \oplus b = xy \\ 0 & \text{else} \end{cases} \tag{5.30}$$

which allow for a CHSH value of 4. The physical reason behind why quantum distributions are more restricted than general no-signalling distributions is a topic of much research.

Regarding *quantum* distributions, there are a few definitions - some of which are provably *not* equivalent. The first construction is all distributions that can be written

$$p(ab|xy) = \langle \phi | M_{a|x} \otimes M_{b|y} | \phi \rangle \tag{5.31}$$

with $|\phi\rangle$ some quantum state, and $\{M_{a|x}\}$ measurement operators satisfying $M_{a|x} M_{a'|x} = \delta_{aa'}$ and $\sum_a M_{a|x} = \mathbb{I}$ (and the same for $\{M_{b|y}\}$) - this may seem like we are dismissing mixed states and POVMs, but we can always convert these into the above form using purification of mixed states and Naimark's theorem. The above definition is ambiguous as to whether to allow infinite dimensional states or not - it was recently shown in [25] the two scenarios are non-equivalent; however for our purposes this distinction is not important.

We shall refer to the set of correlations of this form as $\mathcal{Q}$.

Alternatively, we may define the set of quantum correlations as those that may be expressed

$$p(ab|xy) = \langle\phi| M_{a|x}M_{b|y} |\phi\rangle \tag{5.32}$$

with $M_{a|x}M_{a'|x} = \delta_{aa'}$ and $\sum_a M_{a|x} = \mathbb{I}$ (and for $\{M_{b|y}\}$ as before) but now we condition $[M_{a|x}, M_{b|y}] = 0$. This definition is equivalent to the description (5.31) for finite dimension [102] but is *not* equivalent [96] for infinite dimensional spaces. We refer to this set of distributions as $\mathcal{Q}'$. This distinction will not play a role in the rest of this chapter; however it is worth outlining.

5.3.4.1 Local Distributions

Within the set of allowable quantum probability distributions, there lie the set of probability distributions which *may* be expressed as the consequence of a hidden variable theory. These may be written

$$p(ab|xy) = \int_\Lambda q\left(\lambda\right) p(a|x,\lambda)p(b|y,\lambda)d\lambda \tag{5.33}$$

where λ may be an arbitrary (or indeed infinite) number of variables with domain Λ, and $q(\lambda)$ an arbitrary probability density function. This may seem extremely general but fortunately there exists another, neater way to characterise them. For a given number of inputs/outputs, any local distribution can be expressed as a convex combination of *deterministic distributions*,

$$\{\mathbf{d}_i\} = \left\{\boldsymbol{\pi} \,\middle|\, p(ab|xy) = \delta_{a,a_x}\delta_{b,b_y}\right\} \tag{5.34}$$

i.e. the correlations for which every measurement choice for Alice and Bob will return a specific outcome with certainty. This type of structure, in which every point in a set can be written as a convex combination of a finite set of points is a *polytope*, as we have seen earlier in the chapter. We denote the local polytope as $\mathcal{L}$, with the V-representation of $\mathcal{L}$ exactly the set $\{\mathbf{d}_i\}$. We can see that the structure in Eq. (5.34) gives that there are $k_A^{m_A}k_B^{m_B}$ vertices in this set, for the (m_A, m_B, k_A, k_B) scenario.

We saw that the other way to represent polytopes is using the *half-plane* representation, or H-representation. In this form we may express $\mathcal{L}$ as

$$\mathcal{L} = \left\{\boldsymbol{\pi} \,\middle|\, \mathbf{b}_i^T\boldsymbol{\pi} \leq c_i \,\forall i\right\} \tag{5.35}$$

i.e. a set of linear inequalities. These inequalities provide a criterion for determining locality of probability distributions - they are *Bell inequalities*. The minimal H-representation is exactly the set of inequalities defining the $(d-1)$-dimensional facets of the polytope. For the local polytope, the dimension is equal to that of the no-signalling space, t; and the facets which define the polytope are known as either *facet Bell inequalities* or *tight Bell inequalities*. As the H-representation is minimal, this set of facet Bell inequalities will be sufficient to determine locality/non-locality. Due to the fact that converting between the two representations is a non-trivial problem, and the rapidly growing dimension t, for most (m, k) scenarios the full list of inequalities is not known.

By contrast, the set of all no-signalling distributions forms the *no-signalling polytope* $\mathcal{NS}$. For this polytope, the facet inequalities are known and are simply the positivity conditions, but the extremal points giving $V_{\mathcal{NS}}$ are generally not known, except for $k = 2$.

The set of quantum distributions does *not* form a polytope - whilst a convex set, there are an infinite number of extremal points. This means the problem of determining whether a probability distribution is quantum achievable is a difficult one in general. However, we shall see there does exist a set of necessary and sufficient conditions we can apply.

5.3.5 The Quantum Hierarchy

We saw earlier in this chapter two equations, (5.31) and (5.32), defining the set of quantum distributions. Unfortunately, given a probability distribution, it is very difficult to determine if there exists states/measurements which achieve that distribution. In particular, the possibility of a high - or even infinite - dimensional state means that this search can be almost impossible. Fortunately, a series of neccessary conditions was introduced in [76], which utilise the branch of *semidefinite programming* optimisation we saw earlier in the chapter. The conditions define an infinite series of sets $\mathcal{Q}_1, \mathcal{Q}_2 \ldots$ which in the limit $\lim_{i \to \infty} \mathcal{Q}_i = \mathcal{Q}'$. Thus if one can determine that $\Pi \in \mathcal{Q}_i \, \forall i$, then $\Pi \in \mathcal{Q}'$.

The idea is as follows: first, suppose there exists a quantum distribution of the form in Eq. (5.32). One could then construct a set of operators $\mathcal{O} = \{O_1 \ldots O_n\}$, each of which is a combination of $M_{a|x}$ and $M_{b|y}$ under addition and multiplication. Using the properties of the $M_{a|x}, M_{b|y}$, one may write all possible linear constraints that the elements $\langle \phi | O_i^\dagger O_j | \phi \rangle$

must satisfy, as a function of $p(ab|xy)$. As a simple example, set $O_1 = M_{a_1|x_1}, O_2 = M_{a_2|x_1}, O_3 = M_{b_1|y_1}$. Then three such constraints would be:

$$\langle\phi|\, O_1^\dagger O_2\, |\phi\rangle = 0 \tag{5.36}$$

$$\langle\phi|\, O_1^\dagger O_3\, |\phi\rangle = p(a_1 b_1 | x_1 y_1) \tag{5.37}$$

$$\langle\phi|\, O_1^\dagger O_3\, |\phi\rangle - \langle\phi|\, O_3^\dagger O_1\, |\phi\rangle = 0 \tag{5.38}$$

and so on. The set of all linearly independent constraints of this form can be written

$$F(\mathcal{O}) = \left\{ \sum_{i,j} [F_k]_{ij} \langle\phi|\, O_i^\dagger O_j\, |\phi\rangle = g_k(\Pi) \right\}_k. \tag{5.39}$$

Consider now the question: does there exist a matrix $\Gamma \geq 0$ such that

$$\sum_{i,j} [F_k]_{ij} \Gamma_{ij} = g_k(\Pi),\ \forall k \tag{5.40}$$

are satisfied? If there exists the quantum construction of the form (5.31), then such a matrix *must* exist: we simply set $\Gamma_{ij} = \langle\phi|\, O_i^\dagger O_j\, |\phi\rangle$. This does not preclude however other Γ existing satisfying these conditions. The semidefinite positivity comes from the condition that

$$\mathbf{v}^\dagger \Gamma \mathbf{v} = \sum_{i,j} v_i^* \langle\phi|\, O_i^\dagger O_j\, |\phi\rangle\, v_j = \langle\phi|\, \sum_i v_i^* O_i^\dagger \sum_j v_j O_j\, |\phi\rangle = \langle\phi|\, V^\dagger V\, |\phi\rangle \geq 0. \tag{5.41}$$

5.3.5.1 A Canonical Set of Operators

Given this necessary condition is dependent on choosing a set of operators, it is clear the the strictness of this condition will depend on the set chosen. Fortunately for us; the authors of [76] constructed a family of sets $\mathcal{S}_n$, with the strictness increasing to define the exact quantum set in the limit $n \to \infty$.

Definition 5.3.11 For a given scenario (m_A, m_B, k_A, k_B), the set $\mathcal{S}_n$ is defined as all operators consisting of all non-equivalent products of length $\leq n$ of $M_{a|x}, M_{b|y}$, $a \in \{1\ldots(k_A - 1)\}, b \in \{1\ldots(k_B - 1)\}, x \in \{1\ldots m_A\}, y \in \{1\ldots m_B\}$, excluding null operators. For example

$$\mathcal{S}_0 = \{\mathbb{I}\} \tag{5.42}$$

$$\mathcal{S}_1 = \mathcal{S}_0 \cup \{M_{a|x}\} \cup \{M_{b|y}\} \tag{5.43}$$

$$\mathcal{S}_2 = \mathcal{S}_0 \cup \mathcal{S}_1 \cup \{M_{a|x} M_{a'|x'}\} \cup \{M_{b|y} M_{b'|y'}\} \cup \{M_{a|x} M_{b|y}\} \tag{5.44}$$

Satisfaction of $\sum_{i,j}[F(\mathcal{S}_n)_k]_{ij}\Gamma_{ij} = g_k(\Pi)$, $\forall k$ by some matrix Γ means that $\Pi \in \mathcal{Q}_n$, level n of the quantum hierarchy.

The final outcome measurement operators are omitted as they are expressible as $M_{k_A|x} = \mathbb{I} - \sum_{a=1}^{(k_A-1)} M_{a|x}$ (and similarly for $M_{k_B|y}$) and so do not contribute any new constraints. Similarly sequences $M_{b|y}M_{a|x}$ are omitted as the equivalent operator $M_{a|x}M_{b|y}$ may be used. Finally, null operators (such as $O_i = M_{1|1}M_{2|1}$) are omitted as they only define the trivial relation $0 = 0$. These sets give a series of necessary conditions for a given distribution $\Pi \in \mathcal{Q}'$ - there must exist a $|\mathcal{S}_n| \times |\mathcal{S}_n|$ matrix Γ satisfying all linear conditions $F(\mathcal{S}_n)$, which depend on Π.

A Simple Example: $m = k = 2$.

Consider the simplest example, the $(2,2,2,2)$ scenario at level 1. The matrix Γ must be of the form

O	$\mathbb{I}$	$M_{1|x_1}$	$M_{1|x_2}$	$M_{1|y_1}$	$M_{1|y_2}$
$\mathbb{I}$	1	$p(1|x_1)$	$p(1|x_2)$	$p(1|y_1)$	$p(1|y_2)$
$M_{1|x_1}$		$p(1|x_1)$	α	$p(11|x_1y_1)$	$p(11|x_1y_2)$
$M_{1|x_2}$			$p(1|x_2)$	$p(11|x_2y_1)$	$p(11|x_2y_2)$
$M_{1|y_1}$				$p(1|y_1)$	β
$M_{1|y_2}$					$p(1|y_2)$

where we have omitted the lower quadrant as Γ is Hermitian. The greek letters are free variables (satisfying $\Gamma \geq 0$).

5.3.5.2 Bell Inequalities and Bell Inequality Classes

Suppose we have a Bell inequality $\mathbf{b}_i^T \boldsymbol{\pi} \leq c_i$ - is its representation unique? Clearly we may multiply both sides by a constant, and it remains the same inequality. As with our conditional probability distributions, we may also equivalently express them in matrix form, $\text{Tr}[B_i^T \Pi] \leq c_i$. More importantly though, the no-signalling constraints on our probabilities mean that there remain other, non-trivial invariance operations we may perform. To illustrate this, we will take a particular $(2,2)$ Bell inequality,

$$B = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ \hline 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \quad \text{Tr}\left[B^T \Pi\right] \geq 1. \tag{5.45}$$

Taking the above matrix and applying our knowledge that, for no-signalling distributions, $p(11|11) + p(12|11) = p(11|12) + p(12|12)$, we see that the plane defined by

$$B' = \left(\begin{array}{cc|cc} -1 & 0 & 1 & 2 \\ 1 & 0 & 1 & 0 \\ \hline 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{array} \right), \quad \mathrm{Tr}\left[B'^T \Pi\right] \geq 1 \tag{5.46}$$

within the no-signalling space is entirely equivalent.

We can also use the normalisation condition to define different representations of the same inequality. We know that $p(11|11)+p(12|11)+p(21|11)+p(22|11) = 1$, and therefore can define

$$B'' = \left(\begin{array}{cc|cc} -1 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 \\ \hline 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{array} \right), \quad \mathrm{Tr}\left[B''^T \Pi\right] \geq 0 \tag{5.47}$$

and this too is a representation of the same inequality. Indeed, any transformation $kB+s$, $k \neq 0$ will preserve an inequality. In fact, the three matrices above are all representations of the CHSH inequality in Eq. (5.1), where the expectations have been converted to probabilities, and then such transforms as seen above applied. When referring to a Bell inequality, we are implicitly referring to all such representations. To check whether two matrices B_1, B_2 represent the same Bell inequality, we can evaluate them both at the extremal points of the local polytope, to obtain vectors $\mathbf{c}_1, \mathbf{c}_2$. If there exists a transformation $k\mathbf{c}_1 + s = \mathbf{c}_2$, we can conclude they are representations of the same inequality.

One final point we must consider is that of "nomenclature" - suppose we took the Bell inequality of the form B, and relabelled Bob's measurements, such that measurement 1 is now referred to as measurement 2 and vice versa. We end up with B_R, expressible as:

$$B_R = \left(\begin{array}{cc|cc} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ \hline 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{array} \right), \quad \mathrm{Tr}\left[B_R^T \Pi\right] \geq 1. \tag{5.48}$$

Clearly this is still *a* CHSH inequality, but it does *not* correspond to the same inequality - it defines a different facet of the local polytope. Thus it is vital to distinguish between

inequalities and *inequality classes* - the former refers to unique facets of the local polytope, whilst the latter accounts for all facets that are unique up to relabelling of inputs, outputs and parties. For the specific $(2, 2)$ scenario, there are two Bell inequality classes: the CHSH class consisting of 8 inequalities, and the trivial class, consisting of 16 inequalities of the form $p(ab|xy) \geq 0$.

For a general (m_A, m_B, k_A, k_B) scenario, the possibilities for relabelling grow rapidly. For measurements we may choose any permutation from the symmetry group S_{m_A} and S_{m_B} respectively, whilst for outcomes we may choose a permutation from S_{k_A} and S_{k_B} for *each* measurement of Alice and Bob respectively. This means there are $(k_A!)^{m_A} (k_B!)^{m_B} m_A! m_B!$ possible relabellings, or double this if $m_A = m_B$ and we may swap parties (expressed as transposition in the matrix representation). Despite this, relabelled inequalities are not of much interest to us. This is because any relabelling between inequalities can also be applied to probability distributions, meaning all inequalities in the same class will retain the same properties.

For this reason, when discussing characterising a (m_A, m_B, k_A, k_B) scenario, we are really interested in the full set of inequality *classes* - using these we can generate all the inequalities by running through all relabellings for a representative of each class. It is worth noting though that the size of each class is not an obvious result, since a relabelling may result in the same inequality - we shall discuss how to combat this later in the chapter.

Since the number of extremal local distributions is $n = k_A^{m_A} k_B^{m_B}$ for the (m_A, m_B, k_A, k_B) scenario, we can see that the the problem of facet enumeration for the local polytope scales badly, especially with the number of inputs. One potential loophole to this is that we are only interested in the *inequality classes* - perhaps there is a technique allowing for the quicker generation of just representatives of each inequality class, than of the full list of facets. Some approaches to try and achieve this are detailed in this chapter.

5.3.5.3   The No-Signalling Polytope

As mentioned before, the set of no-signalling distributions is a polytope, which we denote $\mathcal{NS}$. The facet-defining inequalities for this polytope are simple - they are $m_A k_A m_B k_B$ inequalities of the form $p(ab|xy) \geq 0$, one for each possible choice of a, b, x, y.

Generally, the extremal points of $\mathcal{NS}$ are not known, except when the two parties are limited to either 2 measurements or 2 outcomes - we shall focus on the latter. Before we look at those, it is worth noting that the extremal local distributions introduced in section 5.3.4.1 are also extremal no-signalling distributions (and thus we always know this subset). This is important for our picture of the no-signalling space, in which the local subset is contained within the no-signalling one, with the extremal local points being extremal for $\mathcal{NS}$, whilst the facets of $\mathcal{NS}$ are always facets of the local polytope, $\mathcal{L}$.

For the case when $m = k = 2$, all non-local extremal no-signalling points are of the form

$$
\Pi_{PR} = \begin{pmatrix}
\frac{1}{2} & 0 & \frac{1}{2} & 0 \\
0 & \frac{1}{2} & 0 & \frac{1}{2} \\
\frac{1}{2} & 0 & 0 & \frac{1}{2} \\
0 & \frac{1}{2} & \frac{1}{2} & 0
\end{pmatrix}
\tag{5.49}
$$

up to relabellings; these are the *PR boxes* [89], which achieve the maximal no-signalling CHSH value. There are 8 such distributions.

When $k = 2$, the non-local distributions of $\mathcal{NS}$ take the form[10] [65]:

$$
\begin{pmatrix}
S & S & S & \dots & S & L & \dots & L \\
S & A & S/A & \dots & S/A & L & \dots & L \\
S & S/A & S/A & \dots & S/A & L & \dots & L \\
\vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
S & S/A & S/A & \dots & S/A & L & \dots & L \\
K & K & K & \dots & K & M & \dots & M \\
\vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
K & K & K & \dots & K & M & \dots & M
\end{pmatrix}
\tag{5.50}
$$

with the following 2×2 blocks:

$$
S = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \quad
A = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix} \quad
K = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 0 \end{pmatrix} \quad
L = \begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 0 \end{pmatrix} \quad
M = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.
$$

---

[10]Up to relabelling.

For the scenario where Alice has m_A inputs, and Bob has m_B inputs, we must consider all the $2m_A$ by $2m_B$ matrices of the form expressed in Eq. (5.50) - whilst the upper-leftmost $\begin{pmatrix} S & S \\ S & A \end{pmatrix}$ are always fixed, the blocks of outcomes K, L and M may then range from 0 to $m_A - 2$ or $m_B - 2$ occurrences respectively.

5.4 Completely Known Scenarios

For some (m_A, m_B, k_A, k_B) scenarios, all Bell inequality classes are known. We shall present these below.

- $(\mathbf{2}, \mathbf{2}, \mathbf{2}, \mathbf{2})$

  The simplest scenario is which non-locality can be obtained, there are just two classes of facet Bell inequality in this scenario - the trivial class of positivity inequalities which is of size 16, and the CHSH class consisting of 8 inequalities of the form:

$$B_{\text{CHSH}} = \left(\begin{array}{cc|cc} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ \hline 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{array} \right), \quad \text{Tr}\left[B_{\text{CHSH}}^T \Pi \right] \geq 1 \tag{5.51}$$

  and relabellings of this. From this point on in the chapter, we shall <u>omit</u> writing the $\text{Tr}\left[B^T \Pi \right] \geq 1$ - this bound should be assumed unless otherwise stated.

