Entropy power inequalities for qudits

Koenraad Audenaert* Nilanjana Datta† Maris Ozols‡

March 13, 2015

Abstract

Shannon's entropy power inequality (EPI) can be viewed as a statement of concavity of an entropic function of a continuous random variable under a scaled addition rule:

$$f(\sqrt{a}\,X + \sqrt{1-a}\,Y) \geq a f(X) + (1-a) f(Y) \quad \forall\, a \in [0,1].$$

Here, X and Y are continuous random variables and the function f is either the differential entropy or, for $a = 1/2$, the *entropy power*. König and Smith [*IEEE Trans. Inf. Theory. 60(3):1536–1548, 2014*] obtained quantum analogues of these inequalities for continuous-variable quantum systems, where X and Y are replaced by bosonic fields and the addition rule is the action of a beamsplitter with transmissivity a on those fields. In this paper, we similarly establish a class of EPI analogues for d-level quantum systems (i.e. qudits). The underlying addition rule for which these inequalities hold is given by a quantum channel that depends on the parameter $a \in [0,1]$ and acts like a finite-dimensional analogue of a beamsplitter with transmissivity a, converting a two-qudit product state into a single qudit state. We refer to this channel as a *partial swap channel* because of the particular way its output interpolates between the states of the two qudits in the input as a is changed from zero to one. We obtain analogues of Shannon's EPI, not only for the von Neumann entropy and the entropy power for the output of such channels, but for a much larger class of functions as well. This class includes the Rényi entropies and the subentropy. We also prove a qudit analogue of the entropy photon number inequality (EPnI). Finally, for the subclass of partial swap channels for which one of the qudit states in the input is fixed, our EPIs and EPnI yield lower bounds on the minimum output entropy and upper bounds on the Holevo capacity.

1 Introduction

Inequalities between entropic quantities play a fundamental role in information theory and have been employed effectively in finding bounds on optimal rates of various information-processing tasks. Shannon's entropy power inequality (EPI) [Sha48] is one such inequality and it has proved to be of relevance in studying problems not only in information theory, but also in probability theory and mathematical physics [Sta59]. It has been used, for example, in finding upper bounds on the capacities of certain noisy channels (e.g. the Gaussian broadcast channel [Ber74]) and in proving convergence in relative entropy for the Central Limit Theorem [Bar86].

*Department of Mathematics, Royal Holloway University of London, Egham TW20 0EX, U.K. & Department of Physics and Astronomy, Ghent University, S9, Krijgslaan 281, B-9000 Ghent, Belgium

†Statistical Laboratory, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB, UK

‡Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK

Classical EPIs

For an arbitrary random variable X on $\mathbb{R}^d$ with probability density function (p.d.f.) f_X, the *entropy power* of X is the quantity

$$v(X) := \frac{e^{2H(X)/d}}{2\pi e}, \tag{1.1}$$

where $H(X)$ is the *differential entropy* of X,

$$H(X) := -\int_{\mathbb{R}^d} f_X(x) \log f_X(x) dx \tag{1.2}$$

(throughout the paper we use log to represent the natural logarithm). The name "entropy power" is derived from the following fact: if X is a Gaussian random variable on $\mathbb{R}$ with zero mean and variance σ^2, then $H(X) = (1/2)\log(2\pi e\sigma^2)$; hence $v(X)$ is equal to its variance, which is commonly referred to as its *power*. Note that the entropy power of a random variable X is equal to the variance of a Gaussian random variable which has the same differential entropy as X. For X on $\mathbb{R}^d$, we shall henceforth omit the factor $1/2\pi e$ and refer to $e^{2H(X)/d}$ as the entropy power, as in [KS14].

The entropy power satisfies the following *scaling property*: $v(\sqrt{\alpha}X) = \alpha v(X)$. This follows from the scaling property of p.d.f.s: if $f_{\alpha X}$ denotes the p.d.f. of a random variable αX on $\mathbb{R}^d$, where $\alpha > 0$, then $f_{\alpha X}(x) = \alpha^{-d} f_X(x/\alpha)$, $x \in \mathbb{R}^d$, which in turn implies that $H(\alpha X) = H(X) + d\log\alpha$. This shows why the factor $1/d$ in the definition of $v(X)$ has to be there for X on $\mathbb{R}^d$.

Shannon's EPI [Sha48] provides a lower bound on the entropy power of a sum of two independent random variables X and Y on $\mathbb{R}^d$ in terms of the sums of the entropy powers of the individual random variables:

$$v(X+Y) \geq v(X) + v(Y), \tag{1.3}$$

or equivalently,

$$e^{2H(X+Y)/d} \geq e^{2H(X)/d} + e^{2H(Y)/d}. \tag{1.4}$$

Here, $H(X+Y)$ is the differential entropy of the p.d.f. of the sum $Z := X+Y$, which is given by the convolution

$$f_{X+Y}(x) = (f_X * f_Y)(x) := \int_{\mathbb{R}^d} f_X(x') f_Y(x-x') dx', \qquad \forall\, x \in \mathbb{R}^d. \tag{1.5}$$

The inequality eq. (1.3) was proposed by Shannon in [Sha48] as a means to bound the capacity of a non-Gaussian *additive noise channel*, that is, a channel with input X and output $X+Y$, with Y being an independent (non-Gaussian) random variable modeling the noise which is added to the input. Later, Lieb [Lie78] and Dembo, Cover, and Thomas [DCT91] showed that the EPI (1.4) can be equivalently expressed as the following inequality between differential entropies:

$$H\big(\sqrt{a}\,X + \sqrt{1-a}\,Y\big) \geq aH(X) + (1-a)H(Y), \qquad \forall\, a \in [0,1]. \tag{1.6}$$

The above inequality was proved by employing the Rényi entropy [Rén61] and using properties of p-norms on convolutions given by a sharp form of Young's inequality [Bec75].

The form of the EPI in eq. (1.6) motivates the definition of an operation (which following [KS13, KS14] we denote as $\boxplus_a$) on the space of random variables, given by the following *scaled addition rule*:

$$X \boxplus_a Y := \sqrt{a}\,X + \sqrt{1-a}\,Y \qquad \forall\, a \in [0,1]. \tag{1.7}$$

The random variable $X \boxplus_a Y$ can be interpreted as an interpolation between X and Y as a is decreased from 1 to 0. With this notation, the inequality (1.6) can be written as

$$H(X \boxplus_a Y) \geq aH(X) + (1-a)H(Y), \qquad \forall\, a \in [0,1]. \tag{1.8}$$

Using the scaling property of the entropy power, the EPI (1.4) can be expressed as follows:

$$e^{2H(X \boxplus_{1/2} Y)/d} \geq \frac{1}{2} e^{2H(X)/d} + \frac{1}{2} e^{2H(Y)/d}. \tag{1.9}$$

Shannon's EPI (1.4) (and hence also (1.8) and (1.9)) was first proved rigorously by Stam [Sta59] and by Blachman [Bla65], by employing de Bruijn's identity, which couples Fisher information with differential entropy. Since then various different proofs and generalizations of the EPI have been proposed (see e.g. [VG06, Rio11, SS15] and references therein).

It is natural to conjecture that an analogue of Shannon's EPI also holds for discrete random variables e.g. on non-negative integers. This conjecture was first proved by [HV03] for the case of binomial random variables. They proved that if $X_n \sim \mathrm{Bin}(n, p)$, then for $p = 1/2$ (see also [SDM11]):

$$e^{2H(X_n + X_m)} \geq e^{2H(X_n)} + e^{2H(X_m)} \quad \forall\ m, n \geq 1. \tag{1.10}$$

Further, Johnson and Yu [JY10] established a form of the EPI which is valid for arbitrary discrete random variables, whereby the scaling operation of a continuous random variable was suitably replaced by the so-called *thinning operation* introduced by Rényi [Rén56], which is considered to be an analogue of scaling for discrete random variables.

Quantum analogues of EPIs

The discovery of an analogue of the EPI in the quantum setting by König and Smith [KS14] marked a significant advance in quantum information theory. They proposed an EPI which holds for continuous-variable quantum systems that arise, for example, in quantum optics. In this case, the random variables X, Y of the classical EPIs (eq. (1.8) and eq. (1.9)) are replaced by quantum fields, bosonic modes of electromagnetic radiation, described by quantum states ρ_X, ρ_Y, which act on a separable, infinite-dimensional Hilbert space. The differential entropy is accordingly replaced by the *von Neumann entropy* $H(\rho) := -\operatorname{Tr}(\rho \log \rho)$.

A prerequisite for any quantum analogue of the EPI is the formulation of a suitable analogue of the addition rule (1.7) which can be applied to pairs of quantum states. Since the quantum-mechanical analogue of additive Gaussian noise is modelled by the mixing of two bosonic modes at a beamsplitter, König and Smith considered the parameter a in eq. (1.7) to be the transmissivity of a beamsplitter, and the states ρ_X, ρ_Y to be its input modes. The classical addition rule eq. (1.7) is thereby replaced by an analogous *quantum field addition rule* for the field operators. In particular, if the two input signals are n-mode bosonic fields, with annihilation operators $\hat{a}_1, \ldots, \hat{a}_n$ and $\hat{b}_1, \ldots, \hat{b}_n$ respectively, then the output is an n-mode bosonic field with annihilation operators $\hat{c}_1, \ldots, \hat{c}_n$, where

$$\hat{c}_i := \sqrt{a}\, \hat{a}_i + \sqrt{1-a}\, \hat{b}_i. \tag{1.11}$$

In a state space description, the input signals are described by quantum states ρ_X, ρ_Y. This yields an equivalent *quantum state addition rule*, where the beamsplitter converts the incoming state $\rho_X \otimes \rho_Y$ to a state $\rho_X \boxplus_a \rho_Y$ given by

$$(\rho_X, \rho_Y) \mapsto \rho_X \boxplus_a \rho_Y := \mathcal{E}_a(\rho_X \otimes \rho_Y). \tag{1.12}$$

Here, $\mathcal{E}_a$ is a linear, completely positive trace-preserving map defined through the relation

$$\mathcal{E}_a(\rho) := \operatorname{Tr}_Y(U_a \rho U_a^\dagger),$$

with the partial trace being taken over the second mode, and U_a is the unitary operator describing the action of the beamsplitter on state space $\mathcal{H}$. Analogous to the classical case, the state $\rho_X \boxplus_a \rho_Y$ reduces to ρ_X when $a = 1$, and to ρ_Y when $a = 0$.

The above inequalities do not reduce to the classical EPIs eq. (1.8) and eq. (1.9) for commuting states; in other words, they are not quantum generalizations of the Shannon's original EPI in the usual sense, as they do not include the latter as a special case. This is because the addition rule acts at the field operator level and not at the state level. In fact, the dependence of the output state on the parameter a is much more complicated than in the classical case.

König and Smith [KS14] proved that the following quantum analogues of the EPIs (1.8) and (1.9) hold, under the quantum addition rule given by eq. (1.12):

$$H(\rho_X \boxplus_a \rho_Y) \geq aH(\rho_X) + (1-a)H(\rho_Y), \tag{1.13}$$

$$e^{H(\rho_X \boxplus_{1/2} \rho_Y)/n} \geq \frac{1}{2}e^{H(\rho_X)/n} + \frac{1}{2}e^{H(\rho_Y)/n}, \tag{1.14}$$

where n is the number of bosonic modes. The inequality (1.14) corresponds to a 50 : 50 beamsplitter (i.e., a beamsplitter with transmissivity $a = 1/2$). Later, De Palma *et al.* [DMG14] proved that an analogous inequality also holds for *any* beamsplitter (i.e., for any $a \in [0, 1]$) and is given by the following:

$$e^{H(\rho_X \boxplus_a \rho_Y)/n} \geq ae^{H(\rho_X)/n} + (1-a)e^{H(\rho_Y)/n}, \qquad \forall\, a \in [0, 1]. \tag{1.15}$$

Note that the EPI given by eq. (1.14) seems to differ from its classical counterpart (1.9) by a factor of 2 in the exponent. However, one can argue that the dimension of the bosonic phase space is $d = 2n$ (as there are 2 quadratures per mode).

Another inequality, related to the EPI (1.15), was conjectured by Guha *et al*. [GES08] and is known as the *entropy photon number inequality* (EPnI). The thermal state of a bosonic mode with annihilation operator $\hat{a}$ can be expressed as [GSE08]:

$$\rho_T = \sum_{i=0}^{\infty} \frac{N^i}{(N+1)^{i+1}} |i\rangle\langle i|, \tag{1.16}$$

where $N := \operatorname{Tr}\left(\rho_T\, \hat{a}^\dagger a\right)$ is the average photon number of the state ρ_T. Its von Neumann entropy evaluates to $H(\rho_T) = g(N)$, where $g(x) := (1+x)\log(1+x) - x\log x$, and hence $N = g^{-1}(H(\rho_T))$. Correspondingly, the *photon number* of an n-mode bosonic state ρ is defined as $N(\rho) = g^{-1}(H(\rho)/n)$. Guha *et al*. [GES08] conjectured that

$$N(\rho_X \boxplus_a \rho_Y) \geq aN(\rho_X) + (1-a)N(\rho_Y) \quad \forall\ a \in [0, 1], \tag{1.17}$$

where $\boxplus_a$ is again the quantum state addition rule eq. (1.12). This conjecture is of particular significance in quantum information theory since if it were true then it would allow one to evaluate classical capacities of various bosonic channels, e.g. the thermal noise channel [GGL$^+$04], the bosonic broadcast channel [GSE07] and the wiretap channel [GSE08]. It has thus far been proved only for Gaussian states [Guh08].

A natural question to ask is whether quantum EPIs can also be found outside the continuous-variable setting. In this paper, we address this question by formulating an addition rule for d-level systems (qudits) in the form of a quantum channel $\mathcal{E}_a$, which we call the *partial swap channel*, that acts on the two input quantum states. We then prove analogues of the quantum EPIs (1.13) and (1.15) for this addition rule. We also prove similar inequalities for a large class $\mathcal{F}$ of functions, including the Rényi entropies of order $\alpha \in [0, 1)$ and the subentropy [JRW94]. Again these are analogues and not generalizations of the classical EPIs for discrete random variables [HV03, JY10, SDM11] to the non-commutative setting, as the latter do not emerge as special cases for commuting states.

Furthermore, the concept of entropy photon number N has a straightforward generalization to qudit systems via its one-to-one relation with the von Neumann entropy, $H = g(N)$, even though it loses its interpretation as an average photon number. We show that the function g^{-1} is in the class $\mathcal{F}$, and as a result obtain the EPnI for our qudit addition rule.

Finally, we apply our results (EPIs and EPnI) to obtain lower bounds on the minimum output entropy and upper bounds on the Holevo capacity for a class of single-input channels that are formed from the channel $\mathcal{E}_a$ by fixing the second input state.

