Technische Universiteit Delft
Faculteit Elektrotechniek, Wiskunde en Informatica
Faculteit Technische Natuurwetenschappen

Optimizing quantum entanglement distillation

Verslag als onderdeel ter verkrijging

van de graad van

BACHELOR OF SCIENCE
in
TECHNISCHE WISKUNDE &
TECHNISCHE NATUURKUNDE

door

Thomas Schiet (4181484)

Delft, Nederland
December, 2016

BSc verslag TECHNISCHE WISKUNDE

"Optimizing quantum entanglement distillation"

Thomas Schiet

Technische Universiteit Delft

Begeleiders

Prof. S.D.C. Wehner

Dr. D.C. Gijswijt

Overige commissieleden

Dr. J.L.A. Dubbeldam

Dr. M. Blaauboer

Dr. J.A.M. de Groot

December, 2016

Delft

Abstract

Quantum entanglement is a physical resource that is essential for many quantum information processing tasks, such as quantum communication and quantum computing. Although entanglement is essential for practical implementations in those fields, it is hard to create and transmit entanglement reliably. External factors introduce noise which may destroy or weaken the entanglement. Consequently, there is a need for methods to improve entanglement.

Entanglement distillation attempts to solve this problem. Entanglement distillation is a process where probabilistically from a fixed number of copies of noisy entangled states, a smaller number of more strongly entangled states is created. This is done using only local operations and classical communication. Various protocols are known to perform distillation. However, for many it is unknown whether better results are possible. The goal of this thesis is to show whether known protocols are optimal.

The greatest amount of entanglement achievable by entanglement distillation can be expressed as a non-convex optimization problem over separable quantum states. This problem is further relaxed to a semidefinite program which will yield upper bounds on performance of the distillation for specific input states.

The program is applied to various states that occur in experimental setups. Two known protocols are shown to perform on the upper bound, thus, being optimal.

Using a heuristic algorithm, we look for new protocols. Here the optimization is done iteratively over one quantum state at a time. However, this method did not result in any useful protocols.

Contents

Introduction

Entanglement is one of the fundamental properties of quantum mechanics and historically, has been a source of much debate. Its counterintuitive properties, that have no classical counterpart, have puzzled generations of physicists. Initially, the debate on entanglement was of a more philosophical nature, but recently the discussion has shifted to its applications. Entanglement is employed in many promising fields such as quantum communication, dense coding and quantum computing.

Quantum entanglement is particularly resourceful, however, it is difficult to entangle systems. Furthermore, the distribution of entangled states has proven to be challenging. The transmitted states are perturbed due to inevitable noise sources, resulting in a loss of entanglement. For quantum communication to be practical, transmission has to occur over large distances. For example, communication may occur transatlantically or between satellites. The effects of noise accumulate over distance, leading to a possibly crippling loss of entanglement.

Therefore, it is of great importance to increase the entanglement. Protocols exist to increase entanglement in weakly entangled systems. One way of doing this is by using *entanglement distillation*. Here, n known quantum states with some entanglement are transformed into $m < n$ more strongly entangled states. This may be done probabilistically, where measurement on a subsystem indicates success or failure of the distillation.

Because the communication may occur over such long distances, we are interested in simple protocols. Specifically, the protocols should involve just one round of local quantum operations and classical communication. Multiple protocols are known, however for most it has been an open question whether these protocols are optimal or if they can be improved upon.

In this thesis numerical methods are developed to answer this question and the optimality of two protocols is shown. The performance of the protocols is quantified by the fidelity of the output states to the maximally entangled state. The fidelity is expressed as the optimal value of a non-convex optimisation problem over separable quantum states. This problem is then relaxed to a semidefinite program, yielding upper bounds on the fidelity. This program optimizes over states that are positive under partial transposition, a criterion for separable states.

In addition, a method is investigated to search for new protocols. This method uses a heuristic approach to the non-convex problem by iteratively solving SDPs.

Theory

2.1 Fundamentals of quantum mechanics

2.1.1 Bra-ket notation

In quantum mechanics bra-ket notation [17] is widely used. In this form of notation a vector in a vector space V is a "ket" $|\cdot\rangle \in V$ and its dual is a "bra" $\langle\cdot| \in V^*$.

Definition 1. *Let V be a vector space. Its **dual vector space** is defined as $\mathcal{L}(V, \mathbb{R})$, the set of linear functions from V to $\mathbb{R}$.*

As an example of a dual vector space, consider $V = \mathbb{R}^n$ with the standard basis $\{|e_1\rangle, \ldots, |e_n\rangle\}$. The dual vector space V^* has basis $\{\langle x^1|, \ldots, \langle x^n|\}$ which consists linear functions from V to $\mathbb{R}$ such that $\langle x^i| (|e_j\rangle) = \delta_{ij}$. Because every $\langle x^i|$ is linear, note that it is a function that returns the i-th component of its argument. As such, $|\phi\rangle \in V$ has a corresponding dual $\langle\phi| \in V^*$ that is just a map

$$\langle\phi| : V \to \mathbb{R}, |x\rangle \mapsto \langle\langle\phi|, |x\rangle\rangle, \tag{2.1}$$

where $\langle\cdot,\cdot\rangle$ indicates the inner product. In bra-ket notation this inner product is written as $\langle\phi|x\rangle$. In the vector space $\mathbb{C}^n$, the dual of an element is its Hermitian conjugate.

2.1.2 Quantum states

A quantum state is a normalized vector $|\phi\rangle \in \mathcal{H}$ where $\mathcal{H}$ is a Hilbert space [17].

Definition 2. *A **Hilbert space** $\mathcal{H}$ is an inner product space which is complete under the norm induced by its inner product.*

In this thesis, any set denoted by calligraphic 'H', $\mathcal{H}$, will be implied to be a Hilbert space. This is the space that all the quantum states live in. Examples of commonly used Hilbert spaces are $\mathbb{C}^n$ and the square integrable functions L^2. The quantum state describes the state of a system. For example we can describe the spin of an electron measured along the z-axis. An electron, having spin $1/2$, can either be measured to be spin up or spin down. We could say for example that $|\uparrow\rangle = \begin{pmatrix}1\\0\end{pmatrix}$ corresponds to spin up, whereas $|\downarrow\rangle = \begin{pmatrix}0\\1\end{pmatrix}$ corresponds to spin down. Note that $|\uparrow\rangle$ and $|\downarrow\rangle$ form a basis for the Hilbert space $\mathcal{H} = \mathbb{C}^2$. A quantum state may be in a superposition. For example

$$|\phi\rangle = \alpha|\uparrow\rangle + \beta|\downarrow\rangle, \tag{2.2}$$

where $|\alpha|^2 + |\beta|^2 = 1$. States that live in a two-dimensional space are called **qubits**. To describe what would happen if we were to measure the spin of the system $|\phi\rangle$, we need to introduce the observable.

Definition 3. *An **observable** is a linear operator $\hat{A} : \mathcal{H} \to \mathcal{H}$ that is self-adjoint.*

This observable describes a physical property of a system. One of the postulates of quantum mechanics, is that the eigenvalues of $\hat{A}$ are the only values that we can measure. Let λ be an arbitrary eigenvalue of A with normalized eigenvector $|\psi\rangle$. If we measure a system that is in state $|\phi\rangle$, the probability to measure λ is, due to the postulates of quantum mechanics, given by $|\langle\psi|\phi\rangle|^2$. The **Pauli matrices** are well-known observables defined as

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.3}$$

In the case of the electron spin, they relate to the basis or spatial direction the spin in measured in. In this case we see that the eigenspaces of σ_z are precisely span$\{|\uparrow\rangle\}$ with eigenvalue 1 and span$\{|\downarrow\rangle\}$ with eigenvalue -1. So if we do a measurement along the z-axis, the observable σ_z must be used and we can only measure either 1 or -1. If we measure 1, the system after measurement is in the eigenstate $|\uparrow\rangle$. Alternatively if we measure -1 the new state is $|\downarrow\rangle$.

As an example, consider the state

$$|\phi\rangle = \frac{1}{\sqrt{2}}|\uparrow\rangle + \frac{1}{\sqrt{2}}|\downarrow\rangle. \tag{2.4}$$

If we measure $|\phi\rangle$ in the z-direction, the probability to measure 1 equals

$$|\langle\uparrow|\phi\rangle|^2 = \left|\frac{1}{\sqrt{2}}\right|^2 = 1/2, \tag{2.5}$$

and equivalently we measure -1 with probability 1/2. However, if we were to measure $|\phi\rangle$ on the x-axis, we will always measure 1. This is because the eigenvalues of σ_x are 1 and -1 with their respective eigenvectors

$$|+\rangle = \frac{1}{\sqrt{2}}|\uparrow\rangle + \frac{1}{\sqrt{2}}|\downarrow\rangle, |-\rangle = \frac{1}{\sqrt{2}}|\uparrow\rangle - \frac{1}{\sqrt{2}}|\downarrow\rangle. \tag{2.6}$$

And $|\phi\rangle = |+\rangle$, so $|\langle\phi|+\rangle|^2 = 1$.

2.1.3 Density matrix

Due to lack of information it might be unclear what state a system is in, but we may probabilistically describe the state that the system is in. For example, we may have a machine that produces the state $|\phi\rangle$ 90% of the time, but in the other cases the machine produces $|\psi\rangle$. To describe such systems, the density matrix is introduced.

Definition 4. *Let $p_1, p_2, \ldots, p_n$ be the probabilities that a system is in state $|\phi_i\rangle \in \mathcal{H}$ where the probabilities sum up to 1. The **density matrix** is a linear operator on $\mathcal{H}$, $\rho \in \mathcal{L}(\mathcal{H})$ that equals*

$$\rho = \sum_{i=1}^{n} p_i |\phi_i\rangle\langle\phi_i|. \tag{2.7}$$

The set of all density matrices on $\mathcal{H}$ is denoted $D(\mathcal{H})$.

The system is said to be in a **pure state** if ρ can be written as $|\phi\rangle\langle\phi|$ for some state $|\phi\rangle$. In this case, due to eigendecomposition, a pure state ρ has only one non-zero eigenvalue. If the state is not pure, it is called a **mixed state**.

Lemma 1. *A density matrix ρ has the following properties:*

1. $\operatorname{tr}(\rho) = 1$
2. *ρ is Hermitian*
3. *ρ has only non-negative eigenvalues*

Proof. 1.

$$\operatorname{tr}(\rho) = \operatorname{tr}\left(\sum_{i=1}^{n} p_i \left|\phi_i\right\rangle\!\left\langle\phi_i\right|\right) = \sum_{i=1}^{n} p_i \operatorname{tr}(\left|\phi_i\right\rangle\!\left\langle\phi_i\right|) = \sum_{i=1}^{n} p_i \left\langle\phi_i|\phi_i\right\rangle = \sum_{i=1}^{n} p_i = 1$$

2.

$$\left(\sum_{i=1}^{n} p_i \left|\phi_i\right\rangle\!\left\langle\phi_i\right|\right)^{\dagger} = \sum_{i=1}^{n} p_i \left(\left|\phi_i\right\rangle\!\left\langle\phi_i\right|\right)^{\dagger} = \sum_{i=1}^{n} p_i \left|\phi_i\right\rangle\!\left\langle\phi_i\right|$$

3. Let $\sum_i \lambda_i \left|v_{\lambda_i}\right\rangle\!\left\langle v_{\lambda_i}\right|$ be the eigendecomposition of ρ. Notice that the eigenvalues are exactly the probabilities from the definition and must thus be non-negative.

□

The first property $\operatorname{tr}(\rho) = 1$ implies that $D(\mathcal{H})$ is not a vector space, contrary to the state space $\mathcal{H}$ which is in fact a vector space. The second and third properties are equivalent to the density matrix being positive semidefinite.

Definition 5. *A Hermitian matrix $A \in \mathbb{C}^{n\times n}$ is **positive semidefinite** if for all vectors $|x\rangle \in \mathbb{C}^n$*

$$\langle x|A|x\rangle \geq 0. \tag{2.8}$$

The set of all positive semidefinite matrices in $\mathbb{C}^{n\times n}$ is denoted $\mathbb{S}_+^n$.

To see why this is equivalent to the second and third properties, we introduce the following lemma.

Lemma 2. *A Hermitian matrix X is positive semidefinite if and only if all its eigenvalues are non-negative.*

Proof. Let Hermitian $A \in \mathbb{C}^{n\times n}$ have only non-negative eigenvalues and let $|x\rangle$ be any vector in $\mathbb{C}^n$. Because A is Hermitian, its eigenspaces are orthogonal and its eigenvectors span $\mathbb{C}^n$ and form an orthogonal basis. Writing $|x\rangle$ as $|x\rangle = \sum_{i=1}^{n} c_i |v_i\rangle$, we see that Eq (2.8) holds:

$$\langle x|A|x\rangle = \left(\sum_{i=1}^{n} \overline{c_i} \langle v_i|\right) A \sum_{i=1}^{n} c_i |v_i\rangle \tag{2.9}$$

$$= \left(\sum_{i=1}^{n} \overline{c_i} \langle v_i|\right) \sum_{i=1}^{n} \lambda_i c_i |v_i\rangle \tag{2.10}$$

$$= \sum_{i=1}^{n} \lambda_i |c_i|^2 \langle v_i|v_i\rangle \geq 0 \tag{2.11}$$

Alternatively let A be non-negative semidefinite and $|v\rangle$ be any eigenvector. Its eigenvalue must be non-negative:

$$\langle v|A|v\rangle = \lambda \langle v|v\rangle \geq 0 \implies \lambda \geq 0. \tag{2.12}$$

□

Corollary 1. *For every Hermitian matrices A and B such that $A \preceq B$ holds* $\mathrm{tr}(A) \leq \mathrm{tr}(B)$.

Proof. Let $\{\lambda_i\}$ be the eigenvalues of $B - A$. Because the trace of a matrix is the sum of the matrix' eigenvalues, we see that

$$A \preceq B \implies 0 \preceq B - A \tag{2.13}$$

$$\implies 0 \leq \mathrm{tr}(B - A) = \sum_{\lambda_i} \lambda_i \tag{2.14}$$

$$\implies \mathrm{tr}(A) \leq \mathrm{tr}(B). \tag{2.15}$$

□

Positive semidefiniteness gives rise to a partial order [8], which means a binary relation $\succeq$ on $\mathbb{S}_+^n$ exists such that for every $A, B, C \in \mathbb{S}_+^n$ the following properties hold

1. (*reflexive*) $A \succeq A$,
2. (*antisymmetry*) $A \succeq B$ and $B \succeq A$ implies $A = B$,
3. (*transitive*) $A \succeq B$ and $B \succeq C$ implies $A \succeq C$.

Therefore we may write $A \succeq 0$ to denote A as a positive semidefinite matrix, as will be done from now on. Moreover we may write $A \succeq B$ for $A - B \succeq 0$. Bear in mind that due to this being just a partial order, simultaneously $A \not\succeq 0$ and $A \not\preceq 0$ may hold. For example it is easily verified that $A \not\succeq 0$ and $A \not\preceq 0$ hold for

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{2.16}$$

by defining $|0\rangle = (1, 0)$, and $|1\rangle = (0, 1)$. We see $\langle 0|A|0\rangle = 1$, while $\langle 1|A|1\rangle = -1$.

Entangled and separable states

A type of state that is of interest is a separable state. This uses the notion of the tensor product, which is described in Section A.1.

Definition 6. *Let $\rho \in D(\mathcal{H}_0 \otimes \mathcal{H}_1)$ be a density matrix. If ρ can be written as*

$$\rho = \sum_i p_i \rho_i \otimes \sigma_i, \tag{2.17}$$

*where p_i sums up to 1 and $\rho_i \in D(\mathcal{H}_0), \sigma_i \in D(\mathcal{H}_1)$, then ρ is a **separable state**. The set of separable states is denoted SEP. States that are not separable are called **entangled**.*

The distinction between separable and entangled states is very important, but determining whether a given state is separable or not is notoriously hard. In fact, it is an NP-hard problem [16].

Entangled systems exhibit properties that have no classical analogue. Systems that are entangled cannot be described independently from each other, meaning that if a measurement is done on one system, another systems is always affected. As an example, consider the maximally entangled state.

Definition 7. *The D-dimensional **maximally entangled state** $|\Phi_D\rangle$ is given by*

$$|\Phi_D\rangle = \frac{1}{\sqrt{D}} \sum_{i=0}^{D-1} |i\rangle \otimes |i\rangle . \tag{2.18}$$

In the 2-dimensional case, its density matrix equals

$$|\Phi_2\rangle\langle\Phi_2| = \frac{1}{2}(|00\rangle + |11\rangle)(\langle 00| + \langle 11|) \tag{2.19}$$

$$= \frac{|00\rangle\langle 00| + |11\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 11|}{2}. \tag{2.20}$$

Note that we abbreviate $|0\rangle \otimes |0\rangle$ to $|00\rangle$. This state is different from the mixed state ρ where the state might be $|00\rangle$ or $|11\rangle$ with probability $\frac{1}{2}$

$$\rho = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|). \tag{2.21}$$

The effect this has on measurement will become clear in the next subsection. The state (2.21) is known as a **maximally mixed state**. The general form of the maximally mixed states of dimension D is $\rho_D = \mathbb{I}_D/D$.

A more general case of the 2-dimensional maximally entangled state are the **Bell states**. They form an orthonormal basis of $\mathbb{C}^4$.

$$|\Phi_+\rangle = (|00\rangle + |11\rangle)/\sqrt{2},$$
$$|\Phi_-\rangle = (|00\rangle - |11\rangle)/\sqrt{2},$$
$$|\Psi_+\rangle = (|01\rangle + |10\rangle)/\sqrt{2},$$
$$|\Psi_-\rangle = (|01\rangle - |10\rangle)/\sqrt{2}.$$

2.1.4 Measurement

Measurement in quantum mechanics is described by a set of measurement operators $\{M_m\}$. Every operator acts on the state space of the system being measured. The index m refers to the measurement outcome m that may occur in the experiment. The probability of measuring m when measuring on a state $|\phi\rangle$ equals

$$p(m) = \langle\phi|M_m^\dagger M_m|\phi\rangle\,. \tag{2.22}$$

After the measurement the system is changed. If m was measured, the post-measurement state equals

$$|\phi_m\rangle = \frac{M_m\,|\phi\rangle}{\sqrt{\langle\phi|M_m^\dagger M_m|\phi\rangle}}. \tag{2.23}$$

Because all the probabilities must sum to one, the following holds

$$\sum_m M_m^\dagger M_m = I. \tag{2.24}$$

When using density matrices, the following relations hold

$$p(m) = \mathrm{tr}\Big(M_m^\dagger M_m \rho\Big) \tag{2.25}$$

$$\rho_m = \frac{M_m \rho M_m^\dagger}{\mathrm{tr}\Big(M_m^\dagger M_m \rho\Big)} \tag{2.26}$$

If a state is measured, but the measurement outcome is unknown to the observer then the state can be described as an ensemble of the post-measurement states.

$$\rho' = \sum_m p(m)\rho_m = \sum_m \mathrm{tr}\Big(M_m^\dagger M_m \rho\Big)\frac{M_m \rho M_m^\dagger}{\mathrm{tr}\Big(M_m^\dagger M_m \rho\Big)} = \sum_m M_m \rho M_m^\dagger \tag{2.27}$$

A very important type of measurement is the projective measurement. In this case, we describe the measurement using an observable M. Because an observable is Hermitian, it is ensured it has a spectral decomposition. That is, m are the eigenvalues with P_m being projectors on the eigenspace.

$$M = \sum_m m P_m. \tag{2.28}$$

From linear algebra it is known that the projective matrices P_m are idempotent (i.e. $P_m^2 = P_m$), orthogonal $P_i P_j = \delta_{ij} P_m$ and Hermitian $P_m^\dagger = P_m$. If m was measured, the state after measurement is

$$|\phi_m\rangle = \frac{P_m |\phi\rangle}{\sqrt{\langle\phi|P_m|\phi\rangle}}. \tag{2.29}$$

It is clear that after repeated projective measurement the state remains the same due to idempotence and orthogonality. In case of density matrices, there is a probablity

$$p(m) = \mathrm{tr}(\rho P_m) \tag{2.30}$$

of measuring m. If m was measured, then the state after measurement is

$$\rho_m = \frac{P_m \rho P_m}{\mathrm{tr}(P_m \rho)} \tag{2.31}$$

POVM measurements

A positive operator-valued measurement (POVM) is a more general formalism of measurements. Here we define new matrices, E_m, based on the measurement operators,

$$E_m = M_m^\dagger M_m. \tag{2.32}$$

These matrices must sum up to the identity. A matrix E_m is called a POVM element, and the set of POVM elements $\{E_m\}$ is called a POVM.

