Ancilla dimension in quantum channel discrimination

Daniel Puzzuoli and John Watrous

Abstract. Quantum channel discrimination is a fundamental task in quantum information theory. It is well known that entanglement with an ancillary system can help in this task, and furthermore that an ancilla with the same dimension as the input of the channels is always sufficient for optimal discrimination of two channels. A natural question to ask is whether the same holds true for the output dimension. That is, in cases when the output dimension of the channels is (possibly much) smaller than the input dimension, is an ancilla with dimension equal to the output dimension always sufficient for optimal discrimination? We show that the answer to this question is "no" by construction of a family of counterexamples. This family contains instances with arbitrary finite gap between the input and output dimensions, and still has the property that in every case, for optimal discrimination, it is necessary to use an ancilla with dimension equal to that of the input.

The proof relies on a characterization of all operators on the trace norm unit sphere that maximize entanglement negativity. In the case of density operators we generalize this characterization to a broad class of entanglement measures, which we call weak entanglement measures. This characterization allows us to conclude that a quantum channel is reversible if and only if it preserves entanglement as measured by any weak entanglement measure, with the structure of maximally entangled states being equivalent to the structure of reversible maps via the Choi isomorphism. We also include alternate proofs of other known characterizations of channel reversibility.

1. Introduction

The task of quantum channel discrimination is to determine, given a single use, which of two known channels is acting on a system. In the abstract setting, the person performing the task can choose any state to feed through the channels, then perform any measurement on the output to guess which channel acted on the state. In general, it can be useful to probe the channels

using a state which is entangled to some ancillary system, called an *ancilla*, then perform a joint measurement on the output and ancilla systems together.

The optimal success probability of discriminating the channels, given as an optimization over all input states and measurements, is given as a simple expression involving the completely bounded trace norm (also called the diamond norm in quantum information theory). Due to properties of this norm it is possible to conclude that for optimal discrimination of two channels it is always sufficient to use an ancilla system having the same size as the input of the channels. It is also known, in cases when the input and output dimensions are the same, that using an ancilla having the same size as the input is sometimes *necessary* for optimal discrimination. One such example, which we will review, is given by the Werner-Holevo channels, introduced in [1] (and described in [2, Example 3.39], for instance).

By construction of a family of examples we show that, in cases when the output dimension is smaller than the input, an ancilla of size equal to the output is not sufficient in general for optimal channel discrimination. This family is parameterized by two natural numbers $n \geq 2$ and $k \geq 1$, with the input dimension being n^k and the output being nk, and hence the output can be made arbitrarily small compared to the input. Despite this arbitrary gap, we show that for optimal discrimination of these channels it remains necessary to use an ancilla as large as the input. This family is based on the Werner-Holevo channels (and is equivalent to these channels in the $k = 1$ case), and therefore can be viewed as extending them as a demonstration of the general necessity of using an ancilla that is as large as the input.

Due to the relationship between channel discrimination and the completely bounded trace norm, this family can also be viewed as a concrete and direct proof of the fact that for an arbitrary linear map taking matrices to matrices, the completely bounded trace norm does not generically achieve its value with an ancilla equal to the output dimension of the map. An equivalent dual statement in terms of the completely bounded norm was proved by Haagerup in [3] while studying decompositions of completely bounded maps.

Our proof is based on a characterization of operators on the trace norm unit sphere that maximize entanglement negativity [4].[1] When restricting attention to density operators, we generalize this characterization to a class of measures that we call *weak entanglement measures*, which satisfy a subset of properties that many entanglement measures have. We conclude by showing that, when quantified by a weak entanglement measure, a channel is reversible if and only if it preserves entanglement, and if and only if its Choi matrix is maximally entangled. Part of proving this is the observation that the structure of maximally entangled states is equivalent

[1] While the physical concept of "entanglement" only applies to density operators, the entanglement negativity as a function can just as well be applied to any bipartite operator.

to the structure of reversible channels shown in [5]. We also give short proofs of the known facts that a channel being reversible is equivalent to it preserving trace norm, preserving fidelity, and that all complementary channels are necessarily constant on the set of density operators.

2. Background and notation

In this section we set up notation and review some basic concepts in finite dimensional vector spaces and quantum theory. Readers familiar with these topics may wish to skip this section and refer back to it if some notation is unclear.

2.1. Finite dimensional complex vector spaces

In this paper we work in finite dimensional (f.d.) complex Hilbert spaces, which we will always take to be $\mathbb{C}^n$ with the standard inner product $\langle u, v\rangle = \sum_{i=1}^{n} \overline{u_i} v_i$ for $u, v \in \mathbb{C}^n$ (conjugate linear in the first argument). We use the symbols $\mathcal{A}, \mathcal{B}, \mathcal{X}, \mathcal{Y}$, and $\mathcal{Z}$ to denote f.d. complex Hilbert spaces when it is useful to have a label, or when it is not necessary to explicitly refer to the dimension. The unit sphere of $\mathcal{X}$ is denoted $\mathrm{S}(\mathcal{X}) = \{x \in \mathcal{X} : \|x\| = 1\}$. The set of linear operators mapping $\mathcal{X} \to \mathcal{Y}$ is denoted $\mathrm{L}(\mathcal{X}, \mathcal{Y})$, and we use the convention $\mathrm{L}(\mathcal{X}) = \mathrm{L}(\mathcal{X}, \mathcal{X})$. We denote the *standard basis* of elementary vectors for $\mathbb{C}^n$ as $e_1, \ldots, e_n$. For any operator $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$, the operator $A^* \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ denotes the adjoint map to A, the operator $A^{\mathsf{T}} \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ denotes the transpose map to A, and the operator $\overline{A} \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ denotes the entrywise conjugate of A. (Transposition and entrywise complex conjugation are taken with respect to the standard basis.) For $u \in \mathcal{X}$, we also use the notations $u^*, u^{\mathsf{T}} \in \mathrm{L}(\mathcal{X}, \mathbb{C})$ and $\overline{u} \in \mathcal{X}$ by identifying u with an element in $\mathrm{L}(\mathbb{C}, \mathcal{X})$ acting as $\alpha \mapsto \alpha u$. The symbol $\mathbb{1}$ is used to denote the identity map, with subscript specifying what space it acts on (e.g. $\mathbb{1}_{\mathcal{X}} \in \mathrm{L}(\mathcal{X})$ is the identity acting on $\mathcal{X}$).

The Hilbert-Schmidt inner product on $\mathrm{L}(\mathcal{X}, \mathcal{Y})$ is $\langle A, B\rangle = \mathrm{Tr}(A^* B)$ for $A, B \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$, where Tr is the trace. For standard basis elements $e_i \in \mathcal{X}$ and $e_j \in \mathcal{Y}$, $E_{ij} = e_i e_j^* \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ denotes the matrix units. We use special notation for various subsets of $\mathrm{L}(\mathcal{X})$:

- $\mathrm{Herm}(\mathcal{X}) = \{A \in \mathrm{L}(\mathcal{X}) : A^* = A\}$, the set of self-adjoint operators.
- $\mathrm{Pos}(\mathcal{X}) = \{P \in \mathrm{L}(\mathcal{X}) : P \geq 0\} \subset \mathrm{Herm}(\mathcal{X})$, the set of positive semidefinite operators.
- $\mathrm{U}(\mathcal{X}, \mathcal{Y}) = \{A \in \mathrm{L}(\mathcal{X}) : A^* A = \mathbb{1}_{\mathcal{X}}\}$ when $\dim(\mathcal{X}) \leq \dim(\mathcal{Y})$, the set of isometries.

It will sometimes be useful for us to think of vectors in $\mathcal{X} \otimes \mathcal{Y}$ as elements in $\mathrm{L}(\mathcal{Y}, \mathcal{X})$, and vice versa. To do so we use the *vectorization mapping* $\mathrm{vec} : \mathrm{L}(\mathcal{Y}, \mathcal{X}) \to \mathcal{X} \otimes \mathcal{Y}$ defined as $\mathrm{vec}(E_{ij}) = e_i \otimes e_j$, and extended by linearity to all of $\mathrm{L}(\mathcal{Y}, \mathcal{X})$. For general $u \in \mathcal{X}$ and $v \in \mathcal{Y}$, $\mathrm{vec}(uv^*) = u \otimes \overline{v}$. The function vec is an isometric isometry, i.e. it is a linear bijection and

satisfies $\langle \mathrm{vec}(A), \mathrm{vec}(B)\rangle = \langle A, B\rangle$ for all $A, B \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$. An identity we make use of is that

$$\mathrm{vec}(ABC) = \big(A \otimes C^{\mathsf{T}}\big)\mathrm{vec}(B), \tag{1}$$

which holds for any A, B, C for which the product ABC is well defined.

The set of linear maps taking $\mathrm{L}(\mathcal{X}) \to \mathrm{L}(\mathcal{Y})$ is denoted $\mathrm{T}(\mathcal{X}, \mathcal{Y})$, and $\mathrm{T}(\mathcal{X}) = \mathrm{T}(\mathcal{X}, \mathcal{X})$. The set of completely positive maps in $\mathrm{T}(\mathcal{X}, \mathcal{Y})$ is denoted $\mathrm{CP}(\mathcal{X}, \mathcal{Y})$. Throughout this paper we let $T \in \mathrm{T}(\mathcal{X})$ denote the transpose map, so that $T(X) = X^{\mathsf{T}}$. It holds that

$$\big(T \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)(\mathrm{vec}(\mathbb{1}_{\mathcal{X}})\mathrm{vec}(\mathbb{1}_{\mathcal{X}})^*) = W_{\mathcal{X}\mathcal{X}}, \tag{2}$$

where $W_{\mathcal{X}\mathcal{Y}} \in \mathrm{U}(\mathcal{X} \otimes \mathcal{Y}, \mathcal{Y} \otimes \mathcal{X})$ denotes the swap operator, which satisfies $W_{\mathcal{X}\mathcal{Y}}(x \otimes y) = y \otimes x$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. The linear map $J : \mathrm{T}(\mathcal{X}, \mathcal{Y}) \to \mathrm{L}(\mathcal{X} \otimes \mathcal{Y})$, defined as

$$J(\Phi) = (\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Phi)(\mathrm{vec}(\mathbb{1}_{\mathcal{X}})\mathrm{vec}(\mathbb{1}_{\mathcal{X}})^*) \tag{3}$$

for $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$, is a vector space isomorphism. The matrix $J(\Phi)$ is called the *Choi matrix* of Φ [6].

For $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ we use three standard matrix norms, the 1-norm (also called the *trace norm*), 2-norm (also called the *Frobenius norm*), and ∞-norm (also called the *spectral norm* or *operator norm*) defined as

$$\begin{aligned} \|A\|_1 &= \mathrm{Tr}\big(\sqrt{A^*A}\big), \\ \|A\|_2 &= \sqrt{\langle A, A\rangle}, \\ \|A\|_\infty &= \max\{\|Ax\| : x \in \mathrm{S}(\mathcal{X})\}. \end{aligned} \tag{4}$$

For $p \in \{1, \infty\}$ we denote the induced p-norms on $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$

$$\|\Phi\|_p = \max\big\{\|\Phi(X)\|_p : X \in \mathrm{L}(\mathcal{X}), \|X\|_p \le 1\big\} \tag{5}$$

and the completely bounded versions as

$$|||\Phi|||_p = \sup\big\{\big\|\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^m)}\big\|_p : m \in \mathbb{N}\big\}. \tag{6}$$

Iit holds that $|||\Phi|||_1 = \big\|\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big\|_1$ and $|||\Phi|||_\infty = \big\|\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big\|_\infty$ for all $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$.

2.2. Some quantum terminology

A vector $u \in \mathrm{S}(\mathbb{C}^n \otimes \mathbb{C}^m)$ is called *maximally entangled* if, for $r = \min(n, m)$, there exists orthonormal sets $\{x_i\}_{i=1}^r \subset \mathbb{C}^n$ and $\{y_i\}_{i=1}^r \subset \mathbb{C}^m$ for which

$$u = \frac{1}{\sqrt{r}} \sum_{i=1}^{r} x_i \otimes y_i. \tag{7}$$

When $m \le n$, this is equivalent to the statement that there exists an isometry $A \in \mathrm{U}(\mathbb{C}^m, \mathbb{C}^n)$ for which $u = \frac{1}{\sqrt{r}}\mathrm{vec}(A)$. We denote $\tau_{\mathcal{X}} \in \mathrm{D}(\mathcal{X} \otimes \mathcal{X})$ as the *canonical maximally entangled state*, defined as

$$\tau_{\mathcal{X}} = \frac{1}{n}\mathrm{vec}(\mathbb{1}_{\mathcal{X}})\mathrm{vec}(\mathbb{1}_{\mathcal{X}})^*, \tag{8}$$

where according to the vectorization convention $\mathrm{vec}(\mathbb{1}_{\mathcal{X}}) = \sum_{i=1}^{n} e_i \otimes e_i$.

For a quantum system with associated f.d. complex Hilbert space $\mathcal{X}$, the states of the system are elements of $\mathrm{D}(\mathcal{X}) = \{\rho \in \mathrm{Pos}(\mathcal{X}) : \mathrm{Tr}(\rho) = 1\}$, called either the set of *density operators*, *density matrices*, or *quantum states*. Quantum transformations, called *quantum channels*, from a system associated with $\mathcal{X}$ to one associated with $\mathcal{Y}$ are given by the completely positive and trace preserving maps from $\mathrm{L}(\mathcal{X})$ to $\mathrm{L}(\mathcal{Y})$, denoted $\mathrm{C}(\mathcal{X}, \mathcal{Y})$.

