Maximal output purity and capacity for asymmetric unital qudit channels

2005 J. Phys. A: Math. Gen. 38 9785

(http://iopscience.iop.org/0305-4470/38/45/005)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **38** (2005) 9785–9802 doi:10.1088/0305-4470/38/45/005

Maximal output purity and capacity for asymmetric unital qudit channels

Nilanjana Datta[1] **and Mary Beth Ruskai**[2,3]

[1] Statistical Laboratory, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB, UK
[2] Department of Mathematics, Tufts University, Medford, MA 02155, USA

E-mail: N.Datta@statslab.cam.ac.uk and Marybeth.Ruskai@tufts.edu

Received 14 August 2005, in final form 16 September 2005
Published 26 October 2005
Online at stacks.iop.org/JPhysA/38/9785

Abstract

We consider generalizations of depolarizing channels to maps of the form $\Phi(\rho) = \sum_k a_k V_k \rho V_k^\dagger + (1-a)(\mathrm{Tr}\,\rho)\frac{1}{d}I$ with V_k being unitary and $\sum_k a_k = a < 1$. We show that one can construct unital channels of this type for which the input which achieves maximal output purity is unique. We give conditions on V_k under which multiplicativity of the maximal p-norm and additivity of the minimal output entropy can be proved for $\Phi \otimes \Omega$ with Ω arbitrary. We also show that the Holevo capacity need not equal $\log d - S_{\min}(\Phi)$ as one might expect for a convex combination of unitary conjugations.

PACS number: 03.67

1. Introduction

The depolarizing channel Γ_a^{dep} has the form

$$\Gamma_a^{\mathrm{dep}}(\rho) = a\rho + (1-a)(\mathrm{Tr}\,\rho)\frac{1}{d}I \tag{1}$$

with $-\frac{1}{d^2-1} \leqslant a \leqslant 1$. In this paper, we consider channels of the more general form

$$\Phi(\rho) = \sum_k a_k V_k \rho V_k^\dagger + (1-a)(\mathrm{Tr}\,\rho)\frac{1}{d}I \tag{2}$$

with $0 < a_k$, $0 < a = \sum_k a_k < 1$ and V_k being unitary.

We describe and study several subclasses of these channels (2), showing that they can exhibit different types of behaviour. Those with simultaneously diagonal V_k have a high level

[3] Partially supported by the National Security Agency (NSA) and Advanced Research and Development Activity (ARDA) under Army Research Office (ARO) contract number DAAD19-02-1-0065, and by the National Science Foundation under Grant DMS-0314228.

of symmetry and much in common with depolarizing channels. However, we also construct asymmetric channels with a unique state of minimal output entropy and other behaviour more typical of non-unital channels; although additivity can be proved for the minimal output entropy, this does not imply additivity of the capacity because the optimal average output is not $\frac{1}{d}I$.

This paper is organized as follows. Section 2 contains some terminology and notation as well as considerable background material on various types of channels and their behaviour. In section 3, we state and prove some theorems about minimal output purity for the channels we consider. In section 4 we consider a special subclass of channels which satisfy (2) and exhibit behaviour similar to unital qubit channels. In section 5, which is the heart of the paper, we describe several types of asymmetric channels to which our results can be applied. In section 6 we report the results of numerical tests on channel capacity.

2. Background

2.1. General notation and terminology

We restrict attention to finite dimensional spaces $\mathbf{C}^d$ and denote the space of $d \times d$ complex matrices as $M_d = \mathcal{B}(\mathbf{C}^d)$. By a channel Φ we mean a completely positive, trace preserving (CPT) map $\Phi : M_d \mapsto M_d$. Let $\mathcal{D} = \{\rho : \rho \geqslant 0, \mathrm{Tr}\,\rho = 1\}$ denote the set of density matrices in M_d. Let $S(\gamma) = -\mathrm{Tr}\,\gamma \log \gamma$ denote the quantum entropy of a state $\gamma \in \mathcal{D}$. For a CPT map Φ, one can define the maximal output p-norm

$$\nu_p(\Phi) = \sup_{\gamma \in \mathcal{D}} \|\Phi(\gamma)\|_p, \tag{3}$$

the minimal output entropy

$$S_{\min}(\Phi) = \inf_{\rho \in \mathcal{D}} S[\Phi(\rho)], \tag{4}$$

and the Holevo capacity

$$C_{\mathrm{Holv}}(\Phi) = \sup_{\{\pi_j,\rho_j\}} \left(S[\Phi(\rho_{\mathrm{av}})] - \sum_j \pi_j S[\Phi(\rho_j)] \right), \tag{5}$$

where $\rho_{\mathrm{av}} = \sum_j \pi_j \rho_j$, and the supremum is taken over all ensembles $\{\pi_j, \rho_j\}$ with $\rho_j \in \mathcal{D}, \pi_j > 0$ and $\sum_j \pi_j = 1$. Both $S_{\min}(\Phi)$ and $C_{\mathrm{Holv}}(\Phi)$ are conjectured to be additive over tensor products, i.e., to satisfy

$$S_{\min}(\Phi \otimes \Omega) = S_{\min}(\Phi) + S_{\min}(\Omega), \tag{6}$$

and

$$C_{\mathrm{Holv}}(\Phi \otimes \Omega) = C_{\mathrm{Holv}}(\Phi) + C_{\mathrm{Holv}}(\Omega). \tag{7}$$

Shor showed [31] that these conjectures (and several related ones) are equivalent in the global sense that both are either true for all general channels $\Phi : M_d \mapsto M_n$ or both are false. However, they are not necessarily equivalent for individual channels, and we will study them separately for the examples in this paper.

Shor also proved [30] that both (6) and (7) hold for entanglement breaking (EB) channels. King [18] gave an alternative proof based on multiplicativity of $\nu_p(\Phi)$. A CP map Φ is EB if $(I \otimes \Phi)(\rho)$ is separable for all input states ρ. A CPT map which is also EB is denoted as EBT. It was shown in [15] that a CP map is EB if all its Kraus operators can be chosen to have rank one, or if $(I \otimes \Phi)(|\Psi\rangle\langle\Psi|)$ is separable for some maximally entangled $|\Psi\rangle$. Any EBT channel be written as

$$\Phi(\rho) = \sum_k \gamma_k \operatorname{Tr} \rho E_k, \tag{8}$$

with $\{E_k\}$ being a POVM, and each $\gamma_k \in \mathcal{D}$. When $\{|e_k\rangle\}$ is an orthonormal basis for $\mathbf{C}^d$ and $E_k = |e_k\rangle\langle e_k|$ the channel is called CQ (classical-quantum); and when each $\gamma_k = |e_k\rangle\langle e_k|$ it is called QC (quantum-classical).

The following max–min characterizations of $C_{\text{Holv}}(\Phi)$ in terms of the relative entropy $H(\rho, \gamma) = \operatorname{Tr} \rho(\log \rho - \log \gamma)$ are extremely useful. They were obtained independently in [24] and [28].

$$C_{\text{Holv}}(\Phi) = \inf_{\gamma\in\mathcal{D}} \sup_{\omega\in\mathcal{D}} H[\Phi(\omega), \Phi(\gamma)] \tag{9a}$$

$$= \sup_{\omega\in\mathcal{D}} H[\Phi(\omega), \Phi(\rho_{\text{av}})] \tag{9b}$$

$$= H[\Phi(\rho_j), \Phi(\rho_{\text{av}})], \tag{9c}$$

where ρ_{av} is the optimal average input and ρ_j is any input in the optimal signal ensemble. It can be shown [11] that (9b) and (9c) are equivalent to the statement that the points $(\rho_i, S(\rho_i))$ define a supporting hyperplane for the convex optimization problem (5).

2.2. Depolarizing channels

The properties of the depolarizing channel are well-known and can be summarized as follows.