Note that if we use a projective measurement, the POVM elements are just the projective matrices because they are Hermitian and idempotent. So

$$E_m = P_m^\dagger P_m = P_m P_m = P_m. \tag{2.33}$$

2.1.5 Quantum operations

If we perform an operation on a quantum state $\rho \in D(\mathcal{H})$, this can be described by a linear mapping $\Lambda \in \mathcal{L}(D(\mathcal{H}))$. The map Λ is known as a **quantum operation** or **quantum channel**. A channel must have the property that it preserves trace

$$\mathrm{tr}(\rho) = \mathrm{tr}(\Lambda(\rho)) \tag{2.34}$$

and that it is completely positive. A map is a **positive** if for any input state $\rho \succeq 0$, the output state is also positive: $\Lambda(\rho) \succeq 0$. Furthermore, Λ is **completely positive** if for any $n > 0$ the map $\mathbb{I}_n \otimes \Lambda$ is positive.

From the definition of the density matrix it follows that a linear combination of a set of quantum channels $\{\Lambda_i\}$ will correspond to probabilistically applying channel Λ_i with a probability of α_i.

$$\Lambda(\rho) = \sum_i \alpha_i \Lambda_i(\rho) \tag{2.35}$$

From the definition of the channel it follows for any two channels Λ_0, Λ_1 the composition $\Lambda_0 \circ \Lambda_1$ forms a channel too.

Unitary transformations

An example of a channel is a unitary transformation. For some unitary matrix U we have the channel

$$\Lambda(\rho) = U\rho U^\dagger. \tag{2.36}$$

The Pauli matrices introduced above are unitary and form rather interesting channels. The Pauli matrix σ_x swaps qubits:

$$\Lambda(\rho) = \sigma_x \rho \sigma_x^\dagger,\ \Lambda(|0\rangle\langle 0|) = |1\rangle\langle 1|\,,\ \Lambda(|1\rangle\langle 1|) = |0\rangle\langle 0| \tag{2.37}$$

Controlled NOT gate

Another example is the controlled NOT (CNOT) gate, which is one of the fundamental logic quantum channels. It is a unitary transformation where the unitary equals

$$U = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{2.38}$$

The reason this channel is called a controller NOT gate becomes clear when one considers its effect on the following pure state. The first qubit controls whether the last qubit is flipped, the quantum equivalent of applying a NOT gate.

$$|00\rangle \to |00\rangle\,, |01\rangle \to |01\rangle\,, |10\rangle \to |11\rangle\,, |11\rangle \to |10\rangle\,. \tag{2.39}$$

A CNOT gate is visualised as the following circuit

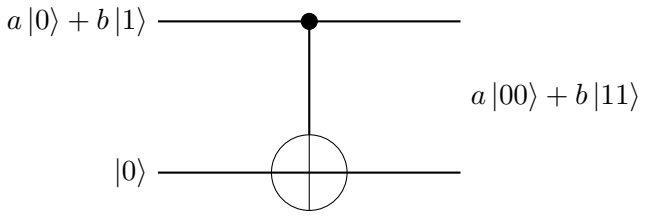


Figure 2.1: *Visual representation of the CNOT gate. As an example, we use the input state* $|\phi\rangle = a|00\rangle + b|10\rangle$*. Using Eq. (2.39) it can be verified that the output state equals* $a|00\rangle + b|11\rangle$

Twirling

Finally we introduce the twirling channel. Here a set of unitaries $T = \{U_i\}$ is defined and with equal probability one of them is applied to the input state. In case of a finite set T, we have

$$\Lambda(\rho) = \frac{1}{|T|} \sum_{U_i \in T} U_i \rho U_i^\dagger \tag{2.40}$$

For an infinite set T we denote twirling as

$$\Lambda(\rho) = \int_T dU U \rho U^\dagger, \tag{2.41}$$

where the integral is a Haar integral [23]. The precise definition of this integral is outside of the scope of this thesis.

Conjugate channel

A useful tool on describing channels is the conjugate channel.

Definition 8. *Let Λ be a channel. There exists a map $\Lambda^\dagger$ called the **conjugate channel** such that for any state σ and ρ*

$$\mathrm{tr}(\sigma\Lambda(\rho)) = \mathrm{tr}\Big(\Lambda^\dagger(\sigma)\rho\Big). \tag{2.42}$$

The property $(\Lambda_0 \circ \Lambda_1)^\dagger = \Lambda_1^\dagger \circ \Lambda_0^\dagger$ holds:

$$\mathrm{tr}(\sigma(\Lambda_0 \circ \Lambda_1)(\rho)) = \mathrm{tr}(\sigma\Lambda_0(\Lambda_1(\rho))) = \mathrm{tr}\Big(\Lambda_1^\dagger(\Lambda_0^\dagger(\sigma))\rho\Big) = \mathrm{tr}\Big((\Lambda_1^\dagger \circ \Lambda_0^\dagger)(\sigma)\rho\Big) \tag{2.43}$$

With the notion of a conjugate channel, we introduce the following theorem [13] which can be useful when dealing with a maximally entangled state.

Theorem 1. *Let $|\Phi_D\rangle\langle\Phi_D|$ be a maximally entangled state as defined in Defintion 7 and T the transposition map $\rho \mapsto \rho^T$. The following holds for every channel Λ*

$$(\Lambda \otimes \mathbb{I})\,|\Phi_D\rangle\langle\Phi_D| = (\mathbb{I} \otimes T \circ \Lambda^\dagger \circ T)\,|\Phi_D\rangle\langle\Phi_D|\,,$$

where $\Lambda^\dagger$ denotes the conjugate channel.

2.1.6 Channel-state duality

A very useful tool is to express quantum channels as matrices. To achieve this we can employ the Choi-Jamiołkowski isomorphism [7][20]

Theorem 2 (Choi-Jamiołkowski isomorphism)**.** *Let $\mathcal{H}_0, \mathcal{H}_1$ be n_0, n_1 dimensional Hilbert spaces. The map*

$$\begin{aligned} \mathcal{J}\colon\ & \mathcal{L}(\mathcal{L}(\mathcal{H}_0), \mathcal{L}(\mathcal{H}_1)) \to \mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_1), \\ & \Lambda \mapsto (\Lambda \otimes \mathbb{I})\,|\Phi_{n_1}\rangle\langle\Phi_{n_1}| \end{aligned} \tag{2.44}$$

is an isomorphism.

The state corresponding to some channel Λ, $J(\Lambda)$, will be referred to as the **Choi state** of that channel. For a channel from system S to R we will denote the Choi state as $C_{RS'}$, where S' is a copy of S. A very useful property of Choi states is the following lemma [30].

Lemma 3. *Let $\rho \in D(\mathcal{H}_S)$ be a density matrix on a Hilbert space $\mathcal{H}_S$ of dimension $|S|$ and let $\Lambda_{S\to R}$ be a channel. For any linear operator M the following holds,*

$$\mathrm{tr}(M_R\Lambda_{S\to R}(\rho_S)) = |S|\,\mathrm{tr}\big((M_R \otimes \rho_{S'}^T)C_{RS'}\big), \tag{2.45}$$

where S' is a copy of system S.

Furthermore, Choi states are completely characterised by the following properties [30]

$$C_{RS'} \succeq 0, \tag{2.46}$$

$$\mathrm{tr}(C_{RS'}) = 1, \tag{2.47}$$

$$\mathrm{tr}_R(C_{RS'}) = \mathbb{I}_{S'}/|S|. \tag{2.48}$$

Here tr_R indicates a partial trace, as described in Section A.3. Notice that the Choi-Jamiołkowski isomorphism this is *not* an isomorphism between channels and states, as not every state is a Choi state.

2.2 Entanglement distillation

The goal of entanglement distillation is to form m maximally entangled states using n other, possibly mixed, known states. The motivation behind this is that it's hard to create maximally entangled pairs which have numerous applications in fields such as quantum communication and quantum computation. In general, to quantify how similar two quantum states are, we use the fidelity. In turn this gives a method to determine how entangled a state is, by calculating the fidelity to the maximally entangled state.

Definition 9. *The **fidelity** F of two states ρ and σ is defined as*

$$F(\rho,\sigma) = \mathrm{tr}\left(\sqrt{\rho^{1/2}\sigma\rho^{1/2}}\right)^2. \tag{2.49}$$

It can be shown [23] that the fidelity is a symmetric function. That is $F(\rho,\sigma) = F(\sigma,\rho)$. When determining the fidelity to a pure state, Eq. (2.49) can be simplified. Assume that ρ is pure, i.e. $\rho = |\rho\rangle\langle\rho|$. Using $\rho = \rho^2$ we see

$$\begin{aligned} F(\rho,\sigma) &= \mathrm{tr}\left(\sqrt{\rho^{1/2}\sigma\rho^{1/2}}\right)^2 && (2.50)\\ &= \mathrm{tr}\left(\sqrt{|\rho\rangle\langle\rho|\sigma|\rho\rangle\langle\rho|}\right)^2 && (2.51)\\ &= \langle\rho|\sigma|\rho\rangle\,\mathrm{tr}\left(\sqrt{|\rho\rangle\langle\rho|}\right)^2 && (2.52)\\ &= \langle\rho|\sigma|\rho\rangle\,\mathrm{tr}(|\rho\rangle\langle\rho|)^2 && (2.53)\\ &= \langle\rho|\sigma|\rho\rangle = \mathrm{tr}(\sigma\rho). && (2.54) \end{aligned}$$

Even though high fidelity to the maximally entangled state implies that a system exhibits entanglement, low fidelity does not imply low entanglement. As an example, consider the Bell state $|\Psi_+\rangle$, which is a maximally entangled state. Its fidelity to the maximally entangled state as defined in Definition 7, $|\Phi_+\rangle = |\Phi_2\rangle$, is $F(|\Psi_+\rangle, |\Phi_+\rangle) = \langle\Psi_+|\Phi_+\rangle = 0$.

2.3 Convex optimisation

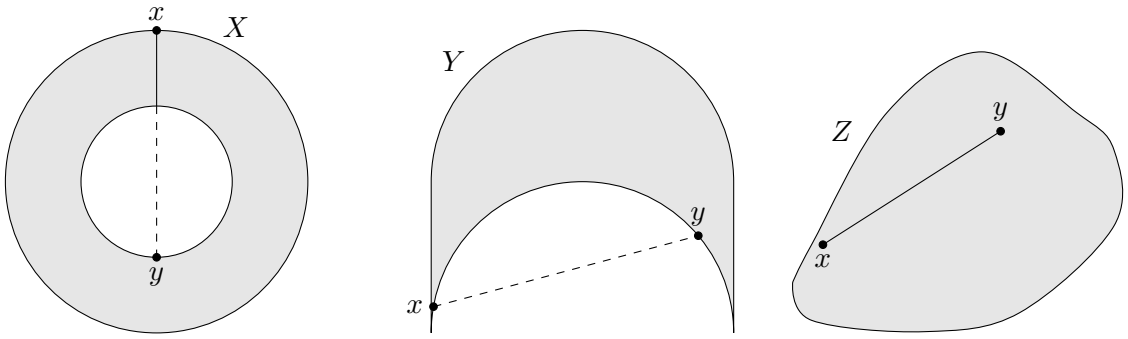
As will become clear in Section 2.4, the essence of the thesis is that we want to solve an optimisation problem. In optimisation, the goal is to maximise or minimise a certain objective function $f : X \to \mathbb{R}$ subject to constraints such that we look for an optimum within a set $Y \subseteq X$, called the feasible set. In general many optimisation problems cannot be solved realistically because their running times grow exponentially as a function of their input size. In particular non-convex problems are notoriously hard to solve, so it is in our interest to pose problems we encounter as a convex one, if this is possible.

For an minimisation problem to be convex, a convex set and a convex function is needed.

Definition 10. *A set X is a **convex set** if for every $x, y \in X$ for every $t \in [0,1]$ holds that*

$$tx + (1-t)y \in X \tag{2.55}$$

Geometrically speaking, this means that every for two points in a convex set, a straight line can be drawn between these points such that the line is completely contained in the convex set. Consequently the set cannot have any 'holes'.



(a) *This annulus is not convex, because the line between x and y is not completely contained in the set X.*

(b) *The set Y also is non-convex.*

(c) *The set Z is convex.*

Figure 2.2: *Examples of convex and non-convex sets.*

Definition 11. *A function $f : X \to \mathbb{R}$ is called a **convex function** if for every $x, y \in X$ for every $t \in [0, 1]$ the following holds:*

$$f(tx + (t-1)y) \leq tf(x) + (1-t)f(y) \tag{2.56}$$

Geometrically a convex function can be seen as 'hollow'. For example the function $x \mapsto x^2$ is convex as can be seen in Figure 2.3.

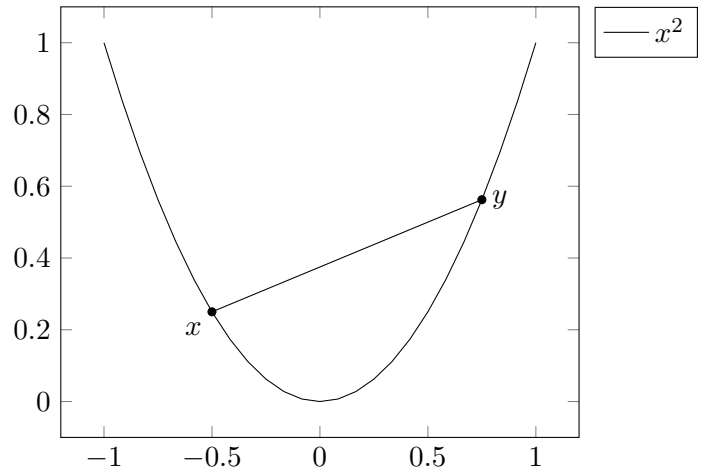


Figure 2.3: *The function $f(x) = x^2$ is convex, because for every x and y the function f is contained below the line segment between x and y.*

If alternatively the objective is the *maximise*, we require the objective function to be concave.

Definition 12. *A function $f : X \to \mathbb{R}$ is called a **concave function** if $-f$ is convex.*

Convexity is a very powerful property, because it guarantees that a local optimum is also a global one. [5]

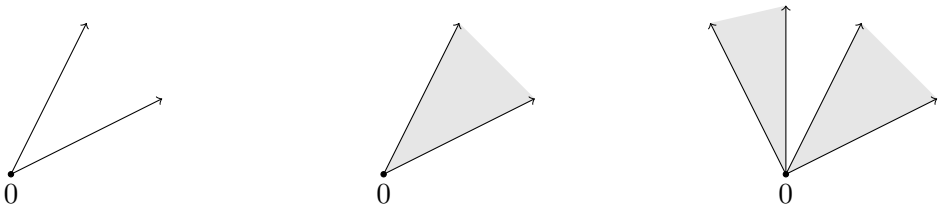
2.3.1 Semidefinite programming

In semidefinite programming we optimise over positive semidefinite matrices. It is a special form of the more general class of conic programming, where the feasible set can be described by a cone.

Definition 13. *Let X be a set. X is a **cone** if for all $x \in X$*

$$x \in X \implies ax \in \overline{X} \quad \forall a \geq 0. \tag{2.57}$$

Where $\overline{X}$ denotes the closure of X.



(a) *Two half-lines originating from 0 form a cone*

(b) *A convex cone*

(c) *The union of two convex cones forms a possibly non-convex cone.*

Figure 2.4: *Examples of various cones*

Since all density matrices are positive semidefinite, there is a strong relation between semidefinite programming and quantum mechanics, and optimisation over these matrices becomes a useful tool in quantum information. These optimisation programs are named semidefinite programs (SDPs). More specifically, we follow [28] by defining an SDP as:

$$\begin{aligned} \underset{X \in \mathbb{S}_+^n}{\text{maximize}} \quad & \operatorname{tr}(CX) \\ \text{subject to} \quad & \Phi(X) = B \\ & X \succeq 0. \end{aligned} \tag{2.58}$$

Where C is a Hermitian matrix in $\mathbb{C}^n$ and B a Hermitian matrix in $\mathbb{C}^m$. Furthermore Φ is a Hermiticity preserving map $\Phi \colon \operatorname{Herm}(\mathbb{C}^n) \to \operatorname{Herm}(\mathbb{C}^m)$. A matrix X satisfying the constraints is called **feasible**. The set of all feasible $X \in \mathbb{S}_n^+$ is called the **feasible set**. If the feasible set is empty, the SDP is called **infeasible**.

Adding more constraints

More constraints may be added to the program (2.59), by observing that the following problems are equivalent.

$$\begin{aligned} \underset{X \in \mathbb{S}_+^n}{\text{maximize}} \quad & \operatorname{tr}(CX) \\ \text{subject to} \quad & \Phi_0(X) = B_0 \\ & \Phi_1(X) = B_1 \\ & X \succeq 0. \end{aligned} \iff \begin{aligned} \underset{X \in \mathbb{S}_+^n}{\text{maximize}} \quad & \operatorname{tr}(CX) \\ \text{subject to} \quad & \Psi(X) = B_0 \oplus B_1 \\ & X \succeq 0. \end{aligned} \tag{2.59}$$

Here we defined $\Psi(X) = \Phi_0(X) \oplus \Phi_1(X)$ and '$\oplus$' refers to the direct sum as described in Section A.2.

Adding more variables

Now, we show that more variables can be added to an SDP, by showing the equivalence between (2.59) and

$$\begin{aligned} \underset{X_0\in\mathbb{S}_+^n, X_1\in\mathbb{S}_+^m}{\text{maximize}} \quad & \operatorname{tr}(C_0X_0)+\operatorname{tr}(C_1X_1) \\ \text{subject to} \quad & \Phi_0(X_0)=B_0 \\ & \Phi_1(X_1)=B_1 \\ & X_0\succeq 0 \\ & X_1\succeq 0. \end{aligned} \tag{2.60}$$

To add a new variable we first define the Hermiticity preserving map Ψ as

$$\Psi\left(\left.\begin{array}{c|c} X_0 & X_{12} \\ \hline X_{12}^\dagger & X_1 \end{array}\right.\right) = \Phi_0(X_0)\oplus\Phi_1(X_1). \tag{2.61}$$

This ensures that X is in the form

$$X=\left(\left.\begin{array}{c|c} X_0 & X_{12} \\ \hline X_{12}^\dagger & X_1 \end{array}\right.\right). \tag{2.62}$$

Note that the following relation holds

$$\begin{aligned} \operatorname{tr}\left[\left(\left.\begin{array}{c|c} C_0 & 0 \\ \hline 0 & C_1 \end{array}\right.\right)\left(\left.\begin{array}{c|c} X_0 & X_{12} \\ \hline X_{12}^\dagger & X_1 \end{array}\right.\right)\right] &= \operatorname{tr}\left(\left.\begin{array}{c|c} C_0X_0 & C_1X_{12} \\ \hline C_0X_{12}^\dagger & C_1X_1 \end{array}\right.\right) && (2.63) \\ &= \operatorname{tr}(C_0X_0)+\operatorname{tr}(C_1X_1). && (2.64) \end{aligned}$$

Lemma 4. *Let $A\in\operatorname{Herm}(\mathbb{C}^n), B\in\operatorname{Herm}(\mathbb{C}^m)$ and $C\in\mathbb{C}^{n\times m}$ such that $\left(\left.\begin{array}{c|c} A & C \\ \hline C^\dagger & B \end{array}\right.\right)\succeq 0$. Then $A\succeq 0$ and $B\succeq 0$.*

Proof. Let for all $|v\rangle\in\mathbb{C}^n, |w\rangle\in\mathbb{C}^m : |z\rangle=[v_1,\ldots,v_n,w_1,\ldots,w_m]^T$.

$$\langle z|\left(\left.\begin{array}{c|c} A & C \\ \hline C^\dagger & B \end{array}\right.\right)|z\rangle = \langle v|A|v\rangle + \langle v|C|w\rangle + \langle w|C^\dagger|v\rangle + \langle w|B|w\rangle \geq 0. \tag{2.65}$$

By setting $|w\rangle=0$ we see that for all $|v\rangle : \langle v|A|v\rangle\geq 0$, thus $A\succeq 0$. Equivalently by setting $|w\rangle$ it is shown that $B\succeq 0$. □

Using Lemma 4 it becomes clear that the following two programs are equivalent.

$$\begin{aligned} \underset{X_0\in\mathbb{S}_+^n, X_1\in\mathbb{S}_+^m}{\text{maximize}} \quad & \operatorname{tr}(C_0X_0)+\operatorname{tr}(C_1X_1) \\ \text{subject to} \quad & \Phi_0(X_0)=B_0 \\ & \Phi_1(X_1)=B_1 \\ & X_0\succeq 0 \\ & X_1\succeq 0. \end{aligned} \quad \Longleftrightarrow \quad \begin{aligned} \underset{X\in\mathbb{S}_+^{(n+m)}}{\text{maximize}} \quad & \operatorname{tr}((C_0\oplus C_1)X) \\ \text{subject to} \quad & \Psi(X)=B_0\oplus B_1 \\ & X\succeq 0. \end{aligned} \tag{2.66}$$

Changing the equality sign

Using a extra variable, we can change the equality sign to a non-strict inequality. This variable is known as a **slack variable**.

$$\begin{aligned} &\underset{X\in\mathbb{S}_+^n}{\text{maximize}} && \operatorname{tr}(CX) \\ &\text{subject to} && \Phi(X) \preceq B \\ &&& X \succeq 0. \end{aligned} \quad \Longleftrightarrow \quad \begin{aligned} &\underset{X\in\mathbb{S}_+^m}{\text{maximize}} && \operatorname{tr}(CX) \\ &\text{subject to} && \Phi(X) - S = B \\ &&& S \succeq 0 \\ &&& X \succeq 0. \end{aligned} \tag{2.67}$$

The inequality sign may be flipped by swapping the minus sign with a plus sign. Finally we may impose more constraints on X using the properties above and the Hermiticity preserving map $X \mapsto (Y - X) \otimes X$

$$\begin{aligned} &\underset{X\in\mathbb{S}_+^n}{\text{maximize}} && \operatorname{tr}(CX) \\ &\text{subject to} && \Phi(X) = B \\ &&& Y \succeq X \succeq 0. \end{aligned} \quad \Longleftrightarrow \quad \begin{aligned} &\underset{X\in\mathbb{S}_+^m}{\text{maximize}} && \operatorname{tr}(CX) \\ &\text{subject to} && \Phi(X) = B \\ &&& (Y - X) \oplus X \succeq 0 \\ &&& X \succeq 0. \end{aligned} \tag{2.68}$$

Together, SDPs form a very general class of convex optimisation programs. For example, it can be shown that every linear optimisation problem and every *convex* quadratic program can be written as an SDP [8].