- $(\mathbf{2}, \mathbf{m}, \mathbf{2}, \mathbf{k})$

  In this scenario Alice is limited to just two inputs and two outputs, whilst Bob may perform any $m \geq 2$ measurements each with $k \geq 2$ outcomes - despite this freedom, all these scenarios retain just two inequality classes, again positivity and CHSH. This scenario is illuminating in how lower dimensional facet inequalities *remain* facet inequalities in higher dimensional scenarios; take for example the $(2, 3, 2, 3)$ CHSH inequality

$$B_{\text{CHSH}_3} = \left(\begin{array}{ccc|ccc|ccc} 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \end{array} \right). \tag{5.52}$$

Notice how there are no coefficients for the additional measurement - it is as if Bob "never chooses" this measurement. For the remaining measurements, we see the coefficients for the third outcome are an exact copy of the second outcome coefficients; Bob is treating these outcomes exactly the same. Thus the above inequality reduces to the $(2, 2, 2, 2)$ CHSH inequality, by ignoring the extra measurement and treating outcome 3 as outcome 2 for each of Bob's measurements, yet the B_{CHSH_3} inequality - and all possible relabellings - are facets of the $(2, 3, 2, 3)$ local polytope. In this chapter, we shall refer the the process of adding measurement outcomes as "lifting" - this term is also used when adding an extra measurement; but we shall explicitly state this when performed.

As facets of lower dimensional scenarios are also facets of higher dimensions, there will be CHSH facets for all scenarios. Moreover, since there are more relabellings in higher dimensions, the size of an inequality class will always increase.

- $(\mathbf{3}, \mathbf{3}, \mathbf{2}, \mathbf{2})$

  This is the first scenario in which a non-trivial, non-CHSH inequality is introduced. Often referred to as I_{3322}, it can be written

$$
I_{3322} = \left(
\begin{array}{cc|cc|cc}
0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 \\
\hline
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
\hline
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1
\end{array}
\right).
\tag{5.53}
$$

  Already this inequality showcases the complexity of quantum theory; not only is the Tsirelson bound not known for this inequality, it is conjectured that to achieve the bound infinite dimensional quantum states are required [78]. This is perhaps the second best understood Bell inequality, yet already our knowledge is limited.

- $(\mathbf{2}, \mathbf{2}, \mathbf{3}, \mathbf{3})$

  Another relatively simple scenario, in which only a single extra output is added.

153

Scenario	Number of Inequality Classes	Number of Facets	Reference
$(2,2,2,2)$	2	24	[23, 42]
$(2,m,2,k)$	2	$2(2^m - 2)(k^2 - 2) + 4km$	[87]
$(3,3,2,2)$	3	684	[27]
$(2,2,3,3)$	3	1116	[27, 28, 66]
$(3,4,2,2)$	6	12,480	[27]
$(3,5,2,2)$	7	71,340	[34]
$(4,4,2,2)$	175	36,391,264	[34]

Table 5.1: A list of the solved scenarios; included in this list are the trivial positivity facets. For the final two a complete list of class representatives was not given, and we do this here.

This too provides a new inequality class, I_{2233}:

$$
I_{2233} = \left(
\begin{array}{ccc|ccc}
1 & \frac{1}{2} & 0 & 1 & \frac{1}{2} & 0 \\
\frac{1}{2} & 0 & 1 & 0 & 1 & \frac{1}{2} \\
0 & 1 & \frac{1}{2} & \frac{1}{2} & 0 & 1 \\
\hline
1 & 0 & \frac{1}{2} & 1 & 0 & \frac{1}{2} \\
\frac{1}{2} & 1 & 0 & 0 & \frac{1}{2} & 1 \\
0 & \frac{1}{2} & 1 & \frac{1}{2} & 1 & 0
\end{array}
\right)
\tag{5.54}
$$

with the other two classes the trivial class and CHSH.

- $(\mathbf{3}, \mathbf{4}, \mathbf{2}, \mathbf{2})$

  There are 6 inequality classes for this scenario, which were presented in [27].

- **Partially known cases**

  For the scenarios $(3,5,2,2)$ and $(4,4,2,2)$ the number of both facets and classes are known, thanks to an isomorphism of two-outcome local polytopes to a class of geometrically interesting "cut polytopes" [33]. The cut polytopes corresponding to these scenarios were solved in [34], and we enumerate a full list of representatives of each class for those local polytopes in this chapter.

For all other bipartite scenarios, the complete lists of Bell inequality classes are *not known* - specific families of Bell inequalities, such as I_{mm22} [27] have been constructed,

but as figure 5.1 shows, the number of classes jumps dramatically, so much so that these constructed classes are but a drop in the ocean.

5.5 An Algorithm for Generating Bell Inequalities

5.5.1 Quantum Application of Linear Programming

Given an arbitrary no-signalling distribution, we would like some measure of "locality" more informative than simply membership or not to the local polytope. One way to give this locality could be using Bell inequalities; however, this raises this issue of which Bell inequality to use, as in higher dimensions the inequality required may not even be known. By contrast, we always know the local deterministic strategies, which correspond to the extremal points of the local polytope; and moreover are a subset of the extremal no-signalling (NS) points. Since our no-signalling space is a convex polytope, we may write any no-signalling distribution as a (not unique) convex combination of extremal points. We can therefore calculate the decomposition which maximises the weight of the local extremal distributions - this problem can be formulated as a linear program, and works even when the extremal non-local no-signalling points are not known.

Definition 5.5.1 ([113]) We define the *local weight* of a no-signalling distribution $\mathbf{q}$ as the solution to the problem:

$$\text{maximise } \sum_i x_i \text{ subject to: } \sum_i x_i \mathbf{d}_i \leq \mathbf{q}, \ x_i \geq 0. \tag{5.55}$$

By defining $\mathbf{c} = \mathbf{1}$ and A as the matrix $(\mathbf{d}_1 \ldots \mathbf{d}_n)$, we can express this as a canonical primal linear programming problem.

If $\mathbf{q}$ is a local distribution, then the maximal value will be 1; similarly if we take an extremal (non-local) NS point, such as a PR box, then the value of this program will be 0. In general, we are able to split a NS distribution $\mathbf{q}$ into two sub-normalised distributions; $A\mathbf{x}^*$, the local distribution which optimises $\mathbf{c}^T\mathbf{x}$, and $(\mathbf{q} - A\mathbf{x}^*)$, a NS distribution with no local part.

Corollary 5.5.2 If $\mathbf{q}$ is non-local, i.e. $\mathbf{c}^T\mathbf{x}^* < 1$, then the solution $\mathbf{y}^*$ to the dual is a Bell inequality.

Proof. Suppose we have that the local weight $\mathbf{c}^T\mathbf{x}^* = k < 1$. The dual problem of the one seen in definition 5.5.1 is

$$\text{minimise } \mathbf{q}^T \mathbf{y} \text{ subject to: } A^T \mathbf{y} \geq \mathbf{1}, \ \mathbf{y} \geq \mathbf{0}. \tag{5.56}$$

By strong duality, we have that $\mathbf{q}^T \mathbf{y}^* = k$. We also have that $\sum_j A_{ji} y_j \geq 1$, $\forall i$. This means the product of $\mathbf{y}$ with each column of A, $\mathbf{d}_i$, is always at least 1. Thus, we have a Bell inequality $\mathbf{y}$ with local bound 1, violated by $\mathbf{q}$, with all positive entries. $\square$

We would like to exploit this result in order to generate *all* facet Bell inequalities for a given scenario. In order to do this, we need to choose suitable $\mathbf{q}$, and be sure it is possible for every facet to be a possible solution for a dual problem of the above form.

For the choice of $\mathbf{q}$, we look at the complementary slackness condition (theorem 5.3.6), which gives us that:

$$(\mathbf{q} - A\mathbf{x}^*)^T \mathbf{y}^* = 0, \tag{5.57}$$

and

$$\left(A^T \mathbf{y}^* - \mathbf{c}\right)^T \mathbf{x}^* = 0. \tag{5.58}$$

The first condition, Eq. (5.57), tells us that the optimum Bell inequality $\mathbf{y}^*$ achieves value 0 at the non-local part of $\mathbf{q}$, $(\mathbf{q} - A\mathbf{x}^*)$. Note that this is the minimal possible value, since all elements of both $\mathbf{y}^*$ and $\mathbf{q} - A\mathbf{x}^*$ are non-negative. Furthermore, this implies each extremal NS point in the decomposition of $\mathbf{q} - A\mathbf{x}^*$ must achieve value 0.

The second condition, Eq. (5.58), implies that for all i either x_i^* or $\left(A^T \mathbf{y}^* - \mathbf{c}\right)_i^T$ is zero, as both values are non-negative. x_i^* is non-zero iff the local distribution $\mathbf{d}_i$ is in the local weight of $\mathbf{q}$ - for these points we are forced to conclude that $\mathbf{d}_i \mathbf{y}^* = A_i \mathbf{y}^* = c_i = 1$ i.e. that the local part of $\mathbf{q}$ saturates the local bound of $\mathbf{y}^*$.

This suggest a suitable $\mathbf{q}$ may be an extremal no-signalling distribution, perhaps in a convex combination with low-weighted local points. The following results support this choice, by providing good evidence we will be able to generate every facet.

Theorem 5.5.3 For every (violatable) facet Bell inequality with non-negative entries and local bound 1, there exists an extremal NS point such that the value of the Bell inequality at that point is 0.

Proof. Let $\boldsymbol{\pi}^T \mathbf{y}^* \geq 1$ be a (violatable) facet Bell inequality. Let us pick a $\mathbf{q}$ such that $\mathbf{q}^T \mathbf{y}^* < 1$ but $\mathbf{q}^T \mathbf{b} \geq 1$ for all other facet Bell inequalities $\mathbf{b} \neq \mathbf{y}^*$. This is always possible since $\mathbf{y}^*$ is in the minimal H-representation of the local polytope. Therefore $\mathbf{y}^*$ is the optimum solution to the dual problem and so the complementary slackness theorem gives $\mathbf{q}_{\mathrm{ns}}^T \mathbf{y}^* = 0$ where $\mathbf{q}_{\mathrm{ns}} = (\mathbf{q} - A\mathbf{x}^*)$, the non-local part of $\mathbf{q}$. By convexity we can write $\mathbf{q}_{\mathrm{ns}}$ as a sum of extremal no-signalling distributions $\{\mathbf{n}_k\}$, each of which much satisfy $\mathbf{n}_k^T \mathbf{y}^* = 0$ by linearity. $\qquad\square$

The above theorem assumes non-negative entries and a local bound 1; however we have the following lemma.

Lemma 5.5.4 All Bell inequalities can be written with non-negative coefficients and local bound 1.

Proof. Suppose our inequality has the form[11] $\mathbf{b}^T \boldsymbol{\pi} \geq k \neq 1$. Then we may transform $\mathbf{b}' = \mathbf{b} + (1 - k) / (m_A m_B) \mathbf{1}$ - this satisfies $\mathbf{b}'^T \boldsymbol{\pi} \geq 1$. Now suppose $\mathbf{b}'$ has negative coefficients, the most negative of which is $-\alpha$. By adding α uniformly to each coefficient, we obtain an inequality $\tilde{\mathbf{b}}$ with local bound $1 + \alpha (m_A m_B) > 1$. Finally we can divide each coefficient of $\tilde{\mathbf{b}}$ by $1 + \alpha (m_A m_B)$, to obtain an inequality $\hat{\mathbf{b}}$ with all non-negative entries, and local bound 1. If $k = 1$, we may skip the first transformation. $\qquad\square$

Based on these results, we apply the following algorithm. We first present the algorithm in the section below, and then break down and explain the steps and choices made. There are two different versions presented, which will be justified.

5.5.1.1 A Linear Programming Algorithm for Bell Inequalities

We begin by choosing a scenario (m_A, m_B, k_A, k_B) we wish to find Bell inequalities for - to take advantage of the no-signalling points, $k_A = k_B$ must be set to 2. Also chosen is a noise value η, with a small positive value.

<div align="center">

Algorithm 1 (**Algorithm 2**)

</div>

1. The first step of the algorithm is to generate the subset of extremal no-signalling distributions of the form (5.50), with only S and A components. We label this set

---

[11]We shall use $\mathbf{b}^T \boldsymbol{\pi}$ and $\boldsymbol{\pi}^T \mathbf{b}$ interchangeably as context requires.

$\tilde{V}_{\mathcal{NS}}$. Also generated is $A = \{\mathbf{d}_1 \ldots \mathbf{d}_n\}$, and the polytope dimension $t = (m_A(k_A - 1) + 1)(m_B(k_B - 1) + 1) - 1$.

(1b) The set of permutations *on the set of deterministic local points* under all possible relabellings, Ω, is generated.

2. A no-signalling point $\mathbf{q}$ is chosen from $\tilde{V}_{\mathcal{NS}}$. The dual problem

$$\text{minimise } \mathbf{q}^T\mathbf{y} \text{ subject to: } A^T\mathbf{y} \geq \mathbf{1}, \ \mathbf{y} \geq \mathbf{0} \tag{5.59}$$

is then solved using the simplex method, giving optimal solution $\mathbf{y}^* = \mathbf{b}$.

3. The vector $\mathbf{c_b} = A^T\mathbf{b}$ gives the value of $\mathbf{b}$ at each local deterministic point. We affinely fix this vector (explained in Section 5.5.6) to obtain vector $\hat{\mathbf{c}}_\mathbf{b}$.

4. The submatrix $A_\mathbf{b}$ is formed with columns $\{\mathbf{d}_i\}$ such that $A_\mathbf{b}^T\mathbf{b} = \mathbf{1}$.

5. The matrix rank of $A_\mathbf{b}$ is calculated. If $\text{rank}[A_\mathbf{b}] = t$, the dimension of the local polytope, then $\mathbf{b}$ is added to our list of facet inequalities, and $\hat{\mathbf{c}}_\mathbf{b}$ is stored with it. Also stored is the tally of $\hat{\mathbf{c}}_\mathbf{b}$ elements, $\mathcal{T}_\mathbf{b}$.

6. Two columns, $\mathbf{d}_\alpha, \mathbf{d}_\beta$ are chosen from $A_\mathbf{b}$, and a new "noisy" distribution

$$\mathbf{q}' = \left(1 - \frac{3\eta}{2}\right)\mathbf{q} + \eta\mathbf{d}_\alpha + \frac{\eta}{2}\mathbf{d}_\beta \tag{5.60}$$

is formed. The problem

$$\text{minimise } \mathbf{q}'^T\mathbf{y} \text{ subject to: } A^T\mathbf{y} \geq \mathbf{1}, \ \mathbf{y} \geq \mathbf{0} \tag{5.61}$$

is then solved to give optimum $\mathbf{y}^* = \mathbf{b}'$.

7. The vector $\mathbf{c}_{\mathbf{b}'} = A^T\mathbf{b}'$ is calculated, and affinely fixed to give $\hat{\mathbf{c}}_{\mathbf{b}'}$.

8. The submatrix $A_{\mathbf{b}'}$ is formed with columns $\{\mathbf{d}_i\}$ such that $A_{\mathbf{b}'}^T\mathbf{b}' = \mathbf{1}$.

9. The matrix rank of $A_{\mathbf{b}'}$ is calculated. If $\text{rank}[A_{\mathbf{b}'}] = t$, then the tally $\mathcal{T}_{\mathbf{b}'}$ of $\hat{\mathbf{c}}_{\mathbf{b}'}$ elements is checked against all stored tallies.

10. If $\mathcal{T}_{\mathbf{b}'}$ is different to all stored tallies, then $\mathbf{b}'$ is added to the facet inequality list. If it is equal, then it is discarded.

158

(10b) If $\mathcal{T}_{\mathbf{b}'}$ is found equal to a stored tally, we permute $\hat{\mathbf{c}}_{\mathbf{b}'}$ from Step 7 under every permutation $\omega \in \Omega$. If it matches a previously stored vector, then we may conclude they are inequalities belonging to the same class, and discard $\mathbf{b}'$. If not, $\mathbf{b}'$ is a new facet inequality class and $\mathbf{b}'$, $\hat{\mathbf{c}}_{\mathbf{b}'}$ and $\mathcal{T}_{\mathbf{b}'}$ are stored.

11. Steps 6-10 (6-10b) are then repeated for all possible pairs $\mathbf{d}_\alpha, \mathbf{d}_\beta$ from $A_{\mathbf{b}}$.

12. Steps 2 to 12 are then repeated for a new $\mathbf{q}$ from $\tilde{V}_{\mathcal{NS}}$, until all no-signalling points have been checked. From here on, Step 5 is also checked in the manner of Steps 9-10 (9-10b).

The simplex algorithm we use is deterministic - although there are many possible optima due to the degeneracy of the problem, rerunning the algorithm will output the same solution. Therefore in order to increase our list of inequalities we employ the technique that, once the full algorithm above has been run, we can change the value of the noise parameter in order to obtain different solutions.

5.5.2 Choice of the Simplex Algorithm

We already saw in corollary 5.5.2 how a non-local $\mathbf{q}$ will mean the solution to the dual problem will be a Bell inequality - so why do we specify a particular solution method? The reasoning is that we are looking to enumerate the *facet* Bell inequality classes of a given scenario. In a linear programming problem, optimisation takes place over a (possible unbounded) polytope - and the dual problem takes place over a corresponding dual polytope. We saw how, for the polar dual, the extremal points of the dual polytope corresponded to the facets of the primal polytope. For the above linear program the primal polytope is the local polytope; the dual polytope given by our dual problem is not the polar duel, but nevertheless many of its extremal points still correspond to facet inequalities of the local polytope. The simplex algorithm explores these extremal points, and thus our solution will often result in a facet inequality. If we were to use the interior point method however; we could end up with a solution corresponding to a convex combination of facet inequalities; from which we would be unable to retrieve the facets. This problem is exacerbated by the fact that an extremal no-signalling point often violates (even maximally) many different inequalities.

159

5.5.2.1 Possible Solutions for the Linear Program

Given the maximum principle (lemma 5.3.4) states that the optimal solution of the linear program will be acheived at an extremal point of the solution space polytope, we wish to know what the extremal points of problem (5.56) are, and are they the facet inequalities?

Due to the computational requirements, we are only able to enumerate these polytopes for two scenarios[12] $(2,2,2,2)$ and $(3,3,2,2)$, and so will have to test our predictions on these scenarios. It should be noted however, that these are not in general representative of $(m_A, m_B, 2, 2)$ scenarios, due to the low number of inequality classes.