For a finite set Σ and some $\mathcal{X}$, a *measurement* with outcomes Σ on a quantum system associated with $\mathcal{X}$ is a function $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ such that $\sum_{a \in \Sigma} \mu(a) = \mathbb{1}_{\mathcal{X}}$. If such a measurement is performed on a quantum state $\rho \in \mathrm{D}(\mathcal{X})$, the probability of outcome $a \in \Sigma$ is given by the inner product $\langle \mu(a), \rho \rangle$. A *projective measurement* is a measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ for which $\mu(a)$ is an orthogonal projection for every $a \in \Sigma$.

3. Channel discrimination

The relevance of the trace and completely bounded trace norms in quantum theory arises in part from their interpretation in terms of quantum state and channel discrimination. These tasks can be formalized in terms of games, where how easy (or difficult) it is to discriminate two states or channels is given by the optimal probability with which this game can be won.

Quantum state discrimination games are single player games which proceed as follows. Descriptions of two quantum states $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})$ and a probability $\lambda \in [0,1]$ are known to the player. A bit $a \in \{0,1\}$ is sampled by the referee according to the distribution $p(0) = \lambda$, $p(1) = 1 - \lambda$. A single copy of the state ρ_a is given to the player, from which they must guess what a was by measuring the system (i.e. guess which of the two states they were given). For a given measurement $\mu : \{0,1\} \to \mathrm{Pos}(\mathcal{X})$, the probability of guessing correctly in a single run of the game is given by the expression

$$\lambda \langle \mu(0), \rho_0 \rangle + (1 - \lambda) \langle \mu(1), \rho_1 \rangle, \tag{9}$$

and hence the optimal success probability is given as the above expression optimized over all choices of two-outcome measurements. The following theorem [7, 8] provides a simple expression for the optimal success probability, which generalizes the expression for the classical version of the game.

Theorem 1 (Holevo-Helstrom theorem). *Let $\mathcal{X}$ be an f.d. complex Hilbert space, let $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})$ be density operators, and let $\lambda \in [0,1]$ be a real number. For every choice of measurement $\mu : \{0,1\} \to \mathrm{Pos}(\mathcal{X})$, it holds that*

$$\lambda \langle \mu(0), \rho_0 \rangle + (1 - \lambda) \langle \mu(1), \rho_1 \rangle \leq \frac{1}{2} + \frac{1}{2} \| \lambda \rho_0 - (1 - \lambda) \rho_1 \|_1. \tag{10}$$

Moreover there exists a projective measurement for which the inequality in this statement can be replaced by an equality.

Hence, the trace norm has an operational interpretation in terms of this discrimination game. A similar discrimination game can be defined

for quantum channels. As in the state case, descriptions of two quantum channels $\Phi_0, \Phi_1 \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ and a probability $\lambda \in [0,1]$ are known to the player. The referee samples a bit $a \in \{0,1\}$ according to the distribution $p(0) = \lambda$, $p(1) = 1 - \lambda$. The player is then given a single use of Φ_a, and must guess a. This game has an additional degree of freedom from the state case, as the player must choose a quantum state to feed into Φ_a. Once this state is chosen the problem reduces to the problem of discriminating $\Phi_0(\rho)$ and $\Phi_1(\rho)$. An additional layer of complexity is that the player may have access to an *ancillary* quantum system with f.d. complex Hilbert space $\mathcal{Z}$, and can choose a state $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Z})$, pass the system associated to $\mathcal{X}$ through Φ_a, then attempt to discriminate the states $\big(\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(\rho)$ and $\big(\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(\rho)$. Hence, by the above theorem, for a choice of $\mathcal{Z}$ and $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Z})$, the optimal success probability of guessing correctly is

$$\frac{1}{2} + \frac{1}{2}\big\|\lambda\big(\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(\rho) - (1-\lambda)\big(\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(\rho)\big\|_1, \tag{11}$$

and the optimal success probability for the game as a whole is given as an optimization of this expression over all choices of $\mathcal{Z}$ and $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Z})$. With this we arrive at the following theorem (see [2, Chapter 3]).

Theorem 2 (Holevo-Helstrom theorem for channels). *Let $\mathcal{X}$ and $\mathcal{Y}$ be finite dimensional complex Hilbert spaces, let $\Phi_0, \Phi_1 \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be channels, and let $\lambda \in [0,1]$ be a real number. For any choice of a positive integer m, a density operator $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathbb{C}^m)$, and a measurement $\mu : \{0,1\} \to \mathrm{Pos}(\mathcal{Y} \otimes \mathbb{C}^m)$, it holds that*

$$\begin{aligned}\lambda\big\langle \mu(0), \big(\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^m)}\big)(\rho)\big\rangle + (1-\lambda)\big\langle \mu(1), \big(\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^m)}\big)(\rho)\big\rangle \\ \leq \frac{1}{2} + \frac{1}{2}\|\!|\lambda\Phi_0 - (1-\lambda)\Phi_1|\!\|_1.\end{aligned} \tag{12}$$

Moreover, if $m \geq \dim(\mathcal{X})$, then there exists a density operator $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathbb{C}^m)$ and projective measurement $\mu : \{0,1\} \to \mathrm{Pos}(\mathcal{Y} \otimes \mathbb{C}^m)$ for which equality in this relation is achieved.

The question we ask in this paper is: does equality necessarily hold in Equation (12) for some state and measurement when $m = \dim(\mathcal{Y})$? In words, is it possible in all cases to optimally discriminate two quantum channels using an ancilla system that is the same size as the channel output? Given the current form of Theorem 2, this question only has relevance when $\dim(\mathcal{Y}) < \dim(\mathcal{X})$.

A more general version of this question is: is it true that

$$\|\!|\Psi|\!\|_1 = \big\|\Psi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big\|_1 \tag{13}$$

for all $\Psi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$? Due to the 1 and ∞ norms being dual to each other, this is equivalent to asking whether

$$\|\!|\Psi|\!\|_\infty = \big\|\Psi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big\|_\infty \tag{14}$$

for all $\Psi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$. It follows from work of Haagerup [3] that this general question has a negative answer. Despite this negative answer, in channel discrimination games we are specifically interested in Ψ of a special form, i.e. $\Psi = \lambda\Phi_0 - (1-\lambda)\Phi_1$ for some $\Phi_0, \Phi_1 \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ and $\lambda \in [0,1]$, and one might be inclined to question whether (13) could still hold for all linear maps of this form. Moreover, Haagerup's proof provides an answer to the general question through a somewhat indirect path, and we believe that it is helpful from the viewpoint of quantum information theory to obtain explicit examples of channels for which equality cannot hold in (12) when $m = \dim(\mathcal{Y})$.

In this paper we construct such examples, thereby answering both of the questions raised above negatively. In particular, we prove the following.

Theorem 3. *For every choice of positive integers $n \geq 2$ and $k \geq 1$ there exist channels*

$$\Gamma_{n,k}^{(0)}, \Gamma_{n,k}^{(1)} \in \mathrm{C}\big(\mathbb{C}^{n^k}, \mathbb{C}^{kn}\big) \tag{15}$$

such that for all real numbers $\lambda \in (0,1)$ it holds that

$$\Big\| \lambda\Gamma_{n,k}^{(0)} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})} - (1-\lambda)\Gamma_{n,k}^{(1)} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})} \Big\|_1 < \Big\| \lambda\Gamma_{n,k}^{(0)} - (1-\lambda)\Gamma_{n,k}^{(1)} \Big\|_1 = 1 \tag{16}$$

for every f.d. complex Hilbert space $\mathcal{Y}$ satisfying $\dim(\mathcal{Y}) < n^k$.

Note that in the setting of channel discrimination, by Theorem 2 the equality

$$\Big\| \lambda\Gamma_{n,k}^{(0)} - (1-\lambda)\Gamma_{n,k}^{(1)} \Big\|_1 = 1 \tag{17}$$

implies that the channels $\Gamma_{n,k}^{(0)}$ and $\Gamma_{n,k}^{(1)}$ can be perfectly discriminated for any $\lambda \in (0,1)$. Also, as the input dimension is n^k, and the output dimension is nk, this family of channels contains instances with arbitrary finite gap between the input and output dimensions.

In the remainder of this section we describe the construction of a family of channels for which the requirements of the above theorem are satisfied. The proof that these channels indeed satisfy these requirements appears in the two sections that follow.

For every integer $n \geq 2$, the *Werner-Holevo channels* are defined as

$$\Phi_n^{(0)} = \frac{1}{n+1}(\Omega + T), \quad \Phi_n^{(1)} = \frac{1}{n-1}(\Omega - T), \tag{18}$$

where $\Omega \in \mathrm{CP}(\mathcal{X})$ is defined as $\Omega(X) = \mathrm{Tr}(X)\mathbb{1}_{\mathcal{X}}$ on all $X \in \mathrm{L}(\mathcal{X})$. Throughout this paper, for any finite sequence of f.d. complex Hilbert spaces $\mathcal{X}_1, \ldots, \mathcal{X}_k$, we will denote the reduction to the i^{th} subsystem as $R_i \in \mathrm{C}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k, \mathcal{X}_i)$. That is, for all $X_1 \in \mathrm{L}(\mathcal{X}_1), \ldots, X_k \in \mathrm{L}(\mathcal{X}_k)$, the channel R_i acts as

$$R_i(X_1 \otimes \cdots \otimes X_k) = \Big(\prod_{j \neq i} \mathrm{Tr}(X_j)\Big) X_i. \tag{19}$$

Now, for integers $n \geq 2$ and $k \geq 1$, assume that $\mathcal{X}_1, \ldots, \mathcal{X}_k$ and $\mathcal{X}$ denote copies of the space $\mathbb{C}^n$. We define the channels

$$\Gamma_{n,k}^{(\alpha)} \in \mathrm{C}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k, \mathbb{C}^k \otimes \mathcal{X}) \tag{20}$$

for all $X \in \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k)$ as

$$\Gamma_{n,k}^{(\alpha)}(X) = \frac{1}{k} \sum_{i=1}^{k} E_{ii} \otimes \Phi_n^{(\alpha)}\big(R_i(X)\big), \tag{21}$$

for each $\alpha \in \{0,1\}$, where each R_i is regarded as a channel of the form $R_i \in \mathrm{C}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k, \mathcal{X})$. Operationally, these channels represent randomly trashing all but one of the input subsystems while keeping a classical record of which is kept, then applying one of the Werner-Holevo channels. It holds that $\Gamma_{n,1}^{(\alpha)} \cong \Phi_n^{(\alpha)}$ under the association $\mathbb{C} \otimes \mathcal{X} \cong \mathcal{X}$, and hence the Werner-Holevo channels themselves are contained in this family.

Similarly, define mappings

$$\Psi_{n,k} \in \mathrm{T}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k, \mathbb{C}^k \otimes \mathcal{X}) \tag{22}$$

for all $X \in \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k)$ as

$$\Psi_{n,k}(X) = \sum_{i=1}^{k} E_{ii} \otimes T\big(R_i(X)\big). \tag{23}$$

For $\lambda_n = \frac{n+1}{2n}$ the following relations hold

$$\frac{1}{n} T = \lambda_n \Phi_n^{(0)} - (1 - \lambda_n)\Phi_n^{(1)}, \tag{24}$$

$$\frac{1}{nk} \Psi_{n,k} = \lambda_n \Gamma_{n,k}^{(0)} - (1 - \lambda_n)\Gamma_{n,k}^{(1)}. \tag{25}$$

The crux of proving Theorem 3 will be to prove that

$$\big\| \Psi_{n,k} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})} \big\|_1 < \|\!| \Psi_{n,k} |\!\|_1 = nk \tag{26}$$

whenever $\dim(\mathcal{Y}) < n^k$, which is equivalent to the desired norm relation of the theorem for the particular probability λ_n. The specific value λ_n is used to make many expressions easier to work with, and the extension of the result from a particular probability to arbitrary $\lambda \in (0,1)$ will be made by a simple argument.

4. Induced 1-norm of partial transpose

For proving the relations in Equation (26) it will be useful to first examine expressions of the form

$$\big\| \big(T \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(X) \big\|_1 \tag{27}$$

for $X \in \mathrm{L}(\mathcal{X} \otimes \mathcal{Y})$ with $\|X\|_1 = 1$. When $X \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ this quantity (up to multiplicative and additive scalars) has been called the *negativity* of the state X [4], and is an easy to compute, though non-faithful entanglement

measure (where "non-faithful" means that there exist entangled states that minimize this quantity). We will abuse terminology by referring to Equation (27) as the negativity of X, even when X is not a state.

We will begin by reviewing some facts about negativity. When X is a rank-1 operator, the expression (27) takes a simple form, as proved in [4, Proposition 8].

Proposition 4 (Vidal and Werner). *Let $\mathcal{X}$ and $\mathcal{Y}$ be f.d. complex Hilbert spaces. For $A, B \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ it holds that*

$$\left\| \left(T \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\right)(\mathrm{vec}(A)\mathrm{vec}(B)^*) \right\|_1 = \|A\|_1 \|B\|_1. \tag{28}$$

Note that [4, Proposition 8] is proven for the case $A = B$, but the above can be reasoned similarly. From this the following known facts can be deduced.