The EPIs in eqs. (1.13) to (1.15) for continuous-variable quantum systems were proved using methods analogous to those used in proving the classical EPIs (1.8) and (1.9), albeit with suitable adaptations to the quantum setting. In contrast, the proof of our EPIs relies on completely different tools, namely, spectral majorization and concavity of functions.

2 Preliminaries

Let $\mathcal{H} \simeq \mathbb{C}^d$ be a finite-dimensional Hilbert space (i.e., a complex Euclidean space), let $\mathcal{L}(\mathcal{H})$ denote the set of linear operators acting on $\mathcal{H}$, and let $\mathcal{D}(\mathcal{H})$ be the set of density operators or *states* on $\mathcal{H}$:

$$\mathcal{D}(\mathcal{H}) := \{\rho \in \mathcal{L}(\mathcal{H}) : \rho \geq 0,\, \operatorname{Tr}\rho = 1\}. \tag{2.1}$$

Moreover, let $\mathcal{U}(\mathcal{H})$ be the set of unitary operators acting on $\mathcal{H}$. We denote the identity operator on $\mathcal{H}$ by I. A quantum channel (or quantum operation) is given by a linear, completely positive, trace-preserving (CPTP) map $\mathcal{N} : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{K})$, with $\mathcal{H}$ and $\mathcal{K}$ being the input and output Hilbert spaces of the channel. For a state $\rho \in \mathcal{D}(\mathbb{C}^d)$ with eigenvalues $\lambda_1, \ldots, \lambda_d$, the *von Neumann entropy* $H(\rho)$ is equal to the Shannon entropy of the probability distribution $\{\lambda_1, \ldots, \lambda_d\}$, i.e., $H(\rho) := -\operatorname{Tr}(\rho \log \rho) = -\sum_{i=1}^d \lambda_i \log \lambda_i$, where we take the logarithms to base e.

The proof of the quantum EPIs that we propose, relies on the concept of majorization (see e.g. [Bha97]). For convenience we recall its definition below, making use of the following notation: for any vector $\vec{u} = (u_1, u_2, \ldots, u_d) \in \mathbb{R}^d$ let $u_1^\downarrow \geq u_2^\downarrow \geq \ldots \geq u_d^\downarrow$ denote the components of $\vec{u}$ arranged in non-increasing order.

Definition 1 (Majorization). *For $\vec{u}, \vec{v} \in \mathbb{R}^d$, we say that $\vec{u}$ is* majorised by *$\vec{v}$ and write $\vec{u} \prec \vec{v}$ if*

$$\sum_{i=1}^k u_i^\downarrow \leq \sum_{i=1}^k v_i^\downarrow, \quad \forall\, k \in \{1, \ldots, d\} \tag{2.2}$$

with equality at $k = d$.

Definition 2. *A function $f : \mathbb{R}^d \to \mathbb{R}$ is called* Schur-concave *[Bha97] if $f(\vec{u}) \geq f(\vec{v})$ whenever $\vec{u} \prec \vec{v}$.*

The notion of majorization can be extended to quantum states as follows. For $\rho, \sigma \in \mathcal{D}(\mathbb{C}^d)$, we write $\rho \prec \sigma$ if $\lambda(\rho) \prec \lambda(\sigma)$, where we use the notation $\lambda(\rho)$ to denote the vector of eigenvalues of ρ, arranged in non-increasing order: $\lambda(\rho) := \big(\lambda_1(\rho), \lambda_2(\rho), \ldots, \lambda_d(\rho)\big)$ with

$$\lambda_1(\rho) \geq \lambda_2(\rho) \geq \cdots \geq \lambda_d(\rho). \tag{2.3}$$

The following class of functions plays an important role in our paper. A canonical example of a function in this class is the von Neumann entropy of a density matrix.

Definition 3. *Let $\mathcal{F}$ denote the class of functions $f : \mathcal{D}(\mathbb{C}^d) \to \mathbb{R}$ satisfying the following properties:*

1. Concavity*: for any pair of states $\rho, \sigma \in \mathcal{D}(\mathbb{C}^d)$ and $\forall\, a \in [0,1]$:*

$$f\big(a\rho + (1-a)\sigma\big) \geq af(\rho) + (1-a)f(\sigma). \tag{2.4}$$

2. Symmetry*: $f(\rho)$ depends only on the eigenvalues of ρ and is symmetric in them; that is, there exists a symmetric (i.e. permutation-invariant) function $\phi_f : \mathbb{R}^d \to \mathbb{R}$ such that $f(\rho) = \phi_f(\lambda(\rho))$.*

By restricting to diagonal states, it follows immediately that for every $f \in \mathcal{F}$ the corresponding function ϕ_f is concave. In turn, this means that ϕ_f is also Schur-concave [Bha97, Theorem II.3.3].

3 Main results

We formulate a finite-dimensional version of the quantum addition rule given by eq. (1.12), which was introduced by König and Smith [KS13, KS14] in the context of continuous-variable quantum systems. Our operation, which we also denote by $\boxplus_a$, is parameterized by $a \in [0,1]$. It combines a pair of d-dimensional quantum states ρ and σ according to the following quantum addition rule:

$$\rho \boxplus_a \sigma := a\rho + (1-a)\sigma - \sqrt{a(1-a)}\, i[\rho, \sigma], \tag{3.1}$$

where $[\rho, \sigma] := \rho\sigma - \sigma\rho$. Note that if $[\rho, \sigma] = 0$ then $\rho \boxplus_a \sigma$ is simply a convex combination of ρ and σ. In Section 4 we prove that the map

$$\boxplus_a : \mathcal{D}(\mathbb{C}^d \otimes \mathbb{C}^d) \mapsto \mathcal{D}(\mathbb{C}^d) \tag{3.2}$$

defines a quantum channel (see eqs. (4.17) and (4.18)). This implies that $\rho \boxplus_a \sigma$ is a valid state of a qudit. Examples of such a channel are given in Section 7. The main motivation behind introducing the map $\boxplus_a$ is that, similar to its analogues (eq. (1.7) and eq. (1.12)) in the continuous-variable classical and quantum and settings, it results in an interpolation between the two states which it combines, as the parameter a is changed from 1 to 0.

We are now ready to summarize our main results, which are given by the following two theorems and corollary.

Theorem 4. *For any $f \in \mathcal{F}$ (see Definition 3), density matrices $\rho, \sigma \in \mathcal{D}(\mathbb{C}^d)$, and any $a \in [0,1]$,*

$$f(\rho \boxplus_a \sigma) \geq af(\rho) + (1-a)f(\sigma).$$

Note that from eq. (3.1) it follows that for commuting states (and hence for diagonal states representing probability distributions) this inequality is equivalent to concavity of the function f.

In analogy with the entropy power of p.d.f.s defined in eq. (1.1), as well as the entropy power and entropy photon number of continuous-variable quantum states, we use the von Neumann entropy of finite-dimensional quantum systems to introduce similar quantities for qudits.

Definition 5. *For any $c \geq 0$, we define the* entropy power *E_c and the* entropy photon number *N_c of $\rho \in \mathcal{D}(\mathbb{C}^d)$ as follows:*

$$E_c(\rho) := e^{cH(\rho)}, \tag{3.3}$$

$$N_c(\rho) := g^{-1}(cH(\rho)) \quad \textit{where} \quad g(x) := (x+1)\log(x+1) - x\log x. \tag{3.4}$$

The function $g(x)$ behaves logarithmically, and is bounded from above and from below as

$$1 + \log(x + 1/e) \leq g(x) \leq 1 + \log(x + 1/2), \tag{3.5}$$

from which it follows that

$$\exp(y - 1) - 1/2 \leq g^{-1}(y) \leq \exp(y - 1) - 1/e. \tag{3.6}$$

Note that the quantity $N_c(\rho)$ does not have any obvious physical interpretation for qudits. It is simply defined in analogy to the continuous-variable quantum setting. Our motivation for looking at this quantity is that it allows us to prove a qudit analogue of the entropy photon number inequality (EPnI), which in the bosonic case remains an open problem.

Here we introduced the scaling parameter c to account for the possibility of having a dependence on dimension or number of modes which is different from that arising in the continuous-variable classical and quantum settings. Recall that the classical EPI (1.9) for continuous random variables on $\mathbb{R}^d$ is stated in terms of $E_{2/d}$, while the quantum EPI (1.15) and the conjectured entropy photon number inequality for n-mode bosonic quantum states involves $E_{1/n}$ and $N_{1/n}$, respectively. Our next theorem establishes concavity of E_c and N_c for a wide range of values of c.

Theorem 6. *For $\rho \in \mathcal{D}(\mathbb{C}^d)$, the following functions are concave:*

- *the entropy power $E_c(\rho)$ for $0 \leq c \leq 1/(\log d)^2$,*
- *the entropy photon number $N_c(\rho)$ for $0 \leq c \leq 1/(d-1)$.*

Since $E_c(\rho)$ and $N_c(\rho)$ depend only on the eigenvalues of ρ and are symmetric in them, the above theorem ensures that E_c and N_c belong to the class of functions $\mathcal{F}$ given in Definition 3. From Theorems 4 and 6, and the concavity of the von Neumann entropy, we obtain the following.

Corollary 7. *For any pair of density matrices $\rho, \sigma \in \mathcal{D}(\mathbb{C}^d)$ and any $a \in [0, 1]$,*

$$H(\rho \boxplus_a \sigma) \geq aH(\rho) + (1 - a)H(\sigma), \tag{3.7}$$

$$e^{cH(\rho \boxplus_a \sigma)} \geq ae^{cH(\rho)} + (1 - a)e^{cH(\sigma)} \qquad \textit{for} \quad 0 \leq c \leq 1/(\log d)^2, \tag{3.8}$$

$$N_c(\rho \boxplus_a \sigma) \geq aN_c(\rho) + (1 - a)N_c(\sigma) \qquad \textit{for} \quad 0 \leq c \leq 1/(d - 1). \tag{3.9}$$

Henceforth, we refer to eqs. (3.7) and (3.8) as qudit EPIs and eq. (3.9) as qudit EPnI.

In addition, Theorem 4 also holds for the Rényi entropy $H_\alpha(\rho)$ of order α [Rén61], for $\alpha \in [0, 1)$, the subentropy $Q(\rho)$ [JRW94, DDJB14], defined as follows:

$$H_\alpha(\rho) := \frac{1}{\alpha - 1} \log(\operatorname{Tr} \rho^\alpha), \tag{3.10}$$

$$Q(\rho) := -\sum_{i=1}^{n} \frac{\lambda_i^n}{\prod_{j \neq i}(\lambda_i - \lambda_j)} \log \lambda_i, \tag{3.11}$$

where $\lambda_1, \ldots, \lambda_d$ denote the eigenvalues of ρ. If some eigenvalues coincide (or are zero), $Q(\rho)$ is defined to be the corresponding limit of the above expression, which is always well-defined and finite.

The above functions are clearly symmetric in the eigenvalues of ρ and are known to be concave. Hence, they belong to the class $\mathcal{F}$ and thus obey the inequality in Theorem 4.

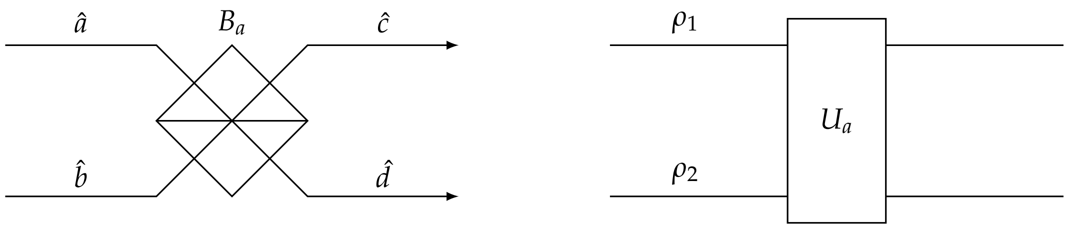

Figure 1: A comparison of a beamsplitter and the partial swap operation.

4 An addition rule for qudit states

In this section we show how we arrive at the quantum addition rule for qudits, (3.1), for which we prove a family of EPIs. This rule is based on a continuous version of the swap operation and it mimics the behavior of a beamsplitter.

4.1 Beamsplitter

Let $\hat{a}, \hat{b}$ denote the annihilation operators of the two bosonic input modes of a beamsplitter and $\hat{b}, \hat{c}$ denote the annihilation operators of the two output modes (see Fig. 1, left). Then the action of a beamsplitter on the input modes is described as follows [KMN$^+$07]:

$$\begin{pmatrix} \hat{c} \\ \hat{d} \end{pmatrix} = B \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix}, \tag{4.1}$$

where B is an arbitrary 2×2 unitary matrix also known as the *scattering matrix*. In particular, let us choose[1]

$$B_a := \begin{pmatrix} \sqrt{a} & i\sqrt{1-a} \\ i\sqrt{1-a} & \sqrt{a} \end{pmatrix}, \tag{4.2}$$

where $a \in [0, 1]$ is the *transmissivity* of the beamsplitter. Note that B_a interpolates between the identity matrix I and $i\sigma_x$ (where σ_x is the Pauli-x matrix) as a changes from 1 to 0. Indeed, we can write

$$B_a = \sqrt{a}\, I + i\sqrt{1-a}\, \sigma_x. \tag{4.3}$$

In particular, up to an unimportant phase, B_0 acts as the swap operation σ_x between the two modes. Thus, for intermediate values of a, we can interpret B_a as an operator that partially swaps the two modes. Following this intuition, in the next section we introduce a partial swap operation for two qudits (see Fig. 1, right) that mimics the action of B_a. It is also described by a unitary matrix that is an interpolation between the identity and a swap operation (up to a phase factor), but with a swap that exchanges two qudits.

4.2 The partial swap operator

Let $\{|i\rangle\}_{i=1}^d$ denote the standard basis of $\mathcal{H} \simeq \mathbb{C}^d$. Then $\{|i,j\rangle\}_{i,j=1}^d$ is an orthonormal basis of $\mathcal{H} \otimes \mathcal{H}$. The qudit *swap operator* $S \in \mathcal{U}(\mathcal{H} \otimes \mathcal{H})$ is defined through its action on the basis vectors $|i,j\rangle$ as follows:

$$S|i,j\rangle = |j,i\rangle \quad \text{for all } i,j \in \{1,\ldots,d\} \tag{4.4}$$

[1]Note that this choice differs from the one corresponding to eq. (1.11)

and can be expressed as

$$S = \sum_{i,j=1}^{d} |i\rangle\langle j| \otimes |j\rangle\langle i|. \tag{4.5}$$

Clearly, $S^\dagger = S$ and $S^2 = I$. In analogy with the beamsplitter scattering matrix eq. (4.3), we define a qudit partial swap operator as a unitary interpolation between the identity and the swap operator.