Performing vertex enumeration of the conditions $\mathbf{d}_i^T \mathbf{y} \geq 1$, $\mathbf{y} \geq \mathbf{0}$ we find the following polytopes:

- $(\mathbf{2, 2, 2, 2})$

    The extremal points of the dual solution polytope are given in table 5.2. We see this is an unbounded polytope, although the rays correspond to non-violatable Bell inequalities. We also see there exist normalisation condition vertices, but the solution will occur at a CHSH vertex - of which there are many redundant vertices corresponding to the same inequality.

Type	#	Inequality Class	Additional Info
Ray	16	Positivity conditions	Each is of the form $y_i = \delta_{ik}, k \in 1\ldots 16$.
Vertex	12	Normalisation condition	Each of these satisfies $\mathbf{y}^T \mathbf{q} = 1$, $\mathbf{q} \in \mathcal{NS}$.
Vertex	16	Positivity conditions	
Vertex	104	CHSH	

Table 5.2: The linear programming variable search space for $(2,2,2,2)$ - note that the space is unbounded, although the problem itself is not.

Although these new and redundant vertices are not ideal, the algorithm does not enumerate the full polytope, and our solution will necessarily be a CHSH (and

---

[12]We may also enumerate $(2,3,2,2)$ and $(2,2,3,3)$ - The first scenario does not provide any new classes, whilst we do not know the extremal no-signalling points for the second.

therefore facet) inequality, since the normalisation and positivity conditions are non-violatable.

- $(\mathbf{3}, \mathbf{3}, \mathbf{2}, \mathbf{2})$

  The extremal points for this scenario are given in table 5.3. We see the number of vertices grows enormously, and importantly we see the addition new "non-facet" vertices, in comparison to the $(2, 2, 2, 2)$ case. These correspond to Bell inequalities defining lower dimensional faces; and are expressible as convex combinations of the facet inequalities. They appear as vertices due to the additional constraints that $y_i \geq 0$.

Type	#	Inequality Class	Additional Info
Ray	36	Positivity conditions	Of the form $y_i = \delta_{ik}, k \in 1 \dots 36$.
Vertex	45	Normalisation condition	
Vertex	144	Positivity conditions	
Vertex	2952	CHSH	
Vertex	248832	I_{3322}	
Vertex	442176	Non-facet inequalities	These vertices correspond to lower dimensional faces of $\mathcal{L}$.

Table 5.3: The linear programming variable search space for $(3, 3, 2, 2)$ - we see here the introduction of extremal points which correspond to non-facet Bell inequalities.

Our constraints that $\mathbf{d}_i^T \mathbf{y} \geq 1$ are intuitive, but the inequalities $y_i \geq 0$ appear simply from the canonical form of the linear program. One could instead consider only the space defined by $\mathbf{d}_i^T \mathbf{y} \geq 1, \ \forall i$. The extremal points of these polytopes are given in table 5.4 and table 5.5 for the $(2, 2, 2, 2)$ and $(3, 3, 2, 2)$ scenarios respectively.

Type	#	Inequality Class	Additional Info
Vertex	1	Normalisation	Our lone vertex is simply the condition $\mathbf{y}^T \mathbf{q} = 1$.
Ray	16	Positivity conditions	
Ray	8	CHSH	

Table 5.4: The alternative $(2, 2, 2, 2)$ search space, allowing negative values.

We see that these solution space look very promising - except for a single vertex, every ray corresponds to a facet inequality, with no degeneracy. We may allow negative values

Type	#	Inequality Class	Additional Info
Vertex	1	Normalisation	Our lone vertex is again the condition $\mathbf{y}^T\mathbf{q} = 1$.
Ray	36	Positivity conditions	
Ray	72	CHSH	
Ray	576	I_{3322}	

Table 5.5: The alternative $(3, 3, 2, 2)$ search space, allowing negative values; we see no non-facet vertices, but our facet vertices have become rays.

of $\mathbf{y}$ in linear programming by defining for each y_k two new variables $y_k = y_k^+ - y_k^-$. We then replace our conditions:

$$\mathbf{d}_i^T\mathbf{y} \geq 1, \ \forall i \qquad \rightarrow \qquad \mathbf{d}_i^T\mathbf{y}_+ - \mathbf{d}_i^T\mathbf{y}_- \geq 1, \ \forall i,$$

$$\mathbf{y} \geq \mathbf{0} \qquad \rightarrow \qquad \mathbf{y}^+ \geq \mathbf{0}, \mathbf{y}^- \geq \mathbf{0}.$$

Unfortunately, this new linear programming problem, with objective function $\mathbf{q}^T\mathbf{y}^+ - \mathbf{q}^T\mathbf{y}^-$ is *unbounded*, as we can no longer bound our objective function by 0 from below. Therefore, this solution space cannot be used for our linear program.

We have seen earlier in the chapter there already exists a bounded polytope, whose vertices are *exactly* the facet Bell inequalities; the polar dual of the local polytope. Although $\mathbf{0} \notin \mathcal{L}$, we can shift the origin to account for this. The natural choice for this is the *uniform distribution*, in which every outcome has equal probability - this distribution is also an equal mixture of every local extremal point. Once more, we use $\mathbf{y} = \mathbf{y}^+ - \mathbf{y}^-$; since the polar dual is bounded we will not obtain an unbounded problem. Despite this, the polar dual is *not* a good choice, as we shall see in the results section.

5.5.3 Omission of the K, L, M components

Choosing no-signalling points of the form seen in Eq. (5.50) already proffers an advantage when searching for Bell inequality classes - all violatable inequalities can be relabelled appropriately to create a class member violated by a no-signalling point of this form. By limiting the choice of $\mathbf{q}$ to these points, we do not remove any classes, but remove many redundant relabelled inequalities. This form does not completely remove all equivalences of no-signalling points under relabelling - this extra pre-processing step could be done to

cut down on redundant inequalities further. One thing we can do with certainty though, is to cut out all no-signalling points with K, L and M components. We prove this in the following lemma:

Lemma 5.5.5 For every positive valued, local bound 1 facet $\mathbf{b}$ of the $(m_A, m_B, 2, 2)$ local polytope, there exists a facet $\hat{\mathbf{b}}$ in the same inequality class, and an extremal no-signalling distribution $\mathbf{q}$ of the form in Eq. (5.50) whose blocks consist solely of S and A, such that $\hat{\mathbf{b}}^T \mathbf{q} = 0$.

Proof. By lemma 5.5.3 every facet inequality has at least one $\boldsymbol{\pi} \in V_{\mathcal{NS}}$ such that $\mathbf{b}^T \boldsymbol{\pi} = 0$. We may always relabel $\boldsymbol{\pi}$ to obtain $\hat{\boldsymbol{\pi}}$ in the form Eq. (5.50), defining a new inequality $\hat{\mathbf{b}}$ with $\hat{\mathbf{b}}^T \hat{\boldsymbol{\pi}} = 0$. Suppose that $\hat{\mathbf{b}}^T \mathbf{q} > 0$ for all $\mathbf{q}$ in the form Eq. (5.50) consisting only of S and A blocks. We define the subsets of measurements $\mathcal{M}_K, \mathcal{M}_L$ such that, for $\hat{\boldsymbol{\pi}}$:

$$\hat{\Pi}_{|xy} = K,\ x \in \mathcal{M}_K,\ y \notin \mathcal{M}_L,$$

$$\hat{\Pi}_{|xy} = L,\ x \notin \mathcal{M}_K,\ y \in \mathcal{M}_L,$$

$$\hat{\Pi}_{|xy} = M,\ x \in \mathcal{M}_K,\ y \in \mathcal{M}_L.$$

By assumption, at least one of $\mathcal{M}_K, \mathcal{M}_L$ must be non-empty. Suppose without loss of generality it is $\mathcal{M}_K$. As $\hat{\mathbf{b}}^T \hat{\boldsymbol{\pi}} = 0$, and the coefficients of $\hat{\mathbf{b}}$ are non-negative, we can therefore conclude

$$\hat{B}(0b|xy) = 0,\ b \in \{0,1\},\ x \in \mathcal{M}_K,\ y \notin \mathcal{M}_L, \tag{5.62}$$

$$\hat{B}(a0|xy) = 0,\ a \in \{0,1\},\ x \notin \mathcal{M}_K,\ y \in \mathcal{M}_L, \tag{5.63}$$

$$\hat{B}(00|xy) = 0,\ x \in \mathcal{M}_K,\ y \in \mathcal{M}_L. \tag{5.64}$$

Consider the subset $D_K \subset V_{\mathcal{L}}$ of deterministic distributions with $p_A(1|x) = 1, x \in \mathcal{M}_K$. Suppose there exists a saturating point $\mathbf{d}_k \in D_K$ i.e. that $\hat{\mathbf{b}}^T \mathbf{d}_k = 1$. Now consider another deterministic distribution $\mathbf{d}'_k$ such that $p_A(0|x) = 1, x \in \mathcal{M}_K,\ p_B(0|y) = 1, y \in \mathcal{M}_L$, but all other probabilities of $\mathbf{d}_k$ are left unchanged. This means that all non-zero probabilities in $\mathbf{d}_k$ are mapped to a probability in $\mathbf{d}'_k$ whose coefficient in $\hat{\mathbf{b}}$ is either unchanged, or 0. Thus we can conclude

$$1 \le \hat{\mathbf{b}}^T \mathbf{d}'_k \le \hat{\mathbf{b}}^T \mathbf{d}_k = 1, \tag{5.65}$$

the lower bound coming from the requirement that $\hat{\mathbf{b}}$ is a Bell inequality. Thus we must

have $\hat{\mathbf{b}}^T \mathbf{d}'_k = \hat{\mathbf{b}}^T \mathbf{d}_k$, and therefore:

$$\hat{B}(10|xy) = 0 \text{ or } \hat{B}(11|xy) = 0, \ x \in \mathcal{M}_K, \forall y. \tag{5.66}$$

with the choice determined by the deterministic choice of $\mathbf{d}_k$. Combining Eq. (5.66) with Eq. (5.62) and Eq. (5.64) respectively implies

$$\hat{B}_{|xy} = \begin{pmatrix} 0 & 0 \\ 0 & \gamma \end{pmatrix} \text{ or } \hat{B}_{|xy} = \begin{pmatrix} 0 & 0 \\ \gamma & 0 \end{pmatrix}, \ x \in \mathcal{M}_K, \ y \notin \mathcal{M}_L, \tag{5.67}$$

$$\hat{B}_{|xy} = \begin{pmatrix} 0 & \gamma_1 \\ 0 & \gamma_2 \end{pmatrix} \text{ or } \hat{B}_{|xy} = \begin{pmatrix} 0 & \gamma_1 \\ \gamma_2 & 0 \end{pmatrix}, \ x \in \mathcal{M}_K, \ y \in \mathcal{M}_L. \tag{5.68}$$

We can see that these coefficients imply that $\hat{\Pi}_{|xy}$, $x \in \mathcal{M}_K$, $y \notin \mathcal{M}_L$ can be replaced with one of either S or A without increasing the value of $\hat{\mathbf{b}}^T \hat{\boldsymbol{\pi}}$. If M_L is empty, then we can thus conclude that no saturating points lie in D_K, as we have contradicted our assumption. If M_L is non-empty, then for our assumption to still hold it must be the case that $\hat{B}_{|xy}$, $x \in \mathcal{M}_K$, $y \in \mathcal{M}_L$ must be of the first form in Eq. (5.68); else we could replace $\hat{\Pi}_{|xy}$ with an S block. Furthermore, both $\gamma_1, \gamma_2 \neq 0$, else we could again replace this block with either S or A.

Consider now the set $D_L \subset V_\mathcal{L}$ of deterministic distributions with $p_B(1|y) = 1, y \in \mathcal{M}_L$. We may conclude that for any point $\mathbf{d}_l \in D_L$,

$$1 \leq \hat{\mathbf{b}}^T \mathbf{d}'_l < \hat{\mathbf{b}}^T \mathbf{d}_l \tag{5.69}$$

where $\mathbf{d}'_l$ is the deterministic point setting $p_A(0|x) = 1, x \in \mathcal{M}_K$, $p_B(0|y) = 1, y \in \mathcal{M}_L$, and leaving all other probabilities in $\mathbf{d}_l$ unchanged. This is because all non-zero probabilities are mapped to probabilities with Bell coefficients either the same or 0, while for $x \in \mathcal{M}_K, y \in \mathcal{M}_L$ they are mapped to Bell coefficients *strictly* less than before. Thus no saturating points of $\hat{\mathbf{b}}$ lie in D_L.

We have thus proved that, for a Bell inequality satisfying our assumption, that the saturating extremal local distributions lie either in $V_\mathcal{L} \backslash D_K$ or $V_\mathcal{L} \backslash D_L$, with a non-empty exclusion. Since all deterministic distributions in this set have for at least one measurement either $p_A(0|x) = 1$ or $p_B(0|y) = 1$, we can conclude

$$\dim\left[\mathcal{A}(V_\mathcal{L} \backslash V_{D_K})\right] \leq \dim\left[(m_A - 1, m_B, 2, 2)\right] < \dim\left[(m_A, m_B, 2, 2)\right] - 1 = \dim\left[\mathcal{L}\right] - 1,$$

$$\dim\left[\mathcal{A}(V_\mathcal{L} \backslash V_{D_L})\right] \leq \dim\left[(m_A, m_B - 1, 2, 2)\right] < \dim\left[(m_A, m_B, 2, 2)\right] - 1 = \dim\left[\mathcal{L}\right] - 1.$$

This contradicts our assumption that $\hat{\mathbf{b}}$ is a facet, as its affine dimension is too low. Thus we may conclude that such a $\hat{\mathbf{b}}$ is impossible, and we have proved our result. $\qquad\square$

5.5.4 Addition of Local Distributions

The simplex algorithm is deterministic: solving the dual problem with the same $\mathbf{q}$ will return us the same Bell inequality candidate. In order to increase the candidates, we add a small amount of local noise to increase the number of input objective functions. Moreover, complementary slackness tells us that $\left(A^T\mathbf{y}^* - \mathbf{c}\right)^T\mathbf{x}^* = 0$. Provided that the choices $\mathbf{d}_\alpha$, $\mathbf{d}_\beta$ give the maximal local weight of $\mathbf{q}'$, then the candidate inequality outputted will be saturated by these local points - this will reduce the number of output solutions corresponding to convex combinations of facets. If these local points do *not* provide the maximal local weight, then different x_i^* will be non-zero, and different local points will saturate the candidate inequality. This still cuts down on on the likelihood on a non-facet solution however.

As the dimension grows, and the number of extremal points corresponding to non-facet inequalities grows, one could potentially add more local points in order to more stringently constrain the resulting output.

5.5.5 Using Matrix Rank to Test a Facet

Once we have obtained a Bell inequality $\mathbf{y}^*$, we wish to test if it is a facet inequality. This means it must define a face of dimension $(t - 1)$, with t the dimension of the local polytope[13]. To check this, we calculate the matrix rank of $A_{\mathbf{y}^*}$, the matrix whose columns are $\mathbf{d}_i$ such that $\mathbf{y}^{*T}\mathbf{d}_i = 1$; exactly the extremal local distributions which lie of the face defined by $\mathbf{y}^*$. The matrix rank gives the dimension of the vector space spanned by the columns of the matrix, and I make the claim this is exactly greater than the dimension of the face defined by $\mathbf{y}^*$ by 1. Let us label the columns of $A_{\mathbf{y}^*}$ as $\{\mathbf{d}_i^*\}_{i=1}^k$. The matrix rank of $A_{\mathbf{y}^*}$ is therefore

$$\text{rank}\,[A_{\mathbf{y}^*}] = \dim\left[\mathcal{V}\left(\{\mathbf{d}_1^*, \mathbf{d}_2^*, \ldots \mathbf{d}_k^*\}\right)\right] \equiv \dim\left[\mathcal{V}\left(\{\mathbf{d}_1^*, \mathbf{d}_2^* - \mathbf{d}_1^*, \ldots \mathbf{d}_k^* - \mathbf{d}_1^*\}\right)\right]. \quad (5.70)$$

---

[13]Given in Eq. (5.29).

We also have that the dimension of the face defined by $\mathbf{y}^*$ is given by the dimension of the affine space $\mathcal{A}_{\mathbf{y}^*} := \mathcal{A}\left(\{\mathbf{d}_1^*, \mathbf{d}_2^* \ldots, \mathbf{d}_k^*\}\right)$. For this space we have that:

$$\dim\left[\mathcal{A}_{\mathbf{y}^*}\right] = \dim\left[\mathcal{V}\left(\{\mathbf{d}_2^* - \mathbf{d}_1^*, \mathbf{d}_3^* - \mathbf{d}_1^*, \ldots, \mathbf{d}_k^* - \mathbf{d}_1^*\}\right)\right]. \tag{5.71}$$

From this, we can see that

$$\operatorname{rank}\left[A_{\mathbf{y}^*}\right] = \begin{cases} \dim\left[\mathcal{A}_{\mathbf{y}^*}\right] & \text{if } \mathbf{d}_1^* \in \mathcal{V}\left(\{\mathbf{d}_2^* - \mathbf{d}_1^*, \mathbf{d}_3^* - \mathbf{d}_1^*, \ldots, \mathbf{d}_k^* - \mathbf{d}_1^*\}\right) \\ \dim\left[\mathcal{A}_{\mathbf{y}^*}\right] + 1 & \text{otherwise} \end{cases} \tag{5.72}$$

using the property that $\dim\left[\mathcal{V} + \mathcal{W}\right] \leq \dim[\mathcal{V}] + \dim[\mathcal{W}]$ for vector spaces.

Suppose $\mathbf{d}_1^* \in \mathcal{V}\left(\{\mathbf{d}_2^* - \mathbf{d}_1^*, \mathbf{d}_3^* - \mathbf{d}_1^*, \ldots, \mathbf{d}_k^* - \mathbf{d}_1^*\}\right)$. Then there exist $\beta_i \in \mathbb{R}$ such that

$$\mathbf{d}_1^* = \sum_{i=2}^{k} \beta_i \left(\mathbf{d}_i^* - \mathbf{d}_1^*\right). \tag{5.73}$$

This implies that

$$\mathbf{0} = \mathbf{d}_1^* - \sum_{i=2}^{k} \beta_i \left(\mathbf{d}_i^* - \mathbf{d}_1^*\right) \in \mathcal{A}_{\mathbf{y}^*}. \tag{5.74}$$

However, this cannot be the case: by the normalisation condition, the elementwise sum of any point in $\mathcal{A}_{\mathbf{y}^*}$ must equal $m_A m_B \neq 0$. Therefore we may conclude $\mathbf{0} \notin \mathcal{A}_{\mathbf{y}^*}$ and therefore $\operatorname{rank}\left[A_{\mathbf{y}^*}\right] = \dim\left[\mathcal{A}_{\mathbf{y}^*}\right] + 1$, and so if $\operatorname{rank}\left[A_{\mathbf{y}^*}\right] = t$ then $\mathbf{y}^*$ is a facet Bell inequality.