Proposition 5. *Let $\mathcal{X} = \mathbb{C}^n$, $\mathcal{Y} = \mathbb{C}^m$. For $u, v \in \mathrm{S}(\mathcal{X} \otimes \mathcal{Y})$, it holds that*

$$\left\| \left(T \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\right)(uv^*) \right\|_1 \leq \min(n, m), \tag{29}$$

with equality if and only if both u and v are maximally entangled. In particular this implies

$$\left\| T \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})} \right\|_1 = \min(n, m). \tag{30}$$

Proof. For $u, v \in \mathrm{S}(\mathcal{X} \otimes \mathcal{Y})$, let $A, B \in \mathrm{L}(\mathcal{Y} \otimes \mathcal{X})$ satisfy $u = \mathrm{vec}(A)$ and $v = \mathrm{vec}(B)$. By Proposition 4,

$$\begin{aligned} \left\| \left(T \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\right)(\mathrm{vec}(A)\mathrm{vec}(B)^*) \right\|_1 &= \|A\|_1 \|B\|_1 && (31) \\ &\leq \min(n, m) \|A\|_2 \|B\|_2 && (32) \\ &= \min(n, m), && (33) \end{aligned}$$

where the inequality follows from the inequality $\|A\|_1 \leq \sqrt{\min(n, m)} \|A\|_2$, with equality if and only if either A or A^* is a scalar multiple of an isometry. Hence, we have the inequality in Equation (29), with equality holding if and and only if u and v are maximally entangled.

Equation (30) follows as the induced 1-norm can be written as an optimization restricted to operators of the form uv^* for $u, v \in \mathrm{S}(\mathcal{X} \otimes \mathcal{Y})$. □

1 We remark that the equality condition for Equation (29), when $u = v$, is the well known fact that the only pure states which maximize negativity are maximally entangled. We also remark that Equation (30) was proved in [9, Theorem 1.2], where it was proved that $\left\| T \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})} \right\|_\infty = n$, and because partial transposition is self-adjoint, $\left\| T \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})} \right\|_\infty = \left\| T \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})} \right\|_1$.

Proposition 5 implies, for $n \geq 2$, $m \geq 1$, and $\lambda_n = \frac{n+1}{2n}$, that

$$\begin{aligned} \max \left\{ \left\| \lambda_n \Phi_n^{(0)}(\rho) - (1 - \lambda_n) \Phi_n^{(1)}(\rho) \right\|_1 : \rho \in \mathrm{D}(\mathbb{C}^n \otimes \mathbb{C}^m) \right\} \\ = \frac{1}{n} \left\| T \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^m)} \right\|_1 = \frac{1}{n} \min(n, m). \end{aligned} \tag{34}$$

Hence, for an ancilla of dimension m, the optimal success probability of a channel discrimination game for the Werner-Holevo channels with probability λ_n is

$$\frac{1}{2} + \frac{1}{2n} \min(n, m). \tag{35}$$

In particular, this implies that this channel discrimination game can be won with certainty, and furthermore, it can be won with certainty if and only if $m \geq n$.

To prove Theorem 3 it will be useful to generalize Proposition 5 to a full characterization of when $\left\| (T \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})(X) \right\|_1 = n$ for (a not-necessarily rank-1) $X \in \mathrm{L}(\mathcal{X} \otimes \mathcal{Y})$ with $\|X\|_1 = 1$. First we prove a proposition about equality conditions in the triangle inequality for the trace norm for sets of orthogonal operators, which requires two facts. The first is that for $A \in \mathrm{L}(\mathcal{X})$, it holds that

$$\|A\|_1 = \max\{|\langle U, A\rangle| : U \in \mathrm{U}(\mathcal{X})\}, \tag{36}$$

and the second is that $\mathrm{Tr}(A) = \|A\|_1$ if and only if $A \geq 0$.

Proposition 6. *Let $\{A_i\}_{i=1}^r \subset \mathrm{L}(\mathcal{X}, \mathcal{Y})$ be an orthogonal set. If*

$$\left\| \sum_{i=1}^r A_i \right\|_1 = \sum_{i=1}^r \|A_i\|_1, \tag{37}$$

then it holds that $A_i A_j^* = 0$ and $A_i^* A_j = 0$ for all $i \neq j$.

Proof. Assume first that $\mathcal{Z}$ is an arbitrary f.d. complex Hilbert space, and $B, C \in \mathrm{L}(\mathcal{Z})$ are orthogonal operators for which the equality $\|B + C\|_1 = \|B\|_1 + \|C\|_1$ holds. Let $U \in \mathrm{U}(\mathcal{Z})$ be a unitary operator satisfying

$$\langle U, B + C\rangle = \|B + C\|_1. \tag{38}$$

It follows that $\langle U, B\rangle = \|B\|_1$ and $\langle U, C\rangle = \|C\|_1$, and therefore $U^*B = B^*U$ and $U^*C = C^*U$ are both positive semidefinite operators. We have

$$\langle B^*U, U^*C\rangle = \langle U^*B, C^*U\rangle = \langle B, C\rangle = 0, \tag{39}$$

and therefore $(B^*U)(U^*C) = 0$ and $(U^*B)(C^*U) = 0$, as orthogonal positive semidefinite operators have product equal to zero. It follows that $B^*C = 0$ and $BC^* = 0$.

Now choose $i, j \in \{1, \ldots, r\}$ with $i \neq j$. The equality (37) implies that $\|A_i + A_j\|_1 = \|A_i\|_1 + \|A_j\|_1$. Defining $B, C \in \mathrm{L}(\mathcal{X} \oplus \mathcal{Y})$ as

$$B = \begin{pmatrix} 0 & 0 \\ A_i & 0 \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} 0 & 0 \\ A_j & 0 \end{pmatrix}, \tag{40}$$

we find that B and C are orthogonal operators satisfying $\|B + C\|_1 = \|B\|_1 + \|C\|_1$, and therefore $B^*C = 0$ and $BC^* = 0$ from the argument above. This implies that $A_i A_j^* = 0$ and $A_i^* A_j = 0$ as required. □

We remark that the converse of the above proposition holds as well. With this in hand we can generalize Proposition 5.

Theorem 7. *Let $\mathcal{X} = \mathbb{C}^n$ and $\mathcal{Y} = \mathbb{C}^m$. For $X \in \mathrm{L}(\mathcal{X} \otimes \mathcal{Y})$ with $\|X\|_1 \leq 1$, the following are equivalent.*

1. $\big\|\big(T \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(X)\big\|_1 = n$.
2. *$m \geq n$, and there exists a choice of $r \in \{1, \ldots, \lfloor m/n \rfloor\}$, $\sigma \in \mathrm{D}(\mathbb{C}^r)$, and $U, V \in \mathrm{U}(\mathcal{X} \otimes \mathbb{C}^r, \mathcal{Y})$ for which*

$$X = (\mathbb{1}_{\mathcal{X}} \otimes U)(\tau_{\mathcal{X}} \otimes \sigma)(\mathbb{1}_{\mathcal{X}} \otimes V^*), \tag{41}$$

   where $\tau_{\mathcal{X}} \in \mathrm{D}(\mathcal{X} \otimes \mathcal{X})$ is the canonical maximally entangled state.

When $X \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ the above equivalence holds with $V = U$.

Proof. The fact that statement 2 implies statement 1 follows by a direct computation together with Proposition 5.

Now suppose that statement 1 holds, and observe that Proposition 5 immediately implies $m \geq n$. Let

$$X = \sum_{i=1}^{r} s_i x_i y_i^* \tag{42}$$

be a singular value decomposition of X, where $r = \mathrm{rank}(X)$. By Proposition 5 all of the x_i and y_i must be maximally entangled, as the triangle inequality would otherwise allow one to conclude that

$$\big\|\big(T \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(X)\big\|_1 < n. \tag{43}$$

Hence, for each i there exist isometries $A_i, B_i \in \mathrm{U}(\mathcal{X}, \mathcal{Y})$ for which

$$x_i = \frac{1}{\sqrt{n}}\mathrm{vec}(A_i^{\mathsf{T}}) \quad \text{and} \quad y_i = \frac{1}{\sqrt{n}}\mathrm{vec}(B_i^{\mathsf{T}}). \tag{44}$$

Now, note that

$$\big(T \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(X) = \frac{1}{n} W_{\mathcal{X}\mathcal{Y}} \sum_{i=1}^{r} s_i A_i \otimes B_i^*, \tag{45}$$

so that

$$\begin{aligned} n = \big\|\big(T \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(X)\big\|_1 &= \frac{1}{n}\Big\| \sum_{i=1}^{r} s_i A_i \otimes B_i^* \Big\|_1 \\ &\leq \frac{1}{n} \sum_{i=1}^{r} s_i \big\| A_i \otimes B_i^* \big\|_1 = n, \end{aligned} \tag{46}$$

where the the last equality follows from the A_i and B_i being isometries, and therefore

$$\|A_i \otimes B_i^*\|_1 = n^2 \tag{47}$$

for every i. Hence, we have equality in the triangle inequality for these operators (which are orthogonal as they arise from a singular value decomposition), and so Proposition 6 implies

$$(A_i \otimes B_i^*)^*(A_j \otimes B_j^*) = A_i^* A_j \otimes B_i B_j^* = 0, \tag{48}$$

and

$$(A_i \otimes B_i^*)(A_j \otimes B_j^*)^* = A_i A_j^* \otimes B_i^* B_j = 0 \tag{49}$$

for all $i \neq j$. As these are isometries, $B_i B_j^* \neq 0$, so the first expression above gives $A_i^* A_j = 0$, and likewise the second implies $B_i^* B_j = 0$ for all $i \neq j$. Hence the A_i (and respectively the B_i) embed $\mathcal{X}$ into r mutually orthogonal n-dimensional subspaces of $\mathcal{Y}$, giving $rn \leq m$.

Lastly, to get the particular form of X, define $U, V \in \mathrm{U}(\mathcal{X} \otimes \mathbb{C}^r, \mathcal{Y})$ as

$$U = \sum_{i=1}^{r} A_i \otimes e_i^* \quad \text{and} \quad V = \sum_{i=1}^{r} B_i \otimes e_i^*, \tag{50}$$

where the fact that U and V are isometries follows from $A_i^* A_j = 0 = B_i^* B_j$ for $i \neq j$. Defining

$$\sigma = \sum_{i=1}^{r} s_i E_{ii} \in \mathrm{D}(\mathbb{C}^r), \tag{51}$$

we see that

$$\begin{aligned} X &= \frac{1}{n} \sum_{i=1}^{r} s_i \mathrm{vec}(A_i^{\mathsf{T}}) \mathrm{vec}(B_i^{\mathsf{T}})^* && (52) \\ &= (\mathbb{1}_{\mathcal{X}} \otimes U) \Big(\sum_{i=1}^{r} \frac{s_i}{n} \mathrm{vec}(\mathbb{1}_{\mathcal{X}}) \mathrm{vec}(\mathbb{1}_{\mathcal{X}})^* \otimes E_{ii} \Big) (\mathbb{1}_{\mathcal{X}} \otimes V^*) && (53) \\ &= (\mathbb{1}_{\mathcal{X}} \otimes U)(\tau_{\mathcal{X}} \otimes \sigma)(\mathbb{1}_{\mathcal{X}} \otimes V^*), && (54) \end{aligned}$$

as required.

When $X \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$, in the above $B_i = A_i$, and hence $V = U$. □

5. Proof of counterexamples

In this section we will prove Theorem 3 by generalizing Theorem 7. We show that, for any $\mathcal{X}_1, \ldots, \mathcal{X}_k, \mathcal{Y}$, and $X \in \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k \otimes \mathcal{Y})$ with $\|X\|_1 = 1$,

$$\big\| \big(T_{\mathcal{X}_i} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)\big(\big(R_i \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(X)\big) \big\|_1 = \dim(\mathcal{X}_i), \tag{55}$$

for all $1 \leq i \leq k$ if and only if

$$\big\| \big(T_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(X) \big\|_1 = \prod_{i=1}^{k} \dim(\mathcal{X}_i) = \dim(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k), \tag{56}$$

where we are using subscripts on the transpose map to be explicit about which space it is acting on. In other words, all of the $\mathcal{X}_i$ subsystems are maximally entangled with $\mathcal{Y}$ (as measured by negativity) if and only if $\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k$ is maximally entangled with $\mathcal{Y}$. This equivalence is given in Theorem 10, which is essentially induction applied to Theorem 7. Figure 1 gives a visual presentation of the structure of the operators. By applying this equivalence, we conclude that, for $\mathcal{X}_1, \ldots, \mathcal{X}_k, \mathcal{X}$ denoting copies of $\mathbb{C}^n$, and $X \in \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k \otimes \mathcal{Y})$ with $\|X\|_1 = 1$,

$$\big\| \big(\Psi_{n,k} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(X) \big\|_1 = nk \tag{57}$$

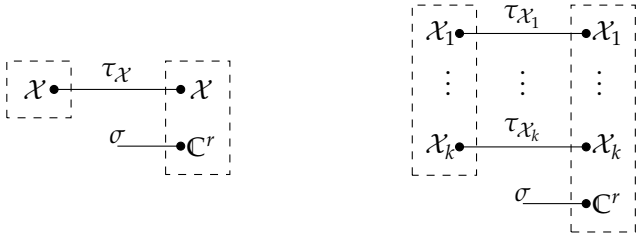
if and only if

$$\left\| \left(T_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})} \right)(X) \right\|_1 = n^k \tag{58}$$

for all i, and hence Equation (57) is only possible if

$$\dim(\mathcal{Y}) \geq n^k = \dim(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k). \tag{59}$$

From this Theorem 3 will follow.



(A) Theorem 7 (B) Theorem 10

FIGURE 1. This is a diagrammatic representation of the structures given in Theorem 7 and Theorem 10. In Theorem 7 the ancilla system factorizes into $\mathcal{X} \otimes \mathbb{C}^r$, and the operator X looks like something maximally entangled across the $\mathcal{X}$ systems with σ left over. In Theorem 10, this factorization-and-maximally-entangled structure is repeated k-times, again, potentially with some σ left over.