Theorem 1. *The depolarizing channel (1) satisfies*

(a) $\Gamma_a^{\text{dep}}(I)$ *is unital, i.e.,* $\Gamma_a^{\text{dep}}(I) = I$.
(b) The output $\Gamma_a^{\text{dep}}(|\psi\rangle\langle\psi|)$ *for any pure state* $|\psi\rangle\langle\psi|$ *has eigenvalues* $\left[a+\frac{1-a}{d}, \frac{1-a}{d}, \ldots \frac{1-a}{d}\right]$.
(c) For any CPT map Ω, $\nu_p\left(\Gamma_a^{\text{dep}} \otimes \Omega\right) = \nu_p\left(\Gamma_a^{\text{dep}}\right)\nu_p(\Omega) \quad \forall\, p \geqslant 1$.
(d) For any CPT map Ω, $S_{\min}\left(\Gamma_a^{\text{dep}} \otimes \Omega\right) = S_{\min}\left(\Gamma_a^{\text{dep}}\right) + S_{\min}(\Omega)$.
(e) $C_{\text{Holv}}\left(\Gamma_a^{\text{dep}}\right) = \log d - S_{\min}\left(\Gamma_a^{\text{dep}}\right)$.
(f) The capacity $C_{\text{Holv}}\left(\Gamma_a^{\text{dep}}\right)$ *can be achieved using d orthogonal input states.*
(g) The optimal average input is $\frac{1}{d}I$.
(h) For any CPT map Ω, $C_{\text{Holv}}\left(\Gamma_a^{\text{dep}} \otimes \Omega\right) = C_{\text{Holv}}\left(\Gamma_a^{\text{dep}}\right) + C_{\text{Holv}}(\Omega)$.
(i) When $a \leqslant \frac{1}{d+1}$, *the channel* Γ_a^{dep} *is EBT.*

The mutiplicativity (c) was proved by King [17] for any depolarizing map, including those with negative a; he also showed that properties (d) and (h) follow. Properties (d) and (h) were proved independently by Fujiwara and Hashizumé [8] for maps with $a > 0$ and $\Omega = \Gamma_a^{\text{dep}}$; they used a majorization argument which also implies (c). Properties (a), (b) and (e) are well-known and easily verified. Property (j) can be verified by computing the Choi matrix $\left(I \otimes \Gamma_a^{\text{dep}}\right)(|\beta\rangle\langle\beta|)$ for a maximally entangled state $|\beta\rangle$ and using theorem 4 of [15].

It is useful to introduce the generalized Pauli operators X_d and Z_d defined on the standard basis so that $X_d|e_\ell\rangle = |e_{\ell+1}\rangle$ with the addition in the subscript taken mod d and $Z_d|e_\ell\rangle = \mathrm{e}^{2\pi \mathrm{i}\ell/d}$. Then for any $d \times d$ matrix A,

$$\frac{1}{d^2}\sum_{m=0}^{d-1}\sum_{n=0}^{d-1} X_d^m Z_d^n A\left(Z_d^\dagger\right)^n\left(X_d^\dagger\right)^m = (\operatorname{Tr} A)\frac{1}{d}I, \tag{10}$$

and

$$\Gamma_a^{\mathrm{dep}}(\rho) = \left[a + \frac{1-a}{d^2}\right] I\rho I + (1-a)\frac{1}{d^2}\sum_{\substack{m=0 \\ m,n\neq 0,0}}^{d-1}\sum_{n=0}^{d-1} X_d^m Z_d^n \rho \left(Z_d^\dagger\right)^n \left(X_d^\dagger\right)^m. \tag{11}$$

Cortese [4] considered channels of the form

$$\Phi(\rho) = \sum_{m=0}^{d-1}\sum_{n=0}^{d-1} c_{mn} X_d^m Z_d^n \rho \left(Z_d^\dagger\right)^n \left(X_d^\dagger\right)^m \tag{12}$$

with $c_{mn} \geqslant 0$ and $\sum_{mn} c_{mn} = 1$, and showed that

$$C_{\mathrm{Holv}}(\Phi) = \log d - S_{\min}(\Phi). \tag{13}$$

A simplified proof of this result was given by Holevo [13], who showed that (13) holds for channels satisfying the covariance condition

$$\Phi\left(U_g \rho U_g^\dagger\right) = U_g' \Phi(\rho) [U_g']^\dagger \qquad \forall\, g \in \mathcal{G} \tag{14}$$

when $\{U_g\}$ and $\{U_g'\}$ are irreducible representations of a group $\mathcal{G}$. Case (12) is called 'Weyl covariance'.

By using (10) to rewrite the second term in (2) and the fact that $\sum_k a_k = a$, one sees that such channels can be expressed as a convex combination of unitary conjugations. We write them in the form (2) because we exploit their relationship to the depolarizing channel. However, (13) need not hold for all channels of the form (2); in section 5 we give examples which show that they can exhibit very different behaviour.

2.3. Qubit channels

As discussed in appendix B, a unital qubit channel can be written (after rotation of bases) [22] as

$$\Phi(\rho) = \sum_{k=0}^{3} \alpha_k \sigma_k \rho \sigma_k. \tag{15}$$

It is also useful to recall that any qubit density matrix can be written as $\rho = \frac{1}{2}[I + \mathbf{w}\cdot\sigma]$, where σ denotes the vector of Pauli matrices and $\mathbf{w} \in \mathbf{C}^3$; then the channel (15) can be written as

$$\Phi(\rho) = \frac{1}{2}\left[I + \sum_{j=1}^{3} \lambda_j w_j \sigma_j\right]. \tag{16}$$

The relations between the parameters $\{\alpha_k\}$ and $\{\lambda_j\}$ are discussed in appendix B.

The following theorem was proved by King in [16].

Theorem 2. *Let Φ be a unital qubit channel and $a = \max_{k=1,2,3} |\lambda_k| = \max_{i\neq j\in 0,1,2,3} (\alpha_i + \alpha_j)$. Then parts (c) to (h) of theorem 1 hold, with Γ_a^{dep} replaced by Φ. In addition, for those k with $|\lambda_k| = a$, the inputs $\frac{1}{2}[I \pm \sigma_k]$ yield outputs with eigenvalues $\frac{1}{2}(1 \pm a)$ and, hence, have the same entropy as the corresponding qubit depolarizing channel.*

This implies that all unital qubit channels for which the image ellipsoid of the Bloch sphere touches, but lies within, the sphere of radius a (which is the image of a depolarizing channel) have the same capacity and minimal output entropy behaviour. A unital qubit channel is EBT [26] if and only if $\sum_k |\lambda_k| \leqslant 1$ or, equivalently, if $\alpha_k \leqslant \frac{1}{2}$ for all k.

A non-unital qubit channel can be written (after rotation of bases) [22] in the form

$$\Phi : \frac{1}{2}[I + \mathbf{w} \cdot \sigma] \mapsto \frac{1}{2}\left[I + \sum_{k=1}^{3} (t_k + \lambda_k w_k)\sigma_k \right]. \tag{17}$$

The conditions imposed on t_k and λ_k by the CPT requirement are given in [27] and summarized in [26]. (The special case $t_1 = t_2 = 0$ was considered earlier in [7].) One expects the generic behaviour of non-unital qubit channels to be quite different from that of unital ones.

(A) Non-unital qubit channels typically have a unique state of optimal output purity. This always holds when $t_k \neq 0$ in the direction for which the ellipsoid axis $|\lambda_k|$ is longest. If $t_k \neq 0$ only in direction(s) orthogonal to the longest axis, then one typically has two non-orthogonal states of optimal output purity (although these can coalesce into one, as for extreme amplitude damping channels, and can come from orthogonal inputs for a CQ channel) [5, 20].

(B) $C_{\text{Holv}}(\Phi) < \log d - S_{\min}(\Phi)$ for all non-unital qubit maps.

(C) In general, the capacity $C_{\text{Holv}}(\Phi)$ can not be achieved using d orthogonal input states [5, 11, 20, 28].

There are, however, a number of exceptions. Two of these are CQ maps which take $\frac{1}{2}[I + \mathbf{w} \cdot \sigma] \mapsto \frac{1}{2}[I + t_1\sigma_1 + \lambda_3 w_3 \sigma_3]$ and QC maps which take $\frac{1}{2}[I + \mathbf{w} \cdot \sigma] \mapsto \frac{1}{2}[I + (t_3 + \lambda_3 w_3)\sigma_3]$. The QC channels are included in the larger class of channels for which $t_k \neq 0$ only for the largest $|\lambda_k|$; then $C_{\text{Holv}}(\Phi)$ can be achieved with a pair of orthogonal inputs [9, 20].