5.5.6 Affine Fixing

Once an inequality $\mathbf{y}^*$ has been determined to be a facet, we then wish to check whether a representative of the facet inequality class has already been found. To do this, we use a method called *affine fixing*, done in the following way.

We first calculate the vector $\mathbf{c}_{\mathbf{y}^*} = A^T \mathbf{y}^*$, which gives the value of $\mathbf{y}^*$ at each local extremal distribution. As some extremal points saturate the inequality, the minimum element of this vector will be 1. We label the next lowest element $\gamma > 1$. We then perform the affine transformation

$$\hat{\mathbf{c}}_{\mathbf{y}^*} = \frac{1}{\gamma - 1} \mathbf{c}_{\mathbf{y}^*} + \frac{\gamma - 2}{\gamma - 1} \tag{5.75}$$

which maps the lowest element (1) to 1, and the second lowest element (γ) to 2. As Bell inequalities are invariant under transformations of the form $k\mathbf{b} + s$, $k \neq 0$, this vector will

be identical for all representations of $\mathbf{y}^*$. It also gives us a representation

$$\hat{\mathbf{y}}^* = \frac{1}{\gamma - 1}\mathbf{y}^* + \frac{\gamma - 2}{\gamma - 1}\frac{1}{m_A m_B}\mathbf{1} \tag{5.76}$$

which we can be certain has local bound 1 and next lowest local extremal value 2. This can be a useful standardisation to compare properties of Bell inequalities.

Now consider another facet inequality $\mathbf{z}$ belonging to the same class as $\mathbf{y}^*$. As they belong to the same class, there exists a relabelling ℓ which, when applied to $\mathbf{z}$, gives a representation of $\mathbf{y}^*$; say $\tilde{\mathbf{y}}^*$. We could then take the vector $\mathbf{c}_{\tilde{\mathbf{y}}^*} = A^T\tilde{\mathbf{y}}^*$, and apply the affine fix to obtain $\hat{\mathbf{c}}_{\tilde{\mathbf{y}}^*}$. Clearly $\hat{\mathbf{c}}_{\tilde{\mathbf{y}}^*} = \hat{\mathbf{c}}_{\mathbf{y}^*}$. Instead of this, we could have affine fixed $\mathbf{c}_\mathbf{z}$ to obtain $\hat{\mathbf{c}}_\mathbf{z}$, then instead applied the relabelling ℓ to the local distributions, inducing a permutation ω_ℓ. This would permute the elements of $\hat{\mathbf{c}}_\mathbf{z}$, and as the relabelling $\ell : \mathbf{z} \to \tilde{\mathbf{y}}$, we would obtain the vector $\hat{\mathbf{c}}_{\tilde{\mathbf{y}}^*} = \hat{\mathbf{c}}_{\mathbf{y}^*}$. Thus if two facet inequalities $\mathbf{y}^*$ and $\mathbf{z}$ belong to the same class, their affine vectors $\hat{\mathbf{c}}_{\mathbf{y}^*}$ and $\hat{\mathbf{c}}_\mathbf{z}$ will be reorderings of each other.

In <u>algorithm 1</u>, we simply tally the elements of $\hat{\mathbf{c}}_{\mathbf{y}^*}$ and compare it to the tallies of all stored facet inequalities, storing the inequality if it is non-equal to any already obtained, and discarding it otherwise. If the tally does not coincide with the tally of any stored inequalities, $\hat{\mathbf{c}}_{\mathbf{y}^*}$ cannot be a permutation of any previously stored $\hat{\mathbf{c}}_\mathbf{z}$ and thus cannot be of the same inequality class. Non-equality of the tally is not a *necessary* condition to belong to a different class however; the two $(4, 4, 2, 2)$ inequalities B_1, B_2 both have the same "affine tally", $\mathcal{T}_{B_1} = \mathcal{T}_{B_2}$, but do *not* belong to the same class. These inequalities have been transformed in the manner described[14] by Eq. (5.76). $\mathcal{T}_{B_1}$ is given in table 5.6.

---

[14]For presentation, rather than removing the constant uniformly, we have taken it only from the $(1, 1)$ measurement pair.

Local Value	$\#\mathbf{d}_i$
1	48
2	96
3	64
4	32
5	16

Table 5.6: The affine tally of both B_1 and B_2.

$$
B_1 = \begin{pmatrix}
-\frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\
-\frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\
0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\
0 & 0 & 0 & 1 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\
1 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0
\end{pmatrix}
\tag{5.77}
$$

$$
B_2 = \begin{pmatrix}
-\frac{1}{2} & -\frac{1}{2} & 0 & 1 & 0 & 0 & 0 & 1 \\
\frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\
0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\
0 & 1 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\
0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2}
\end{pmatrix}
\tag{5.78}
$$

It is exactly because of this that we have the variation algorithm 2, to avoid throwing away new facet inequalities. Before the search begins, the set of possible permutations of the local points by relabelling, Ω, is generated and stored. Then when a facet inequality $\mathbf{y}^*$ is found, $\mathcal{T}_{\mathbf{y}^*}$ is first compared against that of the classes already found as before, but if found to be equal to a stored tally $\mathcal{T}_{\mathbf{z}}$, rather than throwing the inequality away, we instead apply all permutations $\omega \in \Omega$ to $\hat{\mathbf{c}}_{\mathbf{y}^*}$ - if $\omega\left(\hat{\mathbf{c}}_{\mathbf{y}^*}\right) = \hat{\mathbf{c}}_{\mathbf{z}}$ we may conclude $\mathbf{y}^*, \mathbf{z}$ are in

the same inequality class and discard $\mathbf{y}^*$, and if no such ω is obtained then we conclude $\mathbf{y}^*$ is a new class, and store $\mathbf{y}^*$, $\hat{\mathbf{c}}_{\mathbf{y}^*}$ and $\mathcal{T}_{\mathbf{y}^*}$.

The size of Ω is $(k_A!)^{m_A} (k_B!)^{m_B} m_A! m_B!$, so the storage space required grows very rapidly. However, once the permutation set is generated, the variations can be checked equal/non-equal quite rapidly.

5.5.7 Results - and Limitations

One of the most pertinent limitations to this algorithm is that is does *not* terminate once all facet Bell inequalities have been found. In theory, it is possible for a variation of the algorithm to guarantee finding all facet inequality classes: by cycling through every no-signalling point adding every set of t linearly independent local deterministic points as a small noise factor, one would guarantee finding every facet inequality - since one would eventually hit a no-signalling point with value 0 at that inequality, with the deterministic noise set a subset of the saturating deterministic points. These conditions are sufficient for the facet to be the unique optimal solution. This would be an extremely unfeasible algorithm to run, however. This means that in general we are running the algorithm with no idea when to stop, or how close the number of inequality classes obtained is to the true number of classes. With the exception of the $(4, 4, 2, 2)$ and $(3, 5, 2, 2)$ scenarios, where the number of classes was enumerated in [34], the results presented below are necessarily lower bounds only. We can obtain an upper bound to the number of facets from [75], but it is orders of magnitude larger than our lower bounds, and thus provides little illumination of the true number of classes - especially because the relation between number of facets and number of Bell inequality classes remains obscure.

- $(\mathbf{4}, \mathbf{4}, \mathbf{2}, \mathbf{2})$

  For the $(4, 4, 2, 2)$ scenario, we have enumerated all 175 inequivalent classes of Bell inequality (including the trivial positivity inequality). Previously the most comprehensive list was 129 non-trivial inequalities given in [79], which employs an alternative linear-program based method of generation. Our generation of these inequalities was performed on a standard desktop computer within a day or two - first running algorithm 1, which generated 165 inequalities - after finding no new inequalities for 10000 iterations, we swapped to algorithm 2, which generated the remaining 9 (we

Size of Class	Number of Classes
64	1
288	1
9216	2
18432	4
24576	1
36864	4
49152	2
73728	8
98304	2
147456	61
294912	89
36391264	175

Table 5.7: The size of each facet class for the $(4, 4, 2, 2)$ local polytope. The totals at the bottom coincide with those presented in [34], as expected.

added the trivial inequality manually). A full list of these classes can be found at `http://www-users.york.ac.uk/~tpwc500`.

In order to gain a clearer picture of this polytope, we also provide a table giving the *size* of each class. This is given in table 5.7.

- $(\mathbf{3}, \mathbf{5}, \mathbf{2}, \mathbf{2})$

  For this scenario, it was given in [34] that there are 7 classes of facet. In [27], 6 classes were given for the $(3, 4, 2, 2)$ scenario - which we know will be facet classes for $(3, 5, 2, 2)$ also. We enumerated the 7 classes of this scenario using algorithm 1 and compared to the 6 known, finding the new inequality class to be:

$$
I_{3522} = \begin{pmatrix}
0 & \frac{2}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} \\
0 & 0 & \frac{2}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 \\
0 & 0 & 0 & 0 & \frac{2}{3} & 0 & \frac{2}{3} & 0 & 0 & 0 \\
\frac{2}{3} & 0 & 0 & \frac{2}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \frac{2}{3} & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \\
0 & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{2}{3} & 0 & \frac{1}{3}
\end{pmatrix}
\tag{5.79}
$$

We can also affine fix this to obtain:

$$
\hat{I}_{3522} = \begin{pmatrix}
-\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\
-\frac{1}{2} & -\frac{1}{2} & 1 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\
0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 1 & 0 & \frac{1}{2}
\end{pmatrix}
\tag{5.80}
$$

- $(\mathbf{4}, \mathbf{5}, \mathbf{2}, \mathbf{2})$

  For the $(3, 4, 2, 2)$ scenario, there are 6 inequality classes. For $(4, 4, 2, 2)$, there are 175 classes - an extraordinary jump! However, this pales in comparison to the addition of a further measurement choice. Running <u>algorithm 1</u> for the $(4, 5, 2, 2)$ scenario, we have found a staggering $16{,}642$ inequality classes - and this is only a lower bound. One cannot think of this even as a close lower bound, as we stopped this scenario due to memory constraints.

- $(\mathbf{3}, \mathbf{3}, \mathbf{3}, \mathbf{3})$

  For this case, we do *not* know the full set of no-signalling extremal points - and thus cannot use them as our objective function for the linear program. Instead, we generate random quantum distributions. To do this, we generate a vector of 3 real elements α_i; from this we define a normalised vector $\hat{\alpha}_i = \alpha_i/\|\boldsymbol{\alpha}\|$ - such that $\sum_i \hat{\alpha}_i{}^2 = 1$. We take this to be the Schmidt coefficients of some pure entangled (since there exists multiple non-zero coefficients) state, $|\phi\rangle = \sum_{i=1}^{3} \alpha_i |ii\rangle$. We then generate 6 random unitaries $\left\{ U^{A_1}, U^{A_2}, U^{A_3}, U^{B_1}, U^{B_2}, U^{B_3} \right\}$, each unitary corresponding to a measurement. Since the columns of each U^i are orthonormal, we can define projection operators $P_k^i := |[U^i]_k^T\rangle \langle [U^i]_k^T|$ satisfying $\sum_k P_k^i = \left(U^i\right)^\dagger U_i = \mathbb{I}_3$. Thus, we obtain the probability distribution:

  $$
  p(ab|xy) = \langle \phi | \left(P_a^{A_x} \otimes P_b^{B_y} \right) | \phi \rangle .
  \tag{5.81}
  $$

  We may then use this as the objective function. Using this, we were able to find 10143 inequality classes[15]. Once more this can be seen as a loose lower bound, being stopped at an arbitrary point. Moreover, in the above generation we have limited

---

[15]In [92] a similar linear programming technique was used to provide 19 classes, of which 3 were found here.

Scenario	# Extremal S/A points	Min facet %	Max facet %	Mean %
$(2,2,2,2)$	1	100%	100%	100%
$(3,3,2,2)$	8	44.3%	80.6%	65.25%
$(4,4,2,2)$	256	33.9%	66.7%	57.50%

Table 5.8: A survey of % success of obtaining a facet-defining solution.

ourselves to projective measurements, in dimension 3. We could easily generalise this to higher dimension (one could take a $d \times d$ unitaries, and partition the columns into three sets) or generate POVM elements instead. However, one should be aware that projective measurements often result in a local distribution - and it is likely POVM elements would increase the chance of this happening.

5.5.7.1 Facet Density

As mentioned previously, not every output is a facet inequality. Unfortunately it is too difficult to enumerate the full polytopes in order to determine the exact proportion of these, except in the low dimensional cases, so instead we shall employ an "operational" test. In table 5.8, there is an analysis of how often a facet Bell inequality was obtained, by performing 1000 runs over each S/A extremal no-signalling point. Note this makes no reference to how often a *new* facet class was obtained, just to how often a facet was outputted. It appears that the % success decreases as the dimension increases.

5.5.7.2 Use of the Polar Dual

Earlier in this chapter, we discussed the possibility of using the polar dual of $\mathcal{L}$ as the solution space to our linear program. As the vertices correspond exactly to facet Bell inequalities, this would guarantee our solutions would be facet inequalities by the maximum principle (lemma 5.3.4). In order to use the polar dual, we need to shift the origin, so that it lies in the interior of $\mathcal{L}$. The natural choice is the uniform distribution $\boldsymbol{\pi}_u$, where $p(ab|xy) = \frac{1}{k_A k_B}, \forall a, b, x, y$. For example, in the $(2,2,2,2)$ scenario this would map the

extremal no-signalling point:

$$\Pi_{PR} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \rightarrow \overrightarrow{\Pi}_{PR} = \begin{pmatrix} \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \end{pmatrix}. \tag{5.82}$$

We can see that we are now allowed negative values in our distributions, and in order to keep the dual polytope as our solution space we must also relax our constraint that $\mathbf{y} \geq 0$. Therefore our new linear program will be

$$\text{maximise } \overrightarrow{\mathbf{q}}^T \mathbf{y} \text{ subject to: } \overrightarrow{A}^T \mathbf{y} \leq \mathbf{1} \tag{5.83}$$

where $\overrightarrow{A}$ is the matrix whose columns are the extremal local distributions with shifted origins, $(\overrightarrow{\mathbf{d}}_1 \ldots \overrightarrow{\mathbf{d}}_n)$. Notice the inequality sign is flipped from Eq. (5.59) in order to be of the correct form for the polar dual, and consequently our optimisation have become a maximisation. As the local polytope is bounded, so too is the dual polytope, and so the objective function $\overrightarrow{\mathbf{q}}^T \mathbf{y}$ is bounded. However, our conversion of the problem into the form in Eq. (5.83) means we no longer have a known bound on the objective function[16]. To illustrate the problem with this, we perform the optimisation in Eq. (5.83), for all extremal no-signalling points which are non-local $\overrightarrow{\mathbf{q}} \in V_{\overrightarrow{\mathcal{NS}}} \backslash V_{\overrightarrow{\mathcal{L}}}$. For the $(4, 4, 2, 2)$ scenario, the greatest solution obtained is $\frac{12}{5}$, and the least solution is 2. Unlike Eq. (5.59), the optimal value varies with the no-signalling point used.

We now consider a particular $(4, 4, 2, 2)$ Bell inequality of the form:

$$\overrightarrow{B}_{\text{ex}} = \begin{pmatrix} 0 & -\frac{4}{5} & 0 & 0 & 0 & 0 & 0 & -\frac{4}{5} \\ 0 & 0 & -\frac{4}{5} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\frac{4}{5} & 0 & 0 \\ -\frac{4}{5} & 0 & 0 & -\frac{4}{5} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{4}{5} & 0 & 0 & -\frac{4}{5} & 0 \\ 0 & 0 & 0 & 0 & -\frac{4}{5} & 0 & 0 & 0 \end{pmatrix}, \text{Tr}\left[\overrightarrow{B}_{\text{ex}}^T \overrightarrow{\Pi}\right] \leq 1. \tag{5.84}$$

---

[16]In the original form, we had that $\mathbf{q}^T \mathbf{y} \geq 0$.

and a corresponding linear programming problem:

$$\text{maximise } \mathbf{q}^T \overrightarrow{\mathbf{b}}_{\text{ex}} \text{ subject to: } \mathbf{q} \in \overrightarrow{\mathcal{NS}}. \tag{5.85}$$

We may do this because the no-signalling constraints are linear equalities, whilst we know the H-representation of $\mathcal{NS}$ is the set of positivity conditions, which may be shifted to define $\overrightarrow{\mathcal{NS}}$. Performing this optimisation we find the maximal value to be $\frac{9}{5} < 2$. We can therefore conclude there is *no* extremal no-signalling point that when used as the objective function for Eq. (5.83), will give the solution $\mathbf{y}^* = \overrightarrow{\mathbf{b}}_{\text{ex}}$ - by using the dual polytope, we are no longer able to generate all facet Bell inequalities using extremal no-signalling points. However, one could still use the dual polytope by generating quantum distributions and measurements, as seen for the $(3, 3, 3, 3)$ scenario.

5.6 The Detection Loophole

In the introduction to this chapter we discussed Einstein et al.'s postulate that there existed "hidden variables", which somehow determined quantum behaviour. Moreover, we saw how Bell's theorem gave an inequality for all models of this form - one which is violated by quantum theory. This inequality takes in correlations[17] between joint measurements. This is slightly disingenuous - we cannot perform all such joint measurements on a single state. To test the inequality, we require many joint measurements over several copies of the same state. This in itself not a problem, since any observation of a violation will still discount a hidden variable model across these states. What is a concern however, is that experimentally we cannot guarantee a successful measurement with 100% certainty. If we simply discount these failed detections, then it is possible [80] to obtain correlations which violate a Bell inequality, but if the whole distribution (including detection failures) is considered, admits a local hidden variable model.

It is perhaps reasonable to dismiss the concept that, within the machinations of the universe, particles are behaving in such a way that when detected they are reliably appearing to follow quantum mechanics, whilst in fact being determined by a different model, as paranoid. Indeed, the important experiment detailed in [54] put paid to this idea, performing an experiment free of this (and other) "loopholes" - this particular problem being

---

[17]Or joint probabilities, in alternative forms.

referred to as the "detection loophole". What is not paranoid however, is to consider this problem in the context of quantum cryptography. One of the recent advancements quantum cryptography offers is the concept of "device independent cryptography" [26, 74] - in which a secret key is established between parties using *black box* devices, where they know nothing about the inner workings. In classical cryptography to trust such a device would be impossible; they could simply be preprogrammed beforehand to spit out a seemingly random string which is completely known to a malicious party. In quantum cryptography however, the two parties may sacrifice some of their generated string in order to Bell test the devices, and ensure the boxes are behaving in a quantum manner, excluding the possibility of a pre-programmed (i.e. hidden variable) scenario.

If the black boxes are allowed to offer no output however, then once more the attacker may proffer a string which appears to violate a Bell inequality, but in fact has been generated by only outputting outcomes desirable to the attacker. This means the failure to output needs to be considered when Bell testing the device. This is especially pertinent given that current commercial quantum cryptography devices, and likely future ones, operate using quantum optics, where the practicalities of sending and detecting single photons inevitably lead to detection failures. The user will want to ensure that any failures to output are genuine errors in the protocol, rather than deliberately pre-programmed.