Solving SDPs

Several algorithms exist to solve SDPs. The performance of these solvers can be highly dependent on the specific problem it tries to solve. We make the distinction between first-order methods such as the alternating direction method of multipliers, and second-order methods like the well known interior point method. [29]

The difference between these classes is that first-order methods only use the first derivative whereas second-order implies that also the second derivative is used. Consequentially, second-order methods usually achieve better results but the iterations are more expensive because it requires larger systems of equations to be solved. This causes second-order methods do perform worse for large problems as they require a lot of memory.

2.3.2 Solving complex SDPs

The SDP solvers that are used in this thesis only solve SDPs that involve real matrices. However complex SDPs can be converted into real SDPs using the following bijection between real and complex matrices

$$\phi(Z) = \begin{pmatrix} \operatorname{Re} Z & -\operatorname{Im} Z \\ \operatorname{Im} Z & \operatorname{Re} Z \end{pmatrix}. \tag{2.69}$$

Not only does this preserve multiplication and addition, but more importantly it also preserves semidefiniteness [1]. That is, $Z \succeq 0 \iff \phi(Z) \succeq 0$. This bijection should be used with caution, as not all operations are respected. For example, $\phi(Z)^T = \begin{pmatrix} \operatorname{Re} Z^T & \operatorname{Im} Z^T \\ -\operatorname{Im} Z^T & \operatorname{Re} Z^T \end{pmatrix} = \phi(Z^\dagger) \neq \phi(Z^T)$.

2.3.3 Duality

A very powerful concept in optimisation is *duality*. The notion of duality is widespread throughout the field, but here only the duality of SDPs will be discussed. For every SDP we will define another program which we will call its *dual*. If we have an SDP in standard form which we will call the *primal*

$$\begin{aligned} \underset{X \in \mathbb{S}_+^n}{\text{maximize}} \quad & \operatorname{tr}(CX) \\ \text{subject to} \quad & \Phi(X) = B \\ & X \succeq 0. \end{aligned} \tag{2.70}$$

We define, as in [28], its dual as

$$\begin{aligned} \underset{Y \in \operatorname{Herm}(\mathbb{C}^m)}{\text{minimize}} \quad & \operatorname{tr}(BY) \\ \text{subject to} \quad & \Phi^\dagger(Y) \succeq C. \end{aligned} \tag{2.71}$$

Where $\Phi^\dagger$ is the conjugate of Φ. Note that Y isn't necessarily a positive semidefinite matrix anymore and its size doesn't have to match the size of X, instead it matches B. Although Y may not be positive semidefinite, the dual still is an SDP as it's equivalent in the following manner.

$$\begin{aligned} \underset{Y \in \operatorname{Herm}(\mathbb{C}^m)}{\text{minimize}} \quad & \operatorname{tr}(BY) \\ \text{subject to} \quad & \Phi^\dagger(Y) \succeq C. \end{aligned} \iff \begin{aligned} \underset{Y,Z \in \mathbb{S}_+^m}{\text{minimize}} \quad & \operatorname{tr}(BY) \\ \text{subject to} \quad & \Phi^\dagger(Y - Z) \succeq C \\ & Y \succeq 0 \\ & Z \succeq 0. \end{aligned} \tag{2.72}$$

The dual and primal are related, as the optimal value of the primal is guaranteed to be lower than the optimal value of the dual. This is known as the **weak duality theorem**. The difference between the optimal values is referred to as the **duality gap**.

Theorem 3 (Weak duality theorem)**.** *For every SDP, the optimal primal is less than the optimal dual.*

Proof. Let the primal and dual be defined as in 2.70 and 2.71. Due to Cholesky decomposition, we may factorize any positive semidefinite matrix as $X = LL^\dagger$.

$$\operatorname{tr}\Big(\Phi^\dagger(Y)X\Big) - \operatorname{tr}(CX) = \operatorname{tr}\Big(\Phi^\dagger(Y)LL^\dagger\Big) - \operatorname{tr}\Big(CLL^\dagger\Big) \tag{2.73}$$

$$= \operatorname{tr}\Big(L^\dagger\Phi^\dagger(Y)L\Big) - \operatorname{tr}\Big(L^\dagger CL\Big) \tag{2.74}$$

$$= \operatorname{tr}\Big(L^\dagger(\Phi^\dagger(Y) - C)L\Big) \geq 0. \tag{2.75}$$

Where the inequality in Eq. (2.75) holds by observing that for every $|v\rangle$ and every matrix L

$$\langle v|L^\dagger(\Phi^\dagger(Y) - C)L|v\rangle = \langle w|\Phi^\dagger(Y) - C)|w\rangle \geq 0 \tag{2.76}$$

holds. Thus $L^\dagger(\Phi^\dagger(Y) - C)L \succeq 0$. We conclude

$$\operatorname{tr}(CX) \leq \operatorname{tr}\Big(\Phi^\dagger(Y)X\Big) = \operatorname{tr}(Y\Phi(X)) = \operatorname{tr}(YB). \tag{2.77}$$

□

Under certain condition the duality gap may close, in which case we say the SDP exhibits a **strong duality**. Slaters' conditions are sufficient to show strong duality. [28]

Theorem 4 (Slater's theorem for semidefinite programs)**.** *If one of the two conditions hold, the duality gap of an SDP is 0.*

1. *The primal is feasible and there exists a Hermitian operator Y for which $\Phi^\dagger(Y) \succ C$.*
2. *The dual is feasible and there exists a Hermitian operator X for which $\Phi(X) \succ B$*

Duality is a powerful tool that is employed by SDP solver to determine convergence to the optimal solution. Furthermore, we can use the formulation of the dual to find an analytical solution to certain SDPs. This is beyond the scope of this thesis, but in the upcoming paper optimality is shown of the EPL protocol, which is described in Section 3.3.

2.4 Problem

We will consider distillation protocols performed by two parties, which will be called Alice (A) and Bob (B). Alice and Bob have access to known, possibly mixed, input states. They perform local operations on their systems first. Then, they perform a measurement on a subsystem, which we will call the *flags*. More specifically, Alice will perform a map $\Lambda_{A\to\hat{A}F_A}$, where $\hat{A}$ should resemble Alice's subsystem of a maximally entangled state and F_A is Alice's flag. Analogously, Bob performs a map $\Lambda_{B\to\hat{B}F_B}$. The flags indicate success or failure of the distillation.

Different approaches exist, in some protocols the flags indicate success if and only if A and B measure '1'. In this case we say the protocol uses 'local flags'. The other approach we will investigate, is where the protocol is successful if and only if A and B measure the same. That is, they both measure '0' or they both measure '1'. In this case we say the protocol uses 'non-local flags'.

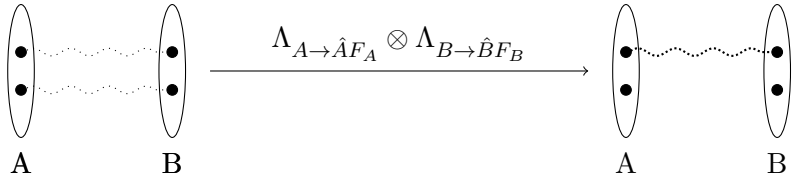


Figure 2.5: *Quantum distillation of two entangled systems. Alice (A) and Bob (B) perform separate operations on their own systems. Their measurement on the second pair destroys entanglement. If the flags indicate success they use the newly obtained more strongly entangled state. Otherwise, they discard the output state.*

Since the protocol behaves probabilistically, there is an inherent trade-off between probability of success and the output fidelity. The goal is to determine what is the best output fidelity we can obtain for a fixed probability of success. The state after Alice and Bob applied their operations equals

$$\sigma_{\hat{A}F_A\hat{B}F_B} = \Lambda_{A\to\hat{A},F_A} \otimes \Lambda_{B\to\hat{B},F_B}(\rho_{AB}). \tag{2.78}$$

Using a projective measurement on the flags, Alice and Bob obtain the following state if they both measure 1

$$\sigma^{\text{succ}}_{\hat{A}F_A\hat{B}F_B} = \frac{(\mathbb{I}_{\hat{A},\hat{B}} \otimes P_{F_A,F_B})\sigma_{\hat{A}F_A\hat{B}F_B}(\mathbb{I}_{\hat{A},\hat{B}} \otimes P_{F_A,F_B})}{p_{\text{succ}}}, \tag{2.79}$$

where $P_{F_A,F_B} = |11\rangle\langle 11|$ are projectors on the flags and

$$p_{\text{succ}} = \text{tr}\Big((\mathbb{I}_{\hat{A},\hat{B}} \otimes P_{F_A,F_B})\sigma_{\hat{A}F_A\hat{B}F_B}\Big) \tag{2.80}$$

normalises the output state. In addition p_{succ} is equal to the probability that distillation is considered successful. The indices indicate what systems are involved. So if one were to numerically compute p_{succ}, it is important to order the systems correctly. For example, a permutation matrix can be used on $\sigma_{\hat{A}F_A\hat{B}F_B}$ to obtain $\sigma_{\hat{A}\hat{B}F_AF_B}$, or we might equivalently use

$$p_{\text{succ}} = \operatorname{tr}\Big((\mathbb{I}_{\hat{A}} \otimes P_{F_A} \otimes \mathbb{I}_{\hat{B}} \otimes P_{F_B})\sigma_{\hat{A}F_A\hat{B}F_B}\Big). \tag{2.81}$$

We want to solve the following optimization problem over the maps $\Lambda_{A\to\hat{A},F_A}$ and $\Lambda_{B\to\hat{B},F_B}$, where we constrain the probability of success to δ.

$$\begin{aligned} &\text{maximize} && \operatorname{tr}\Big(|\Omega_D\rangle\langle\Omega_D|\,\sigma^{\text{succ}}_{\hat{A}\hat{B}}\Big) \\ &\text{subject to} && p_{\text{succ}} = \delta \\ & && \sigma^{\text{succ}}_{\hat{A}F_A\hat{B}F_B} = \frac{\mathbb{I}_{\hat{A},\hat{B}} \otimes P_{F_A,F_B}\left(\Lambda_{A\to\hat{A},F_A} \otimes \Lambda_{B\to\hat{B},F_B}\left(\rho_{AB}\right)\right)\mathbb{I}_{\hat{A},\hat{B}} \otimes P_{F_A,F_B}}{p_{\text{succ}}} \end{aligned} \tag{2.82}$$

Note that in the objective function we use $\sigma^{\text{succ}}_{\hat{A}\hat{B}} = \operatorname{tr}_{F_A,F_B}(\sigma^{\text{succ}}_{\hat{A}F_A\hat{B}F_B})$, because we are not interested in the flags after measurement. The convention that leaving out indices means tracing out those systems will be used from now on for brevity.

The problem can be written more concisely as follows

$$\begin{aligned} &\text{maximize} && \delta^{-1}\operatorname{tr}\Big(|\Omega_D\rangle\langle\Omega_D|_{\hat{A},\hat{B}} \otimes |11\rangle\langle 11|_{F_AF_B}\,\sigma_{\hat{A}F_A\hat{B}F_B}\Big) \\ &\text{subject to} && p_{\text{succ}} = \delta \\ & && \sigma_{\hat{A}F_A\hat{B}F_B} = \Lambda_{A\to\hat{A},F_A} \otimes \Lambda_{B\to\hat{B},F_B}(\rho_{AB}) \end{aligned}$$

The feasible set is non-convex, because the set of channels in this tensor product is not convex. Therefore we will consider the larger set of convex combinations of channels that correspond to probabilities. More specifically, we will look at output states

$$\sigma_{\hat{A}F_A\hat{B}F_B} = \sum_j p_j \Lambda_{j,A\to\hat{A},F_A} \otimes \Lambda_{j,B\to\hat{B},F_B}(\rho_{AB}). \tag{2.83}$$

Physically this is very easy to implement, because this corresponds to randomly applying the channel $\Lambda_{j,A\to\hat{A},F_A} \otimes \Lambda_{j,B\to\hat{B},F_B}$ with probability p_j. The randomness can be distributed ahead of time between Alice and Bob.

Quantum channels can be expressed as states due to the Choi-Jamiołkowski isomorphism, which is a duality between quantum channels and quantum states. The Choi state reads

$$C_{\hat{A}F_A\hat{B}F_B,A'B'} = \sum_j p_j \left(\Lambda_{j,A\to\hat{A},F_A} \otimes \mathbb{I}_{A'}\right) \otimes \left(\Lambda_{j,B\to\hat{B},F_B} \otimes \mathbb{I}_{B'}\right)(\Phi_{AA'} \otimes \Phi_{BB'}) \tag{2.84}$$

$$= \sum_j p_j C_{j,\hat{A}F_A,A'} \otimes C_{j,\hat{B}F_B,B'}\ . \tag{2.85}$$

The set of these separable Choi states is denoted SEP-C which is a subset of the set containing all separable states SEP.

Using Eq. (2.45) we may express the objective function using the Choi state as

$$\delta^{-1}|A||B|\operatorname{tr}\Big([|\Omega_D\rangle\langle\Omega_D|_{\hat{A},\hat{B}} \otimes |11\rangle\langle 11|_{F_AF_B} \otimes \rho^T]C_{\hat{A}F_A\hat{B}F_B,A'B'}\Big), \tag{2.86}$$

and the probability of success

$$p_{\text{succ}} = |A||B|\operatorname{tr}\Big((\mathbb{I}_{\hat{A},\hat{B}} \otimes P_{F_A,F_B} \otimes \rho^T)C_{\hat{A}F_A\hat{B}F_B,A'B'}\Big). \tag{2.87}$$

Furthermore, Eq. (2.46–2.48) must hold to ensure that $C_{\hat{A}F_A\hat{B}F_B,A'B'}$ is a Choi state. This brings us to the SDP

$$\begin{aligned} \text{maximize} \quad & \delta^{-1}|A||B|\,\mathrm{tr}\Big([|\Omega_D\rangle\langle\Omega_D|_{\hat{A},\hat{B}} \otimes |11\rangle\langle 11|_{F_AF_B} \otimes \rho^T]C_{\hat{A}F_A\hat{B}F_B,A'B'}\Big) \\ \text{subject to} \quad & p_{\text{succ}} = \delta \\ & C_{\hat{A}F_A\hat{B}F_B,A'B'} \succeq 0 \\ & C_{A'B'} = \mathbb{I}_{A'B'}/|A||B| \end{aligned} \tag{2.88}$$

2.4.1 Getting rid of flags

It is possible to make this problem smaller, namely by removing the flags. Considering F_A and F_B are both 2 dimensional, this is a significant reduction.

Let $C^*_{\hat{A}F_A,A'}$ and $C^*_{\hat{B}F_B,B'}$ be optimal solutions to the optimisation problem defined in Eq. (2.88). Because the flags are always measured, we can assume that these states are classical quantum-states. That is,

$$\tilde{C}_{\hat{A}F_A,A'} = \sum_{f\in\{0,1\}} \hat{C}_{f,\hat{A},A'} \otimes |j\rangle\langle j|_{F_A} \ , \tag{2.89}$$

$$\tilde{C}_{\hat{B}F_B,B'} = \sum_{f\in\{0,1\}} \hat{C}_{f,\hat{B},B'} \otimes |j\rangle\langle j|_{F_B} \ . \tag{2.90}$$

Hence the following states are also optimal:

$$\tilde{C}_{\hat{A}F_A,A'} = \sum_{f\in\{0,1\}} \mathbb{I}_{\hat{A}A'} \otimes |j\rangle\langle j|_{F_A} C^*_{\hat{A}F_A,A'} \mathbb{I}_{\hat{A}A'} \otimes |j\rangle\langle j|_{F_A} \ , \tag{2.91}$$

$$\tilde{C}_{\hat{B}F_B,B'} = \sum_{f\in\{0,1\}} \mathbb{I}_{\hat{B}B'} \otimes |j\rangle\langle j|_{F_B} C^*_{\hat{B}F_B,B'} \mathbb{I}_{\hat{B}B'} \otimes |j\rangle\langle j|_{F_B} \ . \tag{2.92}$$

Because we are only interested in the case of success, the problem reduces to

$$\begin{aligned} \text{maximize} \quad & \delta^{-1}|A||B|\,\mathrm{tr}\left(\Phi_{D,\hat{A},\hat{B}} \otimes \rho^T_{A'B'}\left(\hat{C}_{1,\hat{A},A'} \otimes \hat{C}_{1,\hat{B},B'}\right)\right) \\ \text{subject to} \quad & |A||B|\,\mathrm{tr}\left(\rho^T_{A'B'}\left(\hat{C}_{1,A'} \otimes \hat{C}_{1,B'}\right)\right) = \delta \\ & \hat{C}_{1,\hat{A},A'} \succeq 0, \hat{C}_{1,\hat{B},B'} \succeq 0 \\ & \hat{C}_{1,A'} \preceq \frac{\mathbb{I}_{A'}}{|A|}, \hat{C}_{1,B'} \preceq \frac{\mathbb{I}_{B'}}{|B|}. \end{aligned} \tag{2.93}$$

Where the last inequality follows from the condition that

$$\hat{C}_{0,A'} + \hat{C}_{1,A'} = \frac{\mathbb{I}_{A'}}{|A|} \ , \tag{2.94}$$

and analagously for Bob.

Non-local flags

In some protocols we encounter non-local flags. This means that distillation is designated successful if Alice and Bob have the same outcome measurement. That is, $|00\rangle$ and $|11\rangle$ are

considered to be successful flags. In this case we get

$$\begin{aligned}
\text{maximize} \quad & \frac{|A||B|}{\delta} \operatorname{tr}\left(\Phi_{D,\hat{A},\hat{B}} \otimes \rho^T_{A'B'} \left(\hat{C}_{1,\hat{A},A'} \otimes \hat{C}_{1,\hat{B},B'} + \hat{C}_{0,\hat{A},A'} \otimes \hat{C}_{0,\hat{B},B'}\right)\right) \\
\text{subject to} \quad & |A||B| \operatorname{tr}\left[\rho^T_{A'B'} \left(\hat{C}_{1,A'} \otimes \hat{C}_{1,B'} + \hat{C}_{0,\hat{A},A'} \otimes \hat{C}_{0,\hat{B},B'}\right)\right] = \delta \\
& \hat{C}_{1,\hat{A},A'}, \hat{C}_{1,\hat{B},B'}, \hat{C}_{0,\hat{A},A'}, \hat{C}_{0,\hat{B},B'} \succeq 0 \\
& \hat{C}_{1,A'} + \hat{C}_{0,A'} = \frac{\mathbb{I}_{A'}}{|A|}, \hat{C}_{1,B'} + \hat{C}_{0,A'} = \frac{\mathbb{I}_{B'}}{|B|}.
\end{aligned} \tag{2.95}$$

2.5 Solving the problem

Thus far, solving the problem would yield the actual optimal output fidelity. Unfortunately, this problem cannot be solved in any reasonable time by a computer. This is due to the fact that our problem can't be posed as an SDP, as the objective function is neither convex nor concave.