Before beginning we introduce an implicit permutation notation. At points in the section we will be working with operators that act on a tensor product space, where the ordering of the tensor factors for which it is convenient to specify the operator is not the same as the ordering used in the context that the operator appears. This primarily occurs for operators of product form. For example, given $A \in \mathrm{L}(\mathcal{X} \otimes \mathcal{Z})$, and $B \in \mathrm{L}(\mathcal{Y})$, the operator $A \otimes B \in \mathrm{L}(\mathcal{X} \otimes \mathcal{Z} \otimes \mathcal{Y})$ has a simple form, but if our spaces are naturally ordered as $\mathcal{X} \otimes \mathcal{Y} \otimes \mathcal{Z}$, then we must write

$$(\mathbb{1}_{\mathcal{X}} \otimes W_{\mathcal{Z},\mathcal{Y}})(A \otimes B)(\mathbb{1}_{\mathcal{X}} \otimes W^*_{\mathcal{Z},\mathcal{Y}}) \tag{60}$$

to specify it as an operator in $\mathrm{L}(\mathcal{X} \otimes \mathcal{Y} \otimes \mathcal{Z})$, which can become clunky.

To avoid this, we introduce the following notation. For some finite list of f.d. Hilbert spaces $\mathcal{Z}_1, \ldots, \mathcal{Z}_k$, a permutation $\sigma : \{1, \ldots, k\} \to \{1, \ldots, k\}$, and an operator $X \in \mathrm{L}(\mathcal{Z}_1 \otimes \cdots \otimes \mathcal{Z}_k)$, we write

$$\underbrace{X}_{\in \mathrm{L}(\mathcal{Z}_{\sigma(1)} \otimes \cdots \otimes \mathcal{Z}_{\sigma(k)})} = PXP^*, \tag{61}$$

where $P \in \mathrm{U}(\mathcal{Z}_1 \otimes \cdots \otimes \mathcal{Z}_k, \mathcal{Z}_{\sigma(1)} \otimes \cdots \otimes \mathcal{Z}_{\sigma(k)})$ is the isometry which permutes the subsystems as given in the definition. For the example in the

preceding paragraph, this notation gives

$$\underbrace{A \otimes B}_{\in \mathrm{L}(\mathcal{X}\otimes\mathcal{Y}\otimes\mathcal{Z})} = (\mathbb{1}_{\mathcal{X}} \otimes W_{\mathcal{Z},\mathcal{Y}})(A \otimes B)(\mathbb{1}_{\mathcal{X}} \otimes W^*_{\mathcal{Z},\mathcal{Y}}). \tag{62}$$

Note as well that for f.d. complex Hilbert spaces $\mathcal{A}$ and $\mathcal{B}$, it holds that

$$\tau_{\mathcal{A}\otimes\mathcal{B}} = \underbrace{\tau_{\mathcal{A}} \otimes \tau_{\mathcal{B}}}_{\in \mathrm{L}(\mathcal{A}\otimes\mathcal{B}\otimes\mathcal{A}\otimes\mathcal{B})}. \tag{63}$$

In the above there is a potential ambiguity as multiple copies of the same space appear, so it is not necessarily well defined. In this case however, the operator is invariant under swapping the order of these copies, and so there is no real ambiguity.

Lemma 8. *Let $X \in \mathrm{L}(\mathcal{X} \otimes \mathcal{Y})$ with $\|X\|_1 = 1$. If $\mathrm{Tr}_{\mathcal{Y}}(X) = uv^*$ for some $u, v \in \mathrm{S}(\mathcal{X})$, then there exists $\sigma \in \mathrm{D}(\mathcal{Y})$ for which $X = uv^* \otimes \sigma$.*

Proof. First consider the case in which X is positive semidefinite, and therefore a density operator by the condition $\|X\|_1 = 1$. The partial trace is a positive map, from which it follows that $v = u$. Define a projection operator $\Pi = \mathbb{1}_{\mathcal{X}} - uu^*$, and observe that $\langle \Pi \otimes \mathbb{1}_{\mathcal{Y}}, X\rangle = \langle \Pi, \mathrm{Tr}_{\mathcal{Y}}(X)\rangle = 0$. As X and $\Pi \otimes \mathbb{1}_{\mathcal{Y}}$ are both positive semidefinite, it follows that $(\Pi \otimes \mathbb{1}_{\mathcal{Y}})X = X(\Pi \otimes \mathbb{1}_{\mathcal{Y}}) = 0$, and therefore

$$\begin{aligned} X &= (uu^* \otimes \mathbb{1}_{\mathcal{Y}} + \Pi \otimes \mathbb{1}_{\mathcal{Y}})X(uu^* \otimes \mathbb{1}_{\mathcal{Y}} + \Pi \otimes \mathbb{1}_{\mathcal{Y}}) && (64)\\ &= (uu^* \otimes \mathbb{1}_{\mathcal{Y}})X(uu^* \otimes \mathbb{1}_{\mathcal{Y}}) && (65)\\ &= uu^* \otimes \sigma, && (66)\end{aligned}$$

where $\sigma = (u^* \otimes \mathbb{1}_{\mathcal{Y}})X(u \otimes \mathbb{1}_{\mathcal{Y}}) \in \mathrm{D}(\mathcal{Y})$.

For the general case, let $U \in \mathrm{U}(\mathcal{X})$ be a unitary operator satisfying $Uu = v$. It follows that

$$\|(U \otimes \mathbb{1}_{\mathcal{Y}})X\|_1 = 1 = \mathrm{Tr}((U \otimes \mathbb{1}_{\mathcal{Y}})X), \tag{67}$$

and therefore $(U \otimes \mathbb{1}_{\mathcal{Y}})X$ is positive semidefinite. Substituting X with the operator $(U \otimes \mathbb{1}_{\mathcal{Y}})X$ in the case considered above yields $(U \otimes \mathbb{1}_{\mathcal{Y}})X = vv^* \otimes \sigma$, and therefore $X = uv^* \otimes \sigma$, for some choice of $\sigma \in \mathrm{D}(\mathcal{Y})$, which completes the proof. □

Lemma 9. *Let $X \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$, and let $\Pi_1 \in \mathrm{L}(\mathcal{Y})$ and $\Pi_2 \in \mathrm{L}(\mathcal{X})$ be orthogonal projections. If*

$$\|\Pi_1 X \Pi_2\|_1 = \|X\|_1, \tag{68}$$

then it holds that $\Pi_1 X \Pi_2 = X$.

Proof. Let $X = \sum_{i=1}^r s_i u_i v_i^*$ be a singular value decomposition of X. Then, we have that

$$\begin{aligned} \sum_{i=1}^r s_i = \|X\|_1 = \|\Pi_1 X \Pi_2\|_1 &= \Big\| \sum_{i=1}^r s_i \Pi_1 u_i v_i^* \Pi_2 \Big\|_1 \\ &\leq \sum_{i=1}^r s_i \|\Pi_1 u_i v_i^* \Pi_2\|_1 \leq \sum_{i=1}^r s_i. \end{aligned} \tag{69}$$

Hence, all inequalities are equalities, which implies $1 = \|\Pi_1 u_i v_i^* \Pi_2\|_1 = \|\Pi_1 u_i\| \|\Pi_2 v_i\|$ for all $1 \leq i \leq r$, implying that $\Pi_1 u_i = u_i$ and $\Pi_2 v_i = v_i$ for all i, and hence $\Pi_1 X \Pi_2 = X$. □

Theorem 10. *Let $\mathcal{X}_1 = \mathbb{C}^{n_1}, \ldots, \mathcal{X}_k = \mathbb{C}^{n_k}$, $\mathcal{Y} = \mathbb{C}^m$, $N = \prod_{i=1}^k n_i = \dim(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k)$, and let $X \in \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k \otimes \mathcal{Y})$ with $\|X\|_1 = 1$. The following are equivalent:*

1. $\left\| \left(T_{\mathcal{X}_i} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\right)\left((R_i \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})(X)\right) \right\|_1 = n_i$, *for all* $1 \leq i \leq k$.
2. $\left\| \left(T_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\right)(X) \right\|_1 = N$.
3. $m \geq N$, *and there is some* $r \in \{1, \ldots, \lfloor m/N \rfloor\}$, $\sigma \in \mathrm{D}(\mathbb{C}^r)$, *and* $U, V \in \mathrm{U}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k \otimes \mathbb{C}^r, \mathcal{Y})$ *for which*

$$X = (\mathbb{1}_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k} \otimes U)(\tau_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k} \otimes \sigma)(\mathbb{1}_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k} \otimes V^*), \tag{70}$$

where $\tau_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k} \in \mathrm{D}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k \otimes \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k)$ is the canonical maximally entangled state.

If $X \in \mathrm{D}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k \otimes \mathcal{Y})$ the above equivalence holds with $V = U$.

Proof. The equivalence of statements 2 and 3 is the content of Theorem 7, and from this we also retrieve the statement that if X is a density operator, then we can take $V = U$ in statement 3. That statement 3 implies statement 1 follows by a direct computation, along with the observation in Equation (63). When $k = 1$, statements 1 and 2 are the same, so in this case there is nothing to prove. When $k = 2$ we will show that statement 1 implies statement 3 (in which case we will have the full equivalence for $k = 2$), then use induction to directly show that statement 1 is equivalent to statement 2 for $k > 2$.

For statement 1 implies statement 3 in the $k = 2$ case, to simplify notation we denote $\mathcal{A} = \mathcal{X}_1$, $\mathcal{B} = \mathcal{X}_2$, $a = n_1$, and $b = n_2$, and hence $N = ab$. By Theorem 7 it follows from $\left\| \left(T_{\mathcal{A}} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\right)(\mathrm{Tr}_{\mathcal{B}}(X)) \right\|_1 = a$ that $a \leq m$, and there exists $s \in \{1, \ldots, \lfloor m/a \rfloor\}$, $\nu \in \mathrm{D}(\mathbb{C}^s)$, and isometries $A, B \in \mathrm{U}(\mathcal{A} \otimes \mathbb{C}^s, \mathcal{Y})$ for which

$$\mathrm{Tr}_{\mathcal{B}}(X) = (\mathbb{1}_{\mathcal{A}} \otimes A)(\tau_{\mathcal{A}} \otimes \nu)(\mathbb{1}_{\mathcal{A}} \otimes B^*). \tag{71}$$

This implies that

$$\mathrm{Tr}_{\mathcal{B} \otimes \mathbb{C}^s}((\mathbb{1}_{\mathcal{A} \otimes \mathcal{B}} \otimes A^*) X (\mathbb{1}_{\mathcal{A} \otimes \mathcal{B}} \otimes B)) = \tau_{\mathcal{A}}. \tag{72}$$

Note that

$$1 = \|\tau_{\mathcal{A}}\|_1 = \|\mathrm{Tr}_{\mathcal{B} \otimes \mathbb{C}^s}((\mathbb{1}_{\mathcal{A} \otimes \mathcal{B}} \otimes A^*) X (\mathbb{1}_{\mathcal{A} \otimes \mathcal{B}} \otimes B))\|_1 \tag{73}$$

$$\leq \|(\mathbb{1}_{\mathcal{A} \otimes \mathcal{B}} \otimes A^*) X (\mathbb{1}_{\mathcal{A} \otimes \mathcal{B}} \otimes B)\|_1 \leq \|X\|_1 = 1, \tag{74}$$

giving $\|(\mathbb{1}_{\mathcal{A} \otimes \mathcal{B}} \otimes A^*) X (\mathbb{1}_{\mathcal{A} \otimes \mathcal{B}} \otimes B)\|_1 = 1$, and so Lemma 8 implies that there exists $\eta \in \mathrm{D}(\mathcal{B} \otimes \mathbb{C}^s)$ for which

$$(\mathbb{1}_{\mathcal{A} \otimes \mathcal{B}} \otimes A^*) X (\mathbb{1}_{\mathcal{A} \otimes \mathcal{B}} \otimes B) = \underbrace{\tau_{\mathcal{A}} \otimes \eta}_{\in \mathrm{L}(\mathcal{A} \otimes \mathcal{B} \otimes \mathcal{A} \otimes \mathbb{C}^s)}, \tag{75}$$

and hence

$$(\mathbb{1}_{\mathcal{A}\otimes\mathcal{B}} \otimes AA^*)X(\mathbb{1}_{\mathcal{A}\otimes\mathcal{B}} \otimes BB^*) = (\mathbb{1}_{\mathcal{A}\otimes\mathcal{B}} \otimes A) \underbrace{(\tau_{\mathcal{A}} \otimes \eta)}_{\in \mathrm{L}(\mathcal{A}\otimes\mathcal{B}\otimes\mathcal{A}\otimes\mathbb{C}^s)} (\mathbb{1}_{\mathcal{A}\otimes\mathcal{B}} \otimes B^*). \quad (76)$$

As the above operator has trace norm 1, and $\mathbb{1}_{\mathcal{A}\otimes\mathcal{B}} \otimes AA^*$ and $\mathbb{1}_{\mathcal{A}\otimes\mathcal{B}} \otimes BB^*$ are both orthogonal projections, Lemma 9 implies

$$X = (\mathbb{1}_{\mathcal{A}\otimes\mathcal{B}} \otimes A) \underbrace{(\tau_{\mathcal{A}} \otimes \eta)}_{\in \mathrm{L}(\mathcal{A}\otimes\mathcal{B}\otimes\mathcal{A}\otimes\mathbb{C}^s)} (\mathbb{1}_{\mathcal{A}\otimes\mathcal{B}} \otimes B^*). \quad (77)$$