(D) Properties (c), (d) and (h) of theorem 1 are conjectured to hold for non-unital qubit maps; however, a proof is known only for (c) in the case $p = 2$.

2.4. *Some channels for $d > 2$*

When Φ maps a larger space into qubit density matrices, it is possible to have $C_{\text{Holv}}(\Phi) = \log d - S_{\min}(\Phi)$, even when the optimal input $\rho_{\text{av}} \neq \frac{1}{d} I$. This is the case for Shor's extended channel in section 9 of [31]. In that case, the original map Φ is extended to Φ_{ext} for which the optimal average input is $R_{\text{av}} = \rho_{\min} \otimes \frac{1}{d^2} I$, with $\rho_{\min}$ achieving $S_{\min}(\Phi)$ for the original channel. Then $\Phi_{\text{ext}}(R_{\text{av}}) = \frac{1}{d} I$. Note that one also has $\Phi_{\text{ext}}(I_d \otimes I_{d^2}) = I_d$ so that Φ_{ext} is unital. Moreover, if $S_{\min}(\Phi)$ is achieved for more than one state, then the optimal average input is not unique, although the optimal average output is unique.

For qubits, a channel is unital if and only if it can be written as a convex combination of unitary conjugations [22]. It is well-known that this result does not extend to $d > 2$. One well-known example is the Werner–Holevo channel [32] for which the Kraus operators can be written as partial isometries. This example does satisfy (13) as well as (6) and (7), although it has only been shown to satisfy (20) when $1 \leqslant p \leqslant 2$ [2] and is known to violate (20) for large p.

For $d = 3$, Fuchs *et al* [6] found a unital channel which satisfies (13) but for which the optimal inputs are not orthogonal. This channel is given by equation (19) of [15].

The asymmetric examples in section 5 appear to be the first for which a unital channel does not satisfy (13).

It is natural to look for classifications of unital channels which include a type whose behaviour is similar to that of unital qubit channels. The results presented here show that there are channels which can be written as convex combinations of unitary conjugations which do not exhibit this behaviour. Thus we are left with the conjecture that channels of the form (12) behave like unital qubit channels and, hence, satisfy (c) to (h) of theorem 1 with Γ replaced by Φ, as in theorem 2.

2.5. Majorization

We will use the notation $[x_1, x_2, \ldots, x_n] \succ [y_1, y_2, \ldots, y_n]$ to indicate that both sets are non-negative and arranged in non-increasing order $x_1 \geqslant x_2 \geqslant x_3 \cdots \geqslant 0$ and satisfy the majorization condition $\sum_{i=1}^{k} x_i \geqslant \sum_{i=1}^{k} y_i$ for $k = 1 \ldots n-1$ and $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$. It is well-known [14, 23] that this implies

$$\sum_{j=1}^{n} x_j^p \geqslant \sum_{j=1}^{n} y_j^p \tag{18}$$

for all $p \geqslant 1$. Therefore, whenever ρ and γ are density matrices for which the eigenvalues of ρ majorize those of γ, $\|\rho\|_p > \|\gamma\|_p$ and $S(\rho) < S(\gamma)$.

When only an inequality holds for $k = n$, we use the term submajorize, and observe that the same conclusions follow by extending both sets with $x_{n+1} = 0$ and y_{n+1} chosen to give equality.

3. Results on minimal output purity

In this section we state and prove some theorems on the minimal output purity of certain subclasses of the channels defined by (2).

Theorem 3. *Let Φ be a channel of the form (2) for which all of the unitary operators V_k have a common eigenvector $|\psi\rangle$. Then for any CPT map Ω*

$$\text{(a)}\quad \|\Phi(|\psi\rangle\langle\psi|)\|_p = \nu_p(\Phi) = \nu_p\big(\Gamma_a^{\text{dep}}\big) \quad \forall\, p \geqslant 1 \tag{19}$$

$$\text{(b)}\quad \nu_p(\Phi \otimes \Omega) = \nu_p(\Phi)\nu_p(\Omega) \quad \forall\, p \geqslant 1 \tag{20}$$

$$\text{(c)}\quad S[\Phi(|\psi\rangle\langle\psi|)] = S_{\min}(\Phi) = S_{\min}\big(\Gamma_a^{\text{dep}}\big) \tag{21}$$

$$\text{(d)}\quad S_{\min}(\Phi \otimes \Omega) = S_{\min}(\Phi) + S_{\min}(\Omega). \tag{22}$$

Proof. First, observe that

$$\begin{aligned}\Phi(\rho) &= \sum_k \frac{a_k}{a} V_k \left[a\rho + (1-a)(\operatorname{Tr}\rho)\frac{1}{d}I \right] V_k^\dagger \\ &= \sum_k \frac{a_k}{a} V_k \Gamma_a^{\text{dep}}(\rho) V_k^\dagger.\end{aligned} \tag{23}$$

is a convex combination of conjugation with V_k composed with the depolarizing channel. Therefore, for any density matrix ρ

$$\begin{aligned}\|\Phi(\rho)\|_p &\leqslant \sum_k \frac{a_k}{a} \big\| V_k \Gamma_a^{\text{dep}}(\rho) V_k^\dagger \big\|_p \\ &\leqslant \sum_k \frac{a_k}{a} \nu_p\big(\Gamma_a^{\text{dep}}\big) = \nu_p\big(\Gamma_a^{\text{dep}}\big).\end{aligned} \tag{24}$$

Now consider $\rho = |\psi\rangle\langle\psi|$ with $|\psi\rangle$ being the common eigenvector of V_k. Then

$$\|\Phi(|\psi\rangle\langle\psi|)\|_p = \left\| a|\psi\rangle\langle\psi| + \frac{1-a}{d} I \right\|_p = \Gamma_a^{\text{dep}}(|\psi\rangle\langle\psi|) = \nu_p\big(\Gamma_a^{\text{dep}}\big),$$

where we used part (b) of theorem 1. Therefore, $\nu_p(\Phi)$ is at least as big as $\nu_p\big(\Gamma_a^{\text{dep}}\big)$. Combining this with (24), proves part (a).

To prove (b), we proceed similarly, using (23), to see that

$$\|(\Phi \otimes \Omega)(\rho_{12})\|_p \leqslant \sum_k \frac{a_k}{a} \left\| \left(\Gamma_a^{\text{dep}} \otimes \Omega\right)(\rho_{12}) \right\|_p \tag{25}$$

$$\leqslant \sum_k \frac{a_k}{a} \nu_p\left(\Gamma_a^{\text{dep}}\right) \nu_p(\Omega) \tag{26}$$

$$= \nu_p\left(\Gamma_a^{\text{dep}}\right) \nu_p(\Omega) = \nu_p(\Phi) \nu_p(\Omega), \tag{27}$$

where the last step used part (a). Since we can achieve $\nu_p(\Phi)\nu_p(\Omega)$ using a product state, this proves (b). Parts (c) and (d) then follow by the established technique [3] of taking the right derivative at $p = 1$. □

By choosing all $V_k = W^k$ with W being a unitary matrix which generates a cyclic group of order d, one can construct channels with precisely d input states whose outputs have optimal purity. Additional channels with d states of optimal output purity are discussed in section 4. Channels for which each V_k has the form $\sum_{j=1}^m |f_j\rangle\langle f_j| \oplus W_k$ with $|f_j\rangle$ being a set of m mutually orthonormal vectors and W_k being unitary operators on $[\text{span}\{|f_j\rangle\}]^\perp$ are more interesting. Several classes of examples are discussed in detail in section 5. When the W_k have no common eigenvectors, it follows from theorem 4 below that these channels have precisely m mutually orthogonal states of optimal purity. One can construct channels with $m = 1, 2, \ldots d-2$; however, if the V_k have $d-1$ common eigenvectors, then they have d common eigenvectors, precluding the possibility that $m = d-1$.