Instead, we will look at a larger set of feasible solutions that contains the set of separable Choi states, SEP-C. On this larger set, the objective function is convex. This will lead to an answer that is potentially greater than the actual optimal, because the state we find, might actually not be a separable Choi state. Clearly, it is useful to approximate SEP-C as closely as possible while maintaining convexity.

After a state has been found, it is too hard to check whether it is in SEP-C. This is namely an NP-hard problem. However, there are certain criteria to check if a state is separable. The following theorem [18] makes it easy to find new criteria.

Theorem 5. *A state ρ is separable if and only if $(\mathbb{I}\otimes\Lambda)\rho \succeq 0$ for all positive but not completely positive maps Λ.*

This means that for every positive but not completely positive map, we can introduce a new criterion.

2.5.1 PPT criterion

The criterion which we will investigate is the positive partial transpose (PPT) criterion, also known as the Peres–Horodecki criterion. Note that the map $\rho \mapsto \rho^T$ is positive because transposition preserves eigenvalues. On the other hand it is not completely positive. Consider the matrix

$$A = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}. \tag{2.96}$$

Applying the map $\mathbb{I}_2 \otimes T$, where T is the transposition map $\rho \mapsto \rho^T$, yields

$$(\mathbb{I}_2 \otimes T)A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.97}$$

This matrix has eigenvalues $1, 1, 1, -1$ and is thus not positive semidefinite. It follows that the transposition map is not a completely positive map.

Thus from Theorem 5 it directly follows that for every separable state $(\mathbb{I}\otimes T)\rho \succeq 0$. This criterion is sufficient to show separability for small matrices [18], but will fail for the higher dimensional states we will consider.

The map $\mathbb{I}_A \otimes T_B$ is known as the **partial transpose**. Here the indices indicate the system the maps work on. We will denote the partial transpose as

$$(\mathbb{I}_A \otimes T_B)\rho_{AB} = \rho_{AB}^{\Gamma_B}. \tag{2.98}$$

The set of all states that are positive under partial transposition will be denoted 'PPT'.

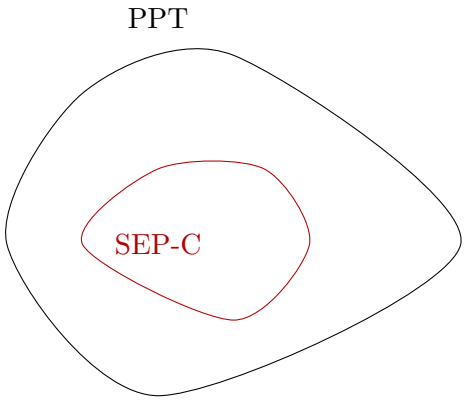


Figure 2.6: *The set of separable Choi states, SEP-C, is contained in the larger set of PPT matrices. On this set, the objective function is convex.*

Where before there was $\hat{C}_{1,\hat{A},A'} \otimes \hat{C}_{1,\hat{B},B'}$ in our program, this will now be replaced by the potentially non-separable $\hat{C}_{1,\hat{A},A',\hat{B},B'}$ and the PPT criterion will be added as a constraint. The expressions $\hat{C}_{1,\hat{A},A'}$ and $\hat{C}_{1,\hat{B},B'}$ now refer to the partial traces of $\hat{C}_{1,\hat{A},A',\hat{B},B'}$.

$$\begin{aligned}
\text{maximize} \quad & \delta^{-1}|A||B| \operatorname{tr}\left(|\Phi\rangle\langle\Phi|_{D,\hat{A},\hat{B}} \otimes \rho^T_{A'B'}\left(\hat{C}_{1,\hat{A},A',\hat{B},B'}\right)\right) \\
\text{subject to} \quad & |A||B| \operatorname{tr}\left(\rho^T_{A'B'}\hat{C}_{1,A',B'}\right) = \delta \ , \\
& \hat{C}_{1,\hat{A},A'} \succeq 0,\ \hat{C}_{1,\hat{B},B'} \succeq 0 \ , \\
& \hat{C}_{1,A'} \preceq \tfrac{\mathbb{I}_{A'}}{|A|},\ \hat{C}_{1,B'} \preceq \tfrac{\mathbb{I}_{B'}}{|B|} \ , \\
& \hat{C}^{\Gamma}_{1,\hat{A},A',\hat{B},B'} \succeq 0 \ .
\end{aligned}$$

By observing that $\operatorname{tr}, \rho \mapsto \rho^{\Gamma_B}$ and tr_A are Hermiticity preserving maps, we see that the problem above is a valid SDP.

Non-local flags

We see that in the PPT program, the protocols with local and non-local flags can be relaxed to the same programme. To see this more clearly, let us demonstrate that this holds for the last condition for which it might not be completely obvious. Firstly let us examine the case with local flags, then $\hat{C}_{1,\hat{A},A',\hat{B},B'} = \hat{C}_{1,\hat{A},A'} \otimes \hat{C}_{1,\hat{B},B'}$. Hence:

$$\hat{C}_{1,A',B'} = \hat{C}_{1,A'} \otimes \hat{C}_{1,B'} \preceq \frac{\mathbb{I}_{A'}}{|A|} \otimes \frac{\mathbb{I}_{B'}}{|B|} = \frac{\mathbb{I}_{A'B'}}{|A||B|}. \tag{2.99}$$

For the case with non-local flags we have $\hat{C}_{1,\hat{A},A',\hat{B},B'} = \hat{C}_{1,\hat{A},A'} \otimes \hat{C}_{1,\hat{B},B'} + \hat{C}_{0,\hat{A},A'} \otimes \hat{C}_{0,\hat{B},B'}$. Hence:

$$\begin{aligned}
\hat{C}_{1,A',B'} &= \hat{C}_{1,A'} \otimes \hat{C}_{1,B'} + \hat{C}_{0,A'} \otimes \hat{C}_{0,B'} && (2.100)\\
&= \hat{C}_{1,A'} \otimes \hat{C}_{1,B'} + \left(\frac{\mathbb{I}_{A'}}{|A|} - \hat{C}_{1,A'}\right) \otimes \left(\frac{\mathbb{I}_{B'}}{|B|} - \hat{C}_{0,A'}\right) && (2.101)\\
&= \frac{\mathbb{I}_{A'B'}}{|A||B|} + \hat{C}_{1,A'} \otimes \hat{C}_{1,B'} + \hat{C}_{1,A'} \otimes \hat{C}_{1,B'} - \hat{C}_{1,A'} \otimes \frac{\mathbb{I}_{B'}}{|B|} - \frac{\mathbb{I}_{A'}}{|A|} \otimes \hat{C}_{1,B'} && (2.102)\\
&\preceq \frac{\mathbb{I}_{A'B'}}{|A||B|} + \hat{C}_{1,A'} \otimes \frac{\mathbb{I}_{B'}}{|B|} + \frac{\mathbb{I}_{A'}}{|A|} \otimes \hat{C}_{1,B'} - \hat{C}_{1,A'} \otimes \frac{\mathbb{I}_{B'}}{|B|} - \frac{\mathbb{I}_{A'}}{|A|} \otimes \hat{C}_{1,B'} && (2.103)\\
&= \frac{\mathbb{I}_{A'B'}}{|A||B|}. && (2.104)
\end{aligned}$$

In conclusion, we have the following PPT program

$$\begin{aligned} \text{maximize} \quad & \delta^{-1}|A||B| \operatorname{tr}\left(|\Phi\rangle\langle\Phi|_{D,\hat{A},\hat{B}} \otimes \rho^T_{A'B'}\left(\hat{C}_{1,\hat{A},A',\hat{B},B'}\right)\right) \\ \text{subject to} \quad & |A||B| \operatorname{tr}\left(\rho^T_{A'B'}\hat{C}_{1,A',B'}\right) = \delta \\ & \hat{C}_{1,A',B'} \preceq \frac{\mathbb{I}_{A'B'}}{|A||B|} \\ & \hat{C}^{\Gamma}_{1,\hat{A},A',\hat{B},B'} \succeq 0 \\ & \hat{C}_{1,\hat{A},A',\hat{B},B'} \succeq 0. \end{aligned} \tag{2.105}$$

Symmetry reduction

In this program a symmetry reduction is possible. An analagous symmetry reduction has been done before by Rains [26]. However, in Rains' derivation no probability of success was included, which will be included here. We state the result of the symmetry reduction in the following lemma.

Lemma 5. *The program with PPT constraints,*

$$\begin{aligned} \textit{maximize} \quad & \delta^{-1}|A||B| \operatorname{tr}\left(|\Phi\rangle\langle\Phi|_{D,\hat{A},\hat{B}} \otimes \rho^T_{A'B'}\left(\hat{C}_{1,\hat{A},A',\hat{B},B'}\right)\right) \\ \textit{subject to} \quad & |A||B| \operatorname{tr}\left(\rho^T_{A'B'}\hat{C}_{1,A',B'}\right) = \delta \\ & \hat{C}_{1,A',B'} \preceq \frac{\mathbb{I}_{A'B'}}{|A||B|} \\ & \hat{C}^{\Gamma}_{1,\hat{A},A',\hat{B},B'} \succeq 0 \\ & \hat{C}_{1,\hat{A},A',\hat{B},B'} \succeq 0. \end{aligned} \tag{2.106}$$

yields the same optimal value as the following program

$$\begin{aligned} \textit{maximize} \quad & \delta^{-1}|A||B| \operatorname{tr}\big(\rho^T_{A'B'}M_{A'B'}\big) \\ \textit{subject to} \quad & |A||B| \operatorname{tr}\big(\rho^T_{A'B'}(M_{A'B'} + E_{A'B'})\big) = \delta \\ & M_{A'B'} + E_{A'B'} \preceq \frac{\mathbb{I}_{A'B'}}{|A'||B'|} \\ & M^{\Gamma}_{A'B'} + \frac{1}{D+1}E^{\Gamma}_{A'B'} \succeq 0 \\ & -M^{\Gamma}_{A'B'} + \frac{1}{D-1}E^{\Gamma}_{A'B'} \succeq 0 \\ & M_{A'B'} \succeq 0 \\ & E_{A'B'} \succeq 0. \end{aligned} \tag{2.107}$$

We will show that this reduction is possible by proving that there is a symmetry in the maximally entangled state. Following from this symmetry we will see that a similar symmetry holds for the optimal solution of the PPT program. Due to this symmetry the optimal solution can be decomposed into two smaller variables. Finally, with this decomposition the equivalent program will be constructed.

Symmetry in maximally entangled state The maximally entangled state is invariant under unitaries of the form $U_A \otimes U_B^*$, where U is any unitary matrix, the indices indicate the system they work on and * denotes complex conjugation and † the Hermitian transpose. Using Theorem 1 we have

$$(U \otimes U^*)\Phi_D = (UU^\dagger \otimes \mathbb{I})\Phi_D = \Phi_D, \tag{2.108}$$

where we write $\Phi_D = |\Phi_D\rangle\langle\Phi_D|$ for brevity. Because of this invariance, it is invariant under twirling over all unitaries

$$\int \mathrm{d}U (U_{\hat{A}} \otimes U_{\hat{B}}^*)\Phi_D(U_{\hat{A}} \otimes U_{\hat{B}}^*)^\dagger = \Phi_D. \tag{2.109}$$

This twirling operation will be denoted τ:

$$\tau(\rho) = \int \mathrm{d}U (U_{\hat{A}} \otimes U_{\hat{B}}^*)\rho(U_{\hat{A}} \otimes U_{\hat{B}}^*)^\dagger. \tag{2.110}$$

Symmetry in the optimal From this symmetry follows that the optimal solution has the same type of symmetry.

$$\mathrm{tr}\left(\Phi_{D,\hat{A},\hat{B}} \otimes \rho_{A'B'}^T \hat{C}_{1,\hat{A},A',\hat{B},B'}\right) \tag{2.111}$$

$$= \mathrm{tr}\left(\left(\int dU \left(U_{\hat{A}} \otimes U_{\hat{B}}^*\right) \Phi_{D,\hat{A},\hat{B}} (U_{\hat{A}} \otimes U_{\hat{B}}^*)^\dagger \otimes \rho_{A'B'}^T\right) \hat{C}_{1,\hat{A},A',\hat{B},B'}\right) \tag{2.112}$$

$$= \mathrm{tr}\left(\left(\Phi_{D,\hat{A},\hat{B}} \otimes \rho_{A'B'}^T\right)\left(\int dU \left(U_{\hat{A}} \otimes U_{\hat{B}}^* \otimes \mathbb{I}_{A'B'}\right) \hat{C}_{1,\hat{A},A',\hat{B},B'} (U_{\hat{A}} \otimes U_{\hat{B}}^* \otimes \mathbb{I}_{A'B'})^\dagger\right)\right) \tag{2.113}$$

The optimal solution $\hat{C}_{1,\hat{A},A',\hat{B},B'}$ will be denoted below by $\hat{C}_1$ for brevity. This shows that if $\hat{C}_1$ is optimal, the twirled state

$$\tau(\hat{C}_1) = \int \mathrm{d}U (U_A \otimes U_B^* \otimes \mathbb{I})\hat{C}_1(U_A \otimes U_B^* \otimes \mathbb{I})^\dagger \tag{2.114}$$

is optimal. To see that $\tau(\hat{C}_1)$ is feasible, the following statements have to be true for it to meet the constraints in Eq. (2.106).

$$|A||B|\,\mathrm{tr}\Big((\mathbb{I}_{\hat{A}\hat{B}} \otimes \rho_{A'B'}^T)\tau(\hat{C}_1)\Big) = \delta \tag{2.115}$$

$$\tau(\hat{C}_1) \succeq 0 \tag{2.116}$$

$$\tau(\hat{C}_1)^\Gamma \succeq 0 \tag{2.117}$$

$$\mathrm{tr}_{\hat{A},\hat{B}}(\tau(\hat{C}_1)) \preceq \frac{\mathbb{I}_{A'B'}}{|A||B|} \tag{2.118}$$

First note that the conjugate channel of τ equals τ (i.e. $\tau = \tau^\dagger$):

$$\begin{aligned}\mathrm{tr}\Big(\tau(\rho_{\hat{A},\hat{B}})\sigma_{\hat{A},\hat{B}}\Big) &= \mathrm{tr}\left(\int dU \rho_{\hat{A},\hat{B}}(U_{\hat{A}} \otimes U_{\hat{B}}^*)\sigma_{\hat{A},\hat{B}}(U_{\hat{A}} \otimes U_{\hat{B}}^*)^\dagger\right) \\ &= \mathrm{tr}\left(\int dU (U_{\hat{A}} \otimes U_{\hat{B}}^*)^\dagger \rho_{\hat{A},\hat{B}}(U_{\hat{A}} \otimes U_{\hat{B}}^*)\sigma_{\hat{A},\hat{B}}\right) \\ &= \mathrm{tr}\Big(\rho_{\hat{A},\hat{B}}\tau(\sigma_{\hat{A},\hat{B}})\Big)\end{aligned}$$

Thus for (2.115) holds:

$$|A||B|\operatorname{tr}\Big((\mathbb{I}_{\hat{A}\hat{B}}\otimes\rho^T_{A'B'})\tau(\hat{C}_1)\Big) = \operatorname{tr}\Big(\tau(\mathbb{I}_{\hat{A}\hat{B}}\otimes\rho^T_{A'B'})\hat{C}_1\Big) \tag{2.119}$$

$$= |A||B|\operatorname{tr}\Big((\tau(\mathbb{I}_{\hat{A}\hat{B}})\otimes\rho^T_{A'B'})\hat{C}_1\Big) \tag{2.120}$$

$$= |A||B|\operatorname{tr}\Big((\mathbb{I}_{\hat{A}\hat{B}}\otimes\rho^T_{A'B'})\hat{C}_1\Big) = \delta \tag{2.121}$$

Where $\tau(\mathbb{I}_{\hat{A}\hat{B}}\otimes\rho^T_{A'B'}) = \tau(\mathbb{I}_{\hat{A}\hat{B}})\otimes\rho^T_{A'B'}$, because the twirling operation acts on the systems $\hat{A},\hat{B}$. By the same argument, this shows that Eq. (2.118) holds. Moreover because τ is a channel, it is also completely positive. Hence $\tau(\hat{C}_1)\succeq 0$. Following [19] we see that τ is an LOCC channel, and therefore preserves the PPT property of states. Consequently, Eq. (2.117) holds. We conclude that $\tau(\hat{C}_1)$ is feasible.

Decomposition of the optimal Analogous to [26] we will decompose C_1. To this end, we express the Choi state C_1 in terms of the Choi state of the twirling operation C_T. First, we decompose the twirling operation. Consider the group $\{U\otimes U^*\}$ under multiplication, where U indicates a unitary matrix. It is known from representation theory[15] that this group has two inequivalent irreducible representations. The first spanned by the maximally entangled state: $V_0 = \operatorname{span}\{|\Phi_D\rangle\}$. The second irreducible representation is the space orthogonal to V_0. Following [3], we can decompose the twirling operation τ to

$$\tau(\rho) = \operatorname{tr}\left(\tau(\rho)\Phi_{D,\hat{A},\hat{B}}\right)\Phi_{D,\hat{A},\hat{B}} + \operatorname{tr}\left(\tau(\rho)\left(\mathbb{I}_{\hat{A},\hat{B}} - \Phi_{D,\hat{A},\hat{B}}\right)\right)\frac{\mathbb{I}_{\hat{A},\hat{B}} - \Phi_{D,\hat{A},\hat{B}}}{D^2-1}. \tag{2.122}$$

Specifically we use the two Schur's lemmas as stated in [3].