Next, it holds that

$$\big\|\big(T_{\mathcal{B}} \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^s)}\big)(\eta)\big\|_1 = \big\|\big(T_{\mathcal{B}} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(\mathrm{Tr}_{\mathcal{A}}(X))\big\|_1 = b, \quad (78)$$

and so again by Theorem 7, $b \leq s$, and there exists $r \in \{1, \ldots, \lfloor s/b \rfloor\}$, $\sigma \in \mathrm{D}(\mathbb{C}^r)$, and an isometry $S \in \mathrm{U}(\mathcal{B} \otimes \mathbb{C}^r, \mathbb{C}^s)$ for which

$$\eta = (\mathbb{1}_{\mathcal{B}} \otimes S)(\tau_{\mathcal{B}} \otimes \sigma)(\mathbb{1}_{\mathcal{B}} \otimes S^*). \quad (79)$$

Hence, letting $U = A(\mathbb{1}_{\mathcal{A}} \otimes S)$ and $V = B(\mathbb{1}_{\mathcal{A}} \otimes S)$ we get that

$$X = (\mathbb{1}_{\mathcal{A}\otimes\mathcal{B}} \otimes A) \underbrace{\big[\tau_{\mathcal{A}} \otimes (\mathbb{1}_{\mathcal{B}} \otimes S)(\tau_{\mathcal{B}} \otimes \sigma)(\mathbb{1}_{\mathcal{B}} \otimes S^*)\big]}_{\in \mathrm{L}(\mathcal{A}\otimes\mathcal{B}\otimes\mathcal{A}\otimes\mathbb{C}^s)} (\mathbb{1}_{\mathcal{A}\otimes\mathcal{B}} \otimes B^*) \quad (80)$$

$$= (\mathbb{1}_{\mathcal{A}\otimes\mathcal{B}} \otimes U) \underbrace{(\tau_{\mathcal{A}} \otimes \tau_{\mathcal{B}} \otimes \sigma)}_{\in \mathrm{L}(\mathcal{A}\otimes\mathcal{B}\otimes\mathcal{A}\otimes\mathcal{B}\otimes\mathbb{C}^r)} (\mathbb{1}_{\mathcal{A}\otimes\mathcal{B}} \otimes V^*) \quad (81)$$

$$= (\mathbb{1}_{\mathcal{A}\otimes\mathcal{B}} \otimes U)(\tau_{\mathcal{A}\otimes\mathcal{B}} \otimes \sigma)(\mathbb{1}_{\mathcal{A}\otimes\mathcal{B}} \otimes V^*), \quad (82)$$

and $ab \leq as \leq m$, and $r \leq s/b \leq m/ab$, as required.

Lastly, we show that statement 1 is equivalent to statement 2 for all k by induction. So, assuming the equivalence holds for some $k \geq 2$, we show it holds for $k+1$. Note that

$$\big\|\big(T_{\mathcal{X}_i} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)\big((R_i \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})(X)\big)\big\|_1 = n_i \quad (83)$$

for all $1 \leq i \leq k$, by the induction hypothesis, is equivalent to

$$\big\|\big(T_{\mathcal{X}_1\otimes\cdots\otimes\mathcal{X}_k} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(\mathrm{Tr}_{\mathcal{X}_{k+1}}(X))\big\|_1 = \prod_{i=1}^{k} n_i, \quad (84)$$

which, together with $\big\|\big(T_{\mathcal{X}_{k+1}} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)\big((R_{k+1} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})(X)\big)\big\|_1 = n_{k+1}$, again by the induction hypothesis, is equivalent to

$$\big\|\big(T_{\mathcal{X}_1\otimes\cdots\otimes\mathcal{X}_{k+1}} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(X)\big\|_1 = \prod_{i=1}^{k+1} n_i, \quad (85)$$

as required. □

The content of Figure 1 follows by the above theorem along with the observation

$$\tau_{\mathcal{X}_1\otimes\cdots\otimes\mathcal{X}_k} = \underbrace{\tau_{\mathcal{X}_1} \otimes \cdots \otimes \tau_{\mathcal{X}_k}}_{\in \mathrm{L}(\mathcal{X}_1\otimes\cdots\otimes\mathcal{X}_k\otimes\mathcal{X}_1\otimes\cdots\otimes\mathcal{X}_k)}. \quad (86)$$

For the case $n_1 = \cdots = n_k = n$, by noting that $\left\|\left(\Psi_{n,k} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\right)(X)\right\|_1 = nk$ if and only if

$$\left\|\left(T \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\right)\left(\left(R_i \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\right)(X)\right)\right\|_1 = n \tag{87}$$

for all $1 \leq i \leq k$, we arrive at the following.

Corollary 11. *Let $\mathcal{X}, \mathcal{X}_1, \ldots, \mathcal{X}_k$ denote copies of $\mathbb{C}^n$, and let $\mathcal{Y} = \mathbb{C}^m$. For $X \in \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k \otimes \mathcal{Y})$ with $\|X\|_1 = 1$, the following are equivalent.*

1. $\left\|\left(\Psi_{n,k} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\right)(X)\right\|_1 = nk$.
2. $\left\|\left(T_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\right)(X)\right\|_1 = n^k$.
3. *$m \geq n^k$, and there is some $r \in \{1, \ldots, \lfloor m/n^k \rfloor\}$, $\sigma \in \mathrm{D}(\mathbb{C}^r)$, and $U, V \in \mathrm{U}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k \otimes \mathbb{C}^r, \mathcal{Y})$ for which*

$$X = (\mathbb{1}_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k} \otimes U)(\tau_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k} \otimes \sigma)(\mathbb{1}_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k} \otimes V^*), \tag{88}$$

where $\tau_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k} \in \mathrm{D}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k \otimes \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k)$ is the canonical maximally entangled state.

When $X \in \mathrm{D}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k \otimes \mathcal{Y})$ the above equivalence holds with $V = U$.

In the setting of channel discrimination, the above corollary gives that a state $\rho \in \mathrm{D}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k \otimes \mathcal{Y})$ can be used to perfectly discriminate $\Gamma_{n,k}^{(0)}$ and $\Gamma_{n,k}^{(1)}$ with probability λ_n if and only if it can be used to perfectly discriminate $\Phi_{n^k}^{(0)}$ and $\Phi_{n^k}^{(1)}$ with probability λ_n. We are now in a position to prove Theorem 3.

Proof of Theorem 3. Fix $n \geq 2$ and $k \geq 1$, and let $\mathcal{X}_1, \ldots, \mathcal{X}_k$, and $\mathcal{X}$ denote copies of $\mathbb{C}^n$. For our examples we identify $\mathbb{C}^{n^k} \cong \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k$ and $\mathbb{C}^{kn} \cong \mathbb{C}^k \otimes \mathcal{X}$.

Let $\Gamma_{n,k}^{(0)}, \Gamma_{n,k}^{(1)}, \Psi_{n,k} \in \mathrm{T}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k, \mathbb{C}^k \otimes \mathcal{X})$ be as defined in Section 3. First we show that

$$\begin{aligned} &\left\|\lambda_n \Gamma_{n,k}^{(0)} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})} - (1-\lambda_n)\Gamma_{n,k}^{(1)} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\right\|_1 \\ &\qquad < \left|\left|\left|\lambda_n \Gamma_{n,k}^{(0)} - (1-\lambda_n)\Gamma_{n,k}^{(1)}\right|\right|\right|_1 = 1, \end{aligned} \tag{89}$$

whenever $\dim(\mathcal{Y}) < n^k$, where $\lambda_n = \frac{n+1}{2n}$. The above is equivalent to showing that

$$\left\|\Psi_{n,k} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\right\|_1 < |||\Psi_{n,k}|||_1 = nk \tag{90}$$

whenever $\dim(\mathcal{Y}) < n^k$.

By Corollary 11, for $\tau_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k} \in \mathrm{D}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k \otimes \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k)$ it holds that

$$\left\|\left(\Psi_{n,k} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k)}\right)(\tau_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k})\right\|_1 = nk, \tag{91}$$

and hence $|||\Psi_{n,k}|||_1 = nk$. Furthermore, for any f.d. complex Hilbert space $\mathcal{Y}$ with $\dim(\mathcal{Y}) < n^k$ and $X \in \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k \otimes \mathcal{Y})$ with $\|X\|_1 = 1$, the above

corollary implies that

$$\big\| \big(\Psi_{n,k} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(X) \big\|_1 < nk, \tag{92}$$

giving that

$$\big\| \Psi_{n,k} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})} \big\|_1 < nk. \tag{93}$$

This completes the proof of Equation (90).

Lastly, we need to show Equation (89) holds for any $\lambda \in (0,1)$, not just the particular choice λ_n. To do this we require the following fact: for $A, B \in \mathrm{L}(\mathcal{Z})$ with $\|A\|_1 \leq 1$ and $\|B\|_1 \leq 1$, if $\|\alpha A - (1-\alpha)B\|_1 = 1$ for a particular $\alpha \in (0,1)$, then it holds that $\|\lambda A - (1-\lambda)B\|_1 = 1$ for all $\lambda \in (0,1)$. To see this, note that the assumption is equivalent to the existence of a unitary $U \in \mathrm{L}(\mathcal{Z})$ for which

$$\langle U, \alpha A - (1-\alpha)B \rangle = \alpha \langle U, A \rangle + (1-\alpha)\langle U, -B \rangle = 1. \tag{94}$$

As $|\langle U, A \rangle| \leq \|A\|_1 \leq 1$ and $|\langle U, -B \rangle| \leq \|B\|_1 \leq 1$, the above equality implies that $\langle U, A \rangle = \langle U, -B \rangle = 1$. Thus, for any $\lambda \in (0,1)$, we have

$$1 = \langle U, \lambda A - (1-\lambda)B \rangle \leq \|\lambda A - (1-\lambda)B\|_1 \leq 1. \tag{95}$$

Thus, as there exists $X \in \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k \otimes \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k)$ with trace norm 1 for which

$$\Big\| \lambda_n \big(\Gamma^{(0)}_{n,k} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k)}\big)(X) - (1-\lambda_n)\big(\Gamma^{(1)}_{n,k} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k)}\big)(X) \Big\|_1 = 1 \tag{96}$$

it follows by the above paragraph that the above equation must hold for all $\lambda \in (0,1)$, and therefore

$$\Big\| \lambda \Gamma^{(0)}_{n,k} - (1-\lambda)\Gamma^{(1)}_{n,k} \Big\|_1 = 1 \tag{97}$$

for all $\lambda \in (0,1)$. By a similar argument, for $\mathcal{Y}$ with $\dim(\mathcal{Y}) < n^k$, if

$$\Big\| \lambda \Gamma^{(0)}_{n,k} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})} - (1-\lambda)\Gamma^{(1)}_{n,k} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})} \Big\|_1 = 1 \tag{98}$$

for some $\lambda \in (0,1)$, then the above equation would also hold for λ_n, which we have already shown is not the case. □

6. Weak entanglement measures and reversible quantum channels

Theorem 7 provides an alternative characterization of the set of operators $X \in \mathrm{L}(\mathcal{X} \otimes \mathcal{Y})$ whose trace norm equals 1 and whose negativity is maximized. In this section we prove a generalization of this result, albeit for the restricted case in which X must be a density operator, in which the negativity can be replaced by any member of a class of entanglement measures that we call *weak entanglement measures*. Many well-known measures of entanglement fall into this class.

Once the structure of density operators that maximize weak entanglement measures is established, we will apply it to the question of when a quantum channel is *reversible*, meaning that it has a left-inverse that is also a channel. We prove that a channel is reversible if and only if it preserves entanglement as measured by any weak entanglement measure, and equivalently, if and only if its Choi matrix is maximally entangled as measured by any weak entanglement measure.

6.1. Structure of states that maximize weak entanglement measures

We will begin by defining a class of entanglement measures that we call *weak entanglement measures*.

Definition 12. A *weak entanglement measure* is a family of functions

$$\{\mathrm{E}_{n,m} \,:\, n, m \in \mathbb{N},\, 1 \leq n \leq m\}, \tag{99}$$

each of which takes the form

$$\mathrm{E}_{n,m} : \mathrm{D}(\mathbb{C}^n \otimes \mathbb{C}^m) \to \mathbb{R}, \tag{100}$$

for which the following properties hold:

1. There exists a function $g : \mathbb{N} \to \mathbb{R}$ for which
$$\max_{\rho \in \mathrm{D}(\mathbb{C}^n \otimes \mathbb{C}^m)} \mathrm{E}_{n,m}(\rho) = g(n). \tag{101}$$
That is, we assume that the maximum exists and that it is a function only of the minimum of the two dimensions. We call g the *maximum function* for the family $\{\mathrm{E}_{n,m}\}$.
2. For any unit vector $u \in \mathrm{S}(\mathbb{C}^n \otimes \mathbb{C}^m)$, it holds that $\mathrm{E}_{n,m}(uu^*) = g(n)$ if and only if u is maximally entangled (in the sense given in Equation (7)).
3. The measure is monotonically decreasing under quantum channels acting on the second subsystem. That is, for all density operators $\rho \in \mathrm{D}(\mathbb{C}^n \otimes \mathbb{C}^m)$ and channels $\Phi \in \mathrm{C}(\mathbb{C}^m, \mathbb{C}^k)$ for $k \geq n$, it holds that
$$\mathrm{E}_{n,k}((\mathbb{1}_{\mathrm{L}(\mathbb{C}^n)} \otimes \Phi)(\rho)) \leq \mathrm{E}_{n,m}(\rho). \tag{102}$$
4. Each function $\mathrm{E}_{n,m}$ is *pure state convex*: for any set $\{u_1, \ldots, u_N\} \subset \mathrm{S}(\mathbb{C}^n \otimes \mathbb{C}^m)$ and probability vector $(p_1, \ldots, p_N)$, it holds that
$$\mathrm{E}_{n,m}\left(\sum_{i=1}^{N} p_i u_i u_i^*\right) \leq \sum_{i=1}^{N} p_i \, \mathrm{E}_{n,m}\left(u_i u_i^*\right). \tag{103}$$

A few comments on this definition are in order. First, pure state convexity may seem an odd axiom (as opposed to general convexity), but there may exist entanglement measures that are pure state convex and not generally convex. (For example, distillable entanglement is known to be pure-state convex [10, Lemma 25], but may not be generally convex [11].) Second, it is generally desired that entanglement measures satisfy stronger versions

of the third condition (e.g., monotonicity with respect to any LOCC channel between both subsystems). Furthermore entanglement measures usually treat the two subsystems symmetrically, and Property 3 is asymmetric in that it only applies to the second subsystem. In our proof the subsystems are treated asymmetrically, and we only need monotonicity to hold with respect to the second system (and hence this result can be applied to functions like the coherent information).