First note that since S is Hermitian, we can view it as a Hamiltonian. The evolution for time $t \in \mathbb{R}$ under its action is given by the following unitary operator, where we used the fact that $S^2 = I$:

$$\exp(itS) = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} S^n \tag{4.6}$$

$$= I \cos t + iS \sin t. \tag{4.7}$$

In particular, $\exp(i(\pi/2)S) = iS$, so $\exp(itS) = (iS)^{2t/\pi}$. Thus, as t changes from 0 to $\pi/2$, this unitary operator is an interpolation between I and iS, the swap gate up to a global phase.[2] This leads to the following definition of a partial swap operator as an incompletely evolved version of iS:

Definition 8. *For $a \in [0,1]$, with $a = \cos^2 t$, the* partial swap operator $U_a \in \mathcal{U}(\mathbb{C}^d \otimes \mathbb{C}^d)$ *is the unitary operator*

$$U_a := (iS)^{2t/\pi} = \sqrt{a}\, I + i\sqrt{1-a}\, S. \tag{4.8}$$

Note that $U_1 = I$ while $U_0 = iS$ acts as the qudit swap operation under conjugation: $U_0(\rho_1 \otimes \rho_2)U_0^\dagger = \rho_2 \otimes \rho_1$.

Example (Qubit case: $d = 2$)**.** *The matrix representation of the partial swap operator for qubits is*

$$U_a = \begin{pmatrix} \sqrt{a} + i\sqrt{1-a} & 0 & 0 & 0 \\ 0 & \sqrt{a} & i\sqrt{1-a} & 0 \\ 0 & i\sqrt{1-a} & \sqrt{a} & 0 \\ 0 & 0 & 0 & \sqrt{a} + i\sqrt{1-a} \end{pmatrix}. \tag{4.9}$$

4.3 The partial swap channel

Consider a family of CPTP maps $\mathcal{E}_a : \mathcal{D}(\mathbb{C}^d \otimes \mathbb{C}^d) \to \mathcal{D}(\mathbb{C}^d)$ parameterized by $a \in [0,1]$ and defined in terms of the partial swap operator U_a given in eq. (4.8). For any $\rho_{12} \in \mathcal{D}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ with $\mathcal{H}_1, \mathcal{H}_2 \simeq \mathbb{C}^d$, let

$$\mathcal{E}_a(\rho_{12}) := \mathrm{Tr}_2\big(U_a \rho_{12} U_a^\dagger\big), \tag{4.10}$$

where we trace out the second system. We are particularly interested in the case in which the input state ρ_{12} is a product state, i.e., $\rho_{12} = \rho_1 \otimes \rho_2$ for some $\rho_1, \rho_2 \in \mathcal{D}(\mathbb{C}^d)$. When $\mathcal{E}_a$ is applied on such states, it combines the two density matrices ρ_1 and ρ_2 in a non-trivial manner, which mimics the action of a beamsplitter [KS13, KS14]. To wit, $\mathcal{E}_0(\rho_1 \otimes \rho_2) = \rho_2$ and $\mathcal{E}_1(\rho_1 \otimes \rho_2) = \rho_1$, while for general $a \in [0,1]$ equation eq. (4.12) below shows that the output of $\mathcal{E}_a(\rho_1 \otimes \rho_2)$ lies on an elliptical path from ρ_1 to ρ_2. The following lemma provides an explicit expression for the resulting state.

[2] We are interested only in how this matrix acts under conjugation, so the global phase can be ignored.

Lemma 9. *Let $\mathcal{E}_a$ denote the map defined in eq.* (4.10)*, with $\sqrt{a} = \cos t$, and $[\rho_1, \rho_2] := \rho_1\rho_2 - \rho_2\rho_1$. Then for $\rho_1, \rho_2 \in \mathcal{D}(\mathbb{C}^d)$,*

$$\mathcal{E}_a(\rho_1 \otimes \rho_2) = a\rho_1 + (1-a)\rho_2 - \sqrt{a(1-a)}\, i[\rho_1, \rho_2] \tag{4.11}$$

$$= \frac{\rho_1 + \rho_2}{2} + \cos 2t\, \frac{\rho_1 - \rho_2}{2} - \sin 2t\, \left(\frac{i}{2}[\rho_1, \rho_2]\right). \tag{4.12}$$

Proof. Using eq. (4.8) we get

$$\begin{aligned} U_a(\rho_1 \otimes \rho_2)U_a^\dagger &= \big(\sqrt{a}\, I + i\sqrt{1-a}\, S\big)(\rho_1 \otimes \rho_2)\big(\sqrt{a}\, I - i\sqrt{1-a}\, S\big) \\ &= a\, \rho_1 \otimes \rho_2 + (1-a)\, \rho_2 \otimes \rho_1 + i\sqrt{a(1-a)}\, \big(S(\rho_1 \otimes \rho_2) - (\rho_1 \otimes \rho_2)S\big). \end{aligned} \tag{4.13}$$

After tracing out the second system, the first two terms of the above expression give the first two terms of eq. (4.11). To get the last term of eq. (4.11), note that

$$\begin{aligned} \mathrm{Tr}_2\big((\rho_1 \otimes \rho_2)S\big) &= \sum_{k=1}^{d} (I \otimes \langle k|)\left((\rho_1 \otimes \rho_2) \sum_{i,j=1}^{d} |i\rangle\langle j| \otimes |j\rangle\langle i|\right)(I \otimes |k\rangle) \\ &= \sum_{i,j,k=1}^{d} \rho_1|i\rangle\langle j| \otimes \langle k|\rho_2|j\rangle\langle i|k\rangle \\ &= \rho_1 \sum_{i,j=1}^{d} \langle i|\rho_2|j\rangle|i\rangle\langle j| = \rho_1\rho_2 \end{aligned} \tag{4.14}$$

and similarly $\mathrm{Tr}_2\big(S(\rho_1 \otimes \rho_2)\big) = \rho_2\rho_1$. Hence,

$$\mathrm{Tr}_2\big(S(\rho_1 \otimes \rho_2) - (\rho_1 \otimes \rho_2)S\big) = \rho_2\rho_1 - \rho_1\rho_2 = [\rho_2, \rho_1], \tag{4.15}$$

which yields the last term of eq. (4.11). □

One can check that the action of the channel $\mathcal{E}_a$ on an arbitrary state $\rho \in \mathcal{D}(\mathbb{C}^d \otimes \mathbb{C}^d)$ (i.e. not necessarily a product state) can be expressed as $\mathcal{E}_a(\rho) = \sum_{k=1}^{d} A_k \rho A_k^\dagger$ with the Kraus operators A_k given by

$$A_k := \sqrt{a}\, I \otimes \langle k| + i\sqrt{1-a}\, \langle k| \otimes I \quad \text{for} \quad k \in \{1, \ldots, d\}. \tag{4.16}$$

Using Lemma 9, we introduce a qudit addition rule which combines two $d \times d$ density matrices.

Definition 10 (Qudit addition rule)**.** *For any $a \in [0, 1]$ and any $\rho_1, \rho_2 \in \mathcal{D}(\mathbb{C}^d)$, we define*

$$\rho_1 \boxplus_a \rho_2 := \mathcal{E}_a(\rho_1 \otimes \rho_2) = \mathrm{Tr}_2\big(U_a(\rho_1 \otimes \rho_2)U_a^\dagger\big) \tag{4.17}$$

$$= a\rho_1 + (1-a)\rho_2 - \sqrt{a(1-a)}\, i[\rho_1, \rho_2]. \tag{4.18}$$

This operation is bilinear under convex combinations and obeys $\rho_1 \boxplus_0 \rho_2 = \rho_2$ and $\rho_1 \boxplus_1 \rho_2 = \rho_1$.

Example (Qubit case: $d = 2$)**.** *Let $\vec{r}$, $\vec{r}_1$, $\vec{r}_2$ denote the Bloch vectors (see Appendix A.1) of states $\rho_1 \boxplus_a \rho_2$, ρ_1, ρ_2, respectively. Using the properties of Pauli matrices, one can show that eq.* (4.18) *is equivalent to*

$$\vec{r} = a\vec{r}_1 + (1-a)\vec{r}_2 + \sqrt{a(1-a)}\, \vec{r}_1 \times \vec{r}_2, \tag{4.19}$$

where $\vec{r}_1 \times \vec{r}_2$ denotes the cross product of $\vec{r}_1$ and $\vec{r}_2$.

4.4 Partial swap vs. mixing

It is natural to ask whether there is any other CPTP map $\widetilde{\mathcal{E}}_a$ for which the EPIs that we prove also hold. The answer to this question is "yes". A simple example is the CPTP map $\widetilde{\mathcal{E}}_a$ that acts on product states by mixing the two factors, i.e. for which

$$\widetilde{\mathcal{E}}_a(\rho \otimes \sigma) := \rho \,\widetilde{\boxplus}_a\, \sigma = a\rho + (1-a)\sigma. \tag{4.20}$$

It has the $2d$ Kraus operators: $A_k := \sqrt{a}\, I \otimes \langle k|$, $B_k := \langle k| \otimes \sqrt{1-a}\, I$, for $k = 1, \ldots, d$, and requires an ancillary qubit. Note, however, that for this choice of $\widetilde{\mathcal{E}}_a$ (and hence $\boxplus_a$) the EPI (3.7) is simply the concavity of the von Neumann entropy.

In contrast, the partial swap channel $\mathcal{E}_a$ has the following features: (i) it yields non-trivial EPIs (that are not simply a statement of concavity), and (ii) it does not require an ancillary qubit, so it has only d Kraus operators, which is the minimal number required for tracing out a d-dimensional system.

5 Proof of Theorem 4

In this section we prove Theorem 4, our main result. Due to the very different setup as compared to the work of König and Smith, with our addition rule acting at the level of states rather than at the level of field operators, our mathematical treatment is entirely different from theirs and bears no obvious similarity with the classical case either. Instead of proceeding via quantum generalizations of Young's inequality, Fisher information and de Bruijn's identity, the main ingredient in our proof is the following majorization relation relating the spectrum of the output state to the spectra of the input states.

Theorem 11. *For any pair of density matrices $\rho, \sigma \in \mathcal{D}(\mathbb{C}^d)$ and any $a \in [0,1]$,*

$$\lambda(\rho \boxplus_a \sigma) \prec a\lambda(\rho) + (1-a)\lambda(\sigma). \tag{5.1}$$

Remark. In the case of the addition rule for fields corresponding to the action of a beamsplitter, a closely related inequality holds when the incoming quantum fields are both Gaussian. For such fields, the field addition rule translates to linearly combining the covariance matrices γ [KS14]:

$$\gamma(\rho \boxplus_a \sigma) = a\gamma(\rho) + (1-a)\gamma(\sigma). \tag{5.2}$$

Denoting by $\nu(A)$ the symplectic eigenvalues of a covariance matrix A, Hiroshima [Hir06] has shown that for any $A, B \geq 0$,

$$\nu(A+B) \prec_w \nu(A) + \nu(B), \tag{5.3}$$

where $\prec_w$ stands for *weak majorization*, which is ordinary majorization (see Definition 1) but without equality for $k = d$. Applied to $\gamma(\rho)$ and $\gamma(\sigma)$, this inequality can be used to derive an EPI for Gaussian fields in a similar way as we have done for qudits.

We will first show how our main result follows from Theorem 11, as this is straightforward, and then proceed with the proof of the latter, which is the bulk of the work. We restate Theorem 4 here, for convenience.

Theorem 4. *For any $f \in \mathcal{F}$ (see Definition 3), density matrices $\rho, \sigma \in \mathcal{D}(\mathbb{C}^d)$, and any $a \in [0,1]$,*

$$f(\rho \boxplus_a \sigma) \geq af(\rho) + (1-a)f(\sigma).$$

Proof. Assume Theorem 11 has been established. Let $\tilde{\rho}, \tilde{\sigma} \in \mathcal{D}(\mathbb{C}^d)$ be diagonal states whose entries are the eigenvalues of ρ and σ (respectively), arranged in non-increasing order. Since $\lambda(\tilde{\rho}) = \lambda(\rho)$ and $\lambda(\tilde{\sigma}) = \lambda(\sigma)$, eq. (5.1) can be equivalently written as

$$\begin{aligned} \lambda(\rho \boxplus_a \sigma) &\prec a\lambda(\tilde{\rho}) + (1-a)\lambda(\tilde{\sigma}), \\ &= \lambda\big(a\tilde{\rho} + (1-a)\tilde{\sigma}\big). \end{aligned} \tag{5.4}$$

For any function $f \in \mathcal{F}$ (see Definition 3) eq. (5.4) implies that

$$\begin{aligned} f(\rho \boxplus_a \sigma) &\geq f\big(a\tilde{\rho} + (1-a)\tilde{\sigma}\big) && (5.5)\\ &\geq af(\tilde{\rho}) + (1-a)f(\tilde{\sigma}) && (5.6)\\ &= af(\rho) + (1-a)f(\sigma), && (5.7) \end{aligned}$$

where the first inequality follows by Schur-concavity, the second inequality follows from concavity, and the last line follows by symmetry. Thus, we have arrived at the statement of Theorem 4. □

It remains to prove Theorem 11. For this we will need the following two lemmas.

Lemma 12 (von Neumann [vN50, p. 55])**.** *Let $\mathcal{L}$ and $\mathcal{M}$ be two subspaces of a vector space and let $P(\mathcal{L})$ and $P(\mathcal{M})$ denote the corresponding projectors. Then*

$$P(\mathcal{L} \cap \mathcal{M}) = \lim_{n\to\infty} \big(P(\mathcal{L})P(\mathcal{M})\big)^n. \tag{5.8}$$

Lemma 13. *For $0 \leq x, y \leq 1$, the following inequality holds:*

$$xy + \sqrt{x(1-x)y(1-y)} \geq \min\{x, y\}. \tag{5.9}$$

Proof. Without loss of generality, we can assume that $0 \leq x \leq y \leq 1$, so we need to show that

$$x \leq xy + \sqrt{x(1-x)y(1-y)}. \tag{5.10}$$

Since $x \leq y$, we have $x - xy \leq y - yx$, or $x(1-y) \leq y(1-x)$. By the above assumption, each side is non-negative. Taking the geometric mean of each side with $x(1-y)$ then yields

$$x(1-y) \leq \sqrt{x(1-y)y(1-x)}, \tag{5.11}$$

which is equivalent to what we had to prove. □

Now we are ready to prove Theorem 11.