Lemma 6 (Schur's first)**.** *If $T(g)$ is an irreducible representation of the group G on a Hilbert space $\mathcal{H}$, then any operator satisfying $T(g)AT^\dagger(g) = A$ for all $g\in G$ is a multiple of the identity on $\mathcal{H}$.*

Lemma 7 (Schur's second)**.** *If $T_1(g)$ and $T_2(g)$ are inequivalent representations of G, then $T_1(g)AT_2^\dagger(g) = A$ for all $g\in G$ implies $A = 0$.*

The expression in Eq. (2.122) can be simplified by removing the twirling from the arguments of the trace functions. Since τ is a trace-preserving operation, so $\operatorname{tr}(\tau(\rho)) = \operatorname{tr}(\rho)$ and $\tau^\dagger = \tau$, Eq. (2.122) reduces to

$$\tau(\rho_{\hat{A},\hat{B}}) = \operatorname{tr}\left(\rho\Phi_{D,\hat{A},\hat{B}}\right)\Phi_{D,\hat{A},\hat{B}} + \operatorname{tr}\left(\rho\left(\mathbb{I}-\Phi_D\right)\right)\frac{\mathbb{I}_{\hat{A},\hat{B}} - \Phi_{D,\hat{A},\hat{B}}}{D^2-1}. \tag{2.123}$$

Before we can show that the Choi state of the optimal solution can be expressed in terms of the Choi state of the twirling, we will need some tools. The Choi state of the twirling operation is given by

$$\hat{C}_\tau = \frac{1}{D^2}\left((\Phi_{D,\hat{A},\hat{B}}\otimes\Phi_{DA',B'} + \frac{1}{D^2-1}\left(\mathbb{I}_{\hat{A},\hat{B}} - \Phi_{D,\hat{A},\hat{B}}\right)\otimes\left(\mathbb{I}_{A',B'} - \Phi_{DA',B'}\right)\right). \tag{2.124}$$

This is verified by substitution into

$$\tau(\rho) = |A||B|\operatorname{tr}_{\hat{A}',\hat{B}'}\left[\left(\mathbb{I}_{\hat{A},\hat{B}}\otimes\rho^T_{\hat{A}',\hat{B}'}\right)C_\tau\right]. \tag{2.125}$$

We see:

$$|A||B|\,\mathrm{tr}_{\hat{A}',\hat{B}'}\left[\left(\mathbb{I}_{\hat{A},\hat{B}}\otimes\rho^T_{\hat{A}',\hat{B}'}\right)C_\tau\right]=\mathrm{tr}_{\hat{A}',\hat{B}'}\left[\left(\mathbb{I}_{\hat{A},\hat{B}}\otimes\rho^T_{\hat{A}',\hat{B}'}\right)\Phi_{D,\hat{A},\hat{B}}\otimes\Phi_{D\,\hat{A}',\hat{B}'}\right] \tag{2.126}$$

$$+\,\mathrm{tr}_{\hat{A}',\hat{B}'}\left[\left(\mathbb{I}_{\hat{A},\hat{B}}\otimes\rho^T_{A'B'}\right)\frac{1}{D^2-1}\left(\mathbb{I}_{\hat{A},\hat{B}}-\Phi_{D,\hat{A},\hat{B}}\right)\otimes\left(\mathbb{I}_{\hat{A}',\hat{B}'}-\Phi_{D\,\hat{A}',\hat{B}'}\right)\right] \tag{2.127}$$

$$=\mathrm{tr}\left(\rho^T\Phi_D\right)\Phi_{D,\hat{A},\hat{B}}+\mathrm{tr}\left(\rho^T\left(\mathbb{I}-\Phi_D\right)\right)\frac{\mathbb{I}_{\hat{A},\hat{B}}-\Phi_{D,\hat{A},\hat{B}}}{D^2-1} \tag{2.128}$$

$$\tag{2.129}$$

The first tool we use is the identity $(T\circ\tau)=(\tau\circ T)=\tau$ where T indicates the transposition map $\rho\mapsto\rho^T$. To see why this holds consider for any operator σ

$$(T\circ\tau)\sigma_{\hat{A},\hat{B}}=\mathrm{tr}\left[\sigma\Phi_D\right]\Phi_{D\,\hat{A},\hat{B}}^T+\mathrm{tr}\left[\sigma\left(\mathbb{I}-\Phi_D\right)\right]\frac{\mathbb{I}^T_{\hat{A},\hat{B}}-\Phi_{D\,\hat{A},\hat{B}}^T}{D^2-1} \tag{2.130}$$

$$=\mathrm{tr}\left[\sigma\Phi_D\right]\Phi_{D,\hat{A},\hat{B}}+\mathrm{tr}\left[\sigma\left(\mathbb{I}-\Phi_D\right)\right]\frac{\mathbb{I}_{\hat{A},\hat{B}}-\Phi_{D,\hat{A},\hat{B}}}{D^2-1} \tag{2.131}$$

and

$$(\tau\circ T)\sigma_{\hat{A},\hat{B}}=T(\sigma^T_{\hat{A},\hat{B}}) \tag{2.132}$$

$$=\mathrm{tr}\left[\sigma^T\Phi_D\right]\Phi_{D,\hat{A},\hat{B}}+\mathrm{tr}\left[\sigma^T\left(\mathbb{I}-\Phi_D\right)\right]\frac{\mathbb{I}_{\hat{A},\hat{B}}-\Phi_{D,\hat{A},\hat{B}}}{D^2-1} \tag{2.133}$$

$$=\mathrm{tr}\left[\sigma^T{\Phi_D}^T\right]\Phi_{D,\hat{A},\hat{B}}+\mathrm{tr}\left[\sigma^T\left(\mathbb{I}^T-{\Phi_D}^T\right)\right]\frac{\mathbb{I}_{\hat{A},\hat{B}}-\Phi_{D,\hat{A},\hat{B}}}{D^2-1} \tag{2.134}$$

$$=\mathrm{tr}\left[\sigma\Phi_D\right]\Phi_{D,\hat{A},\hat{B}}+\mathrm{tr}\left[\sigma\left(\mathbb{I}-\Phi_D\right)\right]\frac{\mathbb{I}_{\hat{A},\hat{B}}-\Phi_{D,\hat{A},\hat{B}}}{D^2-1}. \tag{2.135}$$

Furthermore, recall that the conjugate map of the twirling is itself, i.e. $\tau=\tau^\dagger$. Also recall the identity stated in Theorem 1:

$$(\Lambda\otimes\mathbb{I})\Phi=(\mathbb{I}\otimes T\circ\Lambda^\dagger\circ T)\Phi \tag{2.136}$$

Finally, we have denote the map Ψ to be the actual distilling operator $\Lambda_{AB\to\hat{A}\hat{B}F}$ performed by Alice and Bob, followed by the post-selection and measurement of the flags. Due to this measurement and post-selection, We define Ψ as

$$\Psi_{AB\to\hat{A}\hat{B}}(\rho_{AB})=\mathrm{tr}_F\left((\mathbb{I}_{\hat{A}\hat{B}}\otimes P_F)\Lambda_{AB\to\hat{A}\hat{B}F}(\rho_{AB})\right) \tag{2.137}$$

where $P_F=|1\rangle\langle 1|_F$. Note that Ψ is a completely positive linear map but is not a valid quantum operation because the projection on the success flag is not trace preserving. However, it can be shown that Eq. (2.136) still holds.

Using these tools we can express the optimal state $\hat{C}_1$ in terms of C_τ:

$$\hat{C}_1=(\tau_{\hat{A},\hat{B}}\circ\Psi_{AB\to\hat{A}\hat{B}}\otimes\mathbb{I}_{A',B'})\Phi_{A,B,A',B'} \tag{2.138}$$

$$=(\mathbb{I}_{\hat{A},\hat{B}}\otimes T_{A'B'}\circ(\tau_{\hat{A}'\hat{B}'}\circ\Psi_{\hat{A}'\hat{B}'\to A'B'})^\dagger\circ T_{\hat{A}'\hat{B}'})\Phi_{\hat{A},\hat{B},\hat{A}'\hat{B}'} \tag{2.139}$$

$$=(\mathbb{I}_{\hat{A},\hat{B}}\otimes T_{A'B'}\circ\Psi^\dagger_{\hat{A}'\hat{B}'\to A'B'}\circ\tau_{\hat{A}'\hat{B}'}\circ T_{\hat{A}'\hat{B}'})\Phi_{\hat{A},\hat{B},\hat{A}'\hat{B}'} \tag{2.140}$$

$$=(\mathbb{I}_{\hat{A},\hat{B}}\otimes T_{A'B'}\circ\Psi^\dagger_{\hat{A}'\hat{B}'\to A'B'}\circ\tau_{\hat{A}'\hat{B}'})\Phi_{\hat{A},\hat{B},\hat{A}'\hat{B}'} \tag{2.141}$$

$$=(T_{\hat{A},\hat{B}}\circ T^\dagger_{\hat{A},\hat{B}}\circ T_{\hat{A},\hat{B}}\otimes T_{A'B'}\circ\Psi^\dagger_{\hat{A}'\hat{B}'\to A'B'})\Phi_{\hat{A},\hat{B},\hat{A}'\hat{B}'} \tag{2.142}$$

$$=(\tau_{\hat{A},\hat{B}}\otimes T_{A'B'}\circ\Psi^\dagger_{\hat{A}'\hat{B}'\to A'B'})\Phi_{\hat{A},\hat{B},\hat{A}'\hat{B}'} \tag{2.143}$$

$$=(\mathbb{I}_{\hat{A},\hat{B}}\otimes T_{A'B'}\circ\Psi^\dagger_{\hat{A}'\hat{B}'\to A'B'})C_\tau \tag{2.144}$$

Thus by reviewing the Choi state for the twirl (2.124) and substituting into (2.144), we have the expression

$$\hat{C}_1 = \Phi_{D,\hat{A},\hat{B}} \otimes M + \frac{1}{D^2-1}\left(\mathbb{I}_{\hat{A},\hat{B}} - \Phi_{D,\hat{A},\hat{B}}\right) \otimes E, \tag{2.145}$$

where M and E are defined as

$$M = \frac{(\Psi^\dagger(\Phi_{DA',B'}))^T}{D^2} \tag{2.146}$$

$$E = \frac{\left(\Psi^\dagger\left(\mathbb{I}_{A',B'} - \Phi_{D,A',B'}\right)\right)^T}{D^2}. \tag{2.147}$$

Since $\Psi^\dagger$ is a completely positive map and the transposition map is a positive map, it follows that

$$M, E \succeq 0. \tag{2.148}$$

These will be the new variables that will be optimised over.

Adding PPT constraints We can add the PPT constraint on Eq. (2.145) by noting that Φ_D^Γ is a swap operator.

$$\Phi_D = \frac{1}{D}\sum_{1\le i,j\le D} |i\rangle\langle j| \otimes |i\rangle\langle j| \tag{2.149}$$

$$\Phi_D^\Gamma = \frac{1}{D}\sum_{1\le i,j\le D} |i\rangle\langle j| \otimes |j\rangle\langle i| \tag{2.150}$$

To see that this is a swap operator, let any bipartite state $|u\rangle \otimes |v\rangle$ with $|u\rangle$ and $|v\rangle$ elements of the standard basis. We have

$$\Phi_D^\Gamma(|u\rangle \otimes |v\rangle) = \frac{1}{D}\sum_{1\le i,j\le D} |i\rangle\langle j| \otimes |j\rangle\langle i| \left(|u\rangle \otimes |v\rangle\right) \tag{2.151}$$

$$= \frac{1}{D}\sum_{1\le i,j\le D} |i\rangle\, \delta_{j,u} \otimes |j\rangle\, \delta_{i,v} = |v\rangle \otimes |u\rangle\,. \tag{2.152}$$

Since we can represent any bipartite state in the standard basis and $|x\rangle \mapsto \Phi_D^\Gamma |x\rangle$ is linear, it follows that Φ_D^Γ is a swap operator. We say that a state $|\varphi\rangle$ is symmetric under the swap operation if $\Phi_D^\Gamma |\varphi\rangle = |\varphi\rangle$ and antisymmetric if $\Phi_D^\Gamma |\varphi\rangle = -|\varphi\rangle$. Furthermore, note that any bipartite state has a symmetric and antisymmetric part.

$$|\alpha\rangle\,|\beta\rangle = \underbrace{\frac{|\alpha\rangle\,|\beta\rangle + |\beta\rangle\,|\alpha\rangle}{2}}_{\text{symmetric}} + \underbrace{\frac{|\alpha\rangle\,|\beta\rangle - |\beta\rangle\,|\alpha\rangle}{2}}_{\text{antisymmetric}}. \tag{2.153}$$

We can express a swap operator as $\Phi_D^\Gamma = P_s - P_a$, where P_s and P_a are projectors on the symmetric and antisymmetric subspaces respectively. Using this expression, we see

$$C_1^\Gamma = \Phi_D^\Gamma \otimes M_{A'B'}^\Gamma + \frac{(\mathbb{I} - \Phi_D)^\Gamma}{D^2-1} \otimes E_{A'B'}^\Gamma \tag{2.154}$$

$$= \frac{1}{D}(P_S - P_A) \otimes M_{A'B'}^\Gamma + \frac{\left(1-\frac{1}{D}\right)P_S + \left(1+\frac{1}{D}\right)P_A}{D^2-1} \otimes E_{A'B'}^\Gamma \tag{2.155}$$

$$= P_S \otimes \left(\frac{1}{D}M_{A'B'}^\Gamma + \frac{1-\frac{1}{D}}{D^2-1}E_{A'B'}^\Gamma\right) + P_A \otimes \left(-\frac{1}{D}M_{A'B'}^\Gamma + \frac{1+\frac{1}{D}}{D^2-1}E_{A'B'}^\Gamma\right) \succeq 0. \tag{2.156}$$

Since P_A and P_S are orthogonal, we must have

$$M^{\Gamma}_{A'B'} + \frac{1}{D+1} E^{\Gamma}_{A'B'} \succeq 0, \tag{2.157}$$

$$- M^{\Gamma}_{A'B'} + \frac{1}{D-1} E^{\Gamma}_{A'B'} \succeq 0. \tag{2.158}$$

Probability of success constraint Recall the constraint on probability of success to equal

$$|A||B| \operatorname{tr}\left(\rho^T_{A'B'} C_{1,A',B'}\right) = \delta. \tag{2.159}$$

We can express $C_{1,A',B'}$ as follows

$$\begin{aligned} C_{1,A',B'} &= \operatorname{tr}_{\hat{A},\hat{B}}(C_1) && (2.160)\\ &= \operatorname{tr}_{\hat{A},\hat{B}} \left(\Phi_{D,\hat{A},\hat{B}} \otimes M + \frac{1}{D^2-1} \left(\mathbb{I}_{\hat{A},\hat{B}} - \Phi_{D,\hat{A},\hat{B}} \right) \otimes E \right) && (2.161)\\ &= \operatorname{tr}\Big(\Phi_{D,\hat{A},\hat{B}} \Big) M + \operatorname{tr}\left(\frac{\mathbb{I}_{\hat{A},\hat{B}} - \Phi_{D,\hat{A},\hat{B}}}{D^2-1} \right) E && (2.162)\\ &= M + E. && (2.163) \end{aligned}$$

This means we can express the constraint in terms of the new variables M and E as

$$|A||B| \operatorname{tr}\big(\rho^T_{A'B'}(M+E)\big) = \delta. \tag{2.164}$$

Choi state constraints Recall the constraint for the Choi state is

$$C_{1,A',B'} \preceq \frac{\mathbb{I}}{|A||B|}. \tag{2.165}$$

As we have just seen, $C_{1,A',B'} = M + E$, thus the following constraint holds

$$M + E \preceq \frac{\mathbb{I}}{|A||B|}. \tag{2.166}$$

Objective function In this decomposition the objective function becomes, using $\Phi_D^2 = \Phi_D$:

$$\begin{aligned} &\operatorname{tr}\Big(\Phi_{D,\hat{A},\hat{B}} \otimes \rho^T_{A'B'} \left(\hat{C}_{1,\hat{A},A',\hat{B},B'} \right) \Big) && (2.167)\\ &= \operatorname{tr}\left(\Phi_{D,\hat{A},\hat{B}} \otimes \rho^T_{A'B'} \left(\Phi_{D,\hat{A},\hat{B}} \otimes M + \frac{\mathbb{I}_{\hat{A},\hat{B}} - \Phi_{D,\hat{A},\hat{B}}}{D^2-1} \otimes E \right) \right) && (2.168)\\ &= \operatorname{tr}\left(\Phi_{D,\hat{A},\hat{B}} \otimes \left(\rho^T_{A'B'} M + \frac{1}{D^2-1} \left(\Phi_{D,\hat{A},\hat{B}} - \Phi_{D,\hat{A},\hat{B}} \right) \otimes E \right) \right) && (2.169)\\ &= \operatorname{tr}\Big(\Phi_{D,\hat{A},\hat{B}} \Big) \operatorname{tr}\big(\rho^T_{A'B'} M\big) = \operatorname{tr}\big(\rho^T_{A'B'} M\big). && (2.170) \end{aligned}$$

Program with symmetry reduction Now that we have expressed the objective function and all the constraints of the original PPT program in terms of the new variables M and E, we

conclude the new program:

$$\begin{aligned} \text{maximize} \quad & \frac{|A||B|}{\delta} \operatorname{tr}\big(\rho^T_{A'B'} M_{A'B'}\big) \\ \text{subject to} \quad & |A||B| \operatorname{tr}\big(\rho^T_{A'B'}(M_{A'B'} + E_{A'B'})\big) = \delta \\ & M_{A'B'} + E_{A'B'} \preceq \frac{\mathbb{I}_{A'B'}}{|A'||B'|} \\ & M^\Gamma_{A'B'} + \frac{1}{D+1} E^\Gamma_{A'B'} \succeq 0 \\ & - M^\Gamma_{A'B'} + \frac{1}{D-1} E^\Gamma_{A'B'} \succeq 0 \\ & M_{A'B'} \succeq 0, E_{A'B'} \succeq 0. \end{aligned} \tag{2.171}$$

Real input states

Here we will show that for real input states ρ we may restrict our optimisation program to real matrices. First, define the set $\mathbb{S}$ for some symmetric $A, B \in \mathbb{R}^{n\times n}$ and $c \in \mathbb{R}$.

$$\mathbb{S} = \left\{ X \in \mathbb{C}^{n\times n} \mid 0 \preceq X \preceq A, \operatorname{tr}(XB) = c, X^\Gamma \succeq 0 \right\}. \tag{2.172}$$

Because $X \succeq 0$, it is implied that X is Hermitian. Therefore X's eigenvalues are real and thus invariant under complex conjugation. This implies $0 \preceq X^* \preceq A$. Furthermore, observe that $\operatorname{tr}(XB) = \operatorname{tr}(XB)^* = \operatorname{tr}(X^*B) = c$. Finally let $X = \sum_{ijkl} p_{ijkl} |i\rangle\langle j| \otimes |k\rangle\langle l|$ for i, j, k, l from 1 to n. Then $X^\Gamma = \sum_{ijkl} p_{ijkl} |i\rangle\langle j| \otimes |l\rangle\langle k|$ and $X^* = \sum_{ijkl} p^*_{ijkl} |i\rangle\langle j| \otimes |k\rangle\langle l|$. The partial transpose map and conjugation map commute:

$$(X^\Gamma)^* = (X^*)^\Gamma = \sum_{ijkl} p^*_{ijkl} |i\rangle\langle j| \otimes |l\rangle\langle k|. \tag{2.173}$$

Therefore

$$X^\Gamma \succeq 0 \iff (X^\Gamma)^* \succeq 0 \iff (X^*)^\Gamma \succeq 0. \tag{2.174}$$

Together this entails

$$X \in \mathbb{S} \iff X^* \in \mathbb{S}. \tag{2.175}$$

Then because $\mathbb{S}$ is a convex set, we see that the real part of any $X \in \mathbb{S}$ is also in $\mathbb{S}$:

$$X \in \mathbb{S} \implies \frac{X + X^*}{2} = \operatorname{Re}(X) \in \mathbb{S}. \tag{2.176}$$

Now consider the optimisation program for some positive semidefinite $D \in \mathbb{R}^{n\times n}$

$$\begin{aligned} \text{maximize} \quad & \operatorname{tr}(XD) \\ \text{subject to} \quad & X \in \mathbb{S} \end{aligned} \tag{2.177}$$

And let X_{opt} be an optimal solution of the program with optimal value $y_{\text{opt}} \in \mathbb{R}$. Then we can show that $\operatorname{Re}(X_{\text{opt}})$ is also an optimal solution

$$y_{\text{opt}} = \operatorname{tr}(X_{\text{opt}} D) = \operatorname{tr}\big(X^*_{\text{opt}} D\big) = \frac{1}{2}\left[\operatorname{tr}(X_{\text{opt}} D) + \operatorname{tr}\big(X^*_{\text{opt}} D\big)\right] = \operatorname{tr}(\operatorname{Re}(X_{\text{opt}}) D). \tag{2.178}$$

Thus the following program

$$\begin{aligned} \text{maximize} \quad & \operatorname{tr}(XD) \\ \text{subject to} \quad & X \in \mathbb{S} \\ & X \in \mathbb{R}^{n\times n} \end{aligned} \tag{2.179}$$

is equivalent to Eq. (2.177). This shows that the PPT programs are equivalent to that program restricted to the real matrices if ρ is real.

2.5.2 k-extension

Separable states have the property that they can be infinitely extended [12][11]. This mean that if we have a state $C_{\hat{A}_1,A'_1,\hat{B},B'}$ we can attach $k-1$ new systems such that the state becomes $C_{\hat{A}_1,A'_1,\ldots,\hat{A}_k,A'_k,\hat{B},B'}$ satisfying

$$\forall \pi \in S_k,\ \pi\left(\hat{C}_{(\hat{A}_1,A'_1),\ldots,(\hat{A}_k,A'_k),\hat{B},B'}\right)\pi^\dagger = \hat{C}_{(\hat{A}_1,B'_1),\ldots,(\hat{A}_k,A'_k),\hat{B},B'}, \tag{2.180}$$

$$\forall j \in \{1,\ldots,k\},\ \hat{C}_{(\hat{A}_1,A'_1),\hat{B},B'} = \hat{C}_{(\hat{A}_j,A'_j),\hat{B},B'}, \tag{2.181}$$

where S_k denotes the permutation group. It can be proven that this hierarchy is *complete*, in the sense that for every non-separable state there is a certain k for which the state is not k-extendable [12]. Even though this is a better approximation of the set SEP-C, the size of the matrix increases exponentially. Specifically, the dimension of the Choi state grows by a factor $|\hat{B}||B'|$ per extension. Therefore, it becomes an intractable problem very quickly.