5.6.1 How to Treat Detection Failures

In this section, we shall focus on conditional probability distributions, rather than correlations. We need to know how to treat probability distributions in the event of detection failures. Before we do this it is worth mentioning that study of the detection loophole is often split into two disciplines - asymmetric detection failure, where one party has a flawless detector whilst the other detects with probability η, and symmetric detection failure, where both parties have the same detection efficiency η. We shall focus on the latter here - though there exist some scenarios where the asymmetric treatment may be appropriate; perhaps Alice is using a diamond cavity, which has a high chance of successful detection, then sending optically an entangled photon which Bob measures with a higher chance of failure. A more realistic scenario is that both Alice and Bob's devices both fail to detect for some runs, succeeding with respective probabilities η_1, η_2 - by taking $\eta = \min\{\eta_1, \eta_2\}$ we may simplify the maths while allowing us to "play it safe".

Mathematically, we treat the detection loophole in the following way: we begin with a perfect (m_A, m_B, k_A, k_B) scenario with a conditional probability distribution $\boldsymbol{\pi} = \{p(ab|xy)\}$. Alice and Bob choose their measurement choice x, y respectively, and we assume that the probability of successful detection for each of them is η. One of the following four things occurs:

- *Two successful detections.*

  With probability η^2, both detections are successful, and Alice and Bob obtain respective outcomes a, b, according to distribution $\boldsymbol{\pi}$. The overall probability of this outcome is $\eta^2 p(ab|xy)$.

- *Alice's detection fails and Bob's succeeds.*

  This occurs with probability $\eta(1 - \eta)$; η for Bob's success, $(1 - \eta)$ for Alice's failure, and Bob's outcome b is determined by his conditional distribution $p(b|xy) \equiv p(b|y)$. Therefore the overall probability is $\eta(1 - \eta)p(b|y)$.

- *Alice's detection succeeds and Bob's fails.*

  This also occurs with probability $\eta(1 - \eta)$, with Alice's outcome a determined by her conditional distribution $p(a|xy) \equiv p(a|x)$. The probability of this outcome is $\eta(1 - \eta)p(a|x)$.

- *Both detections fail.*

  This occurs with probability $(1 - \eta)^2$, with no dependence on $\boldsymbol{\pi}$.

With this in mind, we can define a new $(m_A, m_B, k_A + 1, k_B + 1)$ probability distribution $\boldsymbol{\pi}_\eta$ with an extra outcome N. $\boldsymbol{\pi}_\eta$ is constructed:

$$\boldsymbol{\pi}_\eta(ab|xy) = \eta^2 \boldsymbol{\pi}(ab|xy),$$

$$\boldsymbol{\pi}_\eta(Nb|xy) = \eta(1 - \eta)\boldsymbol{\pi}(b|y),$$

$$\boldsymbol{\pi}_\eta(aN|xy) = \eta(1 - \eta)\boldsymbol{\pi}(a|x),$$

$$\boldsymbol{\pi}_\eta(NN|xy) = (1 - \eta)^2.$$

Notice that this in not a *comprehensive* treatment of all detection failure scenarios, as it assumes the two devices fail independently. However, because the two devices are spatially separated, this is the distribution we would expect if our experimental devices were working "naturally". If we saw significant correlation between detection failures of each

party, we should definitely suspect the influence of a malicious third party. The above structure is designed to probe distributions which appear natural, but instead could have been pre-programmed.

Now that we have a $(m_A, m_B, k_A + 1, k_B + 1)$ probability distribution, we know exactly how to test it - using a $(m_A, m_B, k_A + 1, k_B + 1)$ Bell inequality! If our distribution violates an inequality, we know the underlying system is inherently non-local, and thus contains randomness we can exploit. There are typically two schools of thought when it comes to the Bell inequality used to test a particular inefficient distribution. The first is to take an inequality that is a "lifting" of a (m_A, m_B, k_A, k_B) inequality - this can be thought of operationally as treating every failure as one of the valid outcomes - mapping N to 0, for example[18]. This was the approach taken by Clauser and Horne, who came up with the CH representation [22] of the CHSH inequality to tackle this problem. By contrast, we may also choose an inequality from a class that only appears in the $(m_A, m_B, k_A + 1, k_B + 1)$ scenario. This is treating the failure outcome as fundamentally a separate outcome. At this point in time, it is not clear which is the better treatment generally, but in this chapter we shall focus on using lifted inequalities.

Lemma 5.6.1 $\boldsymbol{\pi}_\eta$ is non-local only if $\boldsymbol{\pi}$ is.

Proof. The distribution $\boldsymbol{\pi}_\eta$ can be thought of as a convex combination of the following distributions:

$$\boldsymbol{\pi},$$
$$\boldsymbol{\pi}_{b|y} := \{p(ab|xy)|\ p(ab|xy) = \delta_{aN}\boldsymbol{\pi}(b|y)\},$$
$$\boldsymbol{\pi}_{a|x} := \{p(ab|xy)|\ p(ab|xy) = \delta_{bN}\boldsymbol{\pi}(a|x)\},$$
$$\boldsymbol{\pi}_N := \{p(ab|xy)|\ p(ab|xy) = \delta_{aN}\delta_{bN}\},$$

in the following way:

$$\boldsymbol{\pi}_\eta = \eta^2\boldsymbol{\pi} + \eta(1-\eta)\boldsymbol{\pi}_{a|x} + \eta(1-\eta)\boldsymbol{\pi}_{b|y} + (1-\eta)^2\boldsymbol{\pi}_N. \tag{5.86}$$

---

[18]Note this can be measurement dependent - e.g. treating N as 0 for measurement 1, and N as 1 for measurement 2.

Distribution $\boldsymbol{\pi}_N$ is local deterministic, whilst $\boldsymbol{\pi}_{a|x}$ and $\boldsymbol{\pi}_{b|y}$ are clearly local distributions, as there exist no correlations between parties. Thus if $\boldsymbol{\pi}$ is local also, then $\boldsymbol{\pi}_\eta$ can be written as a convex combination of local distributions and therefore lies within $\mathcal{L}$. $\qquad\square$

Corollary 5.6.2 For $\eta_1 \geq \eta_2$, $\boldsymbol{\pi}_{\eta_2} \notin \mathcal{L}$ implies $\boldsymbol{\pi}_{\eta_1} \notin \mathcal{L}$.

Proof. Suppose $\boldsymbol{\pi}_{\eta_1} \in \mathcal{L}$. As $\boldsymbol{\pi}_{\eta_2}$ can be expressed as the convex combination

$$\boldsymbol{\pi}_{\eta_2} = \frac{\eta_2^2}{\eta_1^2}\boldsymbol{\pi}_{\eta_1} + \left(\eta_2\left(1-\eta_2\right) - \frac{\eta_2^2}{\eta_1}\left(1-\eta_1\right)\right)\boldsymbol{\pi}_{a|x} + \left(\eta_2\left(1-\eta_2\right) - \frac{\eta_2^2}{\eta_1}\left(1-\eta_1\right)\right)\boldsymbol{\pi}_{b|y}$$
$$+ \left(\left(1-\eta_2\right)^2 - \frac{\eta_2^2}{\eta_1^2}\left(1-\eta_1\right)^2\right)\boldsymbol{\pi}_N \tag{5.87}$$

we have that if $\boldsymbol{\pi}_{\eta_1} \in \mathcal{L}$, then $\boldsymbol{\pi}_{\eta_2} \in \mathcal{L}$ also. This gives our result. $\qquad\square$

5.6.2 The Best Inequalities (so Far)

The question of which Bell inequality is best to tackle the detection loophole is somewhat ambiguous, and depends on your purpose. For example, do you want the secret key verification rate to be robust against detection loophole noise? Or perhaps you prioritise low dimensional states and measurements, since they are easier to create experimentally. One undeniably important question is that of the *detection threshold* of a Bell inequality.

Definition 5.6.3 Given a Bell inequality $\mathbf{b}$ such that $\mathbf{b}^T\boldsymbol{\pi} \geq k$, $\forall \boldsymbol{\pi} \in \mathcal{L}$, the detection threshold η_c is

$$\eta_c := \inf_\eta \exists\, \boldsymbol{\pi} \in \mathcal{Q},\ \mathbf{b}^T\boldsymbol{\pi}_\eta < k \tag{5.88}$$

i.e. the lowest detection efficiency for which $\mathbf{b}$ can distinguish a non-local distribution.

We also may use the following, equivalent definition.

Definition 5.6.4 Given a Bell inequality $\mathbf{b}$ such that $\mathbf{b}^T\boldsymbol{\pi}_L \geq k$, $\forall \boldsymbol{\pi} \in \mathcal{L}$, the detection threshold η_c is

$$\eta_c := \sup_\eta \forall\, \boldsymbol{\pi} \in \mathcal{Q},\ \mathbf{b}^T\boldsymbol{\pi}_\eta \geq k \tag{5.89}$$

i.e. the highest detection efficiency for which $\mathbf{b}$ cannot distinguish a non-local distribution.

The first threshold obtained was for the CHSH inequality: Garg and Mermin showed [44] that, for maximally entangled states, so long as the efficiency was above $2/\left(\sqrt{2}+1\right)$ ($\approx$

82.8%, non-locality could be observed. As these were the states that achieved the maximal violation for the CHSH inequality, it seemed reasonable that these would be the states that could tolerate the lowest efficiency. This turned out *not* to be the case, as Eberhard constructed [38] a state and measurements that gave CHSH violations down to an efficiency of 2/3. Perhaps even more surprisingly, the optimal state achieving this threshold is $|\phi_\varepsilon\rangle = \sqrt{1 - \varepsilon^2}\,|00\rangle + \varepsilon\,|11\rangle$, with $\varepsilon \to 0$ as $\eta \to 2/3$ - far from being the maximally entangled state, the state achieving the lowest tolerable efficiency is almost unentangled! In this scenario, a lifting of CHSH is being used; no-click (N) is being mapped to 1. Moreover, the limit $\eta \to 2/3$ is achieved using qubit states and measurements.

Another significant result is given in [104], which looked at a particular $(4, 4, 2, 2)$ inequality, I^4_{4422}, once again considering a lifting[19]. For maximally entangled (ququart) states, they were able to show the tolerable efficiency was 76.98%, and they introduced explicit ququart states and measurements achieving a minimum tolerable efficiency of $\left(\sqrt{5} - 1\right)/2 \approx 61.80\%$. Interestingly the required state is of the form $|\psi_\varepsilon\rangle = \sqrt{(1 - \varepsilon^2)/3}\,(|00\rangle + |11\rangle + |22\rangle) + \varepsilon\,|33\rangle$, with the threshold achieved as $\varepsilon \to 0$ - showing a similar form to the Eberhard state; a generalisation of the maximally entangled state in which the final dimension has almost 0 weight.

There is also the important result of Serge Massar in [72]. He gave a set of dimension dependent states and measurements for which the corresponding $\boldsymbol{\pi}_\eta$ is non-local for $\eta \geq \eta_c \to 0$ in the limit as $d \to \infty$. However, $\eta_c \leq 2/3$ only when $d \geq 1600$, and the number of measurements required scales with 2^d for both Alice and Bob. This means that implementation of this method is practically impossible.

5.6.3 Limits of a Bell Inequality

Given a Bell inequality, we would like a lower bound on its detection threshold - to prevent situations such as that which occurred with Eberhard, where the threshold is substantially lowered by the discovery of a new state/measurement construction. To do this, we may look at a superset of the set of quantum correlations. Suppose we have a set $\tilde{\mathcal{Q}}$ such that the quantum set $\mathcal{Q}$ satisfies $\mathcal{Q} \subseteq \tilde{\mathcal{Q}}$. Given an η value, if $\forall \boldsymbol{\pi} \in \tilde{\mathcal{Q}}$, $\mathbf{b}^T \boldsymbol{\pi}_\eta \geq k$, then the

---

[19]Their binary outcomes were 1 and -1, mapping failure to -1.

same must be true for $\mathcal{Q}$, and we have obtained a lower bound on the detection threshold of **b**. Furthermore, if one has an η such that, $\forall \boldsymbol{\pi} \in \tilde{\mathcal{Q}}$, $\boldsymbol{\pi}_\eta \in \mathcal{L}$, then one has a lower bound on *all* Bell inequalities' (for that scenario) detection threshold, and for that particular efficiency no distributions are useful.

There are two obvious choices of $\tilde{\mathcal{Q}}$. The first is $\tilde{\mathcal{Q}} = \mathcal{NS}$, the set of all no-signalling correlations. For $k_A = k_B = 2$ we know all extremal NS distributions, and can thus use linear programming in order to determine the lower bound for the detection threshold. This can be done using a *binary chop method*. The method is as follows:

1. Take the set[20] of extremal non-local points $V_{\mathcal{NS}} \backslash V_{\mathcal{L}} = \{\mathbf{n}_k\}$.

2. For each $\mathbf{n}_k$ in turn, construct the inefficient distribution $\mathbf{n}_{k,\eta}$ according to the method in section 5.6.1.

3. Choose two values $\eta_{\max}$ and $\eta_{\min}$ such that $\mathbf{n}_{k,\eta_{\max}} \notin \mathcal{L}$ and $\mathbf{n}_{k,\eta_{\min}} \in \mathcal{L}$; we can always choose $\eta_{\max} = 1$ and $\eta_{\min} = 0$ to ensure this.

4. Take $\eta = (\eta_{\max} + \eta_{\min})/2$. Find the local weight w of $\mathbf{n}_{k,\eta}$, using the linear program outlined in definition 5.5.1.

5. If $w = 1$, then $\mathbf{n}_{k,\eta}$ is local and we set $\eta_{\min} = \eta$. If not, then $\mathbf{n}_{k,\eta}$ is non-local and we set $\eta_{\max} = \eta$.

6. Repeating Steps 4-5 we home in on the maximal η such that $\mathbf{n}_{k,\eta} \in \mathcal{L}$; we do this until we are satisfied with the precision. This value we call η_k.

7. Repeat the process 2-6 with each extremal $\mathbf{n}_k$, obtaining a set $\{\eta_k\}$. For $\eta_L := \min\{\eta_k\}$, we can conclude by convexity that $\boldsymbol{\pi}_{\eta_L} \in \mathcal{L}$, $\forall \boldsymbol{\pi} \in \mathcal{NS}$. Thus we have obtained our fundamental lower bound for the detection threshold. The results of this process for various $(m_A, m_B, 2, 2)$ scenarios are presented in table 5.9, and compared to some previous bounds in the literature.

Using this method, we can see that Eberhard's construction cannot be beaten for CHSH - since $2/3$ is the lowest possible for the $(2, 2, 2, 2)$ scenario. It is surprising that the best no-signalling efficiency is quantum achievable in this scenario - this does not hold

---

[20]We can take the subset given in section 5.3.5.3, since local/non-locality is preserved by relabelling.

m_B \ m_A	2	3	4	5	6	$\frac{1}{\sqrt{m_B}}$	$2/(m_B+1)$
2	2/3	2/3	2/3	2/3	2/3	0.707	2/3
3		4/7	5/9	5/9	5/9	0.577	1/2
4			1/2	1/2	1/2	1/2	2/5
5				4/9	*	0.447	1/3

Table 5.9: The maximum detection efficiency such that any no-signalling distribution can be generated classically. $(5, 6, 2, 2)$ was not evaluated due to the high number of non-local extremal points. The two columns are previous bounds given by [73].

in general.

The other sensible choice for $\tilde{\mathcal{Q}}$ is any member of the semidefinite hierarchy $\mathcal{Q}_i$. Given any Bell inequality $\mathbf{b}$ with local bound $\geq k$, we can again use a binary chop method.

1. Choose an $\eta_{\max}$ such that there exists a $\boldsymbol{\pi}_{\eta_{\max}} \in \mathcal{Q}$ with $\mathbf{b}\boldsymbol{\pi}_{\eta_{\max}} < k$, and an $\eta_{\min}$ such that $\forall \boldsymbol{\pi} \in \mathcal{NS}$, $\boldsymbol{\pi}_{\eta_{\min}} \in \mathcal{L}$ - again we may always take $\eta_{\max} = 1$ and $\eta_{\min} = 0$.

2. Set $\eta = (\eta_{\max} + \eta_{\min})/2$. Solve the semidefinite program

$$w = \min_{\Pi \in \mathcal{Q}_i} \mathrm{Tr}\left[B^T \Pi_\eta\right]. \tag{5.90}$$

3. If $w = k$, then there exists no non-local $\boldsymbol{\pi}_\eta$, $\boldsymbol{\pi} \in \mathcal{Q}_i$ distinguishable by $\mathbf{b}$, and we set $\eta_{\min} = \eta$. If $w < k$, then we set $\eta_{\max} = \eta$ and make a note of $\boldsymbol{\pi}$.

4. Repeating Steps 2-3 we home in on the maximal η such that $w = k$; we do this until we are satisfied with the precision. This gives us a lower bound of the detection threshold for $\mathbf{b}$, as well as a candidate $\boldsymbol{\pi}$ for achieving the threshold.

We can increase i to obtain tighter bounds on the detection threshold - if we receive the same lower bound for several i, this implies we may have found the true threshold, in which case we may look for quantum states and measurements achieving it; moreover, they should reproduce distribution $\boldsymbol{\pi}$.

181

Inequality	$\mathcal{Q}_1$	$\mathcal{Q}_2$	$\mathcal{Q}_3$	η_c
CHSH	66.67%	66.67%	66.68%	66.67%
I_{3322}	60.00%	66.51%	66.69%	66.67%
I_{2233}	64.95%	66.67%	66.67%	66.67%
I_{4422}^4	50.03%	61.83%	61.85%	61.80%

Table 5.10: A comparison of detection thresholds, as given by CVX. Note the discrepancy between the true value and the one outputted.

5.6.4 Implementation of The Semidefinite Search, and Results

To implement the search over levels of the semidefinite hierarchy, I used Matlab 2016a/2017a (depending on the machine used), and some Matlab-specific toolboxes. The first is CVX [46, 47] - a convex optimisation toolbox which allows the expression of semidefinite variables, and the automatic conversion of "naturally written" constraints into the formal semidefinite programming constraints. This comes with a built in semidefinite solver SDPT3 [100], however, I instead primarily used the compatible MOSEK solver [2]. Finally, I also used the QETLAB toolbox [64], which allowed me to specify the $\mathcal{Q}_i$ membership constraint. For all of this optimisation, I have used Bell inequalities of the form local bound ≥ 1.

In table 5.10, I present the results for some known cases, from which we can see that our binary chop only appears accurate to 10^{-3}. This is unfortunate, since the CVX precision parameter is set to 1.49×10^{-8}, and our binary chop precision is set[21] to 1.53×10^{-6}. This discrepancy is most likely due to the quantum violation being extremely small just above the detection threshold. This limitation will come to be important later on, and I spent much time trying to improve this. These attempts will be detailed in subsection 5.6.5.