Proof of Theorem 11. The expression $\rho \boxplus_a \sigma = a\rho + (1-a)\sigma - \sqrt{a(1-a)}\, i[\rho, \sigma]$ can be written as follows:

$$\rho \boxplus_a \sigma = a(\rho - \rho^2) + (1-a)(\sigma - \sigma^2) + (\sqrt{a}\,\rho + i\sqrt{1-a}\,\sigma)(\sqrt{a}\,\rho + i\sqrt{1-a}\,\sigma)^\dagger. \tag{5.12}$$

It is convenient to express the state $\rho \boxplus_a \sigma$ as $TT^\dagger$ for some 1×3 block-matrix

$$T = \begin{pmatrix} T_1 & T_2 & T_3 \end{pmatrix}. \tag{5.13}$$

We choose $T := A + iB$ where A and B are the following 1×3 block matrices:

$$A := \sqrt{a}\begin{pmatrix}(\rho-\rho^2)^{1/2} & 0 & \rho\end{pmatrix}, \tag{5.14}$$

$$B := \sqrt{1-a}\begin{pmatrix}0 & (\sigma-\sigma^2)^{1/2} & \sigma\end{pmatrix}. \tag{5.15}$$

Here the operator square roots are well-defined, since $X \geq X^2$ for any matrix $I \geq X \geq 0$. Also, note that all blocks of A and B (and hence of T) are Hermitian. One can easily check that

$$AA^\dagger = a(\rho-\rho^2) + a\rho^2 = a\rho, \tag{5.16}$$

$$BB^\dagger = (1-a)(\sigma-\sigma^2) + (1-a)\sigma^2 = (1-a)\sigma, \tag{5.17}$$

$$\begin{aligned} TT^\dagger &= (A+iB)(A^\dagger - iB^\dagger) \\ &= AA^\dagger + BB^\dagger - i(AB^\dagger - BA^\dagger) \\ &= a\rho + (1-a)\sigma - i\sqrt{a(1-a)}[\rho,\sigma] \\ &= \rho \boxplus_a \sigma. \end{aligned} \tag{5.18}$$

Given these expressions, we can rewrite eq. (5.1) as

$$\lambda(TT^\dagger) \prec \lambda(AA^\dagger) + \lambda(BB^\dagger). \tag{5.19}$$

If A and B had been positive semidefinite, this inequality would have followed straight-away from Theorem 3.29 in [Zha02]. Nevertheless, we can adapt the proof of this theorem to our needs. Note that

$$\operatorname{Tr}(TT^\dagger) = \operatorname{Tr}(AA^\dagger) + \operatorname{Tr}(BB^\dagger) - i\operatorname{Tr}[A,B] = \operatorname{Tr}(AA^\dagger) + \operatorname{Tr}(BB^\dagger), \tag{5.20}$$

since $\operatorname{Tr}[A,B] = 0$ by the cyclicity of the trace. Hence,

$$\sum_{j=1}^{d} \lambda_j(TT^\dagger) = \sum_{j=1}^{d} \lambda_j(AA^\dagger) + \sum_{j=1}^{d} \lambda_j(BB^\dagger). \tag{5.21}$$

From this and Definition 1 we see that eq. (5.19) is equivalent to

$$\sum_{j=d-k+1}^{d} \lambda_j(TT^\dagger) \geq \sum_{j=d-k+1}^{d} \lambda_j(AA^\dagger) + \sum_{j=d-k+1}^{d} \lambda_j(BB^\dagger), \qquad \forall\, k \in \{1,\ldots,d\}. \tag{5.22}$$

The left-hand side of the above inequality can be expressed variationally as follows (see e.g. Corollary 4.3.39 in [HJ12]):

$$\sum_{j=d-k+1}^{d} \lambda_j(TT^\dagger) = \min\bigl\{\operatorname{Tr}(U_k^\dagger TT^\dagger U_k) : U_k \in M_{d,k}, U_k^\dagger U_k = I_k\bigr\}, \tag{5.23}$$

where $M_{d,k}$ denotes the set of $d \times k$ matrices, and $I_k \in M_{k,k}$ is the identity matrix. Note that the constraint $U_k^\dagger U_k = I_k$ is equivalent to U_k being a $d \times k$ matrix consisting of k columns of a $d \times d$ unitary matrix U. We can express U_k as $U_k = UI_{k,d}$, where $I_{k,d} := I_k \oplus 0_{d-k}$, with $0_k \in M_{d-k,d-k}$ being a matrix with all entries equal to zero. Hence, $U_k U_k^\dagger = UI_{k,d}U^\dagger$, which is a projector of rank k.

Clearly, $U_k U_k^\dagger \leq I_d$, so that

$$\begin{aligned}
\mathrm{Tr}(U_k^\dagger T T^\dagger U_k) &= \sum_{l=1}^{3} \mathrm{Tr}(U_k^\dagger T_l T_l^\dagger U_k) \\
&\geq \sum_{l=1}^{3} \mathrm{Tr}(U_k^\dagger T_l U_k\, U_k^\dagger T_l^\dagger U_k) \\
&= \sum_{l=1}^{3} \mathrm{Tr}\big(U_k^\dagger (A_l + iB_l) U_k\, U_k^\dagger (A_l^\dagger - iB_l^\dagger) U_k\big) \\
&= \sum_{l=1}^{3} \mathrm{Tr}(U_k^\dagger A_l U_k)^2 + \sum_{l=1}^{3} \mathrm{Tr}(U_k^\dagger B_l U_k)^2 - i \sum_{l=1}^{3} \mathrm{Tr}[U_k^\dagger A_l U_k, U_k^\dagger B_l U_k], \\
&= \sum_{l=1}^{3} \mathrm{Tr}(U_k^\dagger A_l U_k)^2 + \sum_{l=1}^{3} \mathrm{Tr}(U_k^\dagger B_l U_k)^2
\end{aligned} \tag{5.24}$$

where we used that $A_l^\dagger = A_l$ and $B_l^\dagger = B_l$ for all l.

To complete the proof of eq. (5.22), we will show that $\sum_{l=1}^{3} \mathrm{Tr}(U_k^\dagger A_l U_k)^2 \geq \sum_{j=d-k+1}^{d} \lambda_j(AA^\dagger)$, with a corresponding inequality for B following in the same way. From the definition of A we have

$$\sum_{l=1}^{3} \mathrm{Tr}(U_k^\dagger A_l U_k)^2 = a\Big(\mathrm{Tr}\big(U_k^\dagger (\rho - \rho^2)^{1/2} U_k\big)^2 + \mathrm{Tr}(U_k^\dagger \rho U_k)^2\Big). \tag{5.25}$$

Recall from eq. (5.16) that $AA^\dagger = a\rho$. Therefore, we have to show that

$$\mathrm{Tr}\big(U_k^\dagger (\rho - \rho^2)^{1/2} U_k\big)^2 + \mathrm{Tr}(U_k^\dagger \rho U_k)^2 \geq \sum_{j=d-k+1}^{d} \lambda_j(\rho), \qquad \forall\, k \in \{1, \ldots, d\}. \tag{5.26}$$

Let $\rho = \sum_{i=1}^{d} \lambda_i |\psi_i\rangle\langle\psi_i|$ be the eigenvalue decomposition of ρ, with the eigenvalues λ_i being arranged in non-increasing order:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d. \tag{5.27}$$

Then the right-hand side of eq. (5.26) is $\sum_{j=d-k+1}^{d} \lambda_j$ while the left-hand side is

$$\mathrm{Tr}\left(\sum_{i=1}^{d} \sqrt{\lambda_i(1-\lambda_i)}\, U_k^\dagger |\psi_i\rangle\langle\psi_i| U_k\right)^2 + \mathrm{Tr}\left(\sum_{i=1}^{d} \lambda_i\, U_k^\dagger |\psi_i\rangle\langle\psi_i| U_k\right)^2. \tag{5.28}$$

Expanding the squares gives

$$\sum_{i,j=1}^{d} \Big(\sqrt{\lambda_i(1-\lambda_i)\lambda_j(1-\lambda_j)} + \lambda_i\lambda_j\Big)\, \mathrm{Tr}\Big(U_k^\dagger |\psi_i\rangle\langle\psi_i| U_k U_k^\dagger |\psi_j\rangle\langle\psi_j| U_k\Big). \tag{5.29}$$

Noting that

$$C_{ij} := \mathrm{Tr}\Big(U_k^\dagger |\psi_i\rangle\langle\psi_i| U_k U_k^\dagger |\psi_j\rangle\langle\psi_j| U_k\Big) = \big|\langle\psi_i| U_k U_k^\dagger |\psi_j\rangle\big|^2 \tag{5.30}$$

is a non-negative real quantity, we can use Lemma 13 to show that the expression (5.29), and hence the left-hand side of eq. (5.26), is bounded below by

$$\sum_{i,j=1}^{d} \min\{\lambda_i, \lambda_j\}\, C_{ij}. \tag{5.31}$$

Let Λ be the matrix whose elements are $\Lambda_{ij} := \min\{\lambda_i, \lambda_j\}$:

$$\Lambda = \begin{pmatrix} \lambda_1 & \lambda_2 & \lambda_3 & \cdots & \lambda_d \\ \lambda_2 & \lambda_2 & \lambda_3 & \cdots & \lambda_d \\ \lambda_3 & \lambda_3 & \lambda_3 & \cdots & \lambda_d \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_d & \lambda_d & \lambda_d & \cdots & \lambda_d \end{pmatrix}. \tag{5.32}$$

For $m \in \{1, \dots, d\}$, we define matrices E_m of size $d \times d$ such that

$$(E_m)_{ij} := \begin{cases} 1, & \text{for } 1 \le i, j \le m \\ 0, & \text{otherwise.} \end{cases} \tag{5.33}$$

Then we can write

$$\Lambda = \lambda_d E_d + \sum_{m=1}^{d-1} (\lambda_m - \lambda_{m+1}) E_m. \tag{5.34}$$

Hence,

$$\sum_{i,j=1}^{d} \min\{\lambda_i, \lambda_j\}\, C_{ij} \equiv \sum_{i,j=1}^{d} \Lambda_{ij} C_{ij} = \lambda_d \sum_{i,j=1}^{d} (E_d \circ C)_{ij} + \sum_{m=1}^{d-1} (\lambda_m - \lambda_{m+1}) \sum_{i,j=1}^{d} (E_m \circ C)_{ij}, \tag{5.35}$$

where we use the notation $(A \circ B)_{ij} := A_{ij} B_{ij}$ for $d \times d$ matrices A and B.

If we define $\pi(m) := \sum_{i,j=1}^{d} (E_m \circ C)_{ij} = \sum_{i,j=1}^{m} C_{ij}$, we can write eq. (5.35) as

$$\sum_{i,j=1}^{d} \min\{\lambda_i, \lambda_j\}\, C_{ij} = \lambda_d \pi(d) + \sum_{m=1}^{d-1} (\lambda_m - \lambda_{m+1}) \pi(m). \tag{5.36}$$

Recall from eq. (5.27) that the eigenvalues λ_i are arranged in non-increasing order, so all coefficients λ_d and $\lambda_m - \lambda_{m+1}$ are non-negative, so it only remains to find a lower bound on $\pi(m)$.

Recall from eq. (5.30) that

$$\sum_{i,j=1}^{m} C_{ij} = \sum_{i,j=1}^{m} \mathrm{Tr}\Big(U_k^\dagger |\psi_i\rangle\langle\psi_i| U_k U_k^\dagger |\psi_j\rangle\langle\psi_j| U_k \Big) \tag{5.37}$$

$$= \mathrm{Tr}\big(U_k^\dagger Q_m U_k U_k^\dagger Q_m U_k \big) \tag{5.38}$$

$$= \mathrm{Tr}(P_k Q_m)^2, \tag{5.39}$$

where $P_k := U_k U_k^\dagger$ and $Q_m := \sum_{i=1}^{m} |\psi_i\rangle\langle\psi_i|$ are rank-k and rank-m projectors, respectively. Note that $\mathrm{Tr}(P_k Q_m)^n$ is monotonically decreasing as a function of $n \in \mathbb{N}$, so

$$\mathrm{Tr}(P_k Q_m)^2 \ge \lim_{n \to \infty} \mathrm{Tr}(P_k Q_m)^n = \mathrm{Tr} \lim_{n \to \infty} (P_k Q_m)^n = \mathrm{Tr}\, R \tag{5.40}$$

where $R := \lim_{n \to \infty} (P_k Q_m)^n$. If $\mathcal{S}_k$ and $\mathcal{S}_m$ are the subspaces of $\mathbb{C}^d$ corresponding to projectors P_k and Q_m respectively, then, by Lemma 12, R is the projector onto $\mathcal{S}_k \cap \mathcal{S}_m$. Since $\dim \mathcal{S}_k = k$ and $\dim \mathcal{S}_m = m$, we get

$$\mathrm{Tr}\, R = \dim(\mathcal{S}_k \cap \mathcal{S}_m) \ge \max\{0, k + m - d\}. \tag{5.41}$$

Putting everything together, we obtain

$$\pi(m) = \sum_{i,j=1}^{m} C_{ij} \geq \max\{0, k + m - d\}. \tag{5.42}$$

When we substitute this in eq. (5.36), we get

$$\sum_{i,j=1}^{d} \min\{\lambda_i, \lambda_j\} C_{ij} \geq \lambda_d k + \sum_{m=d-k+1}^{d-1} (\lambda_m - \lambda_{m+1})(k + m - d). \tag{5.43}$$

The right-hand side of the above inequality is simply equal to

$$(\lambda_{d-k+1} - \lambda_{d-k+2}) + 2(\lambda_{d-k+2} - \lambda_{d-k+3}) + \cdots + (k-1)(\lambda_{d-1} - \lambda_d) + k\lambda_d = \sum_{j=d-k+1}^{d} \lambda_j, \tag{5.44}$$

which proves eq. (5.26) and therefore the theorem. □

6 Concavity of entropy power and entropy photon number

In this section we prove Theorem 6, which establishes concavity of the entropy power E_c and the entropy photon number N_c for qudits (see Definition 5). Note that both $E_c(\rho)$ and $N_c(\rho)$ are twice-differentiable and monotonously increasing functions of the von Neumann entropy $H(\rho)$. Hence, our strategy for establishing Theorem 6 is to solve the following more general problem.

Problem. *Let $h : \mathbb{R} \to \mathbb{R}$ be any twice-differentiable and monotonously increasing function. For which values of $c \geq 0$ is $f_c(\rho) := h(cH(\rho))$ concave on the set of d-dimensional quantum states?*

Since $H(\rho)$ is already concave, the function $f_c(\rho) = h(cH(\rho))$ is guaranteed to be concave for any $c \geq 0$ whenever h is monotonously increasing *and* concave. However, there are many more functions h which are not necessarily be concave—in fact, they could even be convex—yet produce a concave function f_c for a limited range of constants c. Our goal is to obtain a condition on pairs (h, c) under which the function f_c is concave.