Theorem 4. *Let Φ be a channel of the form (2) and let ρ be any density matrix other than the projection onto a common pure state eigenvector of all V_k. Then $\|\Phi(\rho)\|_p < \nu_p\left(\Gamma_a^{\text{dep}}\right)$ and $S[\Phi(\rho)] > S_{\min}\left(\Gamma_a^{\text{dep}}\right)$.*

Proof. Under the hypothesis of the theorem,

$$\left\| \sum_k \frac{a_k}{a} V_k \rho V_k^\dagger \right\|_\infty < 1 \tag{28}$$

and one can write the eigenvalues of $\sum_k \frac{a_k}{a} V_k \rho V_k^\dagger$ as $[x_1, x_2, \ldots x_d]$ with $x_1 < 1$. Then the eigenvectors of $\Phi(\rho)$ are

$$\left[ax_1 + \frac{1-a}{d}, ax_2 + \frac{1-a}{d}, \ldots, ax_d + \frac{1-a}{d}\right] \prec \left[a + \frac{1-a}{d}, \frac{1-a}{d}, \ldots \frac{1-a}{d}\right]. \tag{29}$$

Thus, the eigenvalues of $\Phi(\rho)$ are majorized by those of $\Gamma_a^{\text{dep}}(|\psi\rangle\langle\psi|$ for any pure input $|\psi\rangle$). □

Theorem 5. *Let Φ be a channel of the form (2) for which the unitary operators V_k have precisely m mutually orthogonal common eigenvectors with $m < d$. Then $\rho_{\text{av}} \neq \frac{1}{d} I$ and at least $(d-m)$ states in the optimal input ensemble have $S[\Phi(\rho_i)] > S_{\min}(\Phi)$.*

Proof. When the number of common eigenvectors $m < d$, it follows that one can not find a set of d mutually orthogonal pure inputs ρ_i for which $S[\Phi(\rho_i)] = S_{\min}(\Phi)$. Therefore, one can not find an input ensemble such that both $\sum_i \pi_i \rho_i = \frac{1}{d} I$ and $S[\Phi(\rho_i)] = S_{\min}(\Phi) \forall\, i$ hold. Therefore, we must have

$$C_{\text{Holv}}(\Phi) < \log d - S_{\min}(\Phi). \tag{30}$$

Since

$$\sup_{\omega\in\mathcal{D}} H\left[\Phi(\omega), \Phi\left(\frac{1}{d}I\right)\right] = \log d - \inf_{\omega\in\mathcal{D}} S[\Phi(\omega)] = \log d - S_{\min}(\Phi), \tag{31}$$

it follows from (30) and (9) that $\frac{1}{d}I$ is not the optimal average input.

If we know that the optimal signal ensemble has at least d inputs, then at least $d-m$ of them must satisfy $S[\Phi(\rho_i)] > S_{\min}(\Phi)$. □

Although we are primarily interested in channels which are trace preserving, multiplicativity results, e.g., (20) can often be proved using only the CP condition. Moreover, Audenaert and Braunstein [1] showed that multiplicativity of a special class of CP maps would imply superadditivity of entanglement of formation. Therefore, we notice that a weaker version of theorem 3 can be extended to maps of the form (23) in which the V_k are contractions rather than unitary, i.e. $V_k V_k^\dagger \leqslant I$.

Theorem 6. *Let Φ be a CP map of the form*

$$\Phi(\rho) = \sum_k \frac{a_k}{a} V_k \left[a\rho + (1-a)(\operatorname{Tr}\rho)\frac{1}{d}I\right] V_k^\dagger \tag{32}$$

for which all of the operators V_k are contractions with a common eigenvector $|\psi\rangle$ satisfying $V_k|\psi\rangle = \mathrm{e}^{\mathrm{i}\theta_k}|\psi\rangle$. Then for any CP map Ω, (19)–(22) hold.

Proof. The assumption that the eigenvalues of the common eigenvector have $|\mathrm{e}^{\mathrm{i}\theta_k}| = 1$ implies that $\nu_p(\Phi)$ is at least as large as $\nu_p\big(\Gamma_a^{\text{dep}}\big)$. For any contraction V, the eigenvalues of $VAV^\dagger$ are submajorized by those of A, which we write as $[\alpha_1, \alpha_2 \ldots \alpha_d]$. To see this, write $A = UA_DU^\dagger$ with U unitary and A_D the diagonal matrix with elements $\delta_{jk}\alpha_j$. Then $X = VU$ is also a contraction and the diagonal elements of $VAV^\dagger$ are $\sum_j |x_{ij}|^2\alpha_j$ which are submajorized by $[\alpha_1, \alpha_2 \ldots \alpha_d]$. By applying this to $A = a\rho + (1-a)\frac{1}{d}I$, the result follows by the same argument as before. □

4. Diagonal V_k

Before discussing several types of asymmetric channels, we consider channels for which all V_k are simultaneously diagonal, as well as unitary. This includes the case $V_k = W^k$, with $W^d = I$, mentioned earlier. In all these situations, one has precisely d states of minimal output entropy and the capacity is

$$C_{\text{Holv}}(\Phi) = \log d - S_{\min}(\Phi) = \log d - S_{\min}\big(\Gamma_a^{\text{dep}}\big). \tag{33}$$

It then follows from the additivity of $S_{\min}(\Phi)$ in part (d) of theorem 3 that $C_{\text{Holv}}(\Phi)$ is also additive in the sense $C_{\text{Holv}}(\Phi\otimes\Phi) = 2C_{\text{Holv}}(\Phi)$.

The channels considered in this section are, therefore, convex combinations

$$\Phi(\rho) = a\Phi^{\text{diag}}(\rho) + (1-a)(\operatorname{Tr}\rho)\frac{1}{d}I \tag{34}$$

of the completely noisy map and a 'diagonal channel' of the form $\Phi^{\text{diag}}(\gamma) = \sum_k a_k V_k \gamma V_k^\dagger$ with $a_k > 0$. The term diagonal channel was introduced by King [19] for CP maps whose Kraus operators are simultaneously diagonal. King also showed that $\Phi^{\text{diag}}(\gamma) = B * \gamma$ where $*$ denotes the Hadamard product, B is a positive semi-definite matrix, and γ is written in the basis in which the V_k are diagonal. When V_k is unitary, its diagonal elements can be written as $\mathrm{e}^{\mathrm{i}\phi_{km}}$, $m = 1, 2 \ldots d$ and $b_{mn} = \sum_k a_k\, \mathrm{e}^{\mathrm{i}(\phi_{km}-\phi_{kn})}$. If one also requires Φ^{diag} to be trace-preserving, then $\sum_k a_k = 1$ and $b_{mm} = 1\ \forall\ m$. This implies that the states $|m\rangle\langle m|$ are fixed points of Φ^{diag} so that it has d pure state outputs. Hence additivity of both minimal output entropy and Holevo capacity hold trivially for diagonal CPT maps.

In the examples (34) considered here, the corresponding outputs are $\Phi(|m\rangle\langle m|) = a|m\rangle\langle m| + (1-a)\frac{1}{d}I, m = 1, 2, \ldots d$ which yield d states of minimal output entropy. As noted above, this implies that they satisfy (13) and (7) when $\Omega = \Phi$. Since theorem 3 holds, (19)–(22) are also satisfied.

The depolarizing channel (1) satisfies the general covariance condition $\Phi(U\rho U^\dagger) = U\Phi(\rho)U^\dagger$ for arbitrary unitary matrices U, but this does not extend to channels of the form (2). However, when $V_k = W^k$ with $W = UX_dU^\dagger$ and U unitary, the channel satisfies the weaker condition (14) using the generalized Pauli operators of the form $UX_d^mZ_d^nU^\dagger$.

Note that $W = UX_dU^\dagger$ is equivalent to the assumption that W has eigenvalues $\mathrm{e}^{\mathrm{i}2\pi m/d}, m = 0, 1 \ldots d-1$. However, one can have a unitary W with $W^d = I, W^m \neq I$, $m < d$ but $W \neq UX_dU^\dagger$. For example, with $d = 5$, choose W to have eigenvalues $\mathrm{e}^{\mathrm{i}2\pi/5}, \mathrm{e}^{\mathrm{i}2\pi/5}, \mathrm{e}^{\mathrm{i}2\pi 3/5}, 1, 1$.