Symmetric and antisymmetric subspaces

This SDP can be simplified. First, recall that every separable state is in the form

$$\rho_{AB} = \sum_i p_i \left|u_i\right\rangle\!\left\langle u_i\right| \otimes \left|v_i\right\rangle\!\left\langle v_i\right|. \tag{2.182}$$

We can clearly extend this state by any k systems.

$$\rho_{A^kB} = \sum_i p_i \left|u_i\right\rangle\!\left\langle u_i\right|^{\otimes k} \otimes \left|v_i\right\rangle\!\left\langle v_i\right|. \tag{2.183}$$

We say that this state lives in the symmetric subspace of A^k, meaning that if we permute any A system, the state remains unchanged. This can be seen by noting that for any permutation $\pi \in S_k, \pi\left|u_i\right\rangle^{\otimes k} = \left|u_i\right\rangle^{\otimes k}$. Such symmetric extension is called Bose symmetric. [21]

Definition 14. *Let $\rho_{AB} \succeq 0 \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$. We say ρ_{A^kB} is a k **Bose symmetric extension** (BSE) of ρ_{AB} if and only if*

1. $\rho_{A^kB} \succeq 0$,
2. $\mathrm{tr}_{A^{k-1}} = \rho_{AB}$,
3. *ρ_{AB^k} is Bose symmetric, i.e., $\rho_{AB^k}(\mathbb{I}_A \otimes P^k_{sym}) = \rho_{AB^n}$, where P^k_{sym} denotes the symmetric projector of k particles.*

One extension

We can decompose the Choi state with one extension $\hat{C}_{\hat{A}_2,A'_1,\hat{A}_2,A'_2,\hat{B},\hat{B}_1}$ into a symmetric and antisymmetric space.

$$\hat{C}_{\hat{A}_1,A'_1,\hat{A}_2,A'_2,\hat{B},\hat{B}_1} = P(W_s \oplus W_a)P^\dagger, \tag{2.184}$$

where P transforms $W_s \oplus W_a$ into the standard basis. The basis for the symmetric subspace on A_1, A'_1, A_2, A'_2 is denoted $\mathcal{B}_{A,s}$ where

$$\mathcal{B}_{A,s} = V_1 \cup V_2, \tag{2.185}$$

$$V_1 = \{\left|i\right\rangle \otimes \left|i\right\rangle : i \in \{1,\ldots,8\}\} \tag{2.186}$$

$$V_2 = \left\{(\left|i\right\rangle \otimes \left|j\right\rangle + \left|j\right\rangle \otimes \left|i\right\rangle)/\sqrt{2} : i,j \in \{0,\ldots,7\} \text{ and } i \neq j\right\} \tag{2.187}$$

Similarly, the bais of the antisymmetric subspace on A_1, A_1', A_2, A_2' is given by

$$\mathcal{B}_{A,a} = \left\{(|i\rangle \otimes |j\rangle - |j\rangle \otimes |i\rangle)/\sqrt{2} : i, j \in \{0, \ldots, 7\} \text{ and } i \neq j\right\} \tag{2.188}$$

It is easily verified that together these span the space of A_1, A_1', A_2, A_2'. By adding the basis of B's system, we get the basis of the whole space:

$$\mathcal{B}_s = \{|\phi\rangle \otimes |i\rangle : |\phi\rangle \in \mathcal{B}_{A,s}, i \in \{0, \ldots, 7\}\} \tag{2.189}$$
$$\mathcal{B}_a = \{|\phi\rangle \otimes |i\rangle : |\phi\rangle \in \mathcal{B}_{A,a}, i \in \{0, \ldots, 7\}\} \tag{2.190}$$

Using Eq. (2.184) for the Choi state we can write the objective function as follows

$$\frac{|A||B|}{\delta} \operatorname{tr}\left((\mathbb{I}_{\hat{A}_1,A_1'} \otimes |\Phi_D\rangle\langle\Phi_D|_{\hat{A}_2,\hat{B}} \otimes \rho^T_{A_2',B'})(P(W_s \oplus W_a)P^\dagger)_{\hat{A}_1,A_1',\hat{A}_2,A_2',\hat{B},B'}\right) \tag{2.191}$$

Which will be abbreviated as $\operatorname{tr}(X(W_s \oplus W_a))$, where

$$X = \frac{|A||B|}{\delta} P^\dagger(\mathbb{I}_{\hat{A}_1,A_1'} \otimes |\Phi_D\rangle\langle\Phi_D|_{\hat{A}_2,\hat{B}} \otimes \rho^T_{A_2',B'})P. \tag{2.192}$$

Since $W_a = 0$ this can be simplified further to

$$\operatorname{tr}(X_s W_s), \tag{2.193}$$

where X_s is the submatrix of X consisting of its first 288 rows and columns. Similarly we will simplify the constraint on probability of success to

$$\operatorname{tr}(Y_s W_s) = \delta, \tag{2.194}$$

where Y_s is the submatrix of Y consisting of its first 288 rows and columns. We define Y as

$$Y = |A||B|P^\dagger(\mathbb{I}_{\hat{A}_1,A_1',\hat{A}_2,\hat{B}} \otimes \rho^T_{A_2',B'})P. \tag{2.195}$$

Together, we have the following program for one extension:

$$\begin{aligned}
\text{maximize} \quad & \operatorname{tr}(X_s W_s) \\
\text{subject to} \quad & \operatorname{tr}(Y_s W_s) = \delta \\
& \operatorname{tr}_{\hat{A}_1,A_1'}\left(P(W_s \oplus 0)P^\dagger\right)^\Gamma \succeq 0 \\
& \operatorname{tr}_{\hat{A}_1,A_1',\hat{A}_2,\hat{B}}\left(P(W_s \oplus 0)P^\dagger\right) \preceq \frac{\mathbb{I}_{A_2',B'}}{|A||B|} \\
& W_s \succeq 0
\end{aligned} \tag{2.196}$$

2.5.3 Seesaw heuristic

A completely differently approach is the 'seesaw method'. In this method, we consider the original program as defined in Eq. (2.93), but make only the Choi state of Alice a variable.

$$\begin{aligned}
\text{maximize} \quad & \delta^{-1}|A||B| \operatorname{tr}\left(\Phi_{D,\hat{A},\hat{B}} \otimes \rho^T_{A'B'}\left(\hat{C}_{1,\hat{A},A'} \otimes \hat{C}_{1,\hat{B},B'}\right)\right) \\
\text{subject to} \quad & |A||B| \operatorname{tr}\left(\rho^T_{A'B'}\left(\hat{C}_{1,A'} \otimes \hat{C}_{1,B'}\right)\right) = \delta \\
& \hat{C}_{1,\hat{A},A'} \succeq 0 \\
& \hat{C}_{1,A'} \preceq \frac{\mathbb{I}_{A'}}{|A|}.
\end{aligned} \tag{2.197}$$

This program is an SDP. The objective function can be written in the standard form, by observing

$$\mathrm{tr}\left(\Phi_{D,\hat{A},\hat{B}} \otimes \rho^T_{A'B'}\left(\hat{C}_{1,\hat{A},A'} \otimes \hat{C}_{1,\hat{B},B'}\right)\right) = \mathrm{tr}\Big(\left(\Phi_{D,\hat{A}} \otimes \rho_{A'}\right)\hat{C}_{1,\hat{A},A'}\Big)\,\mathrm{tr}\Big(\left(\Phi_{D,\hat{B}} \otimes \rho_{B'}\right)\hat{C}_{1,\hat{B},B'}\Big). \tag{2.198}$$

Since $C_{1,\hat{B},B'}$ is not a variable, $\mathrm{tr}\Big(\left(\Phi_{D,\hat{B}} \otimes \rho_{B'}\right)\hat{C}_{1,\hat{B},B'}\Big)$ is a constant and it follows that Eq. (2.197) is a valid SDP. After the SDP has been solved, we fix Alice's Choi state to the optimal found and optimize over Bob's Choi state. This process is then repeated until a fix point is reached. This will not yield the actual optimal Choi state, but if we start from a known scheme we might find another scheme that is 'nearby'.

Known schemes

3.1 BBPSSW protocol

The BBPSSW protocol [4] is a protocol that performs $2 \to 1$ copy distillation, designed for two copies of the following state

$$\rho(p) = p\,|\Phi_+\rangle\langle\Phi_+| + (1-p)\mathbb{I}_4/4 \tag{3.1}$$

This state is known as a **Werner state**. If the input state is not a Werner state, the first step is to depolarize it to a Werner state. This can be done by performing a twirling operation [14] over $T = \{\mathbb{I}_4, \sigma_x^{\otimes 2}, \sigma_z^{\otimes 2}, \sigma_x^{\otimes 2}\sigma_z^{\otimes 2}\}$, where the σ's denote the Pauli matrices. The resulting state ρ' thus equals

$$\rho' = T(\rho) = \frac{1}{4}\sum_{E\in T} E\rho E^\dagger. \tag{3.2}$$

Subsequent to the twirling, local controlled NOT gates will be applied by Alice and Bob. Alice and Bob both separately measure in the eigenbasis of σ_z and note $\xi_i = 0$ if -1 was measured, and $\xi_i = 1$ otherwise. The protocol is successful if and only if $\xi_A = \xi_B$.

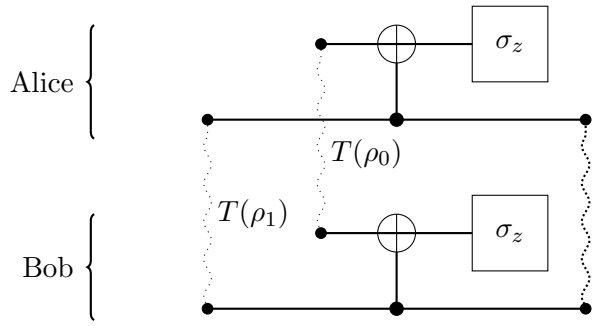


Figure 3.1: *Quantum circuit of the BBPSWW protocol. Alice and Bob share the state $\rho_0 \otimes \rho_1$. First they apply the twirl over $T = \{\mathbb{I}_4, \sigma_x^{\otimes 2}, \sigma_z^{\otimes 2}, \sigma_x^{\otimes 2}\sigma_z^{\otimes 2}\}$. After the twirling, they locally applying the CNOT gate they measure their flags in the σ_z basis. If Alice and Bob have the same outcome measurement in the σ_z basis, the distillation is successful.*

It can be shown [14] that for an input state with fidelity to the maximally entangled state $F = \mathrm{tr}(\rho\,|\Phi_+\rangle\langle\Phi_+|)$, the output state of the protocol has an output fidelity of

$$F' = \frac{F^2 + [(1-F)/3]^2}{F^2 + 2F(1-F)/3 + 5[(1-F)/3]^2}, \tag{3.3}$$

with a probability of success

$$p_{\mathrm{succ}} = F^2 + 2F(1-F)/3 + 5[(1-F)/3]^2. \tag{3.4}$$

3.2 DEJMPS protocol

The DEJMPS protocol [10] is a protocol that performs $2 \to 1$ copy distillation. Alice will perform an unitary operation U_A on both her qubits, defined by

$$|0\rangle \to \frac{1}{\sqrt{2}}(|0\rangle - i\,|1\rangle),\ |1\rangle \to \frac{1}{\sqrt{2}}(|1\rangle - i\,|0\rangle). \tag{3.5}$$

Furthermore, Bob performs in the following operation U_B on both his qubits

$$|0\rangle \to \frac{1}{\sqrt{2}}(|0\rangle + i\,|1\rangle),\ |1\rangle \to \frac{1}{\sqrt{2}}(|1\rangle + i\,|0\rangle). \tag{3.6}$$

Alice and Bob now both apply a CNOT gate on their qubits. They then measure the target qubit. If Alice's and Bob's outcomes coincide, they retain the pair they used for the control.

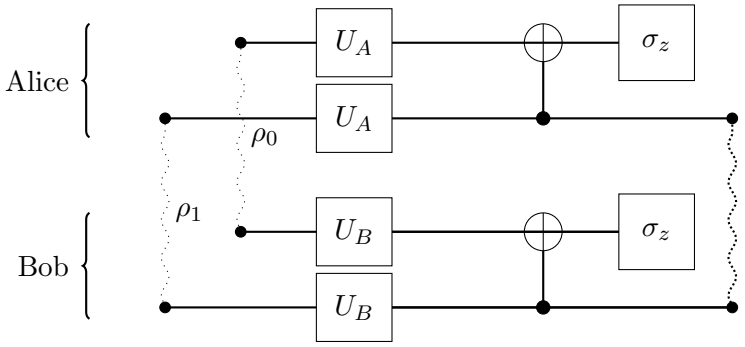


Figure 3.2: *Quantum circuit of the DEJMPS protocol. Alice and Bob share the state $\rho_0 \otimes \rho_1$. First, they perform unitaries U_A and U_B as defined in Eq. (3.5) and Eq. (3.6) respectively. After locally applying the CNOT gate they measure their flags in the σ_z basis. If Alice's and Bob's measurement coincide, the distillation is successful.*

It can be shown [10] that the output fidelity F' of the protocol equals

$$F' = \frac{A_0 A_1 + B_0 B_1}{p_{\text{succ}}}, \tag{3.7}$$

with a probability of success

$$p_{\text{succ}} = (A_0 + B_0)(A_1 + B_1) + (C_0 + D_0)(C_1 + D_1). \tag{3.8}$$

Where we define

$$A_i = \langle \Phi_+ | \rho_i | \Phi_+ \rangle, \tag{3.9}$$
$$B_i = \langle \Phi_- | \rho_i | \Phi_- \rangle, \tag{3.10}$$
$$C_i = \langle \Psi_+ | \rho_i | \Psi_+ \rangle, \tag{3.11}$$
$$D_i = \langle \Psi_- | \rho_i | \Psi_- \rangle, \tag{3.12}$$

for $i = 0, 1$.

3.3 EPL protocol

The Extreme Photon Loss (EPL) protocol [6] is a protocol that performs $2 \to 1$ copy distillation. As its name suggests it was designed to work on systems involving photons, however there is no

reason this protocol can be used on other qubit systems than photons. The circuit that describes this protocol is shown in Figure 3.3.

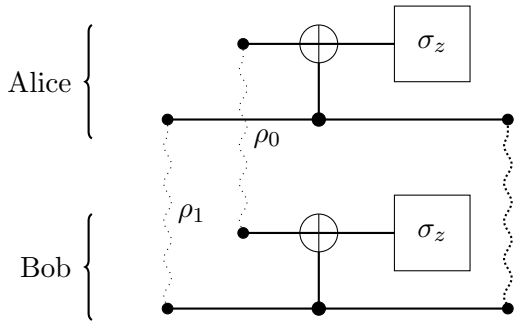


Figure 3.3: *Quantum circuit of the EPL protocol. Alice and Bob share the state $\rho_0 \otimes \rho_1$. After locally applying the CNOT gate they measure their flags in the σ_z basis. If Alice and Bob both measure '1' in the σ_z basis, the distillation is successful.*

Interestingly, this protocol will yield maximally entangled states when it operates on input states of the form

$$\rho = p\,|\varphi\rangle\langle\varphi| + (1-p)\,|11\rangle\langle 11|\,, \tag{3.13}$$

where

$$|\varphi\rangle = \frac{1}{\sqrt{2}}\left(|01\rangle + e^{i\varphi}\,|10\rangle\right). \tag{3.14}$$

These states are prevalent when creating entanglement using a scheme described by Barret and Kok [2]. In this scheme, quantum memories that emit photons are used. The goal is to entangle the quantum memories. These quantum memories can be in the states $|\uparrow\rangle, |\downarrow\rangle, |e\rangle$, where $|e\rangle$ represents an excited state. Due to physical reasons, the transition $|\uparrow\rangle \leftrightarrow |e\rangle$ is forbidden. The quantum memories are first prepared to be in state $(|\downarrow\rangle + |\uparrow\rangle)/\sqrt{2}$. The whole system is now described by

$$|\psi_1\rangle_{AB} = (|\downarrow\rangle + |\uparrow\rangle)_A/\sqrt{2} \otimes (|\downarrow\rangle + |\uparrow\rangle)_B/\sqrt{2}. \tag{3.15}$$

Then a pulse is applied, exciting the memories. This brings the system to the state

$$|\psi_1\rangle_{AB} = (|e\rangle + |\uparrow\rangle)_A/\sqrt{2} \otimes (|e\rangle + |\uparrow\rangle)_B/\sqrt{2}. \tag{3.16}$$

Now the memories might emit a photon, causing a transition $|e\rangle \to |\downarrow\rangle$. In case a photon is emitted we denote the photon system as $|1\rangle$ and if none is emitted the photon system is denoted $|0\rangle$. After emission possibly occured, the complete system, including the photon system equals

$$|\psi_2\rangle_{AB} = (|\downarrow\rangle\,|0\rangle + |\uparrow\rangle\,|1\rangle)_A/\sqrt{2} \otimes (|\downarrow\rangle\,|0\rangle + |\uparrow\rangle\,|1\rangle)_B/\sqrt{2}. \tag{3.17}$$

We then wait for a detection on either of the photon detectors. Since a beam splitter is used it is unknown which quantum memory emitted a photon. If one and only one photon is observed, the operation is successful and the quantum memories are entangled, since we project on the state $|01\rangle + |10\rangle$. Thus, the state on the memories equals

$$|\psi_2\rangle_{AB} = (|\downarrow\uparrow\rangle_{AB} + |\uparrow\downarrow\rangle_{AB})/\sqrt{2}. \tag{3.18}$$

However, it might be the case that two photons were emitted but only one was detected. This might be caused by the photon missing the detector or the photon might have been absorbed.

Then probabilistically, the state of the memories is $|\downarrow\downarrow\rangle$. If we say this happens with a probability of $(1-p)$, the mixed state equals

$$\rho = p\frac{1}{2}(|\downarrow\uparrow\rangle + |\uparrow\downarrow\rangle)(\langle\downarrow\uparrow| + \langle\uparrow\downarrow|) + (1-p)\,|\downarrow\downarrow\rangle\langle\downarrow\downarrow|\,. \tag{3.19}$$

Finally, a phase shift $e^{i\varphi}$ may occur in Eq. (3.18), which can be caused by optical apparatus [22]. Now by relabeling $|\uparrow\rangle \leftrightarrow |0\rangle\,, |\downarrow\rangle \leftrightarrow |1\rangle$, we arrive at Eq. (3.13).

The Barret and Kok scheme provides a way to get rid of the $|\uparrow\uparrow\rangle\langle\uparrow\uparrow|$ noise term and the unknown phase. Without going into further detail, it does this by applying a flip $|\uparrow\rangle \to |\downarrow\rangle$ and $|\downarrow\rangle \to |\uparrow\rangle$. Following this flip, all previously stated steps are repeated. If one detector clicks, the scheme is regarded successful and the quantum memories are maximally entangled. However, the drawback is that twice in a row, we must get one click. This is contrary to the EPL protocol, where you could stop at Eq. (3.18) and In conclusion, if first one detector clicks, then continue with the scheme and again find that one detector clicks, we are ensured that our state is maximally entangled. The drawback of this, is that we are required to successfully get one click twice in a row. With the EPL protocol, one could get one system into the state (3.13) and then fail an arbitrary number of times before obtaining another copy of (3.13).

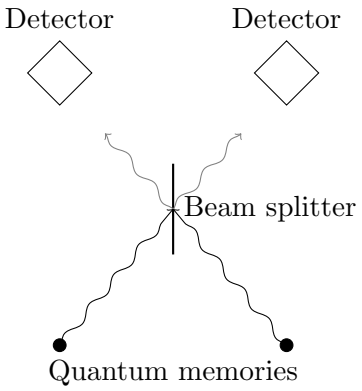


Figure 3.4: *In the Barret and Kok scheme, quantum memories may emit photons which pass through a beam splitter. In this figure, both memories emit a photon. If both photons reach the detectors, the scheme fails. However, if one photon gets lost, the process is deemed successful even though the quantum memories are not entangled.*

Returning to the EPL protocol, we see that the distillation of ρ corresponds in Figure 3.3 to distilling $\rho_1 = \rho_0 = \rho$. To see why this produces maximally entangled states, consider the possible (non-normalized) outputs

$$(|01\rangle + e^{i\varphi}\,|10\rangle)(|01\rangle + e^{i\varphi}\,|10\rangle) \mapsto |01\rangle\,(|00\rangle + e^{i\varphi}\,|11\rangle) + e^{i\varphi}\,|10\rangle\,(|00\rangle + e^{i\varphi}\,|11\rangle) \tag{3.20}$$

$$|11\rangle\,(|01\rangle + e^{i\varphi}\,|10\rangle) \mapsto |11\rangle\,(|10\rangle + e^{i\varphi}\,|01\rangle) \tag{3.21}$$

$$(|01\rangle + e^{i\varphi}\,|10\rangle)\,|11\rangle \mapsto |01\rangle\,|10\rangle + e^{i\varphi}\,|10\rangle\,|01\rangle \tag{3.22}$$

$$|11\rangle\,|11\rangle \mapsto |11\rangle\,|00\rangle\,. \tag{3.23}$$

Here, Alice and Bob measure the 'ket' on the right. So if they have the state $|00\rangle\,|11\rangle$, they will measure '11'. We see that only Eq. (3.20) has these flags. Collecting the terms, we get

$$|01\rangle\,(|00\rangle + e^{i\varphi}\,|11\rangle) + e^{i\varphi}\,|10\rangle\,(|11\rangle + e^{i\varphi}\,|00\rangle) = (|01\rangle + e^{2i\varphi}\,|10\rangle)\,|00\rangle + e^{i\varphi}(|01\rangle + |10\rangle)\,|11\rangle\,. \tag{3.24}$$

It follows that if distillation was successful, the output state is up to a - physically irrelevant - global phase equal to $(|01\rangle+|10\rangle)/\sqrt{2}$. This output state is not equal to the maximally entangled state as defined in Definition 7. However, with the unitary transformation $\mathbb{I}\otimes\sigma_x$, the second qubit is flipped and the state $(|00\rangle+|11\rangle)/\sqrt{2}$ is obtained. Thus for the input state as defined in Eq. (3.13), an output fidelity of $F'=1$ is obtained for any $p>0$.