To prove the concavity of f_c on $\mathcal{D}(\mathbb{C}^d)$, we fix any[3] two states $\rho, \sigma \in \mathcal{D}(\mathbb{C}^d)$ and define a function $v : [0, 1] \to \mathbb{R}$ as follows:

$$v(p) := f_c\big(p\rho + (1-p)\sigma\big) \tag{6.1}$$

(note that $v(p)$ implicitly depends also on c). Our goal now is to determine the range of values of c for which

$$v''(p) \leq 0 \quad \forall\, p \in [0, 1] \text{ and } \forall\, \rho, \sigma \in \mathcal{D}(\mathbb{C}^d). \tag{6.2}$$

This would imply that $v(p)$ is concave and, in particular, that $v(p) \geq pv(1) + (1-p)v(0)$, which by eq. (6.1) is equivalent to concavity of f_c. The following lemma uses this approach to obtain the desired condition on (h, c).

Lemma 14. *Let $h : \mathbb{R} \to \mathbb{R}$ be any twice-differentiable, monotonously increasing function. Then the function $f_c(\rho) := h(cH(\rho))$ with $c \geq 0$ and $\rho \in \mathcal{D}(\mathbb{C}^d)$ is concave on the set of quantum states $\mathcal{D}(\mathbb{C}^d)$ if, for any probability distribution $q = (q_1, \ldots, q_d)$, the following condition is satisfied:*

$$c\,\frac{h''(cH(q))}{h'(cH(q))} \leq \frac{1}{V(q) - H(q)^2} \tag{6.3}$$

where $H(q) = -\sum_{i=1}^{d} q_i \log q_i$ is the Shannon entropy of q and $V(q) := \sum_{i=1}^{d} q_i (\log q_i)^2$.

[3] We assume without loss of generality that ρ and σ have full rank (the general case follows by continuity).

To prove this lemma we employ the following definitions and results from [Aud14]. For operators $A, \Delta \in \mathcal{L}(\mathcal{H})$, where $A > 0$ and Δ is Hermitian, the Fréchet derivative of the operator logarithm is given by the linear, completely positive map $\Delta \mapsto \mathcal{T}_A(\Delta)$ [Aud14], where

$$\mathcal{T}_A(\Delta) := \frac{d}{dt}\bigg|_{t=0} \log(A + t\Delta), \tag{6.4}$$

$$= \int_0^\infty ds (A + sI)^{-1} \Delta (A + sI)^{-1}. \tag{6.5}$$

Here the second line follows from the integral representation of the operator logarithm,

$$\log A = \int_0^\infty ds \left(\frac{1}{1+s} I - (A + sI)^{-1} \right) \quad \text{for any } A > 0, \tag{6.6}$$

and the fact that

$$\frac{d}{dt}(A + t\Delta)^{-1} = -(A + t\Delta)^{-1} \Delta (A + t\Delta)^{-1}. \tag{6.7}$$

When A and Δ commute, the integral in eq. (6.5) can be worked out and we get $\mathcal{T}_A(\Delta) = \Delta A^{-1}$.

It is easy to check that the map $\mathcal{T}_A(\Delta)$ is self-adjoint, i.e., for any $B \in \mathcal{L}(\mathcal{H})$,

$$\operatorname{Tr}\big(B \mathcal{T}_A(\Delta)\big) = \operatorname{Tr}\big(\mathcal{T}_A(B) \Delta\big), \tag{6.8}$$

and that

$$\mathcal{T}_A(A) = I. \tag{6.9}$$

This linear map induces a metric on the space of Hermitian matrices given by

$$M_A(\Delta) := \operatorname{Tr}\big(\Delta \mathcal{T}_A(\Delta)\big). \tag{6.10}$$

This metric is known to be monotone [Aud14]; that is, for any completely positive trace-preserving linear map Λ,

$$M_{\Lambda(A)}\big(\Lambda(\Delta)\big) \le M_A(\Delta). \tag{6.11}$$

Now we are ready to prove Lemma 14. Our proof will proceed in two steps: first we will reduce the problem from general quantum states to commuting ones, and then restate the concavity condition for commuting states in terms of a similar condition for probability distributions.

Proof of Lemma 14. Let $\Delta := \rho - \sigma$ and $\xi := p\rho + (1-p)\sigma = \sigma + p\Delta$. Note that $\xi' := \frac{d}{dp}\xi = \Delta$ and $\xi'' = 0$. Recall from eq. (6.2) that concavity of f_c is equivalent to $v''(p) \le 0$ where

$$v(p) := f_c(\xi) = h(cH(\xi)). \tag{6.12}$$

To compute $v''(p)$, we will need to find the first two derivatives of $H(\xi) = -\operatorname{Tr}(\xi \log \xi)$ with respect to p. Noting that

$$\frac{d}{dp} \log \xi = \mathcal{T}_\xi(\xi') \tag{6.13}$$

and using eq. (6.13), we find that the first derivative of $H(\xi)$ is

$$\begin{aligned} \frac{d}{dp} H(\xi) &= -\operatorname{Tr}\big(\xi' \log \xi\big) - \operatorname{Tr}\big(\xi \mathcal{T}_\xi(\xi')\big) \\ &= -\operatorname{Tr}\big(\xi' \log \xi\big) - \operatorname{Tr}\big(\mathcal{T}_\xi(\xi)\xi'\big) \\ &= -\operatorname{Tr}\big(\xi' \log \xi\big) - \operatorname{Tr} \xi' \\ &= -\operatorname{Tr}\big(\xi' \log \xi\big). \end{aligned} \tag{6.14}$$

In the first line we used the Fréchet derivative of the logarithm as given in eq. (6.13), while the second line follows from the self-adjointness (6.8) of the map $\mathcal{T}_\xi$. The last two lines follow from eq. (6.9) and the fact that $\operatorname{Tr}\xi' = \operatorname{Tr}\Delta = 0$. The second derivative is

$$\frac{d^2}{dp^2}H(\xi) = -\operatorname{Tr}\big(\xi'' \log \xi\big) - \operatorname{Tr}\big(\xi' \mathcal{T}_\xi(\xi')\big) = -M_\xi(\xi'), \tag{6.15}$$

where the first term vanishes since $\xi'' = 0$ while the second term produces $M_\xi(\xi')$ by eq. (6.10).

We are now ready to calculate the second derivative of $v(p) = h(cH(\xi))$ introduced in eq. (6.12). By the chain rule,

$$v'(p) = ch'\big(cH(\xi)\big)\frac{dH(\xi)}{dp}, \tag{6.16}$$

$$v''(p) = c^2 h''\big(cH(\xi)\big)\left[\frac{dH(\xi)}{dp}\right]^2 + ch'\big(cH(\xi)\big)\frac{d^2H(\xi)}{dp^2}. \tag{6.17}$$

Therefore, $v''(p) \leq 0$ is equivalent to

$$ch''\big(cH(\xi)\big)\big[\operatorname{Tr}(\xi' \log \xi)\big]^2 \leq h'\big(cH(\xi)\big)M_\xi(\xi'), \tag{6.18}$$

where we divided by $c > 0$ (the case $c = 0$ is trivial) and substituted the derivatives of $H(\xi)$ from eqs. (6.14) and (6.15). Since we imposed the condition that h is monotonously increasing, we can divide by h' and get the condition

$$c\,\frac{h''\big(cH(\xi)\big)}{h'\big(cH(\xi)\big)} \leq \frac{M_\xi(\Delta)}{\big[\operatorname{Tr}(\Delta \log \xi)\big]^2}. \tag{6.19}$$

By fixing the state ξ and minimizing the right-hand side over all Δ, we get a stronger inequality, which in particular implies eq. (6.19). Consider the dephasing channel $\Lambda := \operatorname{diag}_\xi$ which, when acting on an operator Δ, sets all its off-diagonal elements equal to 0 in any basis in which ξ is diagonal (in particular, in its eigenbasis). Thus, $\operatorname{diag}_\xi(\xi) = \xi$ and

$$M_\xi\big(\operatorname{diag}_\xi(\Delta)\big) \leq M_\xi(\Delta), \tag{6.20}$$

by the monotonicity property (6.11) of the metric $M_\xi(\Delta)$ under CPTP maps. Hence, on replacing Δ by $\operatorname{diag}_\xi(\Delta)$ on the right-hand side of eq. (6.19), the denominator remains the same but the numerator does not increase. Since $[\operatorname{diag}_\xi(\Delta), \xi] = 0$, to obtain the minimum value of the right-hand side of eq. (6.19), it therefore suffices to restrict to those Δ which commute with ξ.

Recall that $\mathcal{T}_\xi(\Delta) = \Delta\xi^{-1}$ for commuting ξ and Δ, so

$$M_\xi(\Delta) = \operatorname{Tr}\big(\Delta\mathcal{T}_\xi(\Delta)\big) = \sum_{i=1}^{d} \delta_i^2/\xi_i \tag{6.21}$$

where ξ_i and δ_i for $i \in \{1, \ldots, d\}$ are the diagonal elements of ξ and Δ in the eigenbasis of ξ (in fact, ξ_i are the eigenvalues of ξ). We can now phrase the problem of minimizing the right-hand side of eq. (6.19) as follows:

$$\text{minimize} \quad \frac{\sum_{i=1}^{d} \delta_i^2/\xi_i}{\big(\sum_{i=1}^{d} \delta_i \log \xi_i\big)^2} \quad \text{subject to} \quad \sum_{i=1}^{d} \delta_i = 0, \tag{6.22}$$

where the condition $\sum_{i=1}^{d} \delta_i = 0$ arises from the fact that $\operatorname{Tr} \Delta = 0$.

Since the objective function in eq. (6.22) is invariant under scaling of all δ_i by the same scale factor, we can convert the minimization problem to the following one:

$$\text{minimize} \quad \sum_{i=1}^{d} \delta_i^2/\xi_i \quad \text{subject to} \quad \sum_{i=1}^{d} \delta_i = 0 \ \text{ and } \sum_{i=1}^{d} \delta_i \log \xi_i = 1. \tag{6.23}$$

Using the method of Lagrange multipliers, we form the Lagrangian

$$\mathcal{L} := \sum_{i=1}^{d} \delta_i^2/\xi_i - 2\lambda \sum_{i=1}^{d} \delta_i - 2\mu \left(\sum_{i=1}^{d} \delta_i \log \xi_i - 1 \right). \tag{6.24}$$

To find its stationary points, we require that $\partial \mathcal{L} / \partial \delta_i = 0$ for all i. This implies

$$\delta_i = \xi_i(\lambda + \mu \log \xi_i). \tag{6.25}$$

To find the Lagrange multipliers λ and μ, we substitute the δ_i back into the constraints of the optimization problem (6.23). We get the following equations:

$$\lambda - \mu H = 0, \qquad -\lambda H + \mu V = 1, \tag{6.26}$$

where $H := -\sum_{i=1}^{d} \xi_i \log \xi_i$ and $V := \sum_{i=1}^{d} \xi_i (\log \xi_i)^2$. Their solution is

$$\mu = \frac{1}{V - H^2}, \qquad \lambda = \frac{H}{V - H^2}. \tag{6.27}$$

Inserting eqs. (6.25) and (6.27) back in the objective function of eq. (6.23) yields

$$\sum_{i=1}^{d} \delta_i^2/\xi_i = \sum_{i=1}^{d} \xi_i(\lambda + \mu \log \xi_i)^2 \tag{6.28}$$

$$= \lambda^2 - 2\lambda\mu H + \mu^2 V \tag{6.29}$$

$$= \frac{1}{V - H^2}. \tag{6.30}$$

Thus, eq. (6.19) is satisfied whenever

$$c \, \frac{h''(cH)}{h'(cH)} \leq \frac{1}{V - H^2}. \tag{6.31}$$

Note that $H = H(q)$ and $V = V(q)$ where $q := (\xi_1, \dots, \xi_d)$ is a probability distribution. Thus, condition (6.3) implies eq. (6.31) and hence the concavity of f_c. □

To find the optimal value of c for which eq. (6.31) holds, we need to minimize the right-hand side of this inequality over all attainable values of the quantity[4] $V - H^2$ for a fixed value of H. In other words, we require the maximum attainable value of $V(q) - H(q)^2$ over all probability distributions q over d elements with a fixed value of the entropy $H(q) = H_0$. We define

$$V_{\max}(H_0) := \max\Big\{ V(q) : H(q) = H_0, \sum_{i=1}^{d} q_i = 1, \text{ and } q_i \geq 0 \text{ for all } i \Big\}. \tag{6.32}$$

To obtain this value and the corresponding optimal distribution q, we employ the following lemma.

[4]The quantity $V - H^2$ is also known as information variance.

Lemma 15. *The maximum of $V(q) := \sum_{i=1}^{d} q_i(\log q_i)^2$ over all probability distributions $q = (q_1, \ldots, q_d)$ with fixed Shannon entropy $H(q) = H_0 \in [0, \log d]$ is achieved by a distribution of the form*

$$q = (x, \ldots, x, y) \text{ for some } 0 \leq x < y \text{ such that } (d-1)x + y = 1. \tag{6.33}$$

If we let $r := d-1$, then the value of $V(q)$ achieved by this distribution is

$$V_{\max}(H_0) = rx(1-rx)\big(\log x - \log(1-rx)\big)^2 + H_0^2. \tag{6.34}$$

Proof. For given $H_0 \in [0, \log d]$, we need to solve the following constrained optimization problem:

$$\text{maximize} \quad \sum_{i=1}^{d} q_i(\log q_i)^2 \quad \text{subject to} \quad \sum_{i=1}^{d} q_i = 1 \;\text{ and }\; -\sum_{i=1}^{d} q_i \log q_i = H_0. \tag{6.35}$$

Since the domain of the logarithm is $\mathbb{R}^+$, we do not have to explicitly impose the condition that $q_i \geq 0$ for all $1 \leq i \leq d$.

We once again use the Lagrangian method to solve this problem. The Lagrangian is given by

$$\mathcal{L} := \sum_{i=1}^{d} q_i(\log q_i)^2 + \lambda\left(\sum_{i=1}^{d} q_i - 1\right) - \mu\left(\sum_{i=1}^{d} q_i \log q_i + H_0\right). \tag{6.36}$$

Requiring that all derivatives $\partial\mathcal{L}/\partial p_i$ be zero yields the equations

$$(\log q_i)^2 + (2-\mu)\log q_i + \lambda - \mu = 0. \tag{6.37}$$

As this is a fixed *quadratic* function of $\log q_i$, and therefore may have at most two solutions, we infer that the stationary points of $\mathcal{L}$ are those distributions q whose elements are either all equal (and hence equal to $1/d$) or equal to two possible values. That is, up to permutations, the distribution q can be uniquely represented as

$$q_{k,x} := (\underbrace{x, \ldots, x}_{k}, \underbrace{y, \ldots, y}_{d-k}) \tag{6.38}$$

for some integer $k \in \{1, \ldots, d\}$ and some probabilities $0 \leq x < y$ such that

$$kx + (d-k)y = 1. \tag{6.39}$$

From this we get in addition that $x \leq 1/d < y$. For $k = d$, there is only one distribution of this form, namely, the uniform distribution $q_{d,x} = (1/d, \ldots, 1/d)$. This distribution has $H(q_{d,x}) = \log d$ and $V(q_{d,x}) = (\log d)^2$, which are independent of x, so there is nothing to optimize in this case.