More generally, of course, one could choose V_k with eigenvalues $\mathrm{e}^{\mathrm{i}\phi_{km}}$ without any rational relationship between eigenvalues for a single V_k or between those for V_j and V_k. Then (13) still holds, despite the absence of any obvious group for which (14) holds. However, we can not completely exclude the possibility of a hidden group.

5. Asymmetric examples

5.1. Qutrit channels

We will now study in detail the case $d = 3$, with

$$V_k = \mathrm{e}^{\mathrm{i}\theta}|e_0\rangle\langle e_0| \oplus \sigma_k = \begin{pmatrix} \mathrm{e}^{\mathrm{i}\theta} & 0 \\ 0 & \sigma_k \end{pmatrix}, \qquad k \in \{0, 1, 2, 3\}, \tag{35}$$

with the convention that $\sigma_0 = I$. As discussed in appendix B we can assume that $a_0 \geqslant a_1$.

It follows from theorems 3 and 5 that Φ has exactly one state of minimal output entropy $|e_0\rangle\langle e_0|$ and two orthogonal states $|e_\pm\rangle\langle e_\pm| = \frac{1}{2}[I \pm \sigma_1]$ whose outputs have eigenvalues $\left[a\frac{1+\lambda_1}{2} + \frac{1-a}{3}, a\frac{1-\lambda_1}{2} + \frac{1-a}{3}, \frac{1-a}{3}\right]$. Here λ_1 is given by (B.4), with $i = 1$. If these states are the optimal inputs ρ_j, symmetry implies that the optimal average input has the form

$$\rho_{\text{av}} = (1-2x)|e_0\rangle\langle e_0| + x|e_{+1}\rangle\langle e_{+1}| + x|e_{-1}\rangle\langle e_{-1}|, \tag{36}$$

for which the optimal average output is

$$\Phi(\rho_{\text{av}}) = \left(a(1-2x) + \frac{1-a}{3}\right)|e_0\rangle\langle e_0| + \left(ax + \frac{1-a}{3}\right)(|e_+\rangle\langle e_+| + |e_-\rangle\langle e_-|). \tag{37}$$

We want to optimize the capacity

$$S[\Phi(\rho_{\text{av}})(x)] - [(1-2x)S[\Phi(\rho_0)] + xS[\Phi(\rho_{+1})] + xS[\Phi(\rho_{-1})]. \tag{38}$$

Since $S[\Phi(\rho_{+1})] = S[\Phi(\rho_{-1})]$, differentiating (38) gives the condition

$$2a\log\left(\frac{1+2a}{3} - 2ax\right) - 2a\log\left(\frac{1-a}{3} + ax\right) = -2S[\Phi(\rho_0)] + 2S[\Phi(\rho_{\pm 1})] \tag{39}$$

or

$$\log\frac{1-a+3ax}{1+2a-6ax} = -\frac{1}{a}\Delta S, \tag{40}$$

where $\Delta S = S[\Phi(\rho_{+1})] - S[\Phi(\rho_0)] > 0$. This has the solution

$$x = \frac{(1+2a)2^{-\Delta S/a} - (1-a)}{3a(1+2^{-\Delta S/a}.2)}. \tag{41}$$

It is easy to verify that $x < \frac{1}{3}$ confirming the intuition that the optimal input will be shifted toward the state $|e_0\rangle$.

Let ρ_x denote the average for the ensemble corresponding to the optimal x (41) and $C^x_{\text{Holv}}(\Phi)$ the corresponding capacity (38). To show that ρ_x is the true optimal average which yields $C_{\text{Holv}}(\Phi)$, we need to verify that $H[\Phi(\omega), \Phi(\rho_x)] \leqslant C^x_{\text{Holv}}(\Phi)$ for all choices of ω. This has been done numerically for a large range of a and λ_1.

5.2. *Doubly depolarizing channels*

We introduce some notation. Let $\{|e_j\rangle\langle e_j|\}$ be an orthonormal basis for $\mathbf{C}^d$, E_m the projection on span$\{|e_1\rangle, |e_2\rangle \ldots |e_m\rangle\}$, and $E_m^\perp$ the projection on the orthogonal complement span$\{|e_m\rangle, |e_{m+1}\rangle \ldots |e_d\rangle\}$.

Now suppose that Φ is a channel of the form (2) in which each V_k has the form $V_k = E_m \oplus W_k = \begin{pmatrix} E_m & 0 \\ 0 & W_k \end{pmatrix}$ where the W_k are chosen to be unitary $(d-m) \times (d-m)$ matrices such that on $E_m^\perp \mathcal{H}$

$$\sum_k \frac{a_k}{a} W_k \rho W_k^\dagger = b\rho + (1-b)\big(\text{Tr}_{E_m^\perp \mathcal{H}} \rho\big) \frac{1}{d-m} E_m^\perp. \tag{42}$$

It suffices to choose W_k to be the generalized Pauli matrices defined before (10) and let $a_k = a(1-b)/(d-m)^2$ for all k except $a_0 = a[b(d-m)^2 + (1-b)]/(d-m)^2$. For the case $d = 4$ and $m = 2$, this reduces to $W_k = \sigma_k$ with $a_0 = a(3b+1)/4$ and $a_j = a(1-b)/4$ for $j = 1, 2, 3$.

The action of Φ is similar to a depolarizing channel when restricted to $E_m\mathcal{H}$ or $E_m^\perp\mathcal{H}$. More precisely,

$$\Phi(|e\rangle\langle e|) = a|e\rangle\langle e| + (1-a)\frac{1}{d} I \qquad \forall\, |e\rangle \in E_m\mathcal{H} \tag{43}$$

$$\Phi(|f\rangle\langle f|) = ab|f\rangle\langle f| + a(1-b)\frac{1}{d-m} E_m^\perp + (1-a)\frac{1}{d} I \qquad \forall\, |f\rangle \in E_m^\perp\mathcal{H}. \tag{44}$$

The case $m = 1, d = 3$ is a special case of the channels in the preceding section.

We expect that capacity can be achieved by a (non-unique) ensemble with d inputs consisting of m orthogonal vectors in $E_m\mathcal{H}$ and $d-m$ orthogonal vectors in $E_m^\perp\mathcal{H}$. (There is no loss of generality in assuming that the optimal inputs can be written as $\rho_j = |e_j\rangle\langle e_j|$.) By symmetry the probabilities for such an optimal ensemble satisfy

$$\pi_j = \begin{cases} t & \text{for} \quad j \leqslant m \\ t^\perp & \text{for} \quad j > m \end{cases}$$

with $mt + (d-m)t^\perp = 1$. Thus $\rho_{\text{av}} = tE_m + t^\perp E_m^\perp$ and

$$\Phi(\rho_{\text{av}}) = atE_m + at^\perp E_m^\perp + (1-a)\frac{1}{d} I, \tag{45}$$

so that $C_{\text{Holv}}(\Phi)$ is the result of optimizing

$$S(\Phi(\rho_{\text{av}})) - mtS[\Phi(|e_1\rangle\langle e_1|)] - (d-m)t^\perp S[\Phi(|e_d\rangle\langle e_d|)]. \tag{46}$$

One finds that the optimal t satisfies

$$a \log \frac{adt^{\perp} + 1 - a}{adt + 1 - a} = -\Delta S, \tag{47}$$

where $\Delta S = S[\Phi(|e_d\rangle\langle e_d|)] - S[\Phi(|e_1\rangle\langle e_1|)] > 0$. This implies that, as expected, the solution will have $t > \frac{1}{d} > t^{\perp}$. It also agrees with (41) when $d = 3, m = 1$ and $x = t^{\perp}$. When $d = 2m$, (47) has the solution

$$t^{\perp} = \frac{1}{ad} \frac{a(1 + 2^{-\Delta S/a}) - (1 - 2^{-\Delta S/a})}{1 + 2^{-\Delta S/a}}. \tag{48}$$

5.3. Successively depolarizing channels

The next example generalizes the qutrit case in a different way. We now choose $V_k = E_1 \oplus W_k$ with $m = 1$ so that

$$\sum_k a_k V_k \rho V_k^{\dagger} = a\left[E_1 \rho E_1 \oplus \left(\sum_k b_k W_k E_1^{\perp} \rho E_1^{\perp} W_k^{\dagger}\right) + (1 - b)(\operatorname{Tr} E_1^{\perp} \rho) \frac{1}{d - 1} E_1^{\perp}\right] \tag{49}$$

with $\sum_k b_k = b$. Equivalently,

$$\Phi(\rho) = aE_1 \rho E_1 + \sum_k ab_k W_k E_1^{\perp} \rho E_1^{\perp} W_k^{\dagger} + a(1 - b)(\operatorname{Tr} E_1^{\perp} \rho) \frac{1}{d - 1} E_1^{\perp} + (1 - a)(\operatorname{Tr} \rho) \frac{1}{d} I. \tag{50}$$

Proceeding in this way, we can inductively construct a channel with the property that the input states $|e_j\rangle\langle e_j|$ have strictly increasing output entropies, with each minimal when Φ is restricted to states on $E_{j-1}^{\perp}$, except that the last pair have equal entropy, i.e., $S[\Phi(|e_{d-1}\rangle\langle e_{d-1}|)] = S[\Phi(|e_d\rangle\langle e_d|)]$.