5.6.4.1 (4,4,2,2) Scenario Test

Given that we have a complete list of the $(4, 4, 2, 2)$ inequality classes, we wished to learn whether there existed a Bell inequality with a lower detection threshold than I_{4422}^4 for this scenario. As we are using $(4, 4, 2, 2)$ Bell inequalities to check $(4, 4, 2, 2)$ probability

---

[21]By the number of iterations.

distributions with an extra detection failure output, we need to lift our Bell inequalities to the $(4, 4, 3, 3)$ scenario. For each measurement choice for both Alice and Bob, we may choose which output we map N to. This means that for every inequality we have $2^8 = 256$ possible liftings to consider. We performed the binary chop method on all 256 liftings for the 174 representatives of each non-trivial inequality class. In order to speed up the process, we introduced a "cut" - before performing the binary chop, we tested each lifting at $\eta = 2/3$ at level 1 of the hierarchy - if $w = 1$, then we can conclude that that particular lifting cannot distinguish non-local distributions better than CHSH (and thus I_{4422}^4), and we move on to the next lifting. As this requires only one semidefinite program at the least constrained hierarchy level, this speeds up the process considerably.

If a lifted Bell inequality gives a value $w < 1$ at $\eta = 2/3$, then we performed the binary chop with level $\mathcal{Q}_2$ of the hierarchy, in order to have a reasonable run time. The results of this test is that I_{4422}^4 *is* the best for this scenario - it achieved 61.83% with the lifting given in [104], again accurate up to 10^{-3}, whilst the next best inequality had a detection threshold of 64.95%. There are two liftings of I_{4422}^4 achieving 61.83% - we present them in figures 5.1 and 5.2. We omit the brackets and partition them into measurement choices, and denote with arrows the output copied for the lifting.

Whilst no $(4, 4, 2, 2)$ inequality was able to provide a lower detection threshold than I_{4422}^4, we do not rule out the possibility of a $(4, 4, 3, 3)$ Bell inequality achieving a lower threshold, by treating N as a separate outcome.

$$
(5.91) \qquad
\begin{array}{l|cc|cc|cc|cc}
 & \downarrow & & \downarrow & & & & \downarrow & \downarrow \\
\hline
\rightarrow & 0 & \tfrac{1}{2} & 0 & 0 & 0 & 0 & 0 & \tfrac{1}{2} \\
 & 0 & 0 & \tfrac{1}{2} & 0 & \tfrac{1}{2} & 0 & 0 & 0 \\
\hline
 & 0 & 0 & \tfrac{1}{2} & 0 & 0 & 0 & \tfrac{1}{2} & 0 \\
\rightarrow & \tfrac{1}{2} & 0 & 0 & 0 & \tfrac{1}{2} & 0 & 0 & 0 \\
\hline
 & 0 & \tfrac{1}{2} & 0 & \tfrac{1}{2} & 0 & 0 & \tfrac{1}{2} & 0 \\
\rightarrow & 0 & 0 & 0 & 0 & 0 & \tfrac{1}{2} & 0 & 0 \\
\hline
 & 0 & \tfrac{1}{2} & 0 & 0 & \tfrac{1}{2} & 0 & 0 & \tfrac{1}{2} \\
\rightarrow & 0 & 0 & 0 & \tfrac{1}{2} & 0 & 0 & 0 & 0 \\
\end{array}
$$

Figure 5.1: The first lifting of I^4_{4422}, achieving the boundary. The arrows denote the choice of coefficients for the third outcome N.

$$
(5.92) \qquad
\begin{array}{l|cc|cc|cc|cc}
 & \downarrow & & & \downarrow & & \downarrow & & \downarrow \\
\hline
 & 0 & \tfrac{1}{2} & 0 & 0 & 0 & 0 & 0 & \tfrac{1}{2} \\
\rightarrow & 0 & 0 & \tfrac{1}{2} & 0 & \tfrac{1}{2} & 0 & 0 & 0 \\
\hline
\rightarrow & 0 & 0 & \tfrac{1}{2} & 0 & 0 & 0 & \tfrac{1}{2} & 0 \\
 & \tfrac{1}{2} & 0 & 0 & 0 & \tfrac{1}{2} & 0 & 0 & 0 \\
\hline
 & 0 & \tfrac{1}{2} & 0 & \tfrac{1}{2} & 0 & 0 & \tfrac{1}{2} & 0 \\
\rightarrow & 0 & 0 & 0 & 0 & 0 & \tfrac{1}{2} & 0 & 0 \\
\hline
 & 0 & \tfrac{1}{2} & 0 & 0 & \tfrac{1}{2} & 0 & 0 & \tfrac{1}{2} \\
\rightarrow & 0 & 0 & 0 & \tfrac{1}{2} & 0 & 0 & 0 & 0 \\
\end{array}
$$

Figure 5.2: The second lifting of I^4_{4422}, achieving the boundary. This lifting is a relabelling of the one used in [104].

5.6.4.2 (3,3,3,3) Scenario Test

By utilising parallelism and a research computing cluster, we were able to perform the binary chop over all 729 liftings of the 10143 Bell inequalities we obtained for the $(3, 3, 3, 3)$ scenario. This testing was also done at level $\mathcal{Q}_2$ of the hierarchy, with an initial cut at $\eta = 2/3$. For this scenario, the best detection threshold achieved was 61.8% - the

same as $(4,4,2,2)$ (up to 10^{-3}). 8 inequality class representatives obtained this threshold, and for these 8, the binary chop at the $\mathcal{Q}_3$ level was performed, and the efficiency was non-increasing for 2. These 2 inequalities we believe will either be able to match the threshold of I_{4422}^4, or even better it by a small amount. These two inequalities are:

$$
I_3^A =
\begin{array}{l|ccc|ccc|ccc}
 & & \downarrow & & & \downarrow & & \downarrow & & \\
\hline
\rightarrow & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\
\hline
 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\
 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\
\rightarrow & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
\hline
\rightarrow & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
\end{array}
$$

$$
I_3^B =
\begin{array}{l|ccc|ccc|ccc}
 & & \downarrow & & & \downarrow & & \downarrow & & \downarrow \\
\hline
 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
\rightarrow & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
\hline
 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
\rightarrow & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\
\hline
 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
\rightarrow & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\end{array}
$$

It is curious to note that for each of these inequalities, there is a measurement with a redundant outcome - for I_3^A, Alice's first measurement has identical coefficients for both outcomes[22] $1,2$, whilst in I_3^B Bob's second measurement treats outcomes $2,3$ identically. Moreover, this redundant outcome is the optimal lifting - I_3^A treats $1,2$ and N as identical for Alice's first measurement, and similarly for I_3^B. Whether this is merely coincidence, or a hint at the structure of inefficiency-robust Bell inequalities, remains unclear.

5.6.4.3 (3,5,2,2) Scenario Test

Since we have also enumerated a single new $(3,5,2,2)$ inequality, it is also worthwhile testing this inequality in order to determine its detection threshold. Searching over all possible liftings, we obtained the values in table 5.11. Although the values in the table

Inequality	$\mathcal{Q}_1$	$\mathcal{Q}_2$	$\mathcal{Q}_3$
I_{3522}	64.62%	66.70%	66.71%

Table 5.11: Comparing the threshold values, it appears this inequality cannot beat CHSH.

---

[22]Labelling outcomes $1,2,3$.

for the higher levels of the hierarchy seem to suggest that this inequality does strictly worse than CHSH, this is simply down to imprecision in the solver. To show this, we can consider I_{3522} with Alice restricted to taking just the first and last measurements, and Bob limited to his third and final measurements. This gives us the restricted inequality:

$$I_{3522}^R = \begin{pmatrix} 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & 0 & \frac{1}{3} \end{pmatrix} \tag{5.93}$$

this is a CHSH inequality and so by restricting their choice of measurements, Alice and Bob can always achieve the CHSH threshold. It is for this same reason that I_{3322} and I_{2233} are able to achieve a threshold of 2/3.

5.6.5 Attempts to Improve Precision

In order to compare the known $(4, 4, 2, 2)$ boundary with our $(3, 3, 3, 3)$ candidates, I attempted to improve the precision of our semidefinite solver, in order that we may trust a higher number of decimal places. This list details some of my attempts to do this:

- **Increase the internal CVX precision.**

  The most natural solution to the problem is to alter the "cvx_precision" parameter, to hope for obtaining more precise results. When set to default, CVX states the problem is "solved" when a precision of 1.49×10^{-8} is achieved, and "inaccurately solved" if a precision of only 1.22×10^{-4} can be reached. By contrast, setting the cvx_precision parameter to "high" these figures become 1.82×10^{-12} and 1.35×10^{-6} respectively. Whilst we have seen that the values do not match up with the precision of our binary chop solutions, our hope is that this more stringent requirement will improve our overall values. The results of this change in parameter are presented in table 5.12. Contrary to the default precision setting, in which for all results presented the "solved" tag was returned, for the high precision setting the status "inaccurate/solved" was often returned - nevertheless, for $\mathcal{Q}_3$ the thresholds obtained matched the known thresholds up to at least 10^{-4}. Our two candidates $I_3^{A|B}$ match the lowest bound, $\left(\sqrt{5} - 1\right)/2$, up to this precision, suggesting that this may too be the detection threshold for these Bell inequalities. We see also that for this level of precision one of our I_{44422}^4 liftings does not match this threshold, suggesting that the lifting given by [104] is unique in achieving that value.

Inequality	$\mathcal{Q}_1$	$\mathcal{Q}_2$	$\mathcal{Q}_3$
CHSH	66.67%	66.67%	66.67%
I_{3322}	60.00%	66.50%	66.67%
I_{2233}	66.67%	66.66%	66.67%
I_{4422}^4 (Lift 1)	50.00%	61.80%	61.86%
I_{4422}^4 (Lift 2)	50.00%	61.80%	61.81%
I_{3522}	64.03%	66.67%	66.67%
I_3^A	50.00%	61.48%	61.80%
I_3^B	50.00%	60.54%	61.80%

Table 5.12: The obtained detection thresholds obtained with the cvx_precision set to "high".

Inequality	$\mathcal{Q}_1$	$\mathcal{Q}_2$	$\mathcal{Q}_3$
I_{4422}^4 (Lift 1)	50.00%	61.80%	61.86%
I_{4422}^4 (Lift 2)	50.00%	61.80%	61.81%
I_3^A	50.00%	61.48%	61.80%
I_3^B	50.00%	60.54%	61.80%

Table 5.13: The obtained detection thresholds obtained with the cvx_precision set to "best".

- **Further increase the internal CVX precision.**

  The third option for increasing cvx_precision is "best" - for this, the tags "solved" and "inaccurately solved" match those of the default setting, the values 1.49×10^{-8} and 1.22×10^{-4}; however instead of stopping once these precisions have been satisfied, instead CVX looks to increase precision by continuing to solve the problem until no progress can be made.

For this precision setting, the results for I_{4422}^4, I_3^A and I_3^B can be found in table 5.13. One can see that the values match those obtained with the "high precision" parametrisation; in fact, they are identical up to machine precision, implying the solver is performing identically in each scenario. Thus we can reasonably conclude

that we learn no more from the "best" option than we do from high precision, and thus the exact relationship between the thresholds of the two inequalities remains unclear.

- **Alternate solvers.**

  Included with CVX are two open-source semidefinite solvers: SDPT3 and SeDuMi. Running the optimal $(4, 4, 2, 2)$ liftings with these solvers, we obtained the results in table 5.14. Both solvers seem to match the accuracy of 10^-3 - but the solver

Solver	Lifting	Bound	Notes
SDPT3	1	61.86%	Solver returns "Inaccurate/Solved" tag
SDPT3	2	61.85%	Solver returns "Inaccurate/Solved" tag
SeDuMi	1	61.82%	Solver returns "Inaccurate/Solved" tag
SeDuMi	2	61.81%	Solver returns "Inaccurate/Solved" tag

Table 5.14: The detection thresholds returned by the pre-packaged CVX solvers; for each case the status "Inaccurate/Solved" is returned.

status is returned as "Inaccurate/Solved" - this corresponds to a precision value of $\sqrt{1.49 \times 10^{-8}} \approx 1.22 \times 10^{-4}$; as this precision is less than that when CVX returns a successful solution tag [23], these solvers only offer *at best* equal precision to MOSEK.

- **Allowance for local errors.**

  Judging by the obtained values, we see that the binary chop is homing in on a value higher than the true bound; we can conclude from this there is a critical $\eta^* > \left(\sqrt{5} - 1 \right)/2$ for which an erroneous solution $\boldsymbol{\pi}_{\eta^*}$ with $\text{Tr} \left[I_{4422}^{4T} \Pi_{\eta^*} \right] \geq 1$ is returned, and thus the binary chop algorithm concludes that $\eta_c \geq \eta^*$, and can never home in on the true value. To account for this, we could instead conclude our distribution is local only if $\text{Tr} \left[I_{4422}^{4T} \Pi_{\eta^*} \right] \geq 1 + \varepsilon$, with ε an error parameter. One could then "tune" this parameter ε, by performing a double binary chop.

  1. First set two bounds of ε; $\varepsilon_{\min} = 0$, $\varepsilon_{\max} = 0.1$. Then set $\varepsilon = (\varepsilon_{\min} + \varepsilon_{\max})/2$.

  2. Perform the standard binary chop algorithm to determine η_c, using the locality criterion $\text{Tr} \left[I_{4422}^{4T} \Pi_{\eta^*} \right] > 1 + \varepsilon$. This will gives us an ε based boundary, η_c^ε.

---

[23]Which occurs when using MOSEK.

3. If $\eta_c^\varepsilon > \eta_c = \left(\sqrt{5} - 1\right)/2$ then we conclude the same problem is still occurring, and thus we set $\varepsilon_{\min} = \varepsilon$. If $\eta_c^\varepsilon < \eta_c$ we conclude we are being "too generous", and set $\varepsilon_{\max} = \varepsilon$.

4. By repeating this process, we home in the most accurate error parameter.

Unfortunately, this tuning method fails consistently. This is because for all $\eta < \eta_c$, the local bound 1 is exactly achievable - thus the solver returns a value of $1 + \delta$ with $\delta \approx 10^{-8}$ (the solver's touted precision) regardless of proximity to η_c. Moreover, for values $\eta_c < \eta < 61.83$ (at level 2) our solver returns a value of the same form. Thus the tuning method homes in an arbitrary $\varepsilon \approx 10^{-8}$, which bears no relation to locality/non-locality.

- **Quadratic description**

  Another attempt to improve the precision was to improve the precision of the variables themselves. Hitherto to now, all calculations have been done using the standard double-precision format - in which each variable is stored using 64 bits. The semidefinite solver SDPT3 has an implementation which allows for quadratic-precision, which uses 128 bits instead. Unfortunately, this implementation is for C++ only - and thus required manual input of the semidefinite hierarchy conditions, as opposed to being able to use QETLAB. The standard double-description C++ failed to solve the semidefinite program due to "numerical inaccuracies", whilst the quadratic description was too slow; failing to converge to a solution on any reasonable time-scale.

- **Replacing CVX with YALMIP**

  YALMIP [70] is an extremely popular package which has been used to much success, in projects such as [104] and [16]. Thus, I decided to try the same problems, using YALMIP rather than CVX. Note that the *solver* choice, MOSEK, does not change; both CVX and YALMIP are toolboxes that process problems to feed to the solver.

  At the time of writing, a bug existed in the code, preventing the use of the QETLAB $\mathcal{Q}_i$ membership function with YALMIP. However, I could utilise the manual constraints generated for the C++ SDPT3. With these constraints, a critical efficiency of 61.92% was obtained for I_{4422}^4 in $\mathcal{Q}_2$, implying a worse precision than CVX (in this instance).

5.6.6 Quantum Realisations of the Detection Threshold

As stated in section 5.6.2, [104] explicitly gives a set of quantum states/measurements achieving the detection loophole threshold for I_{4422}^4 - this is how we may compare the performance of the semidefinite program to the true threshold value. With this in mind, and motivated by the numerical difficulties in implementing the quantum hierarchy, we turned to this method in order to try to determine the true value of the detection threshold for $I_3^{A/B}$.

This search can be implemented by the aid of semidefinite programming. Before we explain the full method, let us begin by explaining how one can use semidefinite programming to find the quantum state and measurements maximally violating a Bell inequality, B, for a (m_A, m_B, k_A, k_B) scenario.

Problem: minimise $\mathrm{Tr}[B^T \Pi]$, **subject to** $p(ab|xy) = \mathrm{Tr}[\rho M_{a|x} \otimes M_{b|y}]$.

1. Fix the dimension of the search. We constrain our search to quantum states $\rho \in \mathcal{H}_{d^2}$, and POVM measurement operators $M_{a|x}, M_{b|y}$ acting on $\mathcal{H}_d$.

2. Generate a random quantum state $\rho \in \mathcal{H}_{d^2}$. This can be done using the RandomDensityMatrix function in QETLAB.

3. Generate m_B random POVMs with k_B outcomes. This can be done using the RandomPOVM function in QETLAB. Note that this gives us $m_B k_B$ operators of size $d \times d$ which satisfy $M_{b|y} \geq 0$, $\sum_b M_{b|y} = \mathbb{I}_d$.

4. We can then formulate a semidefinite problem over Alice's POVM elements, $\{M_{a|x}\}$ - each of these must be positive semidefinite, which occurs iff the block diagonal matrix

$$M_A = \begin{pmatrix} M_{1|1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & M_{k_A|m_A} \end{pmatrix} \geq 0. \qquad (5.94)$$

Since we have fixed ρ, $\{M_{b|y}\}$, the objective $\mathrm{Tr}[B^T \Pi] = \sum_{a,b,x,y} B(ab|xy)p(ax|xy) = \sum_{a,b,x,y} \sum_{i,j} B(ab|xy)[\rho]_{ij}[M_{a|x} \otimes M_{b|y}]_{ji}$ is a linear function of M_A (though we do not explicitly give it here). Our constraints[24] of M_A are simply $\sum_a M_{a|x} = \mathbb{I}_d$, $\forall x$.

---

[24]These can be written in the canonical form by taking the equality elementwise.

5. We now repeat the same process with the elements $\{M_{b|y}\}$, using the $\{M_{a|x}\}$-defining M_A obtained from Step 4, and our state ρ.

6. Finally, we optimise the function $\mathrm{Tr}[B^T\Pi]$ over the density matrix ρ, using the M_A, M_B obtained. Since we require $\rho \geq 0$ in order for ρ to be a valid quantum state, this too forms a semidefinite program, with the constraint $\mathrm{Tr}[\rho] = 1$.