From now on we assume that $k \neq d$ and thus $H_0 < \log d$. Then $y := (1-kx)/(d-k)$ from the normalization constraint (6.39), so we can compute

$$H(k,x) := H(q_{k,x}) = -kx\log x - (d-k)y\log y, \tag{6.40}$$

$$V(k,x) := V(q_{k,x}) = kx(\log x)^2 + (d-k)y(\log y)^2. \tag{6.41}$$

To obtain the global maximum of $V(k,x)$, a further optimization over $k \in \{1, \ldots, d-1\}$ and $x \in [0, 1/d)$ is required. The numerical calculations presented in the diagram in Fig. 2 suggest that $k = d-1$ yields the maximal value of V. To prove that this is actually true we will temporarily remove the restriction that k be an integer and consider the entire range $k \in (0, d)$. Our analysis will show that keeping $H(k,x)$ fixed, $V(k,x)$ increases with k.

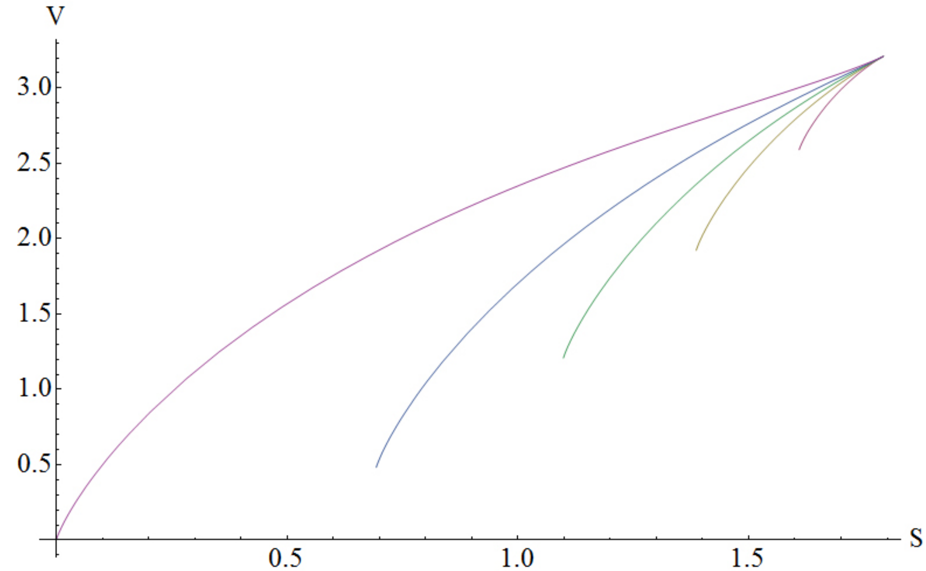

Figure 2: The locus of the points (H, V) as x varies over $[0, 1/d]$, for the case $d = 6$ and for each value of $k \in \{1, \ldots, 5\}$, with k increasing towards the left. These loci are curves with lower end point $H = H(k, 0) = \log(d-k)$, $V = V(k, 0) = (\log(d-k))^2$ and upper end point $H = H(k, 1/d) = \log d$, $V = V(k, 1/d) = (\log d)^2$. The value $k = 0$ yields a single point, just as the $k = d$ case does, coinciding with the upper end point of all other (H, V) curves.

To keep H fixed as k changes, x will have to change as well. For given $H_0 \in [0, \log d)$, let $x(k)$ be the function of k implicitly given by $H(k, x(k)) = H_0$. We would like to know how $V(k, x(k))$ changes as a function of k. Taking the total derivative with respect to k gives

$$\frac{d}{dk} H\big(k, x(k)\big) = \frac{\partial}{\partial k} H\big(k, x(k)\big) + \frac{\partial}{\partial x} H\big(k, x(k)\big)\, x'(k) = 0, \tag{6.42}$$

$$\frac{d}{dk} V\big(k, x(k)\big) = \frac{\partial}{\partial k} V\big(k, x(k)\big) + \frac{\partial}{\partial x} V\big(k, x(k)\big)\, x'(k). \tag{6.43}$$

Solving the first equation for $x'(k)$ and substituting the solution in the second equation gives

$$\frac{d}{dk} V\big(k, x(k)\big) = \frac{\partial}{\partial k} V\big(k, x(k)\big) - \frac{\partial}{\partial x} V\big(k, x(k)\big)\ \frac{\partial}{\partial k} H\big(k, x(k)\big) \Big/ \frac{\partial}{\partial x} H\big(k, x(k)\big) \tag{6.44}$$

$$= \frac{1}{d-k}\left[\Big(1 + (d-2k)x\Big)\Big(\log\frac{1-kx}{d-k} - \log x\Big) - 2(1-dx)\right], \tag{6.45}$$

where the second line follows by substituting the partial derivatives of $H(k, x)$ and $V(k, x)$ defined in eqs. (6.40) and (6.41).

Note that $z \geq \tanh z = (e^{2z} - 1)/(e^{2z} + 1)$ for any $z \geq 0$. By choosing $z := (\log w)/2$ we get $\log w \geq 2(w-1)/(w+1)$ for $w \geq 1$. Next, since $x \leq 1/d$, we can take $w := (1 - kx)/(dx - kx) \geq 1$ which gives

$$\log\frac{1-kx}{d-k} - \log x \geq 2\left(\frac{1-kx}{(d-k)x} - 1\right) \Big/ \left(\frac{1-kx}{(d-k)x} + 1\right) = 2\,\frac{1-dx}{1+(d-2k)x}. \tag{6.46}$$

Inserting this in eq. (6.45) and noting that $1 + (d-2k)x \geq 0$ for $x \leq 1/d$, we conclude that $dV(k, x(k))/dk \geq 0$, so $V(k, x(k))$ is increasing as a function of k just as we intended to show.

Reverting back to integer values of k, we find that, for a fixed value of $H(k,x)$, the value of $V(k,x)$ is maximized when k is the largest integer in the open interval $(0,d)$, namely $d-1$. Then $V_{\max}(H_0)$, the maximum value of $V(k,x)$ subject to $H(k,x)=H_0$, see eq. (6.32), is given by $V(d-1,x)$ where x is such that $H(d-1,x)=H_0$. From eq. (6.41) we infer that

$$V_{\max}(H_0)-H_0^2 = V(d-1,x)-H(d-1,x)^2 = rx(1-rx)\big(\log x-\log(1-rx)\big)^2, \tag{6.47}$$

where $r:=d-1$ and $x\in[0,1/d]$ satisfies $H(d-1,x)=H_0$. □

For $r=d-1$ and any $x\in[0,1/d]$, if $q_{r,x}$ is the probability distribution defined in eq. (6.38), we denote its Shannon entropy and the information variance by

$$s_r(x) := H(q_{r,x}) = -rx\log x-(1-rx)\log(1-rx), \tag{6.48}$$

$$w_r(x) := V(q_{r,x})-H(q_{r,x})^2 = rx(1-rx)\big(\log x-\log(1-rx)\big)^2. \tag{6.49}$$

In terms of these quantities, the condition in Lemma 14, under which a given function of the von Neumann entropy is concave on the set of qudit states, is expressed by the following theorem.

Theorem 16. *Let $h:\mathbb{R}\to\mathbb{R}$ be a twice-differentiable, monotonously increasing function. Then the function $f_c(\rho):=h(cH(\rho))$ with $c\geq 0$ and $\rho\in\mathcal{D}(\mathbb{C}^d)$ is concave on $\mathcal{D}(\mathbb{C}^d)$ if*

$$c\,\frac{h''(cs_r(x))}{h'(cs_r(x))}\leq\frac{1}{w_r(x)}, \tag{6.50}$$

for all $0\leq x\leq 1/d$, where $r:=d-1$ and functions $s_r(x)$ and $w_r(x)$ are defined in eqs. (6.48) *and* (6.49).

6.1 Concavity of Entropy Power

In this section we use Theorem 16 to establish the first item of Theorem 6, namely, that the entropy power $E_c(\rho)=e^{cH(\rho)}$ of a state $\rho\in\mathcal{D}(\mathbb{C}^d)$ is concave for $0\leq c\leq 1/(\log d)^2$.

Proof of Theorem 6 (concavity of E_c). In this case we have $h(x)=\exp(x)$, so the condition (6.50) just translates to

$$c\leq 1/w_{d-1}(x),\quad 0\leq x\leq 1/d. \tag{6.51}$$

Therefore,

$$c\leq\Big(\max_{0\leq x\leq 1/d} w_{d-1}(x)\Big)^{-1} =: c_{\max}. \tag{6.52}$$

From the expression of w_r follows a simple lower bound on the largest allowed value of c. Putting $y=rx$, with $0\leq y\leq(d-1)/d<1$,

$$\begin{aligned} w_r(x) &= y(1-y)\big(-\log r+\log y-\log(1-y)\big)^2 \qquad (6.53)\\ &= y(1-y)(\log r)^2-2y(1-y)\big(\log y-\log(1-y)\big)\log r\\ &\quad + y(1-y)(\log y-\log(1-y))^2. \qquad (6.54)\end{aligned}$$

The coefficients of this quadratic polynomial in $\log r$ are bounded above as $y(1-y)\leq 1/4$, $-2y(1-y)(\log y-\log(1-y))\leq 1/2$ and $y(1-y)(\log y-\log(1-y))^2\leq 1/2$. Hence,

$$w_r(x)\leq(1+\log r)/2+(\log r)^2/4\quad\text{with } r=d-1, \tag{6.55}$$

and we obtain

$$c_{\max}\geq 1/((1+\log(d-1))/2+(\log(d-1))^2/4). \tag{6.56}$$

This bound becomes asymptotically exact in the limit of large d. Note that the right-hand side of eq. (6.56) is smaller than $1/(\log d)^2$. □

From this we can also infer that for any probability distribution p over d elements, the function $E(p) := e^{cH(p)}$ is concave for $0 \le c \le 1/(\log d)^2$.

6.2 Concavity of Entropy Photon Number

In this section we use Theorem 16 to establish the second item of Theorem 6, namely, that the entropy photon number $N_c(\rho)$ of a qudit, defined by eq. (3.4), is concave for $0 \le c \le 1/(d-1)$.

Proof of Theorem 6 (concavity of N_c). In this case the calculations are more complicated because h is not given directly but as the inverse of a function: $h = g^{-1}$, where

$$g(x) = -x\log(x) + (1+x)\log(1+x). \tag{6.57}$$

The derivatives of h are given by

$$h'(x) = \frac{1}{g'(h(x))} = \frac{1}{\log(1+1/h(x))}, \tag{6.58}$$

$$h''(x) = \frac{1}{(h(x)+h^2(x))\left(\log(1+1/h(x))\right)^3}. \tag{6.59}$$

Defining the function

$$k(x) = x(1+x)(\log(x) - \log(1+x))^2, \tag{6.60}$$

we have $h'(x)/h''(x) = k(h(x))$. The function k is monotonously increasing, concave, and ranges from 0 to 1. The condition on c becomes

$$g^{-1}(cs_r(x)) \ge k^{-1}(cw_r(x)). \tag{6.61}$$

If we define the variables y and z according to

$$g(y) = cs_r(x), \quad k(z) = cw_r(x), \tag{6.62}$$

the condition is $y \ge z$. If we now exploit the monotonicity of k, this condition is equivalent to $k(y) \ge k(z) = cw_r(x)$. We therefore require that

$$g(y) = cs_r(x) \ \text{ implies } \ k(y) \ge cw_r(x). \tag{6.63}$$

We will show that this holds for $c \le 1/a = 1/(d-1)$. In Fig. 3 we depict the graph of $k(y)$ versus $g(y)$. The graph seems to indicate that the resulting curve is concave and monotonously increasing; that this is actually true follows from the easily checked fact that the function $k'/g' = (1+2x)(\log(1+x) - \log(x)) - 2$, representing the slope of the curve, is positive and decreasing. The condition eq. (6.63) amounts to the statement that any point $(cs_r(x), cw_r(x))$ lies in the area below this curve. Hence if the condition is satisfied for a certain value of c, then it is also satisfied for any smaller positive value of c. Therefore, we only need to prove eq. (6.63) for $c = 1/(d-1)$.

The formal similarities between g and s_r and between k and w_r allow us to define two interpolating functions $g_1(x,b)$ and $k_1(x,b)$ as a function of the original x and an interpolation parameter b:

$$g_1(x,b) = -x\log x + (1+bx)\frac{\log(1+bx)}{b} \tag{6.64}$$

$$k_1(x,b) = x(1+bx)(\log x - \log(1+bx))^2. \tag{6.65}$$

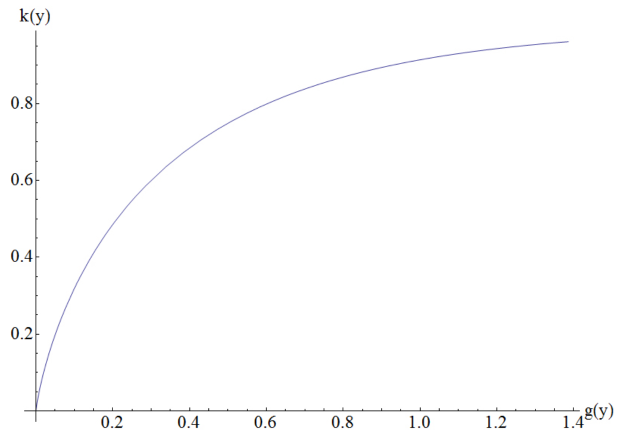

Figure 3: Parametric plot of $k(y)$ versus $g(y)$.