We now make a change of notation so that $x_1 = \sum_k a_k$, $x_2 = \sum_k b_k$, etc. Then

$$\begin{aligned}
\Phi : |e_1\rangle\langle e_1| &\mapsto x_1 |e_1\rangle\langle e_1| + \frac{1 - x_1}{d} I \\
|e_2\rangle\langle e_2| &\mapsto x_1 x_2 |e_2\rangle\langle e_2| + x_1 \frac{1 - x_2}{d - 1} E_1^{\perp} + \frac{1 - x_1}{d} I \\
\vdots \qquad & \qquad \vdots \\
|e_m\rangle\langle e_m| &\mapsto \prod_{j=1}^{m} x_j |e_m\rangle\langle e_m| + \prod_{j=1}^{m-1} x_j \frac{1 - x_m}{d - m + 1} E_m^{\perp} + \cdots + \frac{1 - x_1}{d} I \\
\vdots \qquad & \qquad \vdots \\
|e_{d-1}\rangle\langle e_{d-1}| &\mapsto \prod_{j=1}^{d-1} x_j |e_{d-1}\rangle\langle e_{d-1}| + \prod_{j=1}^{d-2} x_j \frac{1 - x_{d-1}}{2} E_{d-1}^{\perp} \\
&\quad + \prod_{j=1}^{d-3} x_j \frac{1 - x_{d-2}}{3} E_{d-2}^{\perp} + \cdots + \frac{1 - x_1}{d} I \\
|e_d\rangle\langle e_d| &\mapsto \prod_{j=1}^{d-2} x_j (1 - x_{d-1}) |e_d\rangle\langle e_d| + \prod_{j=1}^{d-2} x_j \frac{x_{d-1}}{2} E_{d-1}^{\perp} \\
&\quad + \prod_{j=1}^{d-3} x_j \frac{1 - x_{d-2}}{3} E_{d-2}^{\perp} + \cdots + \frac{1 - x_1}{d} I.
\end{aligned}$$

5.4. Connection with CQ and classical channels

For a channel Φ of the type considered in the preceding sections, define $g_{jk} = \langle e_j|\Phi(|e_k\rangle\langle e_k|)|e_j\rangle$ so that

$$\Phi(|e_k\rangle\langle e_k|) = \sum_j g_{jk}|e_j\rangle\langle e_j|. \tag{51}$$

Explicit expressions for the channels in sections 5.2 and 5.3 are given in appendix C. The matrix G is column stochastic, and the 'successive' minimal entropy outputs are the same as for the CQ channel

$$\Phi_{\mathrm{CQ}}(\rho) = \sum_k \left(\sum_j g_{jk}|e_j\rangle\langle e_j|.\right) \operatorname{Tr}\rho|e_k\rangle\langle e_k|. \tag{52}$$

Under the assumption that the 'successive' minimal entropy inputs form a set of optimal inputs for the Holevo capacity, the optimization problem for the weights in the input ensemble $\{\pi_m, |e_m\rangle\langle e_m|\}$ is the same as for the corresponding CQ channel. Moreover, the bistochastic matrix G defines a classical channel acting on classical probability vectors in $\mathbf{R}^d$. The optimization problem for the Shannon capacity of this channel is the same as that for the Holevo capacity of the CQ channel (52).

We expect the behaviour of the examples in the previous sections to be similar to that of a qubit channel of the form

$$\tfrac{1}{2}[I + \mathbf{w}\cdot\sigma] \mapsto \tfrac{1}{2}\big[I + \lambda_1 w_1\sigma_1 + \lambda_2 w_2\sigma_2 + (t_3 + \lambda_3 w_3)\sigma_3\big] \tag{53}$$

with $\lambda_3 > \lambda_2 = \lambda_1$ so that image is a football and the only non-unital component is a translation along the longest axis. For such channels, it is well-known [9, 22] that the optimal inputs for the capacity C_{Holv} are the orthogonal states $\frac{1}{2}[I \pm \sigma_3]$, and the optimal weights are determined by the corresponding classical problem.

If the *conjecture* for the examples in the preceding sections (that the optimal inputs are orthogonal states which correspond to 'successive' minimal entropy inputs) holds, then, although unital, they behave like the non-unital qubit channel above, i.e., they are closely related to a CQ and a classical problem with the same probability distribution for the optimal ensemble. This has been verified numerically for the qutrit channels of section 5.1 and the double depolarizing channels of section 5.2.

6. Numerical determination of capacity

6.1. Description of the algorithms

Our numerical work is based on the following variant of the max–min principle (9*a*)–(9*c*)

$$C_{\mathrm{Holv}}(\Omega) \leqslant \sup_{\omega\in\mathcal{D}} H[\Omega(\omega), \Omega(\gamma)] \tag{54}$$

with equality if and only if $\Omega(\gamma) = \Omega(\rho_{\mathrm{av}})$. The equality condition follows from the argument in [28] which implies that if $\Omega(\rho_{\mathrm{av}}) \neq \Omega(\gamma)$, then at least one of the inputs ρ_j in an optimal signal ensemble must satisfy

$$H[\Omega(\rho_j), \Omega(\gamma)] \geqslant C_{\mathrm{Holv}}(\Omega) + H[\Omega(\rho_{\mathrm{av}}), \Omega(\gamma)] > C_{\mathrm{Holv}}(\Omega).$$

Note that this also implies that the optimal average output $\Omega(\rho_{\mathrm{av}})$ is unique, a fact which can be proven directly from the strict concavity of the entropy. This uniqueness is implicit in [20] and stated and proved explicitly in [29]. It can happen (as in the first example of section 2.4) that there is more than one optimal signal ensemble or optimal average input; however, the optimal average *output* of a channel is always unique.

Now suppose that we have a candidate for the optimal average output $\Omega(\rho_{\mathrm{av}}^{\star})$ and an associated candidate for the capacity $C_{\mathrm{Holv}}^{\star}(\Omega)$.

(a) If there is a state ω such that $C_{\mathrm{Holv}}^{\star}(\Omega) < H[\Omega(\omega), \Omega(\rho_{\mathrm{av}}^{\star})]$ we can conclude that the candidate is not the true optimal average.
(b) If $C_{\mathrm{Holv}}^{\star}(\Omega) = \sup_{\omega \in \mathcal{D}} H\big[\Omega(\omega), \Omega(\rho_{\mathrm{av}}^{\star})\big]$ we can conclude that we have found the true optimal average and capacity, at least up to the accuracy of the numerical work. Moreover, the states ω which achieve this supremum are the optimal inputs for Ω.

To find the supremum in (54), we used an algorithm based on an optimization principle of Shor[4] which is stated and proved as theorem 7 in appendix A. This algorithm finds relative, rather than absolute, maxima and is applied in situations in which some relative maxima are known (or expected) to satisfy (b) above. Therefore, for each channel tested, it is necessary to use it repeatedly with multiple inputs chosen to ensure that it will find a state satisfying (a) if one exists.