7. Repeat Steps 4-6, until the solution converges to a given precision.

8. Store the solution and optimal value, then return to Step 1 and repeat a set number of times.

This method is known as the "see-saw method" - whilst the objective value is always non-increasing for each given optimisation, it is *not* guaranteed to converge to the true minima - this is because the function $\mathrm{Tr}[B^T\Pi]$ is not *jointly* convex in the variables ρ, M_A, M_B. This means that is possible to get stuck in a local minima, where for each fixed pair of ρ, M_A, M_B no improvement can be made by varying the other variable, yet there exists a better minima which could be achieved by varying at least two of the three simultaneously. Unfortunately, without fixing two of the set one fails to obtain a semidefinite problem, due to the form of $p(ab|xy)$. To counter this, we start the see-saw at many random starting points in the solution space, with the hope that we "fall into" the correct minima for at least one of these starting points. By this same logic, we can never be sure we have truly found the global minima, unless we have a lower bound we may apply.

5.6.6.1 Application the Detection Loophole Problem

Let us return now to our problem - finding the detection threshold for our two Bell inequalities. One thing we wish to know to for these two inequalities is - do they beat the threshold $\eta_c = \left(\sqrt{5} - 1\right)/2$? Equivalently, does there exist a quantum distribution Π such that $\mathrm{Tr}[I_3^{(A/B)T}\Pi_{\eta_c}] < 1$? Since Π_{η_c} is linearly dependent only on Π (having fixed η), we can apply the see-saw method described above with the modified objective function $\mathrm{Tr}[I_3^{(A/B)T}\Pi_{\eta_c}]$.

Performing this using YALMIP[25] for $d = 9$, we were unable to find any solutions significantly different[26] from 1. Bearing in mind the numerical errors previously encountered,

---

[25] With MOSEK.

[26] Greater than machine precision.

we instead decided to increase $\eta = 0.65$ - with the aim of extrapolating the resultant state/measurements to those achieving the boundary value. However, that too failed to result in a non-local state. Thus could be due to four possibilities:

- *Numerical issues* - it is possible that still we are unable to achieve enough precision to obtain the small violation attained by the optimal states.

- *"Bad luck"* - we simply did not choose a good random starting point which will lead to the global minima.

- *Dimension* - it is possible the dimension required to achieve optimal violation is higher than 9 - there is evidence to suggest I_{3322} may require infinite dimensional states for optimal violation, so this possibility is not without precedent.

- *Hierarchy inaccuracy* - Despite the hierarchy appearing to converge on a value of $\approx 61.8\%$, it is possible that the true quantum boundary is higher than this, with violation $< 65\%$ being achieved by non-quantum hierarchy states.

Of these potential issues, only the second is actionable, and thus we tweak the see-saw method, in order to improve the likelihood of "falling down" the correct well.

Method 1: Triple Descent

1. Start with $\eta = 1$. Then perform the standard see-saw method, homing in on (hopefully) optimal ρ^1, M_A^1 and M_B^1.

2. Reduce η by a small step $\delta\eta$. Repeat the see-saw algorithm with this new η, but starting with ρ^1, M_A^1 and M_B^1 as the initial states and measurements.

3. repeat this process, for each η taking $\rho^{\eta+\delta\eta}$, $M_A^{\eta+\delta\eta}$ and $M_B^{\eta+\delta\eta}$ as the initial values, until an $\eta = \eta_L$ such that $\text{Tr}[B^T\Pi_{\eta_L}] \geq 1$ is obtained after convergence. For this run, the allowable efficiency is η_L.

4. Steps 1-3 are then repeated, with the minimal η_L taken over all runs.

The motivation behind this method is that for perfect efficiencies, violation of the inequality is significant, and thus easier for the see-saw to converge to. As the efficiency decreases, and the violation gets smaller, we begin with "good candidate" states and measurements more likely to obtain a violation. At the time of writing, using dimension 9 and $\delta\eta = 0.01$,

the lowest η_L I was able to achieve was 0.67.

    Method 2: Super See-Saw

1. Set $\eta = 1$, and $\rho^1 = |\Phi\rangle_d \langle\Phi|$. See-saw between optimising M_A and M_B, converging on $M_A^{1,1}$ and $M_B^{1,1}$.

2. Reduce η by $\delta\eta$, and see-saw again between M_A and M_B, starting with $M_A^{1,1}$ and $M_B^{1,1}$ and converging on $M_A^{1,\eta}$ and $M_B^{1,\eta}$.

3. Repeat the process, for each η taking $M_A^{1,\eta+\delta\eta}$ and $M_B^{1,\eta+\delta\eta}$ as the initial values, until reaching η_1 such that $\mathrm{Tr}[B^T \Pi_{\eta_1}] \geq 1$ is obtained after convergence.

4. Minimise $\mathrm{Tr}[B^T \Pi_{\eta_1}]$ over ρ, using measurements $M_A^{1,\eta_1+\delta\eta}$ and $M_B^{1,\eta_1+\delta\eta}$ from the previous η value. This gives an optimal state ρ_2.

5. For the new optimal state ρ_2, starting with $M_A^{1,\eta_1+\delta\eta}$ and $M_B^{1,\eta_1+\delta\eta}$ see-saw over the measurements to obtain optimal measurements $M_A^{2,\eta}$ and $M_B^{2,\eta}$.

6. repeat Steps 3-5 until an $\eta = \eta_L$ is achieved such that both the measurement see-saw and state minimisation give local (≥ 1) values.

We begin with the starting state $\rho_1 = |\Phi\rangle_d \langle\Phi|$ due to the results of [38] and [104], in which the states achieving the detection threshold were of the form $\sum_{i=0}^{d-2} \sqrt{(1-\varepsilon^2)/(d-1)} |ii\rangle + \varepsilon |d-1, d-1\rangle$ in the limit as $\varepsilon \to 0$ - the maximally entangled state is also of this form, for $\varepsilon = 1/\sqrt{d}$. Thus, the hope for this method is to "track" this optimal family of states along ε, as the detection efficiency decreases. The lowest η_L obtained with this method was $\eta_L = 0.66$, with $d = 9$ and $\delta\eta = 0.01$.

5.7 Summary

The use of Bell inequalities in practical quantum information is dominated by CHSH. This is because of the difficulty involved in preparing quantum states with dimension higher than 2, and performing non-binary measurements. At our current stage of development, viability is the primary defining factor of an experiment before efficiency. At such a stage, CHSH is the perfect candidate as it is the easiest to implement and allows experimentalists to test violation of locality.

This situation will not last forever though. Our techniques for generating, storing and measuring quantum states will become more and more refined. As this occurs, the trade-off between difficulty in implementation of higher dimensional scenarios, and the benefits they confer will shift. It was with this shift in mind that research such as [104] was performed - investigating the relatively low-dimensional ququart scenario, and showing it outperformed the "benchmark" inequality that is CHSH - at least in terms of the detection loophole.

It is with this low-dimensionality in mind that the research in this chapter has been performed. The inequality generating method that forms the first half of the chapter likely does not scale well as the dimensionality increases, due to it sometimes returning non-facet solutions. Nevertheless, for low dimensions it churns out thousands of inequalities with fundamentally different properties in a reasonable timescale. As the hunt for low-dimensional Bell inequalities with properties suitable to specific quantum information tasks intensifies - as I hope it will with the development of the field - then I believe this method could provide a useful tool.

A natural comparison to draw is with the more rigorous method of obtaining Bell inequalities via the simplex algorithm - by choosing a trivial objective function, one can use the simplex algorithm to pivot around the vertices of the polar dual of the local polytope. At every pivot, a new facet Bell inequality will be found. One can even pivot in such a way that every vertex will be visited exactly once. Yet the extreme degeneracy of the facet classes means that this method is not practical - relabellings of inequalities are of no interest to us. Whilst this method is certainly better in principle, it does not suit our purposes. I believe improving the way in which we search for Bell inequalities would be a reasonable direction for future research - given that one is only concerned with having a representative of each Bell inequality class, enumerating all facets is definitely not necessary. There are several symmetrical properties to be exploited - a simple example is that *every* local deterministic distribution $\mathbf{d}_i$ is a saturating point for at least one representative of *every* class - given a saturating point $\mathbf{d}_j$ for Bell inequality $\mathbf{b}$, the relabelling taking $\mathbf{d}_j \to \mathbf{d}_i$ will define a new inequality $\mathbf{b}'$ with $\mathbf{d}_i$ as a saturating point.

Of course, not all research is done with an entirely practical motivation, and I had

hoped to lower the detection threshold, by providing explicit states and measurements. Unfortunately, I was unable to do this, and only able to present evidence of many Bell inequalities which are likely *not* useful for this problem, as well as two candidate Bell inequalities for which there is evidence that they either satisfy the same threshold, or very slightly lower it. Trying to improve the precision of the semidefinite programming involved may help in determining this, and I plan to contact someone in the field of convex optimisation to achieve this.

There are other directions this work could be extended. All of the analysis in this chapter has been performed using liftings of inequalities, in which detection failure is mapped to a valid output. One could extend this analysis to treating detection failure as a separate outcome. Unfortunately, we can only do this rigorously for the $(2, 2, 2, 2)$ scenario since we know all $(2, 2, 3, 3)$ Bell inequalities. For this scenario we also know that all inefficient distributions $\boldsymbol{\pi}_\eta$ with $\eta = 2/3$ are local - as this is the detection threshold of CHSH, we can be sure that I_{2233}, the only new inequality for $(2, 2, 3, 3)$, will not improve on this. One could use our newly generated $(3, 3, 3, 3)$ inequalities to look at $(3, 3, 2, 2)$ distributions - dealing with binary outcome measurements only may lead to greater precision than our research on $(3, 3, 3, 3)$ distributions.

In summary, this chapter has given an algorithm which generates new facet inequalities to the local polytope. Using this, we have been able to generate thousands of new inequalities in a variety of measurement-output scenarios. We have analysed their efficacy in tackling the cryptographically-relevant problem of the "detection loophole" and shown that, whilst most of the new inequalities are likely worse than the current optimal inequality in the literature, there are two new candidates of a similar dimension for which there is evidence they may be able to marginally reduce the tolerable detection failure threshold.

Chapter 6

Conclusion and Summary

Entanglement is a complicated property to understand. For pure bipartite states it is easy to quantify, but for mixed states, the possibility of convex combinations of states increases the difficulty - so much so that the problem of determining bipartite separability has been shown to be NP-hard in general [50]. There is no one defining measure for entanglement - instead there exists a variety of measures satisfying certain well-reasoned properties. Some of these are operational, such as the distillable entanglement or entanglement cost, and require a difficult optimisation over all possible local protocols taken in the asymptotic limit. Others are algebraic, such as the relative entropy of entanglement (REE), or squashed entanglement, and the relation between these measures and their operational counterparts mean they can be used to bound the operational usefulness of states, or in some cases exactly determine it. These more abstract entanglement measures are also difficult to calculate in general; often requiring optimisation over infinite extensions or asymptotic regularisation, but for some states these problems simplify and values can be found. These problems are compounded when considering multi-partite scenarios. No longer can even a maximally entangled state be agreed on - and many of the most popular bipartite measures of entanglement do not generalise.

This thesis does not claim to solve any of the difficulties in understanding entanglement. Instead, it looks at using entanglement as a tool to investigate capacities of quantum channels. These capacities share many of the problems of entanglement: optimisation over general protocols, the possibility of infinite dimensional states, and asymptotic limits. Unlike with the entanglement of states however, there does not exist the same range of useful abstractions of capacity which we can instead consider. There exist definitions of relative

entropy of entanglement and squashed entanglement for channels, but they are calculated as the supremum taken over all possible output states, over all protocols. The advantage of channel simulation is to connect capacities of quantum channels to entanglement measures on states, in order to bound their value.

The limitation of channel simulation is that a specific LOCC simulation and resource state is required; currently the only protocol which has been explored in great detail is the teleportation protocol. Hopefully the range of protocols which are useful for channel simulation will increase, or even a constructive method discovered where, given a channel $\mathcal{E}$, one can then construct resource state $\sigma_{\mathcal{E}}$ and LOCC protocol Λ simulating the channel, allowing one to use the entanglement properties of $\sigma_{\mathcal{E}}$ to bound the two-way capacities of $\mathcal{E}$. An interesting open question is whether every channel is Choi-simulable, which would give $E_D(\chi_{\mathcal{E}}) = D_2(\mathcal{E})$ for all $\mathcal{E}$ - currently we know this only to be true for teleportation covariant channels [85].

The noisy teleportation protocol introduced in chapter 2 is an attempt to widen the range of simulable channels, by expanding the teleportation protocol with the addition of classical noise. This goal was achieved, allowing for the first time dimension-preserving non-Pauli channels to be simulated over a discrete-variable resource. We proved a no-go theorem on the structure of the resource state in order to allow non-Pauli simulation, and used it in order to choose a specific resource state, for which we completely characterised the set of possible resultant channels - the "Pauli-damping" family. This work could be extended to "noisy qudit teleportation" - however, we saw that for Pauli-damping channels this technique did not provide the best upper bound on secret-key capacity. This highlights the importance of not only showing a channel can be LOCC-simulated, but choosing a resource state which accurately conveys the properties of the channel. It is for this reason that the question of whether channels are Choi-simulable is so important. Furthermore, we were able to more accurately bound Pauli-damping channels due to their composite nature; therefore an investigation into the simulation of irreducible channels would also be productive.

Werner states have proved one of the most fruitful resources for teasing out the unusual features of entanglement. For qubits, Werner states are equivalent under local unitaries

to the isotropic states - a pseduo-mixture of a maximally entangled state and the identity matrix. For $d > 2$ though, the properties of the two classes diverge. Entangled isotropic states are always distillable, and entanglement measures are additive and easy to calculate. By contrast, Werner states exhibit subadditivity of the REE, with the result that for the parameter range $\eta \in [-1, -2/d)$ the regularised REE is not known. Moreover, the conjecture that bound entangled states exist with a negative partial transpose (NPT) has stood for around 20 years, with certain Werner states the primary candidates. There is much circumstantial evidence to support this conjecture but no one has been able to prove it - in part due to the range of possible distillation procedures. In chapter 3 we introduced a new family of states which generalised the Werner states by introducing a phase component. These new states share many similarities with Werner states, and we were able discern some of their entanglement properties to show that they provide new candidates for NPT bound entanglement, which may lead to new insights on the problem.

Connected to Werner states by channel-state duality, the Holevo-Werner channels were shown to be teleportation covariant in chapter 4, and consequently that the optimal parameter estimation may only scale with the shot-noise limit, the limit found in classical estimation theory. We found that entanglement-breaking was not a factor in parameter estimation of these channels, a property likely linked to the fact Werner states are simultaneously diagonalisable and so functions such as the quantum relative entropy simplify to their classical counterpart. Choice of parametrisation greatly affects the influence dimension has on estimation - our favoured choice of the expectation representation was dimension-independent, whilst the α-representation showed that estimation can improve or become more difficult for different regions as dimension increases. We were able to improve the bounds for binary discrimination of Holevo-Werner channels, by providing the analytic form on the quantum Chernoff bound, and extending it to the depolarising channel also.

In the second part of chapter 4, we applied entanglement-based upper bounds to the secret key capacity of Holevo-Werner channels. By considering four different bounds we were able to restrict the capacity more tightly than any one of them individually. One of the entanglement measures used was the two-copy REE, calculated for general Werner states here for the first time. We also used two alternative upper bounds on the squashed

entanglement; a suitable choice as it is additive. Further work in this direction could be to investigate the n-copy REE, $n > 2$ - it has been shown for the extremal Werner state the the regularised REE and RPPT do *not* coincide [21] - but it is not known for which n this first occurs. Another valid line of research would be to try to find the squashed entanglement of entangled Werner states - this is not known even for the extremal Werner state, though the upper bound provided by [21] coincides with the logarithmic negativity and regularised RPPT for even dimensions, and thus is conjectured to be tight. Both of these directions would improve the secret-key capacity bounds presented here. At this time, it is unknown how the secret-key capacity of these channels relates to the conjecture of NPT bound entanglement - although proof of the conjecture would prove the corresponding Holevo-Werner channels have 0 two-way distillation capacity.

In the last chapter we looked at another important aspect of quantum information, non-locality. By exploiting extremal no-signalling distributions - which are non-physical - we were able to enumerate new Bell inequality classes, which may have useful applications in quantum protocols; particularly in the field of device independent cryptography, where Bell inequalities are used to ensure non-locality, and thus the security, of the device. Although the algorithm we presented does return non-facet inequalities, the reduction in relabelling-equivalent solutions means that for lower dimensional cases were able to generate tens of thousands of new non-equivalent Bell inequality classes. A limitation of this algorithm is that it does not terminate when all inequivalent Bell inequalities have been found - meaning we could only provide a complete list of class representatives in scenarios where the number of classes had been previously determined.

Finally, we turned to the problem of the "detection loophole" - a practically motivated problem where detection failure means that outcome distributions cannot be trusted to be non-local. First we provided tighter bounds on the critical efficiency for which below that detection rate no distribution can be trusted, using the extremal no-signalling points to do this. Then, motivated by obtaining the lowest possible efficiency for which non-local distributions may still be determined with a low number of measurements/outcomes, we tested all the enumerated Bell inequalities to see if they could better the current optimal equality - which can discern non-local distributions down to 61.80% detection rate. We found two candidate Bell inequalities which appear likely to match this bound, or perhaps

even beat it by a small value, although were unable to determine the exact threshold for these inequalities, or provide an explicit quantum state/measurements construction achieving it. Since determining the threshold of an inequality is strongly linked to the use of semidefinite programming, I hope to improve this using insight from an expert in the field. We also restricted our testing to the case where detection failure is mapped to a possible successful outcome - treating it as a separate outcome is an avenue of future research.

This thesis has shown new insights into areas of research including channel simulation, entanglement distillation and generation of Bell inequalities. The work presented here may prove useful when considering larger problems such as the non-trivial simulation of *all* quantum channels, whether all NPT states are distillable, and the minimal detection efficiency for which we can close the detection loophole.

Abbreviations

LOCC Local Operations and Classical Communication

PVM Projective Valued Measure

POVM Positive Operator-Valued Measure

PPT Positive Partial Transpose

NPT Negative Partial Transpose

REE Relative Entropy of Entanglement

RPPT Relative entropy of entanglement, with respect to Positive Partial Transpose

CCQ Correlated Classical Quantum

QCB Quantum Chernoff Bound

HW Holevo-Werner

PD Pauli-Damping

LP Linear Program

SDP SemiDefinite Program

EPR Einstein-Podolsky-Rosen

KKT Karush–Kuhn–Tucker

KL Kullback-Leibler

CHSH Clauser-Horne-Shimony-Holt

CH Clauser-Horne

NP Non-deterministic Polynomial time

References

[1] Mafalda Almeida, Stefano Pironio, Jonathan Barrett, Geza Toth, and Antonio Acín. Noise robustness of the nonlocality of entangled quantum states. *Physics Review Letters*, 99:040403, 2007.

[2] MOSEK ApS. *The MOSEK optimization toolbox for Matlab manual. Version 8.1.*, 2017.