Let $\mathcal{S}$ indicate the domain of g_1 and k_1, which is $b \in [1-d, 1]$ and $x \in [0, 1/d]$, as before. To ensure continuity of g_1 at $b = 0$, we define $g_1(x, 0)$ to be its limit value $(-x \log x + x)$. Hence, we have the correspondences

$$s_r(x)/(d-1) = g_1(x, 1-d), \qquad g(x) = g_1(x, 1), \tag{6.66}$$

$$w_r(x)/(d-1) = k_1(x, 1-d), \qquad k(x) = k_1(x, 1). \tag{6.67}$$

The condition eq. (6.63) is therefore satisfied if a continuous path $x(b)$ exists (from $x(1-d) = x$ to $x(1) = y$) such that $g_1(x(b), b)$ remains constant and $k_1(x(b), b)$ increases with b. As in the proof of Lemma 15 this requires the positivity of

$$\frac{d}{db} k_1(x(b), b) \tag{6.68}$$

$$= \frac{\partial}{\partial b} k_1(x(b), b) - \frac{\partial}{\partial x} k_1(x(b), b) \frac{\partial}{\partial b} g_1(x(b), b) / \frac{\partial}{\partial x} g_1(x(b), b) \tag{6.69}$$

$$= \frac{1}{b^2} \Big(bx \left(2 + \log x + bx(\log x)^2\right) \tag{6.70}$$

$$+ (1+bx)^2 \left(\log(1+bx)\right)^2 \tag{6.71}$$

$$- \left(2 + bx + (1 + 2bx + 2b^2x^2) \log x\right) \log(1+bx) \Big). \tag{6.72}$$

Let us introduce the variable $u = 1 + bx$. In $\mathcal{S}$ we have $(1-b)x \le 1$ so that $x \le u$; furthermore, $b \le 1$ and $x \le 1/d$, so that $u \le 1 + 1/d$. The second factor can now be written more succinctly as

$$\begin{aligned}
&(u-1)\left(2 + \log x + (u-1)(\log x)^2\right) + u^2(\log u)^2 - (1 + u + (1 - 2u + 2u^2) \log x) \log u \\
&= (u-1)^2 \log x (\log x - \log u) + u^2 (\log u)^2 \\
&\quad + 2(u-1) - (u+1) \log u - (1 - u + u^2 \log u) \log x.
\end{aligned} \tag{6.73}$$

The first two terms are clearly non-negative. The factor $1 - u + u^2 \log u$ is non-negative too, as can be seen from the inequality $1 - \exp(-v) \le v \le v \exp(v)$ applied to $v = \log u$. Furthermore, $\log x \le \log(1/d) \le \log(1/2) \le -1/2$, so that the last term is bounded below by $(1 - u + u^2 \log u)/2$. It is therefore left to show that $2(u-1) - (u+1)\log u + (1 - u + u^2 \log u)/2$ is non-negative.

For $0 < u \le 1$ we can exploit the inequality $\log u \le 2(u-1)/(u+1)$, so that $2(u-1) - (u+1)\log u \ge 0$. The remaining term is non-negative too, as we have just showed.

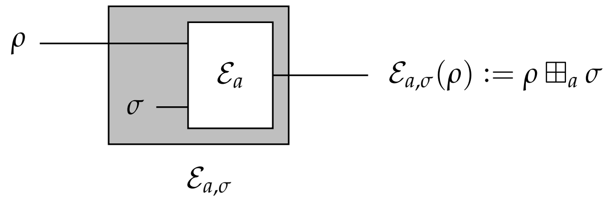

Figure 4: A schematic representation of the channel $\mathcal{E}_{a,\sigma}$ defined in eq. (7.1).

For $1 \leq u \leq 1 + 1/d$ we exploit instead the inequality $\log u \leq u - 1$. Then based on the fact that in this range $(u-1)^2 - 3 < 0$

$$2(u-1) - (u+1)\log u + (1 - u + u^2 \log u)/2 \tag{6.74}$$

$$= \frac{1}{2}(3(u-1) + ((u-1)^2 - 3)\log u) \tag{6.75}$$

$$\geq \frac{1}{2}(3(u-1) + ((u-1)^2 - 3)(u-1)) \tag{6.76}$$

$$= (u-1)^3/2 \geq 0. \tag{6.77}$$

This shows that $k_1(x(b), b)$ indeed increases with b, whence condition (6.63) holds for $c = 1/(d-1)$ and, by a previous argument, for $c \leq 1/(d-1)$. In other words, we have shown that the function $g^{-1}(cH(\rho))$ is concave for $0 < c \leq 1/(d-1)$. As this includes the value $c = 1/d$, the photon number is concave. $\square$

7 Bounds on minimum output entropy and Holevo capacity

As an application of our results we now consider the class of quantum channels $\mathcal{E}_{a,\sigma} : \mathcal{D}(\mathbb{C}^d) \to \mathcal{D}(\mathbb{C}^d)$ obtained from the partial swap channel $\mathcal{E}_a$ from eq. (4.10) by fixing the second input state σ (see Fig. 4). Such channels are parameterized by a variable $a \in [0,1]$ *and* a quantum state $\sigma \in \mathcal{D}(\mathbb{C}^d)$, and act as follows:

$$\mathcal{E}_{a,\sigma}(\rho) := \rho \boxplus_a \sigma. \tag{7.1}$$

For example, for the choice $\sigma = I/d$ (the completely mixed state) the channel $\mathcal{E}_{a,\sigma}$ is just the quantum depolarizing channel with parameter a. If $\sigma = \delta|0\rangle\langle 0| + (1-\delta)|1\rangle\langle 1| \in \mathcal{D}(\mathbb{C}^2)$ for some $\delta \in [0,1]$, then $\mathcal{E}_{a,\sigma}$ is a qubit channel whose output density matrix is

$$\begin{pmatrix} ar_{00} + (1-a)\delta & r_{01}(a - i\sqrt{a(1-a)}(1-2\delta)) \\ r_{01}(a + i\sqrt{a(1-a)}(1-2\delta)) & ar_{11} + (1-a)(1-\delta) \end{pmatrix} \tag{7.2}$$

for any input qubit state $\rho := \sum_{i,j=0}^{1} r_{ij}|i\rangle\langle j|$.

An important characteristic quantity for any quantum channel $\mathcal{E}$ is its *minimum output entropy*, which is defined as

$$H_{\min}(\mathcal{E}) := \min_{\rho} H(\mathcal{E}(\rho)) \tag{7.3}$$

Lower bounds on this quantity for the class of channels $\mathcal{E}_{a,\sigma}$ can be obtained by using our EPIs and EPnI. In fact, the inequalities of Corollary 7 give various lower bounds on the output entropy of the channel $\mathcal{E}_{a,\sigma}$ (i.e. the entropy of any output state) in terms of the entropy $H(\rho)$ of an input state ρ:

$$H(\mathcal{E}_{a,\sigma}(\rho)) \geq aH(\rho) + (1-a)H(\sigma), \tag{7.4}$$

$$H(\mathcal{E}_{a,\sigma}(\rho)) \geq \frac{1}{c}\log\Big[a\exp(cH(\rho)) + (1-a)\exp(cH(\sigma))\Big], \qquad \text{with } c = 1/(\log d)^2, \tag{7.5}$$

$$H(\mathcal{E}_{a,\sigma}(\rho)) \geq \frac{1}{c}g\Big[ag^{-1}(cH(\rho)) + (1-a)g^{-1}(cH(\sigma))\Big], \qquad \text{with } c = 1/(d-1). \tag{7.6}$$

Since the above bounds are of the form $H\left(\mathcal{E}_{a,\sigma}(\rho)\right) \geq G\left(H(\rho)\right)$, for some function G, we have

$$\begin{aligned} H_{\min}\left(\mathcal{E}_{a,\sigma}\right) &\geq \min_{\rho} G\left(H(\rho)\right) \\ &= \min_{0\leq H_0\leq \log d} G(H_0). \end{aligned} \tag{7.7}$$

In Fig. 5 we have plotted the bounds $G(H_0)$ for two illustrative cases, the three curves corresponding to the three choices of the function G as given by the right-hand sides of (7.4)–(7.6).

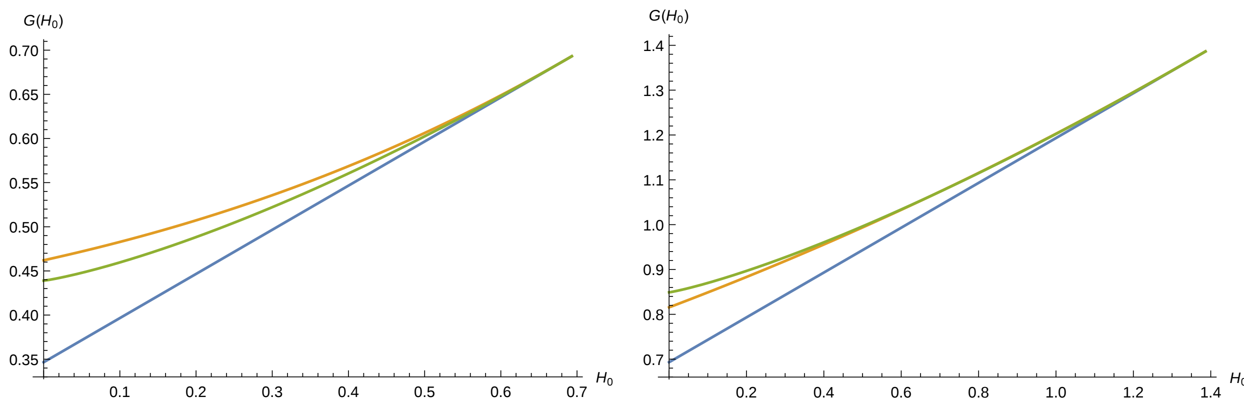

Figure 5: Plots of bounds G for the channel $\mathcal{E}_{1/2,\sigma}(\rho)$, where σ is the maximally mixed state $\sigma = I/d$ in dimensions $d = 2$ (left panel) and $d = 4$ (right panel). The blue curves represent the bound (7.4) obtained from eq. (3.7), the orange curves represent the bound (7.5) obtained from the entropy power inequality eq. (3.8), and the green curves represent the bound (7.6) obtained from the entropy photon number inequality eq. (3.9). For this channel, the numerics suggest that the entropy power inequality yields the sharpest bound for the case $d = 2$, whereas the entropy photon number inequality prevails for $d > 2$.

These bounds imply lower bounds on the minimum output entropy $H_{\min}\left(\mathcal{E}_{a,\sigma}\right)$, which in turn allow us to obtain upper bounds on the *product-state classical capacity* of $\mathcal{E}_{a,\sigma}$. The latter is the capacity evaluated in the limit of asymptotically many independent uses of the channel, under the constraint that the inputs to multiple uses of the channel are necessarily product states. The Holevo-Schumacher-Westmoreland (HSW) [Hol98, SW97] theorem establishes that the product-state capacity of a memoryless quantum channel $\mathcal{E}$ is given by its Holevo capacity $\chi(\mathcal{E})$:

$$\chi(\mathcal{E}) := \max_{\{p_i,\rho_i\}}\left\{H\Big(\sum_i p_i\mathcal{E}(\rho_i)\Big) - \sum_i p_i H\big(\mathcal{E}(\rho_i)\big)\right\}, \tag{7.8}$$

where the maximum is taken over all ensembles $\{p_i, \rho_i\}$ of possible input states ρ_i occurring with probabilities p_i. Using the above expression, and the fact that $H(\omega) \leq \log d$ for any $\omega \in \mathcal{D}(\mathbb{C}^d)$, we obtain the following simple bound:

$$\chi(\mathcal{E}) \leq \log d - \min_{\rho} H\big(\mathcal{E}(\rho)\big), \tag{7.9}$$

where the minimum is taken over all possible inputs to the channel. Applying this bound to the channel $\mathcal{E}_{a,\sigma}$ for any $a \in [0,1]$ and $\sigma \in \mathcal{D}(\mathbb{C}^d)$ and using eq. (7.4) we infer that

$$\begin{aligned}\chi(\mathcal{E}_{a,\sigma}) &\leq \log d - a \min_{\rho} H(\rho) - (1-a)H(\sigma) \\ &= \log d - (1-a)H(\sigma).\end{aligned} \tag{7.10}$$

For the case of the qubit channel introduced above, we thus obtain the bound

$$\chi(\mathcal{E}_{a,\sigma}) \leq \log 2 - (1-a)h(\delta), \tag{7.11}$$

where $h(\delta) := -\delta \log \delta - (1-\delta)\log(1-\delta)$ is the binary entropy. Even sharper bounds are possible by exploiting eqs. (7.5) and (7.6).

8 Summary and open questions

In this paper we establish a class of entropy power inequalities (EPIs) for d-level quantum systems or qudits. The underlying addition rule for which these inequalities hold, is given by a quantum channel acting on the product state $\rho \otimes \sigma$ of two qudits and yielding the state of a single qudit as output. We refer to this channel as a *partial swap channel* since its output interpolates between the states ρ and σ as the parameter a on which it depends is changed from 1 to 0. We establish EPIs not only for the von Neumann entropy and the entropy power, but also for a large class of functions, which include the Rényi entropies and the subentropy. Moreover, for the subclass of partial swap channels for which one of the qudit states in the input is fixed, our EPI for the von Neumann entropy yields an upper bound on the Holevo capacity.

We would like to emphasize that the method that we employ to prove our EPIs is novel, in the sense that it doesn't mimic the proofs of the EPIs in the continuous-variable classical and quantum settings. Instead it relies solely on spectral majorization and concavity of certain functions.

8.1 Open questions

Our results lead to many interesting open questions. Here we briefly mention some of them. For example, can a conditional version of the EPI (see [Koe13]) be proved for qudits? Is it possible to generalize our quantum addition rule (4.18) for combining more than two states? If such a generalization exists, does our EPI admit a multi-mode generalization as well (see [DMLG14])? Is the partial swap channel that we define the unique channel resulting in an interpolation between the input states and yielding a non-trivial EPI (i.e., one that is not simply a statement of concavity)?

In Section 7, we mentioned a simple application of our EPI to quantum Shannon theory. Considering the significance of the classical EPI in information theory and statistics, we expect that our EPIs will also find further applications.

Finally, it would be worth exploring whether our proof of the qudit analogue of the entropy photon number inequality can be generalized to establish the EPnI for the bosonic case (which is known to be an important open problem).

Acknowledgements

We would like to thank Jianxin Chen, Robert König and Will Matthews for useful discussions. KA acknowledges support by an Odysseus Grant of the Flemish FWO. MO acknowledges financial support from European Union under project QALGO (Grant Agreement No. 600700).

References

[Aud14] Koenraad M.R. Audenaert. Quantum skew divergence. *Journal of Mathematical Physics*, 55(11):112202, 2014. `arXiv:1304.5935`, `doi:10.1063/1.4901039`.

[Bar86] Andrew R. Barron. Entropy and the central limit theorem. *The Annals of Probability*, 14(1):336–342, 1986. URL: `http://projecteuclid.org/euclid.aop/1176992632`.

[Bec75] William Beckner. Inequalities in Fourier analysis. *Annals of Mathematics*, 102(1):159–182, Jul 1975. URL: `http://www.jstor.org/stable/1970980`.

[Ber74] Patrick P. Bergmans. A simple converse for broadcast channels with additive white Gaussian noise. *Information Theory, IEEE Transactions on*, 20(2):279–280, Mar 1974. `doi:10.1109/TIT.1974.1055184`.