The set of weak entanglement measures includes the negativity [4], the coherent information [12] (where Property 3, called the *data processing inequality*, is strong sub-additivity), the squashed entanglement [13, 14], entanglement of formation, and distillable entanglement. See [15, Table 1] for a list of commonly used entanglement measures and the properties that they are known to satisfy.

In order to prove the theorem that follows we will make use of the following simple lemma.

Lemma 13. *Let $\mathcal{X}$ and $\mathcal{Y}$ be f.d. complex Hilbert spaces with $\dim(\mathcal{X}) \leq \dim(\mathcal{Y})$, and let $U, V \in \mathrm{U}(\mathcal{X}, \mathcal{Y})$ be orthogonal isometries for which $\alpha U + \beta V$ is proportional to an isometry for all choices of $\alpha, \beta \in \mathbb{C}$. It holds that $U^*V = 0$ (i.e., U and V map $\mathcal{X}$ into orthogonal subspaces of $\mathcal{Y}$).*

Proof. It suffices to consider the pairs $(\alpha, \beta) = (1, 1)$ and $(\alpha, \beta) = (1, i)$. As $U + V$ and $U + iV$ are proportional to isometries, the following operators must be proportional to the identity operator:

$$(U+V)^*(U+V) = 2\mathbb{1} + (U^*V + V^*U), \tag{104}$$

$$(U+iV)^*(U+iV) = 2\mathbb{1} + i(U^*V - V^*U). \tag{105}$$

As U^*V and V^*U are traceless, we conclude that

$$U^*V + V^*U = 0 \quad \text{and} \quad U^*V - V^*U = 0, \tag{106}$$

which implies $U^*V = 0$ as required. □

Theorem 14. *Let $\mathcal{X} = \mathbb{C}^n$ and $\mathcal{Y} = \mathbb{C}^m$ for positive integers n and m satisfying $n \leq m$, and let $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$. The following statements are equivalent:*

1. *For every weak entanglement measure $\{\mathrm{E}_{s,t}\}$ with maximum function g it holds that $\mathrm{E}_{n,m}(\rho) = g(n)$.*
2. *Statement 1 holds for any weak entanglement measure.*
3. *There exists a positive integer $r \leq m/n$, a density operator $\sigma \in \mathrm{D}(\mathbb{C}^r)$, and an isometry $U \in \mathrm{U}(\mathcal{X} \otimes \mathbb{C}^r, \mathcal{Y})$ for which*

$$\rho = (\mathbb{1}_{\mathcal{X}} \otimes U)(\tau_{\mathcal{X}} \otimes \sigma)(\mathbb{1}_{\mathcal{X}} \otimes U^*). \tag{107}$$

Proof. Statement 1 trivially implies statement 2 (as the set of weak entanglement measures is nonempty).

Now assume statement 2 holds: $\mathrm{E}_{n,m}(\rho) = g(n)$ for some weak entanglement measure $\{\mathrm{E}_{s,t}\}$ with maximum function g. By the pure-state

convexity axiom (Property 4), for any pure-state decomposition

$$\rho = \sum_{i=1}^{N} p_i v_i v_i^* \tag{108}$$

(for $p_1, \ldots, p_N$ positive) it holds that

$$g(n) = \mathrm{E}_{n,m}(\rho) \leq \sum_{i=1}^{N} p_i \, \mathrm{E}_{n,m}(v_i v_i^*) \tag{109}$$

and $\mathrm{E}_{n,m}(v_i v_i^*) \leq g(n)$, implying that $\mathrm{E}_{n,m}(v_i v_i^*) = g(n)$, for all $i = 1, \ldots, N$. Hence, by Property 2, every pure state decomposition of ρ necessarily consists only of maximally entangled states. This is equivalent to the statement that every unit vector $v \in \mathrm{Im}(\rho)$ contained in the image of ρ is maximally entangled.

Now consider a spectral decomposition

$$\rho = \sum_{i=1}^{r} p_i v_i v_i^* \tag{110}$$

of ρ, where $r = \mathrm{rank}(\rho)$ and we have restricted the sum to range only over indices corresponding to positive eigenvalues of ρ. By the argument above, one has that each v_i is maximally entangled, so there exists an orthogonal collection of isometries $\{V_1, \ldots, V_r\} \subset \mathrm{U}(\mathcal{X}, \mathcal{Y})$ for which

$$v_i = \frac{1}{\sqrt{n}} \mathrm{vec}(V_i^{\mathsf{T}}) \tag{111}$$

for each $i \in \{1, \ldots, r\}$. For each pair $i \neq j$ we find that

$$\mathrm{vec}\big(\alpha V_i^{\mathsf{T}} + \beta V_j^{\mathsf{T}}\big) \in \mathrm{Im}(\rho), \tag{112}$$

and therefore $\alpha V_i + \beta V_j$ is proportional to an isometry for all $\alpha, \beta \in \mathbb{C}$. By Lemma 13 it holds that $V_i^* V_j = 0$, and hence $rn \leq m$.

Along the same lines as in Theorem 7, define $U \in \mathrm{U}(\mathcal{X} \otimes \mathbb{C}^r, \mathcal{Y})$ and $\sigma \in \mathrm{D}(\mathbb{C}^r)$ as

$$U = \sum_{i=1}^{r} V_i \otimes e_i^* \quad \text{and} \quad \sigma = \sum_{i=1}^{r} p_i E_{ii}, \tag{113}$$

where the fact that U is an isometry follows from $V_i^* V_j = 0$ for $i \neq j$. It follows by direct multiplication that

$$\rho = (\mathbb{1}_{\mathcal{X}} \otimes U)(\tau_{\mathcal{X}} \otimes \sigma)(\mathbb{1}_{\mathcal{X}} \otimes U)^*, \tag{114}$$

and therefore statement 2 implies statement 3.

Finally, assume that statement 3 holds, let $\{\mathrm{E}_{s,t}\}$ be any weak entanglement measure with maximum function g, and define a channel $\Phi \in \mathrm{C}(\mathcal{Y}, \mathcal{X})$ as follows:

$$\Phi(X) = \mathrm{Tr}_{\mathbb{C}^r}(U^* Y U) + \langle \mathbb{1}_{\mathcal{Y}} - UU^*, Y \rangle \eta, \tag{115}$$

for all $Y \in \mathrm{L}(\mathcal{Y})$ and any fixed choice of a density operator $\eta \in \mathrm{D}(\mathcal{X})$. It holds that $(\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Phi)(\rho) = \tau_{\mathcal{X}}$, so by Property 3 one has

$$g(n) = \mathrm{E}_{n,n}(\tau_{\mathcal{X}}) = \mathrm{E}_{n,n}((\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Phi)(\rho)) \leq \mathrm{E}_{n,m}(\rho) \leq g(n). \tag{116}$$

It follows that $\mathrm{E}_{n,m}(\rho) = g(n)$, and so statement 3 implies statement 1. □

Using the above characterization we can arrive at a density operator version of Theorem 10 that holds for any weak entanglement measure.

Corollary 15. *Let $\mathcal{X}_1 = \mathbb{C}^{n_1}, \ldots, \mathcal{X}_k = \mathbb{C}^{n_k}$ and $\mathcal{Y} = \mathbb{C}^m$ for positive integers $n_1, \ldots, n_k$ and m satisfying $n = \prod_{i=1}^k n_i \leq m$, let $\rho \in \mathrm{D}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k \otimes \mathcal{Y})$ be a density operator, and let $\{E_{s,t}\}$ be any weak entanglement measure with maximum function g. The following statements are equivalent:*

1. *It holds that*
$$\mathrm{E}_{n_i,m}((R_i \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})(\rho)) = g(n_i) \tag{117}$$
for all $i = 1, \ldots, k$.
2. *It holds that*
$$\mathrm{E}_{n,m}(\rho) = g(n). \tag{118}$$
3. *There exists a positive integer $r \leq n/m$, a density operator $\sigma \in \mathrm{D}(\mathbb{C}^r)$, and an isometry*
$$U \in \mathrm{U}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k \otimes \mathbb{C}^r, \mathcal{Y}) \tag{119}$$
for which
$$\rho = (\mathbb{1}_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k} \otimes U)(\tau_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k} \otimes \sigma)(\mathbb{1}_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k} \otimes U^*). \tag{120}$$

Proof. The equivalence of the above statements was shown for the negativity in Theorem 10, and Theorem 14 gives that statements 1 and 2 hold for the negativity if and only if they hold for all weak entanglement measures. □

6.2. Reversible channels

A quantum channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ is called *reversible* if there exists a channel $\Psi \in \mathrm{C}(\mathcal{Y}, \mathcal{X})$ for which $\Psi\Phi = \mathbb{1}_{\mathrm{L}(\mathcal{X})}$ (i.e., Φ has a left inverse that is also a channel). We apply Theorem 14 to show that a channel is reversible if and only if it preserves entanglement as measured by any weak entanglement measure. The structure given in Theorem 14 also allows us to re-derive a result from [5], where it was shown that a channel is reversible if and only if it has a certain form. We also add in a couple of other conditions.

Before stating the theorem, let us recall a couple of simple concepts from the theory of quantum information. First, for positive semidefinite operators $P, Q \in \mathrm{Pos}(\mathcal{X})$, the *fidelity* is defined as

$$\mathrm{F}(P, Q) = \left\| \sqrt{P}\sqrt{Q} \right\|_1. \tag{121}$$

Second, for any pair of channels $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ and $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Z})$, it is said that Φ and Ψ are *complementary* if there exists an isometry $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ such that

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}(AXA^*) \quad \text{and} \quad \Psi(X) = \mathrm{Tr}_{\mathcal{Y}}(AXA^*). \tag{122}$$

We will also make use of a couple of simple facts, stated as lemmas as follows. (See, for instance, Corollary 3.24 and Proposition 2.29 in [2].)

Lemma 16. *For any $u, v \in \mathcal{X} \otimes \mathcal{Y}$ it holds that $\mathrm{F}(\mathrm{Tr}_{\mathcal{Y}}(uu^*), \mathrm{Tr}_{\mathcal{Y}}(vv^*)) = \|\mathrm{Tr}_{\mathcal{X}}(uv^*)\|_1$.*

Lemma 17. *For $u \in \mathcal{X} \otimes \mathcal{Y}$ and $P \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Z})$, if $\mathrm{Tr}_{\mathcal{Y}}(uu^*) = \mathrm{Tr}_{\mathcal{Z}}(P)$, then there exists $\Psi \in \mathrm{C}(\mathcal{Y}, \mathcal{Z})$ for which $(\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Psi)(uu^*) = P$.*

Theorem 18. *Let $\mathcal{X} = \mathbb{C}^n$ and $\mathcal{Y} = \mathbb{C}^m$ for positive integers $n \leq m$, let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, and let $\{\mathrm{E}_{s,t}\}$ be any weak entanglement measure with maximum function g. The following statements are equivalent:*

1. *Φ is reversible.*
2. *Φ preserves entanglement with respect to $\{\mathrm{E}_{s,t}\}$, meaning that for all positive integers $k \leq n$ and all density operators $\rho \in \mathrm{D}(\mathbb{C}^k \otimes \mathcal{X})$ it holds that*
$$\mathrm{E}_{k,m}\big((\mathbb{1}_{\mathrm{L}(\mathbb{C}^k)} \otimes \Phi)(\rho)\big) = \mathrm{E}_{k,n}(\rho). \tag{123}$$
3. *It holds that*
$$\mathrm{E}_{n,m}\big(\tfrac{1}{n} J(\Phi)\big) = g(n). \tag{124}$$
4. *There exists a positive integer $r \leq m/n$, a density operator $\sigma \in \mathrm{D}(\mathbb{C}^r)$, and an isometry $U \in \mathrm{U}(\mathcal{X} \otimes \mathbb{C}^r, \mathcal{Y})$ for which*
$$\Phi(X) = U(X \otimes \sigma)U^* \tag{125}$$
for all $X \in \mathrm{L}(\mathcal{X})$.
5. *It holds that*
$$\|\Phi(X)\|_1 = \|X\|_1 \tag{126}$$
for all $X \in \mathrm{L}(\mathcal{X})$.
6. *It holds that*
$$\mathrm{F}(\Phi(\rho), \Phi(\sigma)) = \mathrm{F}(\rho, \sigma) \tag{127}$$
for all $\rho, \sigma \in \mathrm{D}(\mathcal{X})$.
7. *If $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Z})$ is complementary to Φ, then there exists a density operator $\sigma \in \mathrm{D}(\mathcal{Z})$ for which*
$$\Psi(X) = \mathrm{Tr}(X)\sigma \tag{128}$$
for all $X \in \mathrm{L}(\mathcal{X})$ (i.e., all channels which are complementary to Φ are constant on $\mathrm{D}(\mathcal{X})$).