The probability of success is determined by observing that (3.20) occurs with probility p^2, and then the probability of measuring '11' is read of from (3.24) to be 1/2. In conclusion, we have a probability of success equal to

$$p_{\text{succ}}=\frac{1}{2}p^2. \tag{3.25}$$

This raises the question if it's possible to increase the probability of success, without decreasing the output fidelity.

Results

4.1 Used software

The semidefinite programs were solved using Convex.jl and SCS (splitting conic solver) [27][25][24]. Convex.jl formats the program for SCS, which in turn will do the actual solving. In other words, Convex.jl is simply an interface for SCS. Many other interfaces exist, for example JuMP.jl for Julia or Yalmip for MATLAB.

Convex.jl supports other SDP solvers as well, namely MOSEK. It is, in contrast to Convex.jl and SCS, a proprietary package. It was found to be slower than SCS in our cases. Due to the fact that all solvers adhere to a standard called disciplined convex programming (DCP), it is very easy to try out different solvers.

The solutions given by any SDP solver are not exact, but there are well defined bounds for the errors involved. In the following results, there is a bound on the relative error. Meaning that for an output fidelity F, the solver may find an output fidelity $\tilde{F}$, where

$$\left| \frac{\tilde{F} - F}{F} \right| < \epsilon. \tag{4.1}$$

In the following results the upper bound on the relative error ϵ in the objective is 10^{-4} for the k-extensions and 10^{-7} for PPT results.

The code that was used to obtain these results is publicly available at `https://www.github.com/thomasschiet/quantum-entanglement-distillation`.

4.2 Werner state

The Werner state is defined as follows

$$\rho = p \left|\Phi_+\right\rangle\!\left\langle\Phi_+\right| + (1-p)\mathbb{I}_4/4. \tag{4.2}$$

Thus this state is maximally entangled for $p = 1$ and maximally mixed for $p = 0$.

4.2.1 2 to 1 copy distillation

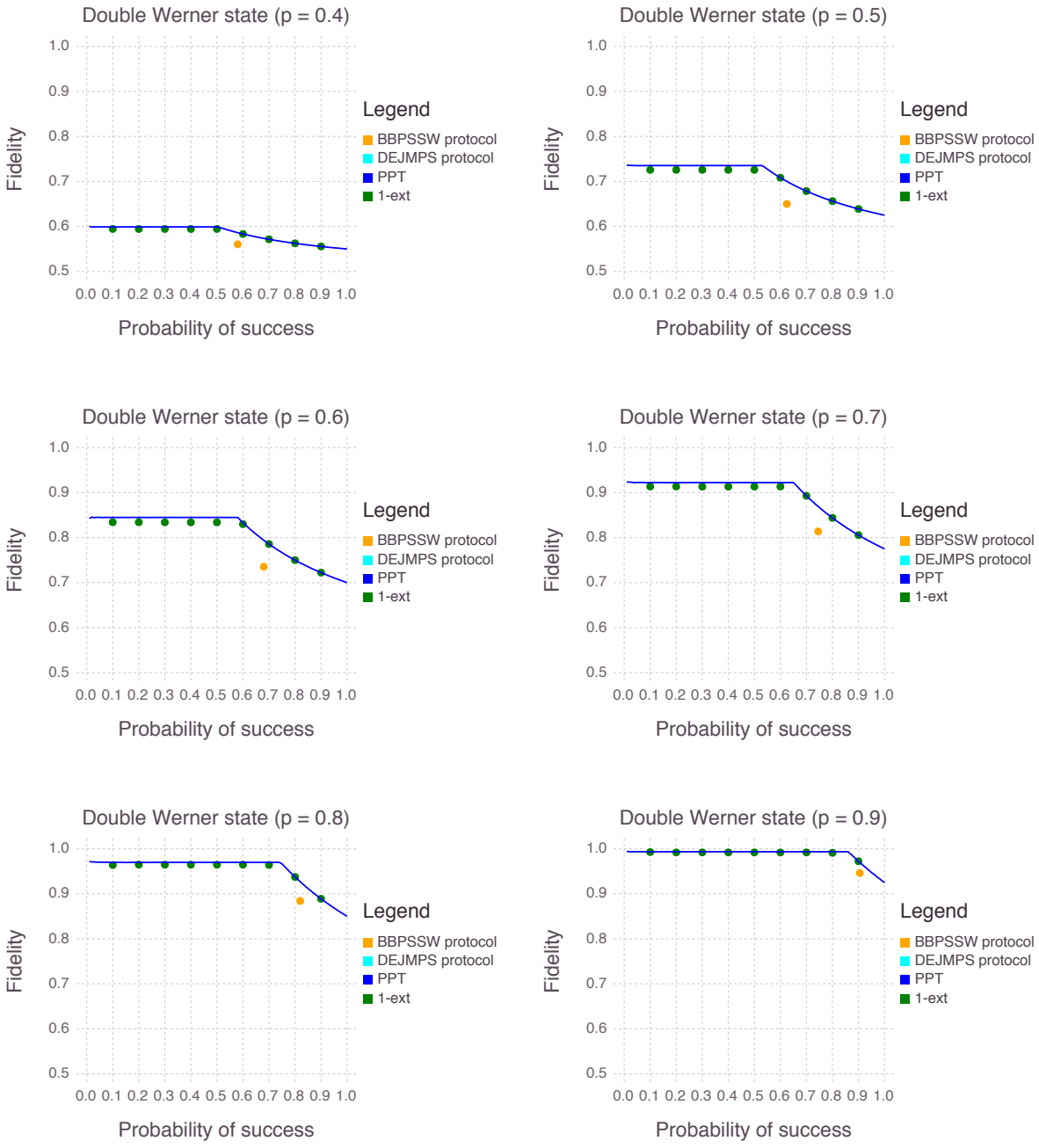


Figure 4.1: *Results of PPT and 1-extension programs. The BBPSSW and DETJMPS protocol perform equally on Werner states and thus overlap. The 1-extension program shows a small improvement up to some p_{succ} as a function of p. Neither protocol performs on the upper bound, thus their optimality for the Werner state remains an open question. The results for $p \leq 1/3$ are excluded, since for those input states the output fidelity is 0.5 for any p_{succ}.*

4.2.2 3 to 1 copy distillation

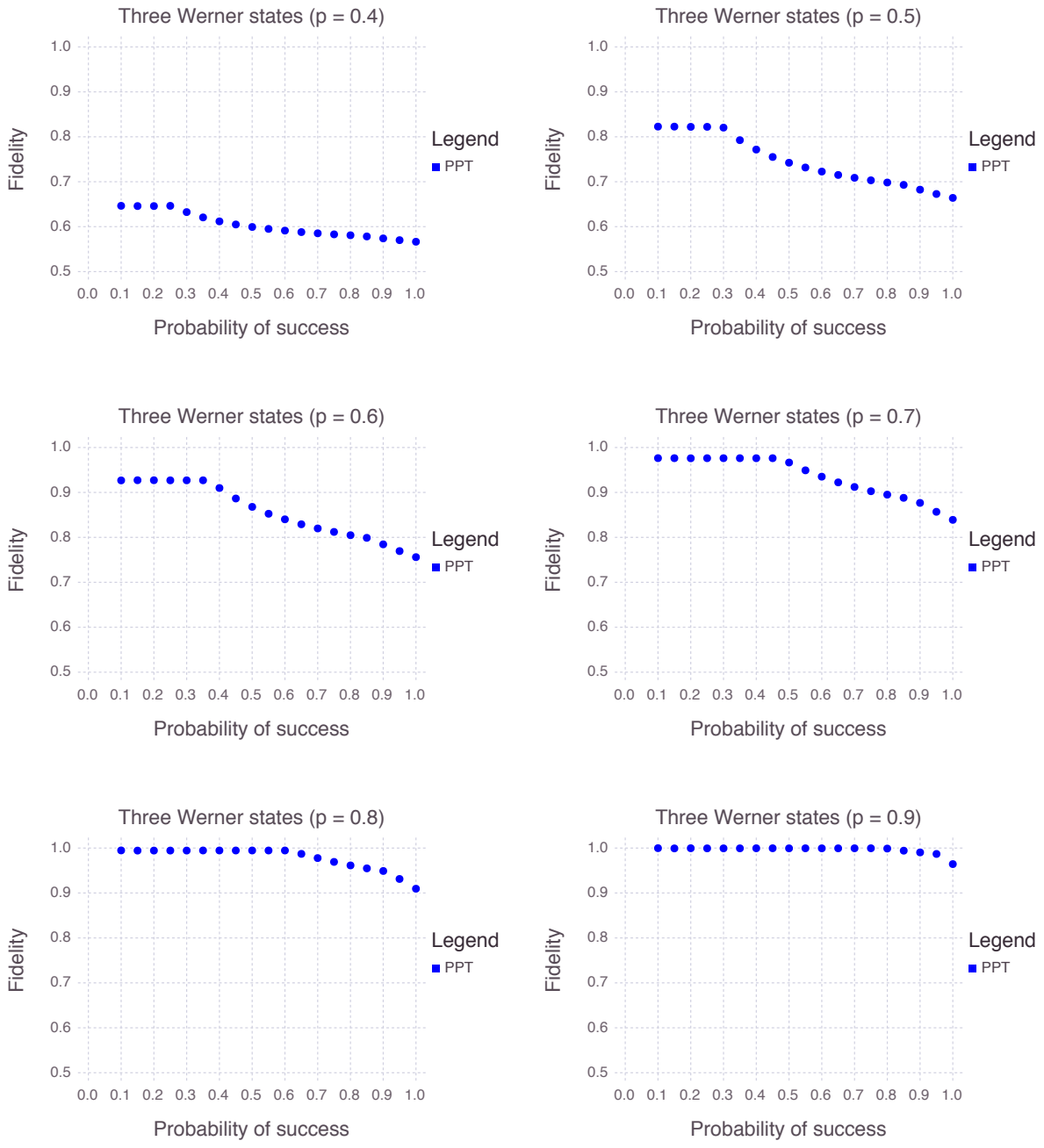


Figure 4.2: *3 to 1 copy distillation of Werner states. Because the DJEMPS and BBPSSW protocols only perform on 2 to 1 copy distillation, they are not included. The results were obtained with a relative accuracy of* $\epsilon = 10^{-4}$.

4.3 Bell state with non-orthogonal noise

This state is defined as follows

$$\rho(p, \varphi) = p\,|\Phi_+\rangle\langle\Phi_+| + (1-p)\,|11\rangle\langle 11| \tag{4.3}$$

We say its noise is non-orthogonal, because

$$\langle \Phi_+ | 11 \rangle = 1/\sqrt{2} \neq 0. \tag{4.4}$$

4.3.1 2 to 1 copy distillation

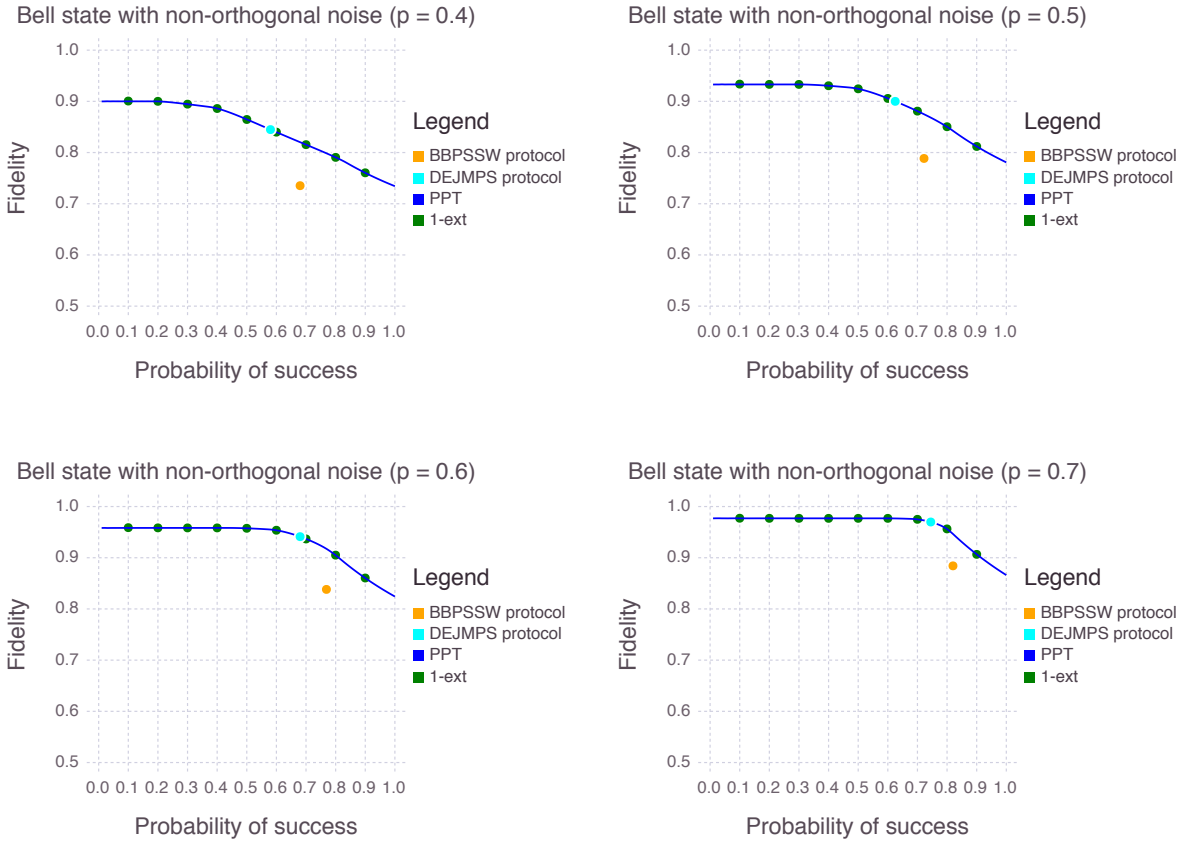


Figure 4.3: *Results of PPT and 1-extension programs. Shown is that the DEJMPS protocol is optimal and that the PPT program is sufficient to show this. Furthermore, the 1-extension showed no improvement over the PPT program. It is highly likely therefore that the results obtained in the PPT program were already 1-extendable.*

4.4 Bell state with orthogonal noise

This state is dependent on two parameters, as there is also a local phase difference φ introduced.

$$|\varphi\rangle = |01\rangle + e^{i\varphi} |10\rangle, \quad \rho_{\text{orth}}(p, \varphi) = p\, |\varphi\rangle\langle\varphi| + (1-p)\, |11\rangle\langle 11|, \tag{4.5}$$

this state frequently arises in experimental setups. As an example, it may arise in the Barret and Kok scheme as is described in Section 3.3.

4.4.1 2 to 1 copy distillation $\varphi = 0$

We can improve the DEJMPS protocol. Following [9], we see that for Bell diagonal states, the protocol performs best on states in the following form

$$\rho = p_{00} \left|\Phi_+\right\rangle\!\left\langle\Phi_+\right| + p_{01} \left|\Psi_+\right\rangle\!\left\langle\Psi_+\right| + p_{10} \left|\Phi_-\right\rangle\!\left\langle\Phi_-\right| + p_{11} \left|\Psi_-\right\rangle\!\left\langle\Psi_-\right| \tag{4.6}$$

such that

$$p_{00} > p_{01} \geq p_{10} \geq p_{11}. \tag{4.7}$$

We can twirl (4.5), such that in the Bell basis all non-diagonal states vanish. After twirling, the state equals

$$\rho_1 = p \left|\Psi_+\right\rangle\!\left\langle\Psi_+\right| + \frac{1-p}{2} \left|\Phi_+\right\rangle\!\left\langle\Phi_+\right| + \frac{1-p}{2} \left|\Phi_-\right\rangle\!\left\langle\Phi_-\right|, \tag{4.8}$$

First, we apply the unitary $\mathbb{I} \otimes \sigma_x$ to obtain

$$\rho_2 = p \left|\Phi_+\right\rangle\!\left\langle\Phi_+\right| + \frac{1-p}{2} \left|\Psi_+\right\rangle\!\left\langle\Psi_+\right| + \frac{1-p}{2} \left|\Psi_-\right\rangle\!\left\langle\Psi_-\right| \tag{4.9}$$

There exists a rotation that we can perform such that we can swap the coefficients of the Bell state. Thus we perform the following swap

$$\left|\Phi_-\right\rangle \leftrightarrow \left|\Psi_-\right\rangle. \tag{4.10}$$

The state now equals

$$\rho_3 = p \left|\Phi_+\right\rangle\!\left\langle\Phi_+\right| + \frac{1-p}{2} \left|\Psi_-\right\rangle\!\left\langle\Psi_-\right| + \frac{1-p}{2} \left|\Phi_-\right\rangle\!\left\langle\Phi_-\right|, \tag{4.11}$$

and the coefficients are in the correct order. As can be seen in Figure 4.4, this is increase the output fidelity, however, it lowers the probability of success.

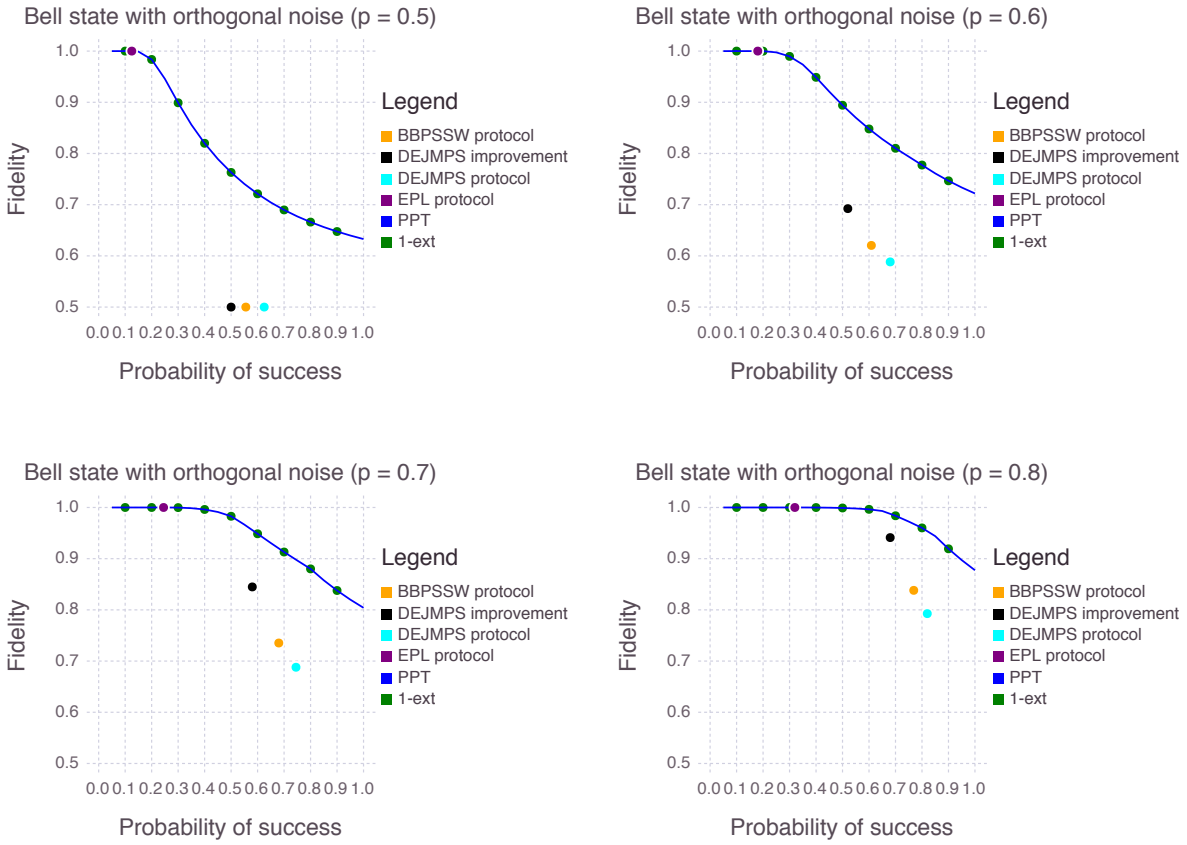


Figure 4.4: *Results of PPT and 1-extension programs. Shown is that the EPL program is optimal. Although the drop in fidelity is too small to be seen in the graph, it is visible in the numerics.*

4.5 Bell state with orthogonal noise and averaged phase

This state is already bipartite and defined as

$$\rho(p) = \frac{1}{2\pi}\int_0^{2\pi} \rho_{\text{orth}}(p,\varphi)\otimes\rho_{\text{orth}}(p,\varphi)\,\mathrm{d}\varphi. \tag{4.12}$$

This state will arise when two copies of Bell state with orthogonal noise are produced, but the phase difference φ is unknown and can be modelled to have a uniform distribution on $[0, 2\pi)$.