[3] Alain Aspect, Philippe Grangier, and Gérard Roger. Experimental tests of realistic local theories via Bell's theorem. *Physical Review Letters*, 47(7):460–463, 1981.

[4] Koenraad Audenaert, John Calsamiglia, Ramon Muñoz-Tapia, Emilio Bagan, Lluís Masanes, Antonio Acín, and Frank Verstraete. Discriminating states: The quantum chernoff bound. *Physical Review Letters*, 98:160501, 2007.

[5] Koenraad Audenaert, Jens Eisert, Enric Jane, Martin Plenio, Shashank Virmani, and Bart de Moor. The asymptotic relative entropy of entanglement. *Physical Review Letters*, 87:217902, 2001.

[6] David Avis and Komei Fukuda. A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra. *Discrete and Computational Geometry*, 8:295–313, 1992.

[7] Jonathan Barrett. Nonsequential positive-operator-valued measurements on entangled mixed states do not always violate a Bell inequality. *Physical Review A*, 65:042302, 2002.

[8] John Bell. On the Einstein Podolsky Rosen paradox. *Physics*, 1:195–200, 1964.

[9] Giuliano Benenti and Giuliano Strini. Computing the distance between quantum channels: Usefulness of the Fano representation. *Journal of Mathematical Physics B*, 43:215508, 2010.

[10] Charles Bennett and Gilles Brassard. Quantum cryptography: Public key distribution and coin tossing. In *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*, volume 175, page 8, 1984.

[11] Charles Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William Wootters. Teleporting an unknown quantum state via dual classical and Einstein–Podolsky–Rosen channels. *Physical Review Letters*, 70:1895, 1993.

[12] Charles Bennett, David DiVincenzo, John Smolin, and William Wootters. Mixed-state entanglement and quantum error correction. *Physical Review A*, 54-5:3824–3851, 1996.

[13] Garry Bowen and Sougato Bose. Teleportation as a depolarizing quantum channel, relative entropy, and classical capacity. *Physical Review Letters*, 87-26:267901, 2001.

[14] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[15] Samuel Braunstein and Carlton Caves. Statistical distance and the geometry of quantum states. *Physical Review Letters*, 72:3439–3443, 1994.

[16] Daniel Cavalcanti, Leonardo Guerini, Rafael Rabelo, and Paul Skrzypczyk. General method for constructing local hidden variable models for entangled quantum states. *Physical Review Letters*, 117:190401, 2016.

[17] Matthias Christandl. *The Structure of Bipartite Quantum States - Insights from Group Theory and Cryptography*. PhD thesis, University of Cambridge, February 2006.

[18] Matthias Christandl, Artur Ekert, Michal Horodecki, Pawel Horodecki, Jonathan Oppenheim, and Renato Renner. Fisher information in quantum statistics. In *Proceedings of the 4th Theory of Cryptography Conference, Lecture Notes in Computer Science*, volume 4392, pages 456–478, 2007.

[19] Matthias Christandl and Roberto Ferrara. Private states, quantum data hiding and the swapping of perfect secrecy. *Physical Review Letters*, 119:220506, 2017.

[20] Matthias Christandl and Andreas Winter. Squashed entanglement - an additive entanglement measure. *Journal of Mathematical Physics*, 45:829, 2003.

[21] Mattias Christandl, Norbert Schuh, and Andreas Winter. Entanglement of the antisymmetric state. *Communications in Mathematical Physics*, 311:397–422, 2012.

[22] John Clauser and Michael Horne. Experimental consequences of objective local theories. *Physical Review D*, 10:526–535, 1974.

[23] John Clauser, Michael Horne, Abner Shimony, and Richard Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23:880–884, 1969.

[24] Valerie Coffman, Joydip Kundu, and William Wootters. Distributed entanglement. *Physical Review A*, 61:052306, 2000.

[25] Andrea Coladangelo and Jalex Stark. Unconditional separation of finite and infinite-dimensional quantum correlations. `https://arxiv.org/abs/1804.05116`, 2018.

[26] Roger Colbeck. *Quantum And Relativistic Protocols For Secure Multi-Party Computation*. PhD thesis, University of Cambridge, 2006.

[27] Daniel Collins and Nicolas Gisin. A relevant two qubit Bell inequality inequivalent to the CHSH inequality. *Journal of Physics A*, 37:1775–1787, 2004.

[28] Daniel Collins, Nicolas Gisin, Noah Linden, Serge Massar, and Sandu Popescu. Bell inequalities for arbitrarily high-dimensional systems. *Physics Review Letters*, 88:040404, 2002.

[29] Thomas Cope, Kenneth Goodenough, and Stefano Pirandola. Converse bounds for quantum and private communication over Holevo-Werner channels. *Journal of Physics A*, 51:494001, 2018.

[30] Thomas Cope, Leon Hetzel, Leonardo Banchi, and Stefano Pirandola. Simulation of non-Pauli channels. *Physical Review A*, 96:022323, 2017.

[31] Thomas Cope and Stefano Pirandola. Adaptive estimation and discrimination of Holevo-Werner channels. *Quantum Measurements and Metrology*, 4:44–52, 2017.

[32] George Dantzig. Maximization of a linear function of variables subject to linear inequalities. *Activity Analysis of Production and Allocation*, page 339–347, 1947.

[33] Michel Deza and Monique Laurent. *Geometry of Cuts and Metrics*, volume 15. Springer-Verlag Berlin Heidelberg, 2010.

References

[34] Michel Deza and Mathieu Sikorić. Enumeration of the facets of cut polytopes over some highly symmetric graphs. *International Transactions in Operational Research*, 23:853–860, 2015.

[35] David DiVincenzo, Peter Shor, John Smolin, Barbara Terhal, and Ashish Thapliyal. Evidence for bound entangled states with negative partial transpose. *Physics Review A*, 61:062312, 2000.

[36] Matthew Donald, Michal Horodecki, and Oliver Rudolph. The uniqueness theorem for entanglement measures. *Journal of Mathematical Physics*, 43:4252–4272, 2002.

[37] Wolfgang Dür, Guifre Vidal, and Juan Ignacio Cirac. Three qubits can be entangled in two inequivalent ways. *Physical Review A*, 62:062314, 2000.

[38] Philippe Eberhard. Background level and counter efficiencies required for a loophole-free Einstein-Podolsky-Rosen experiment. *Physical Review A*, 47:R747–R750, 1993.

[39] Albert Einstein, Boris Podolsky, and Nathan Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, 47:777–780, 1935.

[40] Hamza Fawzi and Omar Fawzi. Efficient optimization of the quantum relative entropy. *Journal of Physics A*, 51:154003, 2018.

[41] Hamza Fawzi, James Saunderson, and Pablo Parrilo. Semidefinite approximations of the matrix logarithm. *Foundations of Computational Mathematics*, 2018. Package cvxquad at `https://github.com/hfawzi/cvxquad`.

[42] Marcel Froissart. Constructive generalization of Bell's inequalities. *Il Nuevo Cimento B*, 64:241–251, 1981.

[43] Raúl García-Patrón, Stefano Pirandola, Seth Lloyd, and Jeffrey Shapiro. Reverse coherent information. *Physical Review Letters*, 102:210501, 2009.

[44] Anupam Garg and Nathaniel Mermin. Detector inefficiencies in the Einstein-Podolsky-Rosen experiment. *Physical Review D*, 35:3831, 1987.

[45] M. Giustina, M. A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, Fabian Steinlechner, Johannes Kofler, Jan-Ake Larsson, Carlos Abellan,

Waldimar Amaya, Valerio Pruneri, Morgan W. Mitchell, Jörn Beyer, Thomas Gerrits, Adriana E. Lita, Lynden K. Shalm, Sae Woo Nam, Thomas Scheidl, Rupert Ursin, Bernhard Wittmann, and Anton Zeilinger. Significant-loophole-free test of Bell's theorem with entangled photons. *Physical Review Letters*, 115:250401, 2015.

[46] Michael Grant and Stephen Boyd. Graph implementations for nonsmooth convex programs. In *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008.

[47] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, 2014.

[48] David Gross, Koenraad Audenaert, and Jens Eisert. Evenly distributed unitaries: on the structure of unitary designs. *Journal Maths Physics*, 48:052104, 2007.

[49] Branko Grünbaum. *Convex Polytopes*. Springer Science and Business Media New York Academic Press, 1967.

[50] Leonid Gurvits. Classical deterministic complexity of Edmonds' problem and quantum entanglement. In *Proceedings of the 35th ACM Symposium on Theory of Computing*, 2003.

[51] Leonid Gurvits and Howard Barnum. Largest separable balls around the maximally mixed bipartite quantum state. *Physics Review A*, 66:062311, 2002.

[52] Matthew Hastings. Superadditivity of communication capacity using entangled inputs. *Nature Physics*, 5:255, 2009.

[53] Carl Helstrom. *Quantum detection and estimation theory*. New York Academic Press, 1976.

[54] B. Hensen, H. Bernien, A. E. Dréau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. L. Vermeulen, R. N. Schouten, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, M. Markham, D. J. Twitchen, D. Elkouss, S. Wehner, T. H. Taminiau, and R. Hanson. Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. *Nature*, 546:682–686, 2015.

[55] Karol Horodecki, Michal Horodecki, Pawel Horodecki, and Jonathan Oppenheim. Secure key from bound entanglement. *Physical Review Letters*, 94:160502, 2005.

[56] Karol Horodecki, Michal Horodecki, Pawel Horodecki, and Jonathan Oppenheim. General paradigm for distilling classical key from quantum states. *IEEE Transactions on Information Theory*, 55:1898, 2009.

[57] Michal Horodecki and Pawel Horodecki. Reduction criterion of separability and limits for a class of protocols of entanglement distillation. *Physical Review A*, 59:4206–4216, 1999.

[58] Michal Horodecki, Pawel Horodecki, and Rsyzard Horodecki. Separability of mixed states: Necessary and sufficient conditions. *Physical Letters A*, 223, 1996.

[59] Michal Horodecki, Pawel Horodecki, and Rsyzard Horodecki. Inseparable two spin-$\frac{1}{2}$ density matrices can be distilled to a singlet form. *Physical Review Letters*, 76:574–577, 1997.

[60] Pawel Horodecki. Separability criterion and inseparable mixed states with positive partial transposition. *Physical Letters A*, 232:333–339, 1997.

[61] Ryszard Horodecki, Pawel Horodecki, Michal Horodecki, and Karol Horodecki. Quantum entanglement. *Review of Modern Physics*, 81, 2009.

[62] Yichen Huang. Computing quantum discord in NP-complete. *New Journal of Physics*, 16:033027, 2014.

[63] Satoshi Ishizaka and Tohya Hiroshima. Asymptotic teleportation scheme as a universal programmable quantum processor. *Physical Review Letters*, 101:240501, 2008.

[64] Nathaniel Johnston. QETLAB: A Matlab toolbox for quantum entanglement, version 0.9. `http://qetlab.com`, 2016.

[65] Nick Jones and Lluiś Masanes. Interconversion of nonlocal correlations. *Physical Review A*, 72:052312, 2005.

[66] Dagomir Kaszlikowski, Leong Kwek, Jing-Ling Chen, Marek Żukowski, and Choo Oh. Clauser-Horne inequality for three-state systems. *Physical Review A*, 65:032118, 2002.

[67] Leonid Khachiyan, Endre Boros, Konrad Borys, Khaled Elbassioni, and Vladimir Gurvich. Generating all vertices of a polyhedron is hard. *Discrete Computation Geometry*, 39:174–190, 2008.

[68] Debbie Leung and William Matthews. On the power of PPT-preserving and non-signalling codes. *IEEE Transactions on Information Theory*, 61:4486–4499, 2015.

[69] Seth Lloyd. Capacity of the noisy quantum channel. *Physical Review A*, 55:1613, 1997.

[70] Johan Löfberg. YALMIP : A toolbox for modeling and optimization in Matlab. In *In Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.

[71] Stefan Lörwald and Gerhard Reinelt. PANDA: a software for polyhedral transformations. *EURO Journal on Computational Optimization*, pages 1–12, 2015.

[72] Serge Massar. Nonlocality, closing the detection loophole, and communication complexity. *Physical Review A*, 65:032121, 2002.

[73] Serge Massar and Stefano Pironio. Violation of local realism vs detection efficiency. *Physical Review A*, 68:062109, 2003.

[74] Dominic Mayers and Andrew Yao. Quantum cryptography with imperfect apparatus. In *IEEE Symposium on Foundations of Computer Science*, 1998.

[75] Peter McMullen and Geoffrey Shephard. *Convex Polytopes and the Upper Bound Conjecture.* Cambridge University Press, 1971.

[76] Miguel Navascues, Stefano Pironio, and Antonio Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New Journal of Physics*, 10:073013, 2008.

[77] Jorge Nocedal and Stephen Wright. *Numerical Optimization.* Springer Series in Operations Research and Financial Engineering, 2006.

[78] Károly Pál and Tamás Vértesi. Maximal violation of the I3322 inequality using infinite dimensional quantum systems. *Physics Review A*, 82:022116, 2010.

[79] Károly Pál and Tamás Vértesi. Quantum bounds on Bell inequalities. *Physics Review Letters*, 104:060401, 2010.

[80] Phillip Pearle. Hidden-variable example based upon data rejection. *Physics Review D*, 2:1418–1425, 1970.

[81] Asher Peres. Seperability criterion for density matrices. *Physical Review Letters*, 77:1413, 1996.

[82] Stefano Pirandola, Samuel Braunstein, Riccardo Laurenza, Carlo Ottaviani, Thomas Cope, Gaetana Spedalieri, and Leonardo Banchi. Theory of channel simulation and bounds for private communication. *Quantum Science and Technology*, 3:035009, 2018.

[83] Stefano Pirandola, Raúl García-Patrón, Samuel Braunstein, and Seth Lloyd. Direct and reverse secret-key capacities of a quantum channel. *Physical Review Letters*, 102:050503, 2009.

[84] Stefano Pirandola, Riccardo Laurenza, and Cosmo Lupo. Fundamental limits to quantum channel discrimination. `https://arxiv.org/abs/1803.02834`, 2018.

[85] Stefano Pirandola, Riccardo Laurenza, Carlo Ottoviani, and Leonardo Banchi. Fundamental limits of repeaterless quantum communication. *Nature Communications*, 8:15043, 2017.

[86] Stefano Pirandola and Cosmo Lupo. Ultimate precision of adaptive noise estimation. *Physical Review Letters*, 118:100502, 2017.

[87] Stefano Pironio. *Aspects of Quantum Non-locality.* PhD thesis, Université Libre de Bruxelles, 2004.

[88] Martin Plenio and Shashank Virmani. An introduction to entanglement measures. *Quantum Information and Computation*, 7:1–51, 2007.

[89] Sandu Popescu and Daniel Rohrlich. Nonlocality as an axiom. *Foundations of Physics*, 24:379–385, 1994.

[90] Ralph Rockafellar. *Convex Analysis.* Princeton University Press, 1970.

[91] Benjamin Schumacher and Michael Nielsen. Quantum data processing and error correction. *Physical Review A*, 54:2629, 1996.

[92] Sacha Schwarz, Bänz Bessire, André Stefanov, and Yeong-Cherng Liang. Bipartite Bell inequalities with three ternary-outcome measurements—from theory to experiments. *New Journal of Physics*, 18:035001, 2016.

[93] Lynden K. Shalm, Evan Meyer-Scott, Bradley G. Christensen, Peter Bierhorst, Michael A. Wayne, Martin J. Stevens, Thomas Gerrits, Scott Glancy, Deny R. Hamel, Michael S. Allman, Kevin J. Coakley, Shellee D. Dyer, Carson Hodge, Adriana E. Lita, Varun B. Verma, Camilla Lambrocco, Edward Tortorici, Alan L. Migdall, Yanbao Zhang, Daniel R. Kumor, William H. Farr, Francesco Marsili, Matthew D. Shaw, Jeffrey A. Stern, Carlos Abellan, Waldimar Amaya, Valerio Pruneri, Thomas Jennewein, Morgan W. Mitchell, Paul G. Kwiat, Joshua C. Bienfang, Richard P. Mirin, Emanuel Knill, and Sae Woo Nam. Strong loophole-free test of local realism. *Physical Review Letters*, 115:250402, 2015.

[94] Claude Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28:656–715, 1949.

[95] Peter Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Review*, 41:303–332, 1997.

[96] William Slofstra. Tsirelson's problem and an embedding theorem for groups arising from non-local games. `https://arxiv.org/abs/1606.03140`, 2016.

[97] Barbara Synak-Radtke and Michal Horodecki. On asymptotic continuity of functions of quantum states. *Journal of Physics A*, 39, 2006.

[98] Tamás Terlaky. A convergent criss-cross method. *Optimization: A Journal of Mathematical Programming and Operations Research*, 16:683–690, 1985.

[99] W. Tittel, J. Brendel, B. Gisin, T. Herzog, H. Zbinden, and N. Gisin. Experimental demonstration of quantum correlations over more than 10 km. *Physical Review A*, 57:3229–3232, 1998.

[100] Kim-Chuan Toh, Michael Todd, and Reha Tutuncu. SDPT3 — a Matlab software package for semidefinite programming. *Optimization Methods and Software*, 11:545–581, 1999.

[101] Géza Tóth and Iagoba Apellaniz. Quantum metrology from a quantum information science perspective. *Journal of Physics A*, 47:424006, 2014.

[102] Boris Tsirelson. Bell inequalities and operator algebras. `https://www.tau.ac.il/~tsirel/download/bellopalg.pdf`, 2006.

References

[103] Vlatko Vedral and Martin Plenio. Entanglement measures and purification procedures. *Physical Review A*, 57:1619–1633, 1998.

[104] Tamás Vértesi, Stefano Pironio, and Nicolas Brunner. Closing the detection loophole in Bell experiments using qudits. *Physics Review Letters*, 104:060401, 2010.

[105] Guifre Vidal and Reinhard Werner. Computable measure of entanglement. *Physical Review A*, 65:032314, 2002.

[106] Karl Vollbrecht and Reinhard Werner. Entanglement measures under symmetry. *Physical Review A*, 64:062307, 2001.

[107] Gordon Walsh. *An Introduction to Linear Programming*. Holt, Rinehart and Winston Ltd., 1971.

[108] Zhe Min Wang. A finite conformal-elimination free algorithm over oriented matroid programming. *Chinese Annals of Mathematics*, 8:120–125, 1987.

[109] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018.

[110] Reinhard Werner. Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model. *Physical Review A*, 40:4277–4281, 1989.

[111] Reinhard Werner. All teleportation and dense coding schemes. *Journal of Physics A*, 34:7081–7094, 2001.

[112] William Wootters and Wojciech Zurek. A single quantum cannot be cloned. *Nature*, 299:802–803, 1982.

[113] Marek Zukowski, Dagomir Kaszlikowski, Adam Baturo, and Jan Åke Larsson. Strengthening the Bell theorem: Conditions to falsify local realism in an experiment. `https://arxiv.org/abs/quant-ph/9910058`, 1999.