6.2. Numerical results

6.2.1. Single use of channel. We first tested our hypothesis that the 'successive' minimal entropy states for the examples in section 5 are optimal inputs for the Holevo capacity. If this hypothesis is correct, the weights for the optimal ensemble are given by the optimization problem of section 5.4. Numerical tests were done only for the qutrit channels of section 5.1 and the double depolarizing channels of section 5.2 in the case $d = 4, m = 2$, with parameter choices similar to those tested for additivity.

For the qutrit case, $\Phi(\rho_{\mathrm{av}}^{\star})$ and $C_{\mathrm{Holv}}^{\star}(\Phi)$ are given by (37) and (38) respectively with x given by (41). The parameters a_k were chosen so that $a_0 > a/2$, and $a_0 \geqslant a_1 \geqslant a_2 \geqslant a_3$ with $a = 0.5, 0.52, 0.54, \ldots, 0.9$ and for each of these $a_0 = a/2 + 0.05, a/2 + 0.1 \ldots$ until a_0 exceeds $a - 0.01$. For each of these pairs, we considered $a_j = (a - a_0)/3$ as well as a selection of parameters with $a_1 > a_2 > a_3$.

For the $d = 4, m = 2$ case, $\Phi(\rho_{\mathrm{av}}^{\star})$ is given by (45) and $C_{\mathrm{Holv}}^{\star}(\Phi)$ by (46) with $d = 4$, $m = 2$ and $t^{\perp}$ given by (48). All pairs of parameters a and b in the set $\{0.5, 0.55, 0.6, \ldots, 0.9\}$ were tested.

The starting inputs used in theorem 7 were chosen as follows. In both cases, for each set of parameters, 50 pure input states $|\psi\rangle\langle\psi|$ were obtained by normalizing the state $|\widetilde{\psi}\rangle = \sum_{k=1}^{d} r_k |k\rangle$ where $\{|k\rangle\}$ denotes the standard basis for $\mathbf{C}^d$ and the complex coefficients r_k were chosen randomly. In both cases, for all choices of parameters, $H[\Phi(\omega), \Phi(\rho_{\mathrm{av}}^{\star})] \leqslant C_{\mathrm{Holv}}^{\star}(\Phi)$ to an accuracy of 10 significant figures.

6.2.2. Additivity. We tested additivity of $C_{\mathrm{Holv}}(\Phi \otimes \Phi)$ for the channels of section 5.1 and those of section 5.2 with $d = 4, m = 2$. In both cases, $\Omega(\rho_{\mathrm{av}}^{\star}) = \Phi(\rho_{\mathrm{av}}) \otimes \Phi(\rho_{\mathrm{av}})$ and $C_{\mathrm{Holv}}^{\star}(\Omega) = 2C_{\mathrm{Holv}}(\Phi)$ with ρ_{av} and $C_{\mathrm{Holv}}(\Phi)$ being the expressions for a single use under the assumption that successively orthogonal minimal entropy inputs are optimal for the capacity. The assumption was tested numerically in the previous section. The results of this section give further support for this conjecture; if it were not true, one could find another pair of products with capacity greater than twice the $C_{\mathrm{Holv}}^{\star}(\Phi)$ from the previous section.

The algorithm in theorem 7 always yields a sequence ω_k for which $H[(\Phi \otimes \Phi)(\omega_k), \Phi(\rho_{\mathrm{av}}) \otimes \Phi(\rho_{\mathrm{av}})]$ is non-decreasing. Although the limiting state ω is stationary

[4] This observation is due to P W Shor. It was communicated to MBR by C King.

in the sense of (A.6), the eigenvalue λ need not equal the supremum in (54). Indeed, when testing additivity, products of optimal inputs will always be stationary states. Therefore, it is important to include starting points which do not automatically converge to these stationary points if others exist.

In choosing the parameters for testing additivity, it is reasonable to exclude values for which some restriction of the channel is entanglement breaking (EBT). Thus, we focus on values *well away* from the EBT regions for the corresponding depolarizing channel, i.e., $a \leqslant 0.25$ for $d = 3$ and $a \leqslant 0.2, b \leqslant \frac{1}{3}$ for $d = 4$ in section 5.2. Similarly, for qutrits, we choose $a_0 > \frac{1}{2}a$. We do not claim that channels with some EBT parameters are EBT or that we can prove additivity. However, it would be quite extraordinary if a channel of the form (43) with parameters in (or near) the EBT regions were super-addditive when those with larger values were not.

Because the double depolarizing examples offer possibilities for entanglement across regions in ways not previously tested numerically, we concentrated on this case. For $d = 4, m = 2$, we considered all pairs of parameters a, b in the set $\{0.5, 0.52, 0.54, \dots 0.98\}$. For each pair, we used the following selection of input states (which are described with the convention that $|k\rangle$ denotes the standard basis in $\mathbf{C}^4$).

(i) 10 random pure states $|\psi\rangle\langle\psi|$, where $|\psi\rangle$ is obtained by normalizing the state

$$|\widetilde{\psi}\rangle = \sum_{j=1}^{4}\sum_{j=1}^{4} r_{jk}|j\rangle \otimes |j\rangle,$$

with complex coefficients r_{jk} chosen randomly.

(ii) 10 maximally entangled input states $|\psi\rangle\langle\psi|$, where

$$|\psi\rangle = c_1|1\rangle \otimes |3\rangle + c_2|2\rangle \otimes |4\rangle + c_3|3\rangle \otimes |2\rangle + c_4|4\rangle \otimes |1\rangle,$$

with $c_k = (1/2)\exp(\mathrm{i}\theta_k)$ and θ_k chosen randomly in $[0, 2\pi]$.

(iii) 10 pure input states $|\psi\rangle\langle\psi|$, where $|\psi\rangle$ is obtained by normalizing the state

$$|\widetilde{\psi}\rangle = \sum_{i=1}^{4} |\phi_i\rangle \otimes |\phi_i\rangle,$$

with each $|\phi_i\rangle$ chosen randomly as in section 6.2.1.

For $d = 3$, the same parameter values were used as in section 6.2.1 with 30 random input pure states chosen as described in (i) above.

In all the situations tested, $C_{\mathrm{Holv}}(\Phi \otimes \Phi)$ agrees with $2C_{\mathrm{Holv}}(\Phi)$ to 10 significant figures.

7. Discussion

We have considered the effect of modifying a depolarizing channel by replacing $a\rho$, the first term in (1), by different convex combinations of unitary conjugations. We have shown that this leads to a rich variety of examples, some of which exhibit behaviour previously associated with non-unital channels. Nevertheless, we prove a number of results, including the additivity of minimal output entropy.

To relate our results to other recent work, let $M(\rho) = \sum_k x_k V_k \rho V_k^\dagger$ with $x_k = \frac{a_k}{a}$ as in (2). Then the channel in (2) can be written as $\Phi = \Gamma_a^{\mathrm{dep}} \circ M$, and Fukuda's lemma [10] can be applied to give an alternate proof of parts (b) and (d) of theorem 3. When the V_k have a common eigenvector, $M(\rho)$ has an output state of rank one so that Fukuda's lemma can be applied to the composition of $M(\rho)$ with other unitarily invariant channels as discussed in [10]. In addition, the channel $T(\rho) = \frac{1}{d-1}[(\mathrm{Tr}\,\rho)I - M(\rho)]$ has an output which is a multiple of a projection. Therefore, the results of Wolf and Eisert [33] imply that additivity (6) and multiplicativity (20) with $1 \leqslant p \leqslant 2$ hold for tensor products of channels $T(\rho)$ in the 'strong' sense defined in [33]. Channels $M(\rho)$ generated from diagonal V_k as in section 4 were considered in [33]; however, using the V_k from the asymmetric examples of section 5 to generate $T(\rho)$ via $M(\rho)$ gives new examples.

Instead of modifying the first term in (1), one could change the second to obtain the channel

$$\Phi(\rho) = a\rho + (1-a)(\mathrm{Tr}\,\rho)\gamma \tag{55}$$

with γ being a fixed density matrix. The simplest such example is the shifted depolarizing channel $\gamma = \frac{1}{d}(1-b)I + b|\psi\rangle\langle\psi|$, for which additivity (6) and multiplicativity (20) for all $p \geqslant 1$ have now been proved by Fukuda [10]. However, the only results which have been proved for the general channel (55) are multiplicativity in the case $p = 2$ [12], and higher integers [21]. Despite recent progress for special cases, resolving the additivity conjectures remains a challenge.