This integral can be evaluated, yielding

$$\rho(p) = \frac{p^2}{4}\left[P_{\text{odd}}\otimes P_{\text{odd}} + (|01\rangle\langle 10|\otimes|10\rangle\langle 01| + |10\rangle\langle 01|\otimes|01\rangle\langle 10|)\right] \tag{4.13}$$

$$+\frac{(1-p)p}{2}\left[|11\rangle\langle 11|\otimes P_{\text{odd}} + P_{\text{odd}}\otimes|11\rangle\langle 11|\right] + (1-p)^2|11\rangle\langle 11|\otimes|11\rangle\langle 11|, \tag{4.14}$$

where $P_{\text{odd}} = |01\rangle\langle 01| + |10\rangle\langle 10|$.

The data shows that EPL is optimal for this state. Not only will it always produce maximally entangled pairs, it is also impossible to design a protocol that has a higher probability of success in doing so.

4.5.1 2 to 1 copy distillation

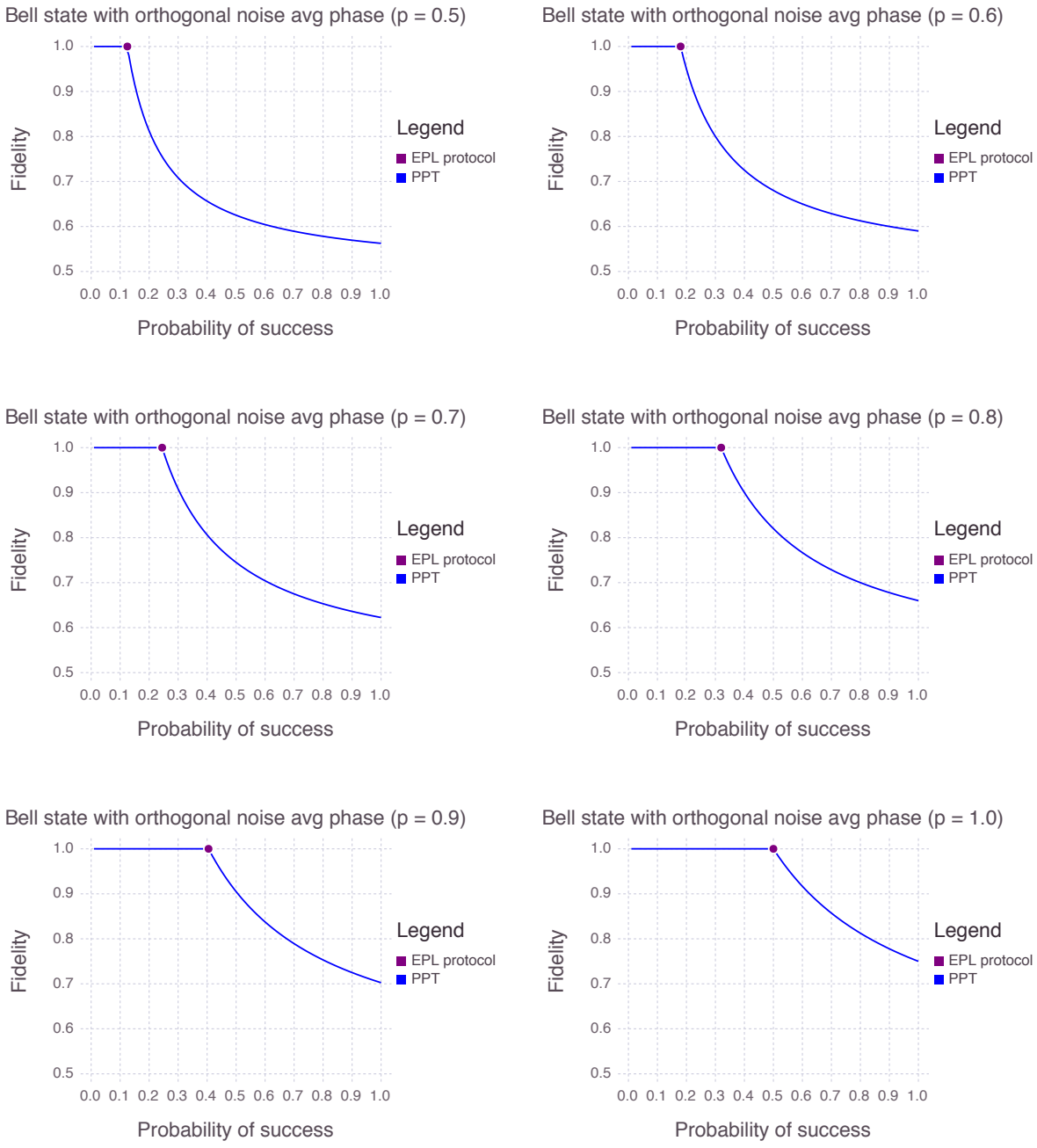


Figure 4.5: *Results of PPT programs. The EPL is shown to be optimal for this state. It is not possible to increase the probability of success while maintaining fidelity at 1, as this would violate the upper bound.*

4.5.2 4 to 1 distillation

It was also attempted to compare the two rounds of distillation with distilling the following state:

$$\rho(p) = \frac{1}{2\pi} \int_0^{2\pi} \rho_{\text{orth}}(p, \varphi)^{\otimes 4} \, \mathrm{d}\varphi. \tag{4.15}$$

Numerically, this proved to be difficult to solve. Therefore we have only been able to show that for $p = 1$, the probability of success where the fidelity starts to drop below 1 is located at $p_{\text{succ}} = 0.875 \pm 10^{-4}$.

4.6 Seesaw heuristic

The seesaw heuristic was tried using the BBPSSW protocol as a starting point, considering it wasn't optimal for any of the states above. Unfortunately this method showed no improvement, suggesting that this protocol is at a local maximum. Furthermore the program was started from various randomly generated density matrices. This only resulted in protocols with an output fidelity of 0.5 which are of no interest.

Conclusion

Upper bounds on entanglement distillation have been found by optimizing over positive partial transpose (PPT) and 1 Bose symmetric extendable (BSE) states for various input states. Furthermore, a heuristic method called the 'seesaw method' was attempted to find new entanglement distillation protocols.

It has been shown that the DEJMPS and EPL protocol achieve the upper bounds for certain input states. As such, these protocols are proven to be optimal. The PPT relaxation has been sufficient to show optimality. In only one case, namely for Werner states, it was found that a BSE improved the upper bound. Furthermore, optimizing over 2 BSE states has been computationally too expensive to implement. Using PPT states we have been able to provide upper bounds on 2, 3 and 4 to 1 copy distillation protocols. In the case of 1 BSE states we have showed upper bounds on 2 to 1 copy distillation. Finally, the seesaw method was unable to improve on the BBPSSW protocol and moreover was unable to produce useful protocols starting from random positions.

These results of great importance, since the input states for which the optimality of the schemes was shown, frequently arise in experimental setups to generate entanglement.

With optimality shown numerically for the two protocols, it is worthwhile to attempt to construct an analytical proof by guessing a solution to the dual SDP. This proves optimality not only for certain parametrisations of the involved states, but to a whole family of states. Using this method, a proof for the EPL protocol will be presented in an upcoming paper.

Linear algebra

A.1 Tensor product

We define the tensor product $\otimes$ as the Kronecker product as follows

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{pmatrix}, \tag{A.1}$$

where a_{ij} represent the elements of A. The tensor product has a more general definition as well, but because in this thesis we are only interested in vectors and matrices, the Kronecker product suffices.

Furthermore we can construct a vector space $V \otimes W$ from two vector spaces V, W over a field K. Let V and W bases be $\mathcal{B}_V, \mathcal{B}_W$. The vector space $V \otimes W$ is spanned by

$$\mathcal{B}_{V\otimes W} = \{|v\rangle \otimes |w\rangle : |v\rangle \in \mathcal{B}_V, |w\rangle \in \mathcal{B}_W\}. \tag{A.2}$$

This entails that $\dim V \otimes W = \dim V \dim W$. Note that this is distinctly different from defining $V \otimes W$ as the tensor product of all elements.

$$V \otimes W \neq \{|v\rangle \otimes |w\rangle : |v\rangle \in V, |w\rangle \in W\}. \tag{A.3}$$

The following relations hold for elements $|v\rangle, |v_1\rangle, |v_2\rangle \in V, |w\rangle, |w_1\rangle, |w_2\rangle \in W, c \in K$:

- $|v_1\rangle \otimes |w\rangle + |v_2\rangle \otimes |w\rangle = (|v_1\rangle + |v_2\rangle) \otimes |w\rangle$
- $|v\rangle \otimes |w_1\rangle + |v\rangle \otimes |w_2\rangle = |v\rangle \otimes (|w_1\rangle + |w_2\rangle)$
- $c(|v\rangle \otimes |w\rangle) = (c|v\rangle) \otimes |w\rangle = |v\rangle \otimes (c|w\rangle)$.

We thus see that $\otimes$ is a bilinear map. Finally we may use the following abbreviation

$$\underbrace{a \otimes ... \otimes a}_{n} = a^{\otimes n}. \tag{A.4}$$

A.1.1 Tensor functions

For *linear* maps $f : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}, g : \mathbb{R}^{m\times m} \to \mathbb{R}^{m\times m}$ we also define the map $f \otimes g$.

$$(f \otimes g)(X) = (f \otimes g)\left(\sum_{ijkl} x_{ijkl} |i\rangle\langle j| \otimes |k\rangle\langle l|\right) \tag{A.5}$$

$$= \sum_{ijkl} x_{ijkl} f(|i\rangle\langle j|) \otimes g(|k\rangle\langle l|). \tag{A.6}$$

As an example, consider the identity map $\mathbb{I}$ with the transposition map T:

$$(\mathbb{I}\otimes T)(X) = (\mathbb{I}\otimes T)\left(\sum_{ijkl} x_{ijkl}\,|i\rangle\langle j|\otimes|k\rangle\langle l|\right) \tag{A.7}$$

$$= \sum_{ijkl} x_{ijkl}\,|i\rangle\langle j|\otimes|l\rangle\langle k| \tag{A.8}$$

This map is also known as the **partial transpose**.

A.2 Direct sum

The direct sum is defined as an operator on two matrices of arbitrary size as follows

$$A\oplus B = \left(\begin{array}{c|c} A & 0 \\ \hline 0 & B \end{array}\right). \tag{A.9}$$

Contrary to the tensor product, the direct sum is not a bilinear map as $c(A\oplus B) = (cA)\oplus(cB)$. We may construct a vector space $V\oplus W$:

$$V\oplus W = \{A\oplus B : A\in V, B\in W\}. \tag{A.10}$$

We see that $\dim V\oplus W = \dim V + \dim W$.

A.3 Partial trace

Consider a state ρ_{AB} shared by two parties, A and B. We can describe the state on A, ρ_A, using the **partial trace**,

$$\rho_A = \mathrm{tr}_B[\rho_{AB}]. \tag{A.11}$$

This operator, also called the reduced density operator, can be defined as the following linear map

$$\mathrm{tr}_B(|a_1\rangle\langle a_2|\otimes|b_1\rangle\langle b_2|) = |a_1\rangle\langle a_2|\,\mathrm{tr}(|b_1\rangle\langle b_2|). \tag{A.12}$$

Equivalently the partial trace can be defined as,

$$\mathrm{tr}_B(\rho_{AB}) = \sum_{i\in\mathcal{B}}(\mathbb{I}_A\otimes\langle i|_B)\rho_{AB}(\mathbb{I}_A\otimes|i\rangle_B) \tag{A.13}$$

for a basis $\mathcal{B}$ for system B. To see that these expressions are equal,

$$\mathrm{tr}_B(\rho_{AB}) = \mathrm{tr}_B\left(\sum_{ijkl} p_{ijkl}\,|i\rangle\langle j|\otimes|k\rangle\langle l|\right) \tag{A.14}$$

$$= \sum_{ijkl} p_{ijkl}\,\mathrm{tr}_B\left(|i\rangle\langle j|\otimes|k\rangle\langle l|\right) \tag{A.15}$$

$$= \sum_{ijkl} p_{ijkl}\,|i\rangle\langle j|\,\mathrm{tr}(|k\rangle\langle l|) \tag{A.16}$$

$$= \sum_{ijkk} p_{ijkk}\,|i\rangle\langle j| \tag{A.17}$$

$$\tag{A.18}$$

$$\sum_{i\in\mathcal{B}}(\mathbb{I}_A\otimes\langle i|_B)\rho_{AB}(\mathbb{I}_A\otimes|i\rangle_B)=\sum_m(\mathbb{I}\otimes\langle m|)\left(\sum_{ijkl}p_{ijkl}\,|i\rangle\langle j|\otimes|k\rangle\langle l|\right)(\mathbb{I}\otimes|m\rangle) \tag{A.19}$$

$$=\sum_m\sum_{ijkl}p_{ijkl}\,|i\rangle\langle j|\otimes\langle m|k\rangle|l\rangle\langle m| \tag{A.20}$$

$$=\sum_{ijm}p_{ijmm}\,|i\rangle\langle j| \tag{A.21}$$

As an example, consider two systems A, B described by the states ρ_A and σ_B. Their joint state state equals $\rho_{AB}=\rho_A\otimes\sigma_B$. From ρ_{AB} we can recover system A by tracing out system B $\mathrm{tr}_B[\rho_{AB}]=\rho_A\,\mathrm{tr}[\sigma_B]=\rho$ and equivalently we can trace out A to get system B $\mathrm{tr}_A[\rho_{AB}]=\sigma_B$.

However, if we consider the 2-dimensional maximally entangled state $|\Phi\rangle\langle\Phi|$ shared by two parties A and B. By tracing out one system we find

$$\mathrm{tr}_B(|\Phi\rangle\langle\Phi|)=\frac{\mathrm{tr}_B\left(|00\rangle\langle 00|+|11\rangle\langle 00|+|00\rangle\langle 11|+|11\rangle\langle 11|\right)}{2} \tag{A.22}$$

$$=\frac{|0\rangle\langle 0|\,\mathrm{tr}(|0\rangle\langle 0|)+|1\rangle\langle 0|\,\mathrm{tr}(|1\rangle\langle 0|)+|0\rangle\langle 1|\,\mathrm{tr}(|0\rangle\langle 1|)+|1\rangle\langle 1|\,\mathrm{tr}(|1\rangle\langle 1|)}{2} \tag{A.23}$$

$$=\frac{|0\rangle\langle 0|+|1\rangle\langle 1|}{2}=\frac{\mathbb{I}_2}{2}. \tag{A.24}$$

This is a maximally mixed state, which means there are no quantum correlations.

A.4 Matrix functions

For an arbitrary function $\tilde{f}$ on the real numbers, we define an analogous matrix function f on Hermitian matrices by diagonalising and applying $\tilde{f}$ on its eigenvalues:

$$f(M)=P^{-1}\begin{pmatrix}\tilde{f}(\lambda_1) & & \\ & \ddots & \\ & & \tilde{f}(\lambda_n)\end{pmatrix}P. \tag{A.25}$$

Bibliography

[1] Christine Bachoc, Dion C. Gijswijt, Alexander Schrijver, and Frank Vallentin. Invariant semidefinite programs, 2010.

[2] Sean D. Barrett and Pieter Kok. Efficient high-fidelity quantum computation using matter qubits and linear optics. *Phys. Rev. A*, 71:060310, Jun 2005.

[3] Stephen D. Bartlett, Terry Rudolph, and Robert W. Spekkens. Reference frames, superselection rules, and quantum information. *Rev. Mod. Phys.*, 79:555–609, Apr 2007.

[4] Charles H. Bennett, Gilles Brassard, Sandu Popescu, Benjamin Schumacher, John A. Smolin, and William K. Wootters. Purification of noisy entanglement and faithful teleportation via noisy channels. *Phys. Rev. Lett.*, 76:722–725, Jan 1996.

[5] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, 2004.

[6] Earl T. Campbell and Simon C. Benjamin. Measurement-based entanglement under conditions of extreme photon loss. *Phys. Rev. Lett.*, 101:130502, Sep 2008.

[7] Man-Duen Choi. Completely positive linear maps on complex matrices. *Linear Algebra and its Applications*, 10(3):285–290, 1975.

[8] Jon Dattorro. *Convex Optimization & Euclidean Distance Geometry*. Meboo Publishing USA, 2011.

[9] Jeroen Dehaene, Maarten van den Nest, Bart de Moor, and Frank Verstraete. Local permutations of products of bell states and entanglement distillation. *Phys. Rev. A*, 67:022310, Feb 2003.

[10] David Deutsch, Artur Ekert, Richard Jozsa, Chiara Macchiavello, Sandu Popescu, and Anna Sanpera. Quantum privacy amplification and the security of quantum cryptography over noisy channels. *Phys. Rev. Lett.*, 77:2818–2821, Sep 1996.

[11] Andrew C. Doherty, Pablo A. Parrilo, and Federico M. Spedalieri. Distinguishing separable and entangled states. *Phys. Rev. Lett.*, 88:187904, Apr 2002.

[12] Andrew C. Doherty, Pablo A. Parrilo, and Federico M. Spedalieri. Complete family of separability criteria. *Phys. Rev. A*, 69:022308, Feb 2004.

[13] Frédéric Dupuis, Omar Fawzi, and Stephanie Wehner. Entanglement sampling and applications. *IEEE Transactions on Information Theory*, 61:1093–1112, 2013.

[14] Wolgang Dür and Hans J Briegel. Entanglement purification and quantum error correction. *Reports on Progress in Physics*, 70(8):1381, 2007.

[15] Willam Fulton and Joseph Harris. *Representation Theory: A First Course*. Graduate Texts in Mathematics. Springer New York, 1991.

[16] Sevag Gharibian. Strong np-hardness of the quantum separability problem. *Quantum Information and Computation*, 10:343–360, 2010.

[17] David J. Griffiths. *Introduction to Quantum Mechanics*. Pearson international edition. Pearson Prentice Hall, 2005.

[18] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Separability of mixed states: necessary and sufficient conditions. *Physics Letters A*, 223(1):1 – 8, 1996.

[19] Ryszard Horodecki, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81:865–942, Jun 2009.

[20] Min Jiang, Shunlong Luo, and Shuangshuang Fu. Channel-state duality. *Phys. Rev. A*, 87:022310, Feb 2013.

[21] Miguel Navascués, Masaki Owari, and Martin B. Plenio. Power of symmetric extensions for entanglement detection. *Phys. Rev. A*, 80:052306, Nov 2009.

[22] Naomi H. Nickerson, Joseph F. Fitzsimons, and Simon C. Benjamin. Freely scalable quantum technologies using cells of 5-to-50 qubits with very lossy and noisy photonic links. *Phys. Rev. X*, 4:041041, Dec 2014.

[23] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge Series on Information and the Natural Sciences. Cambridge University Press, 2000.

[24] Brendan O'Donoghue, Eric Chu, Neal Parikh, and Stephen Boyd. Conic optimization via operator splitting and homogeneous self-dual embedding. *Journal of Optimization Theory and Applications*, 169(3):1042–1068, Jun 2016.

[25] Brendan O'Donoghue, Eric Chu, Neal Parikh, and Stephen Boyd. SCS: Splitting conic solver, version 1.2.6. `https://github.com/cvxgrp/scs`, April 2016.

[26] Eric M. Rains. A semidefinite program for distillable entanglement. *IEEE Transactions on Information Theory*, 47:2921–2933, Nov 2001.

[27] Madeleine Udell, Karanveer Mohan, David Zeng, Jenny Hong, Steven Diamond, and Stephen Boyd. Convex optimization in Julia. *SC14 Workshop on High Performance Technical Computing in Dynamic Languages*, 2014.

[28] John Watrous. Theory of quantum information. `https://cs.uwaterloo.ca/~watrous/TQI/TQI.pdf`, 2016.

[29] Zaiwen Wen. First-order methods for semidefinite programming. `http://www.bicmr.org/~wenzw/paper/SDPThesis.pdf`, 2009.

[30] Michael M. Wolf. Quantum channels and operations. `https://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MichaelWolf/QChannelLecture.pdf`, Jul 2012.