Remark 19. We note that the equivalence of statements 1 and 4 is the content of [5, Theorem 2.1]. In the proof given therein, this equivalence follows from an argument similar to a key step of the proof of Theorem 14 (as well as Theorem 7). A similar argument has also been used to derive conditions under which an error map is correctable [16]. The equivalence of

statements 4 and 6 was proven in [17] for $\mathcal{Y} = \mathcal{X}$, but also for infinite dimensions. Lastly, for the case of the coherent information, the equivalence of statements 1 and 3 is a special case of the result in [12, Section VI], in which it was shown that a channel is reversible on half of a bipartite pure state if and only if the data processing inequality is satisfied with equality.

Proof of Theorem 18. Assume that statement 1 holds, and let $\Psi \in \mathrm{C}(\mathcal{Y}, \mathcal{X})$ be a left-inverse of Φ. By the monotonicity of weak entanglement measures it holds that

$$\begin{aligned} \mathrm{E}_{k,n}(\rho) &= \mathrm{E}_{k,n}\big((\mathbb{1}_{\mathrm{L}(\mathbb{C}^k)} \otimes \Psi\Phi)(\rho)\big) \\ &\leq \mathrm{E}_{k,m}\big((\mathbb{1}_{\mathrm{L}(\mathbb{C}^k)} \otimes \Phi)(\rho)\big) \leq \mathrm{E}_{k,n}(\rho) \end{aligned} \tag{129}$$

for all choices of $k \leq n$ and $\rho \in \mathrm{D}(\mathbb{C}^k \otimes \mathcal{X})$. Hence, statement 1 implies statement 2.

Statement 2 immediately implies statement 3, as statement 3 is equivalent to the particular choice of $k = n$ and $\rho = \tau_{\mathcal{X}}$ in statement 2.

Next, under the assumption that statement 3 holds, one has that the Choi operator of Φ is given by

$$J(\Phi) = (\mathbb{1}_{\mathcal{X}} \otimes U)(\mathrm{vec}(\mathbb{1}_{\mathcal{X}})\mathrm{vec}(\mathbb{1}_{\mathcal{X}})^* \otimes \sigma)(\mathbb{1}_{\mathcal{X}} \otimes U^*), \tag{130}$$

by Theorem 14. This is equivalent to

$$\Phi(X) = U(X \otimes \sigma)U^* \tag{131}$$

for all $X \in \mathrm{L}(\mathcal{X})$. It has therefore been proved that statement 3 implies statement 4.

By well-known properties of the trace norm and the fidelity function, one immediately finds that statement 4 implies both statements 5 and 6.

Now assume that statement 5 holds, and let $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Z})$ be any complementary channel to Φ. For any two unit vectors $u, v \in \mathrm{S}(\mathcal{X})$, Lemma 16 implies that

$$\mathrm{F}(\Psi(uu^*), \Psi(vv^*)) = \|\Phi(uv^*)\|_1 = \|uv^*\|_1 = 1, \tag{132}$$

and therefore $\Psi(uu^*) = \Psi(vv^*)$. From this fact one concludes that Ψ is constant on $\mathrm{D}(\mathcal{X})$, i.e., there exists $\sigma \in \mathrm{D}(\mathcal{Z})$ for which $\Psi(X) = \mathrm{Tr}(X)\sigma$ for all $X \in \mathrm{L}(\mathcal{X})$. Statement 5 therefore implies statement 7.

Along somewhat similar lines, assume that statement 6 holds, and again let $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Z})$ be any complementary channel to Φ. For any choice of orthogonal vectors $u, v \in \mathcal{X}$ it follows by Lemma 16 that

$$\|\Psi(uv^*)\|_1 = \mathrm{F}(\Phi(uu^*), \Phi(vv^*)) = \mathrm{F}(uu^*, vv^*) = 0, \tag{133}$$

and hence $\Psi(uv^*) = 0$. In particular, this implies that for $E_{ij} \in \mathrm{L}(\mathcal{X})$ with $i \neq j$ one has $\Psi(E_{ij}) = 0$. Furthermore, because

$$E_{ii} - E_{jj} = \frac{1}{2}[(e_i + e_j)(e_i - e_j)^* + (e_i - e_j)(e_i + e_j)^*] \tag{134}$$

and $(e_i + e_j) \perp (e_i - e_j)$, it follows that

$$\Psi(E_{ii}) - \Psi(E_{jj}) = \frac{1}{2}\Psi((e_i + e_j)(e_i - e_j)^*) - \frac{1}{2}\Psi((e_i - e_j)(e_i + e_j)^*) = 0. \tag{135}$$

That is, there exists $\sigma \in \mathrm{D}(\mathcal{Z})$ for which $\Psi(E_{ii}) = \sigma$ for all $1 \leq i \leq n$. Hence, we have

$$J(\Psi) = \sum_{i,j=1}^{n} E_{ij} \otimes \Psi(E_{ij}) = \mathbb{1}_{\mathcal{X}} \otimes \sigma, \tag{136}$$

which is equivalent to $\Psi(X) = \mathrm{Tr}(X)\sigma$ for all $X \in \mathrm{L}(\mathcal{X})$. Statement 6 therefore implies statement 7.

Finally, assume that statement 7 holds. Let $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Z})$ be the complementary channel associated with any fixed Stinespring representation $\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}(AXA^*)$ for $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$. Assuming that $\sigma \in \mathrm{D}(\mathcal{Z})$ satisfies $\Psi(X) = \mathrm{Tr}(X)\sigma$ for all $X \in \mathrm{L}(\mathcal{X})$, it holds that $J(\Psi) = \mathbb{1}_{\mathcal{X}} \otimes \sigma$, and hence

$$\mathrm{Tr}_{\mathcal{Y}}(\mathrm{vec}(A^{\mathsf{T}})\mathrm{vec}(A^{\mathsf{T}})^*) = \mathbb{1}_{\mathcal{X}} \otimes \sigma = \mathrm{Tr}_{\mathcal{X}}(\mathrm{vec}(\mathbb{1}_{\mathcal{X}})\mathrm{vec}(\mathbb{1}_{\mathcal{X}})^* \otimes \sigma). \tag{137}$$

By Lemma 17 there exists a channel $\Xi \in \mathrm{C}(\mathcal{Y}, \mathcal{X})$ for which

$$(\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Xi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\mathrm{vec}(A^{\mathsf{T}})\mathrm{vec}(A^{\mathsf{T}})^*) = \mathrm{vec}(\mathbb{1}_{\mathcal{X}})\mathrm{vec}(\mathbb{1}_{\mathcal{X}})^* \otimes \sigma. \tag{138}$$

By tracing out $\mathcal{Z}$ we get

$$J(\Xi\Phi) = (\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Xi)(J(\Phi)) = \mathrm{vec}(\mathbb{1}_{\mathcal{X}})\mathrm{vec}(\mathbb{1}_{\mathcal{X}})^* = J(\mathbb{1}_{\mathrm{L}(\mathcal{X})}), \tag{139}$$

giving $\Xi\Phi = \mathbb{1}_{\mathrm{L}(\mathcal{X})}$. Statement 7 therefore implies statement 1, which completes the proof. □

7. Discussion

We have shown that there exists a family of channel discrimination problems for which a perfect discrimination requires ancilla system with dimension equal to that of the input, even when the output dimension is much smaller. Beyond this it would be nice to have a formula for, or even non-trivial bounds on, $\left\|\Psi_{n,k} \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^m)}\right\|_1$ when $m < n^k$. To serve as a launching ground for future investigations, in Appendix B we have included numerically computed lower bounds for $\left\|\Psi_{n,2} \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^m)}\right\|_1$ for $2 \leq n \leq 6$ and $n \leq m \leq n^2$, computed in MATLAB using QETLAB [18]. More generally, one could try to find non-trivial bounds on

$$\left\|(\lambda\Phi_0 - (1-\lambda)\Phi_1) \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^k)}\right\|_1 \tag{140}$$

for all $\Phi_0, \Phi_1 \in \mathrm{C}(\mathbb{C}^n, \mathbb{C}^m)$ in terms of n, m, k, and $\|\!|\lambda\Phi_0 - (1-\lambda)\Phi_1|\!\|_1$, though this is likely a much more difficult task.

Theorem 10 shows that for $m \geq n^k$ the optimal operators have a special form where the ancilla system factorizes into k copies of $\mathbb{C}^n$. This seems intuitively natural, as in the channel discrimination setting, discriminating

these channels is like playing k separate Werner-Holevo channel discrimination games using a single resource system, where the referee randomly selects which game will be played and throws away the rest of the input systems. In this setting, Theorem 10 says that all optimal strategies are independent, in the sense that the only way of creating an optimal strategy is to stick together k-instances of optimal strategies for discriminating the Werner-Holevo channels. It is thus natural to conjecture that this would be true for $m < n^k$, however this is not the case. For the $k = 2$ case, we show in Proposition 20 in Appendix A that such independent strategies have the optimal value $n + \lfloor m/n \rfloor$ when $n \leq m < n^2$, however, lower bounds on the optimal value computed in Appendix B are well above this.

Another question is whether or not the optimum in the induced 1-norm of $\Psi_{n,k} \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^m)}$ is achieved by some Hermitian operator when $m < n^k$. Even for Hermiticity preserving maps it is known that this does not hold generally [19]. Proposition 5 shows that this holds for the partial transpose map (i.e. the case when $k = 1$), and numerical evidence in Appendix B suggests that this holds when $k = 2$. We conjecture that it holds for all $n \geq 2$ and $k \geq 1$.

Acknowledgements

We thank Gus Gutoski for suggesting the problem, and Vern Paulsen, Nathaniel Johnston, and Marco Piani for helpful discussions. This work was supported by Canada's NSERC and the Ontario Graduate Scholarship.

Appendix A. Optimal value for independent strategies in the $k = 2$ case

To be precise, what we mean by an *independent strategy* for optimizing

$$\begin{aligned}&\left\|\left(\Psi_{n,2} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\right)(X)\right\|_1 \\ &\quad = \left\|\left(T \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\right)\left(\mathrm{Tr}_{\mathcal{X}_2}(X)\right)\right\|_1 + \left\|\left(T \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\right)\left(\mathrm{Tr}_{\mathcal{X}_1}(X)\right)\right\|_1\end{aligned} \tag{141}$$

for $X \in \mathrm{L}(\mathcal{X}_1 \otimes \mathcal{X}_2 \otimes \mathcal{Y})$, is an attempt at optimizing the above expression with an operator of the following form. For $a, b \in \{1, \ldots, \dim(\mathcal{Y})\}$ with $ab \leq \dim(\mathcal{Y})$ and some $U \in \mathrm{U}(\mathbb{C}^a \otimes \mathbb{C}^b, \mathcal{Y})$, X takes the form

$$X = (\mathbb{1}_{\mathcal{X}_1 \otimes \mathcal{X}_2} \otimes U) \underbrace{(Y_1 \otimes Y_2)}_{\in \mathrm{L}(\mathcal{X}_1 \otimes \mathcal{X}_2 \otimes \mathbb{C}^a \otimes \mathbb{C}^b)} (\mathbb{1}_{\mathcal{X}_1 \otimes \mathcal{X}_2} \otimes U^*) \tag{142}$$

for some $Y_1 \in \mathrm{L}(\mathcal{X}_1 \otimes \mathbb{C}^a)$ and $Y_2 \in \mathrm{L}(\mathcal{X}_2 \otimes \mathbb{C}^b)$ with $\|Y_1\|_1 = \|Y_2\|_1 = 1$, and we are again using the implicit permutation notation introduced in Section 5. For an operator of this form we have

$$\begin{aligned}\left\|\left(\Psi_{n,2} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\right)(X)\right\|_1 = &\left\|\left(T \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^a)}\right)(Y_1)\right\|_1 \left\|\mathrm{Tr}_{\mathcal{X}_2}(Y_2)\right\|_1 + \\ &\left\|\left(T \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^b)}\right)(Y_2)\right\|_1 \left\|\mathrm{Tr}_{\mathcal{X}_1}(Y_1)\right\|_1.\end{aligned} \tag{143}$$

Corollary 11 says that when $\dim(\mathcal{Y}) \geq n^2$, optimal operators are *necessarily* of this form. We now give the optimal value for these operators when $n \leq \dim(\mathcal{Y}) < n^2$.

Proposition 20. *Let $\mathcal{X}_1$ and $\mathcal{X}_2$ denote copies of $\mathbb{C}^n$ and let $\mathcal{Y} = \mathbb{C}^m$ with $n \leq m < n^2$. If $X \in \mathrm{L}(\mathcal{X}_1 \otimes \mathcal{X}_2 \otimes \mathcal{Y})$ is of the form given in Equation (142), then*

$$\left\| \left(\Psi_{n,2} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})} \right)(X) \right\|_1 \leq n + \lfloor m/n \rfloor, \tag{144}$$

and furthermore equality is achieved for some operator of this form.