[Bha97] Rajendra Bhatia. *Matrix Analysis*. Graduate Texts in Mathematics. Springer New York, 1997. URL: `https://books.google.com/books?id=F4hRy1F1M6QC`.

[Bla65] Nelson M. Blachman. The convolution inequality for entropy powers. *Information Theory, IEEE Transactions on*, 11(2):267–271, Apr 1965. `doi:10.1109/TIT.1965.1053768`.

[DCT91] Amir Dembo, Thomas M. Cover, and Joy A. Thomas. Information theoretic inequalities. *Information Theory, IEEE Transactions on*, 37(6):1501–1518, Nov 1991. `doi:10.1109/18.104312`.

[DDJB14] Nilanjana Datta, Tony Dorlas, Richard Jozsa, and Fabio Benatti. Properties of subentropy. *Journal of Mathematical Physics*, 55(6):062203, 2014. `arXiv:1310.1312`, `doi:10.1063/1.4882935`.

[DMG14] Giacomo De Palma, Andrea Mari, and Vittorio Giovannetti. A generalization of the entropy power inequality to bosonic quantum systems. *Nature Photonics*, 8(12):958–964, 2014. `arXiv:1402.0404`, `doi:10.1038/nphoton.2014.252`.

[DMLG14] Giacomo De Palma, Andrea Mari, Seth Lloyd, and Vittorio Giovannetti. The multimode quantum entropy power inequality. 2014. `arXiv:1408.6410`.

[GES08] Saikat Guha, Baris I. Erkmen, and Jeffrey H. Shapiro. The entropy photon-number inequality and its consequences. In *Information Theory and Applications Workshop*, pages 128–130, Jan 2008. `arXiv:0710.5666`, `doi:10.1109/ITA.2008.4601037`.

[GGL+04] Vittorio Giovannetti, Saikat Guha, Seth Lloyd, Lorenzo Maccone, Jeffrey H. Shapiro, and Horace P. Yuen. Classical capacity of the lossy bosonic channel: the exact solution. *Phys. Rev. Lett.*, 92(2):027902, Jan 2004. `arXiv:quant-ph/0308012`, `doi:10.1103/PhysRevLett.92.027902`.

[GSE07] Saikat Guha, Jeffrey H. Shapiro, and Baris I. Erkmen. Classical capacity of bosonic broadcast communication and a minimum output entropy conjecture. *Phys. Rev. A*, 76(3):032303, Sep 2007. `arXiv:0706.3416`, `doi:10.1103/PhysRevA.76.032303`.

[GSE08] Saikat Guha, Jeffrey H. Shapiro, and Baris I. Erkmen. Capacity of the bosonic wiretap channel and the entropy photon-number inequality. In *IEEE International Symposium on Information Theory, 2008 (ISIT 2008)*, pages 91–95, Jul 2008. `arXiv:0801.0841`, `doi:10.1109/ISIT.2008.4594954`.

[Guh08] Saikat Guha. *Multiple-user quantum information theory for optical communication channels*. PhD thesis, Dept. Electr. Eng. Comput. Sci., MIT, Cambridge, MA, USA, 2008. URL: `http://hdl.handle.net/1721.1/44413`.

[Hir06] Tohya Hiroshima. Additivity and multiplicativity properties of some Gaussian channels for Gaussian inputs. *Phys. Rev. A*, 73(1):012330, Jan 2006. `arXiv:quant-ph/0511006`, `doi:10.1103/PhysRevA.73.012330`.

[HJ12] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 2nd edition, 2012.

[Hol98] Alexander S. Holevo. The capacity of the quantum channel with general signal states. *Information Theory, IEEE Transactions on*, 44(1):269–273, Jan 1998. `arXiv:quant-ph/9611023`, `doi:10.1109/18.651037`.

[HV03] Peter Harremoës and Christophe Vignat. An entropy power inequality for the binomial family. *Journal of Inequalities in Pure and Applied Mathematics*, 4(5):93, 2003. URL: `http://www.emis.ams.org/journals/JIPAM/article334.html?sid=334`.

[JRW94] Richard Jozsa, Daniel Robb, and William K. Wootters. Lower bound for accessible information in quantum mechanics. *Phys. Rev. A*, 49(2):668–677, Feb 1994. `doi:10.1103/PhysRevA.49.668`.

[JY10] Oliver Johnson and Yaming Yu. Monotonicity, thinning, and discrete versions of the entropy power inequality. *Information Theory, IEEE Transactions on*, 56(11):5387–5395, Nov 2010. `arXiv:0909.0641`, `doi:10.1109/TIT.2010.2070570`.

[KMN+07] Pieter Kok, William J. Munro, Kae Nemoto, Timothy C. Ralph, Jonathan P. Dowling, and Gerard J. Milburn. Linear optical quantum computing with photonic qubits. *Rev. Mod. Phys.*, 79(1):135–174, Jan 2007. `arXiv:quant-ph/0512071`, `doi:10.1103/RevModPhys.79.135`.

[Koe13] Robert Koenig. The conditional entropy power inequality for Gaussian quantum states. 2013. `arXiv:1304.7031`.

[KS13] Robert König and Graeme Smith. Limits on classical communication from quantum entropy power inequalities. *Nature Photonics*, 7(2):142–146, 2013. `arXiv:1205.3407`, `doi:10.1038/nphoton.2012.342`.

[KS14] Robert König and Graeme Smith. The entropy power inequality for quantum systems. *Information Theory, IEEE Transactions on*, 60(3):1536–1548, Mar 2014. `arXiv:1205.3409`, `doi:10.1109/TIT.2014.2298436`.

[Lie78] Elliott H. Lieb. Proof of an entropy conjecture of Wehrl. *Communications in Mathematical Physics*, 62(1):35–41, 1978. URL: http://projecteuclid.org/euclid.cmp/1103904300, doi:10.1007/BF01940328.

[Rén56] Alfréd Rényi. A characterization of Poisson processes. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 1(2):519–527, 1956.

[Rén61] Alfréd Rényi. On measures of entropy and information. In Jerzy Neyman, editor, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pages 547–561, Berkeley, California, 1961. University of California Press. URL: http://projecteuclid.org/euclid.bsmsp/1200512181.

[Rio11] Olivier Rioul. Information theoretic proofs of entropy power inequalities. *Information Theory, IEEE Transactions on*, 57(1):33–55, Jan 2011. arXiv:0704.1751, doi:10.1109/TIT.2010.2090193.

[SDM11] Naresh Sharma, Smarajit Das, and Siddharth Muthukrishnan. Entropy power inequality for a family of discrete random variables. In *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pages 1945–1949. IEEE, Jul 2011. arXiv:1012.0412, doi:10.1109/ISIT.2011.6033891.

[Sha48] Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:623–656, Oct 1948. URL: http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf.

[SS15] Yuri Suhov and Salimeh Y. Sekeh. Entropy-power inequality for weighted entropy. 2015. arXiv:1502.02188.

[Sta59] A. J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control*, 2(2):101–112, Jun 1959. doi:10.1016/S0019-9958(59)90348-1.

[SW97] Benjamin Schumacher and Michael D. Westmoreland. Sending classical information via noisy quantum channels. *Phys. Rev. A*, 56(1):131–138, Jul 1997. doi:10.1103/PhysRevA.56.131.

[VG06] Sergio Verdú and Dongning Guo. A simple proof of the entropy-power inequality. *Information Theory, IEEE Transactions on*, 52(5):2165–2166, May 2006. doi:10.1109/TIT.2006.872978.

[vN50] John von Neumann. *Functional Operators, Volume II: The geometry of orthogonal spaces*. Princeton University Press, 1950.

[Zha02] Xingzhi Zhan. *Matrix Inequalities*, volume 1790 of *Lecture Notes in Mathematics*. Springer, 2002. URL: https://books.google.co.uk/books?id=fvgxt6eYBlOC.

A Entropy power inequality for qubits

For the case of qubits ($d = 2$), there is a simple proof of eq. (3.7) which exploits the Bloch-vector representation of a qubit state.

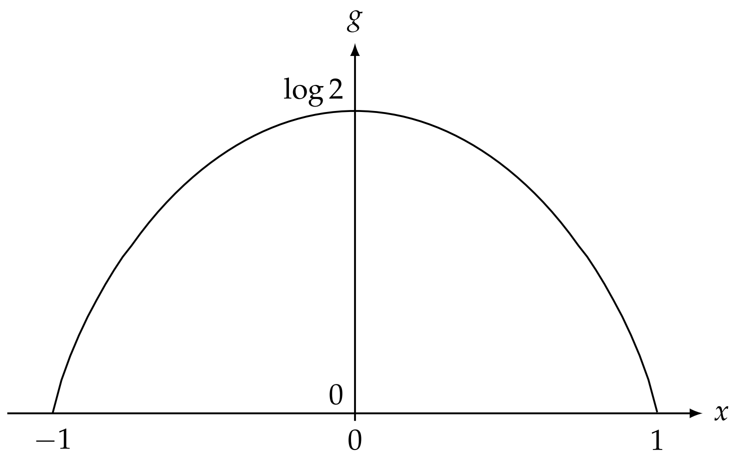

Figure 6: A plot of the function f defined in eq. (A.3).

A.1 Qubit states and the Bloch sphere

It is known that the state ρ of a qubit can be expressed in terms of its Bloch vector $\vec{r}$ as follows:

$$\rho = \frac{1}{2}(I + \vec{r} \cdot \vec{\sigma}) = \frac{1}{2}(I + x\sigma_x + y\sigma_y + z\sigma_z), \tag{A.1}$$

where $\vec{r} := (x, y, z) \in \mathbb{R}^3$ such that $|\vec{r}| := \sqrt{x^2 + y^2 + z^2} \leq 1$. Here $\vec{r} \cdot \vec{\sigma}$ denotes a formal inner product between $\vec{r}$ and $\vec{\sigma} := (\sigma_x, \sigma_y, \sigma_z)$, with σ_x, σ_y and σ_z being the Pauli matrices. Moreover, the eigenvalues of the state ρ can easily be seen to be given by $\frac{1}{2}(1 \pm |\vec{r}|)$. Hence, its von Neumann entropy is simply

$$H(\rho) = h\Big(\tfrac{1}{2}(1 + |\vec{r}|)\Big), \tag{A.2}$$

where $h(p) := -p \log p - (1-p)\log(1-p)$ is the *binary entropy* of $p \in [0,1]$ in nats. For $x \in [-1, 1]$, let us define the function

$$f(x) := h\Big(\tfrac{1}{2}(1 + x)\Big). \tag{A.3}$$

One can easily see that f is symmetric around the vertical axis and verify that

$$f''(x) = -\frac{1}{1 - x^2} \leq 0, \tag{A.4}$$

so f is concave (see Fig. 6). In terms of this function, eq. (A.2) is given by

$$H(\rho) = f(|\vec{r}|). \tag{A.5}$$

A.2 Proof of the qubit EPI

For a pair of qubit states ρ_1 and ρ_2, the first EPI of Corollary 7 is given by

$$H(\rho_1 \boxplus_a \rho_2) \geq aH(\rho_1) + (1-a)H(\rho_2) \qquad \forall\, a \in [0, 1]. \tag{A.6}$$

Below is a simple proof of the above inequality for the special case of qubits.

Proof. Using eq. (A.5), the inequality (A.6) can be expressed in terms of the function f as follows:

$$f(r) \geq af(r_1) + (1-a)f(r_2), \tag{A.7}$$

where $r := |\vec{r}|$, $r_1 := |\vec{r}_1|$, $r_2 := |\vec{r}_2|$, and $\vec{r}$, $\vec{r}_1$, $\vec{r}_2$ denote the Bloch vectors of the states $\rho_1 \boxplus_a \rho_2$, ρ_1 and ρ_2, respectively. Recall from eq. (4.19) that $\vec{r}$ can be expressed in terms of $\vec{r}_1$ and $\vec{r}_2$ as follows:

$$\vec{r} = a\vec{r}_1 + (1-a)\vec{r}_2 + \sqrt{a(1-a)}(\vec{r}_1 \times \vec{r}_2). \tag{A.8}$$

Since $\vec{r}_1$ and $\vec{r}_2$ are both perpendicular to $\vec{r}_1 \times \vec{r}_2$, we get

$$|\vec{r}|^2 = \vec{r} \cdot \vec{r} = a^2|\vec{r}_1|^2 + (1-a)^2|\vec{r}_2|^2 + 2a(1-a)\vec{r}_1 \cdot \vec{r}_2 + a(1-a)|\vec{r}_1 \times \vec{r}_2|^2. \tag{A.9}$$

If we denote by $\gamma \in [0, \pi]$ the angle between vectors $\vec{r}_1$ and $\vec{r}_2$, then $\vec{r}_1 \cdot \vec{r}_2 = |\vec{r}_1||\vec{r}_2| \cos\gamma$ and $|\vec{r}_1 \times \vec{r}_2| = |\vec{r}_1||\vec{r}_2| \sin\gamma$, so eq. (A.9) becomes

$$r^2 = a^2 r_1^2 + (1-a)^2 r_2^2 + a(1-a)(2r_1 r_2 \cos\gamma + r_1^2 r_2^2 \sin^2\gamma). \tag{A.10}$$

Note that the right-hand side of the inequality (A.7) does not depend on the angle γ between the vectors $\vec{r}_1$ and $\vec{r}_2$, so it suffices to prove eq. (A.7) only for those values of γ that *minimize* the left-hand side. Since $f(r)$ is a decreasing function of r for $r \geq 0$ (see Fig. 6), we have to consider only those values of γ that *maximize* r. From eq. (A.10) we have that

$$r = \sqrt{a^2 r_1^2 + (1-a)^2 r_2^2 + a(1-a) r_1 r_2 (2\cos\gamma + r_1 r_2 \sin^2\gamma)} \tag{A.11}$$

where $a, r_1, r_2 \in [0,1]$. To maximize this over γ, we only need to maximize the last term. Note that

$$2\cos\gamma + r_1 r_2 \sin^2\gamma \leq 2\cos\gamma + \sin^2\gamma \leq 2, \tag{A.12}$$

where the last inequality is tight if and only if $\gamma = 0$. This gives a simple upper bound on r:

$$r \leq \sqrt{a^2 r_1^2 + (1-a)^2 r_2^2 + 2a(1-a) r_1 r_2} = ar_1 + (1-a)r_2. \tag{A.13}$$

Since $f(r)$ is monotonically decreasing for $r \geq 0$, we get

$$f(r) \geq f\big(ar_1 + (1-a)r_2\big). \tag{A.14}$$

Note that this lower bound is independent of the parameter γ. Recall from eq. (A.4) that f is concave (see also Fig. 6), so

$$f\big(ar_1 + (1-a)r_2\big) \geq af(r_1) + (1-a)f(r_2). \tag{A.15}$$

By combining the last two inequalities, we get the desired result. □