Proof. First, for such an X the value achieved in Equation (143) can be upper bounded by

$$\begin{aligned} &\left\| \left(T \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^a)} \right)(Y_1) \right\|_1 \left\| \mathrm{Tr}_{\mathcal{X}_2}(Y_2) \right\|_1 + \left\| \left(T \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^b)} \right)(Y_2) \right\|_1 \left\| \mathrm{Tr}_{\mathcal{X}_1}(Y_1) \right\|_1 \\ &\qquad \leq \left\| \left(T \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^a)} \right)(Y_1) \right\|_1 + \left\| \left(T \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^b)} \right)(Y_2) \right\|_1 \\ &\qquad \leq \min(n, a) + \min(n, b), \end{aligned} \tag{145}$$

where the first inequality is monotonicity of the 1-norm under partial trace, and the second is two applications of Proposition 5. Next, observe that for fixed a and b, this value is attained by some choice of Y_1 and Y_2 (again, by Proposition 5), and finally, observe that by virtue of the min functions, there is no reason to consider either $a > n$ or $b > n$. In summary, the optimal value for operators of this form is the same as the optimal value of the following simpler optimization problem

$$\max\{a + b : a, b \in \{1, \ldots, n\}, ab \leq m\} = \alpha. \tag{146}$$

Note that $a = n$ and $b = \lfloor m/n \rfloor$ satisfy the constraints, so $\alpha \geq n + \lfloor m/n \rfloor$.

To see that $\alpha \leq n + \lfloor m/n \rfloor$, consider the relaxed optimization problem

$$\max\{a + b : a, b \in [1, n], ab \leq m\} = \beta \geq \alpha. \tag{147}$$

For a given a the optimal value of b is $\min(n, m/a)$, so

$$\beta = \max\{a + \min(n, m/a) : a \in [1, n]\}. \tag{148}$$

The function $f(a) = a + \min(n, m/a)$ is strictly increasing over the interval $[1, m/n]$, so the optimum is achieved at some point in the interval $[m/n, n]$, on which $f(a) = a + m/a$. f is convex on $[m/n, n]$ as $f''(a) = 2m/a^3 > 0$, so the optimum is achieved at an endpoint, and in this case $f(m/n) = f(n) = n + m/n$. Hence

$$\alpha \leq \beta = n + m/n, \tag{149}$$

and since α is a natural number this implies $\alpha \leq n + \lfloor m/n \rfloor$. □

Appendix B. Numerical tests

For $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$, computing $\|\Phi\|_1$ is hard in general. However, as detailed in [20], there are nice algorithms for computing lower bounds to $\|\Phi\|_1$. For $2 \leq n \leq 6$ and $n \leq m \leq n^2$, Table 1 contains computed lower bounds for $\|\Psi_{n,2} \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^m)}\|_1$, as well as computed lower bounds for $\|\Psi_{n,2} \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^m)}\|_1^H$, where

$$\|\Phi\|_1^H = \max\{\|\Phi(H)\|_1 : H \in \mathrm{Herm}(\mathcal{X}), \|H\|_1 = 1\}. \tag{150}$$

The computations were done in MATLAB using modified versions of the function `InducedSchattenNorm` in the QETLAB [18] package (which uses the algorithm in [20]). For $n = 5$ and $n = 6$, plots ranging over $n \leq m \leq n^2$ are given in Figure 2. The code and data used in this appendix can be found in the GitHub repository at [21].

One feature of the data is that the lower bounds for $\|\Psi_{n,2} \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^m)}\|_1$ and $\|\Psi_{n,2} \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^m)}\|_1^H$ almost always agree (up to stopping precision), and in cases of disagreement the value computed for Hermitian inputs is always the larger of the two. This lends evidence to the conjecture that

$$\left\|\Psi_{n,2} \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^m)}\right\|_1 = \left\|\Psi_{n,2} \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^m)}\right\|_1^H, \tag{151}$$

and the stronger conjecture that

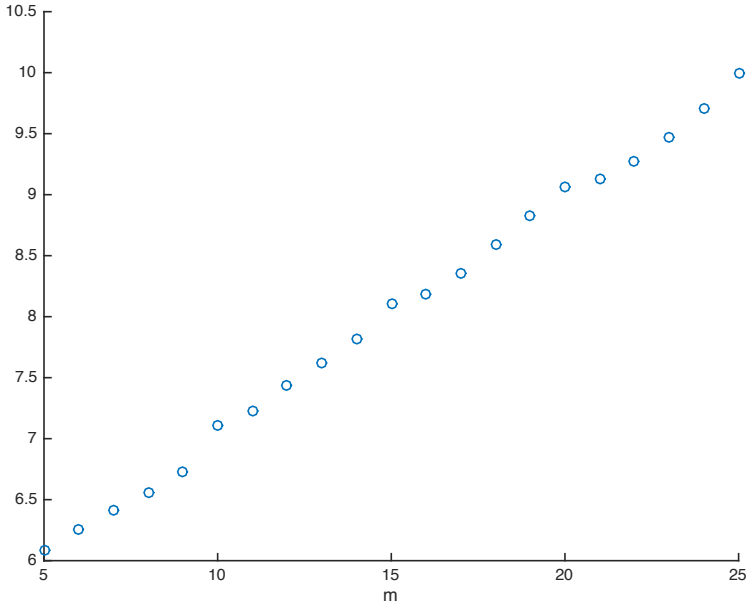
$$\left\|\Psi_{n,k} \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^m)}\right\|_1 = \left\|\Psi_{n,k} \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^m)}\right\|_1^H \tag{152}$$

for all k.

Another curious feature, displayed in Figure 2, is that while seeming to increase roughly linearly in m, there is a bump when m is a multiple of n, with dips between these points. It is unclear whether this is an actual feature of $\left\|\Psi_{n,2} \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^m)}\right\|_1$ or is a peculiarity of the lower bounds found by the algorithm.

FIGURE 2. Plots for the data in Table 1 for $n = 5$ and $n = 6$.

(A) $n = 5$, $5 \leq m \leq 25$.



(B) $n = 6$, $6 \leq m \leq 36$.

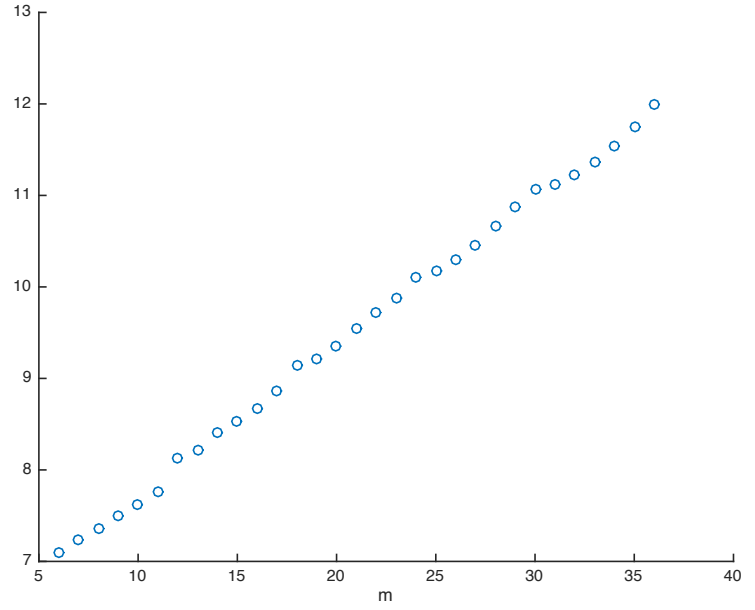


TABLE 1. Lower bounds for $\left\|\Psi_{n,2} \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^m)}\right\|_1$ and $\left\|\Psi_{n,2} \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^m)}\right\|_1^H$ (the columns with '-H') for $2 \leq n \leq 6$ (columns) and $n \leq m \leq n^2$ (rows), computed using 1000 initial guesses and a stopping tolerance of 10^{-5}.

m\n	2	2-H	3	3-H	4	4-H	5	5-H	6	6-H
2	3.0448	3.0448								
3	3.4142	3.4142	4.0656	4.0656						
4	4.0000	4.0000	4.3307	4.3307	5.0777	5.0777				
5			4.6386	4.6386	5.2830	5.2830	6.0857	6.0857		
6			5.0551	5.0551	5.4711	5.4711	6.2527	6.2527	7.0914	7.0914
7			5.2361	5.2361	5.6949	5.6949	6.4100	6.4100	7.2319	7.2319
8			5.5615	5.5616	6.0896	6.0896	6.5593	6.5593	7.3666	7.3666
9			6.0000	6.0000	6.2240	6.2241	6.7331	6.7331	7.4961	7.4961
10					6.4873	6.4873	7.1136	7.1136	7.6209	7.6209
11					6.7635	6.7635	7.2207	7.2209	7.7611	7.7611
12					7.0596	7.0596	7.4396	7.4396	8.1312	8.1312
13					7.1622	7.1623	7.6222	7.6222	8.2202	8.2206
14					7.3722	7.3723	7.8151	7.8152	8.4068	8.4068
15					7.6457	7.6457	8.1023	8.1023	8.5342	8.5342
16					8.0000	8.0000	8.1873	8.1874	8.6700	8.6701
17							8.3605	8.3605	8.8563	8.8564
18							8.5850	8.5850	9.1344	9.1344
19							8.8297	8.8297	9.2058	9.2061
20							9.0623	9.0623	9.3479	9.3480
21							9.1295	9.1296	9.5437	9.5437
22							9.2749	9.2749	9.7192	9.7192
23							9.4641	9.4641	9.8829	9.8830
24							9.7016	9.7016	10.1101	10.1101
25							10.0000	10.0000	10.1708	10.1711
26									10.2970	10.2971
27									10.4621	10.4621
28									10.6717	10.6717
29									10.8717	10.8717
30									11.0639	11.0639
31									11.1145	11.1146
32									11.2170	11.2170
33									11.3589	11.3589
34									11.5311	11.5311
35									11.7416	11.7416
36									12.0000	12.0000

References

[1] Reinhard Werner and Alexander Holevo. Counterexample to an additivity conjecture for output purity of quantum channels. *Journal of Mathematical Physics*, 43(9):4353–4357, September 2002.

[2] John Watrous. Theory of Quantum Information. `https://cs.uwaterloo.ca/~watrous/TQI`, 2015.

[3] Uffe Haagerup. Injectivity and decomposition of completely bounded maps. In Huzihiro Araki, Calvin C. Moore, Şerban-Valentin Stratila, and Dan-Virgil Voiculescu, editors, *Operator Algebras and their Connections with Topology and Ergodic Theory*, number 1132 in Lecture Notes in Mathematics, pages 170–222. Springer Berlin Heidelberg, 1985.

[4] Guifré Vidal and Reinhart Werner. Computable measure of entanglement. *Physical Review A*, 65(3):032314, February 2002.

[5] Ashwin Nayak and Pranab Sen. Invertible Quantum Operations and Perfect Encryption of Quantum States. *Quantum Info. Comput.*, 7(1):103–110, January 2007.

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

[7] Carl Helstrom. Detection theory and quantum mechanics. *Information and Control*, 10(3):254–291, March 1967.

[8] Alexander Holevo. An analog of the theory of statistical decisions in noncommutative probability theory. *Transactions of the Moscow Mathematical Society*, 26:133–149, 1972.

[9] Jun Tomiyama. On the transpose map of matrix algebras. *Proceedings of the American Mathematical Society*, 88(4):635–638, 1983.

[10] Matthew Donald, Micha Horodecki, and Oliver Rudolph. The uniqueness theorem for entanglement measures. *Journal of Mathematical Physics*, 43(9):4252–4272, September 2002.

[11] Peter Shor, John Smolin, and Barbara Terhal. Nonadditivity of Bipartite Distillable Entanglement Follows from a Conjecture on Bound Entangled Werner States. *Physical Review Letters*, 86(12):2681–2684, March 2001.

[12] Benjamin Schumacher and Michael Nielsen. Quantum data processing and error correction. *Physical Review A*, 54(4):2629–2635, October 1996.

[13] Matthias Christandl and Andreas Winter. Squashed entanglement: An additive entanglement measure. *Journal of Mathematical Physics*, 45(3):829–840, March 2004.

[14] Robert Tucci. Quantum Entanglement and Conditional Information Transmission. *arXiv:quant-ph/9909041*, September 1999. arXiv: quant-ph/9909041.

[15] Fernando Brandao, Matthias Christandl, and Jon Yard. Faithful Squashed Entanglement. *Communications in Mathematical Physics*, 306(3):805–830, September 2011. arXiv: 1010.1750.

[16] Emanuel Knill and Raymond Laflamme. Theory of quantum error-correcting codes. *Physical Review A*, 55(2):900–911, February 1997.

[17] Lajos Molnár. Fidelity preserving maps on density operators. *Reports on Mathematical Physics*, 48(3):299–303, December 2001.

[18] Nathaniel Johnston. QETLAB: A MATLAB toolbox for quantum entanglement, version 0.9. http://qetlab.com, January 2016.

[19] John Watrous. Notes on super-operator norms induced by Schatten norms. *Quantum Information and Computation*, 5(1):58–68, January 2005.

[20] Nathaniel Johnston. How to compute hard-to-compute matrix norms. http://www.njohnston.ca/2016/01/how-to-compute-hard-to-compute-matrix-norms, January 2016.

[21] Daniel Puzzuoli. ancilla_dimension. https://github.com/DanPuzzuoli/ancilla_dimension, April 2016.

Daniel Puzzuoli
Department of Applied Mathematics and Institute for Quantum Computing
University of Waterloo
Waterloo, Ontario
Canada
e-mail: dpuzzuol@uwaterloo.ca

John Watrous
School of Computer Science and Institute for Quantum Computing
University of Waterloo
Waterloo, Ontario
Canada
e-mail: watrous@uwaterloo.ca