Acknowledgments

This work began when M B R was a participant in the program on quantum information at the Isaac Newton Institute at Cambridge University in 2004, and benefitted greatly from the stimulating environment there. N D would like to thank Daniel Oi and Alastair Kay for help with Mathematica.

Appendix A. Shor's optimization algorithm

Our numerical results use the following theorem due to Shor (see footnote 4 on page 9797).

Theorem 7. *Let Ω be a CPT map and $\widehat{\Omega}$ its adjoint with respect to the Hilbert–Schmidt inner product. Let ψ be the eigenvector corresponding to the largest eigenvalue of $\widehat{\Omega}[\log\Omega(\rho) - \log A]$. Then $H[\Omega(|\psi\rangle\langle\psi|), A] \geqslant H[\Omega(\rho), A]$.*

Proof. The largest eigenvalue of $\widehat{\Omega}[\log\Omega(\rho) - \log A]$ is

$$\lambda = \sup_{\psi}\langle\psi, \widehat{\Omega}[\log\Omega(\rho) - \log A]\psi\rangle \tag{A.1}$$

$$= \sup_{\psi} \mathrm{Tr}\,|\psi\rangle\langle\psi|\widehat{\Omega}[\log\Omega(\rho) - \log A], \tag{A.2}$$

where the supremum is over vectors ψ with $\|\psi\| = 1$. Let $\gamma = |\psi\rangle\langle\psi|$ for the state which attains this supremum. Then

$$\begin{aligned}\mathrm{Tr}\,\Omega(\gamma)[\log\Omega(\rho) - \log A] &= \mathrm{Tr}\,\gamma\widehat{\Omega}[\log\Omega(\rho) - \log A]\\ &\geqslant \mathrm{Tr}\,\rho\widehat{\Omega}[\log\Omega(\rho) - \log A]\\ &= H[\Omega(\rho), A]\end{aligned} \tag{A.3}$$

so that

$$H[\Omega(\gamma), A] - H[\Omega(\rho), A]$$

$$= H[\Omega(\gamma), \Omega(\rho)] + \mathrm{Tr}\,\Omega(\gamma)[\log\Omega(\rho) - \log A] - H[\Omega(\rho), A] \tag{A.4}$$

$$\geqslant 0. \tag{A.5}$$

□

Given a starting $\rho = |\psi_0\rangle\langle\psi_0|$, let $\gamma_1 = \gamma = |\psi_1\rangle\langle\psi_1|$ be the eigenvector before (A.3), and inductively define $\gamma_{k+1} = |\psi_{k+1}\rangle\langle\psi_{k+1}|$ using the eigenvalue equation for γ_k. This gives a sequence for which $H[\Omega(\gamma_k), \Omega(\rho)]$ increases to a stationary point ω satisfying

$$\widehat{\Omega}[\log\Omega(\rho) - \log A]\omega = \lambda\omega. \tag{A.6}$$

Appendix B. Qubit channel details

It was shown in [22] that any unital qubit channel can be written as

$$\Phi(\rho) = V\left[\sum_{k=0}^{3}\alpha_k\sigma_k(U\rho U^\dagger)\sigma_k\right]V^\dagger \tag{B.1}$$

with U, V being unitary, the $\alpha_k > 0$ with $\sum_k \alpha_k = 1, \sigma_0 = I$ and $\sigma_j, j = 1, 2, 3$ being the three Pauli matrices. There is no loss of generality in assuming that $\alpha_0 \geqslant \alpha_j$ $(j = 1, 2, 3)$; if, instead, α_j is largest, one can factor out σ_j and rewrite Φ in the form (B.1) with $V \to V\sigma_j$. Similarly, one can choose U, V to correspond to rotations in $\mathbf{R}^3$ so that $\alpha_1 \geqslant \alpha_j$ $(j = 2, 3)$. Finally, since the only effect of U, V is to make change of bases which have no effect on the minimal output entropy or the Holevo capacity, we can assume that $U = V = I$. Thus, there is no loss of generality in assuming that Φ has the form (15) with $\alpha_0 \geqslant \alpha_1 \geqslant \alpha_j, j = 2, 3$. If, in addition, $\alpha_0 > \frac{1}{2}$, the channel is *not* EBT [26]. Thus, we often assume that

$$\alpha_0 > \tfrac{1}{2} \geqslant \alpha_1 \geqslant \alpha_j, \qquad j = 2, 3. \tag{B.2}$$

The parameters $\alpha_k, k = 0, 1, 2, 3$ and $\lambda_i, i = 1, 2, 3$, in (15) and (16) are related by the conditions

$$1 = \alpha_0 + \alpha_1 + \alpha_2 + \alpha_3 \tag{B.3}$$

$$\lambda_i = \alpha_0 + \alpha_i - \alpha_j - \alpha_l = 2(\alpha_0 + \alpha_i) - 1 \tag{B.4}$$

with the understanding that i, j, l are distinct. Then the input states $\frac{1}{2}(I \pm \sigma_i)$ have output states $\frac{1}{2}(I \pm \lambda_i\sigma_i)$ whose eigenvalues are

$$\tfrac{1}{2}(1 \pm \lambda_i) = \begin{cases} \alpha_0 + \alpha_i \\ \alpha_j + \alpha_l = 1 - \alpha_0 - \alpha_i. \end{cases} \tag{B.5}$$

The image of the Bloch sphere is an ellipsoid whose axes have lengths $|\lambda_j|, j = 1, 2, 3$, with the output states above at the ends of the axes. Under the order assumption (B.2), all $\lambda_j \geqslant 0$ and the states with optimal output purity satisfy (B.5) with $i = 1$.

In the discussion of section 5.1, $\alpha_k = \frac{a_k}{a}$ and one uses suitably modified forms of equations (B.2)–(B.5).

Appendix C. CQ matrices

For a channel Φ of the type considered in section 5.2, the matrix defined in (51) is given by

$$g_{jk} = \begin{cases} a + \dfrac{1-a}{d} & j = k \leqslant m \\[2ex] \dfrac{1-a}{d} & j \neq k,\ j \leqslant m \text{ or } k \leqslant m \\[2ex] ab + \dfrac{a(1-b)}{d-m} + \dfrac{1-a}{d} & j = k > m \\[2ex] \dfrac{a(1-b)}{d-m} + \dfrac{1-a}{d} & j \neq k,\ j, k \leqslant m. \end{cases} \tag{C.1}$$

For a channel of the type considered in section 5.3, it is

$$g_{jk} = \begin{cases} \frac{1-x_1}{d} & k > 1, j = 1 \\ g_{k,j-1} + \prod_{i=1}^{j-1} x_i \frac{1-x_i}{d-i+1} & k > j > 1 \\ g_{j+1,j} + \prod_{i=1}^{j} x_i & k = j < d \\ g_{jk} & k < j \\ g_{d-1,d-1} & k = j = d. \end{cases} \tag{C.2}$$

References

[1] Audenaert K M R and Braunstein S L 2004 On strong superadditivity of the entanglement of formation *Commun. Math. Phys.* **246** 443–52

[2] Alicki R and Fannes M 2004 Note on multiple additivity of minimal output entropy output of extreme $SU(d)$-covariant channels *Open Systems Inf. Dynam.* **11** 339–42 (*Preprint* quant-ph/0407033)

[3] Amosov G G, Holevo A S and Werner R F 2000 On some additivity problems in quantum information theory *Probl. Inform. Transm.* **36** 305–13 (*Preprint* math-ph/0003002)

[4] Cortese J 2002 The Holevo–Schumacher–Westmoreland channel capacity for a class of qudit unital channels *Preprint* quant-ph/0211093

[5] Fuchs C 1997 Nonorthogonal quantum states maximize classical information capacity *Phys. Rev. Lett.* **79** 1162–5 (*Preprint* quant-ph/9703043)

[6] Fuchs C, Shor P, Smolin J and Terhal B unpublished work mentioned at the end of [5]

[7] Fujiwara A and Algoet P 1999 One-to-one parametrization of quantum channels *Phys. Rev.* A **59** 3290–4

[8] Fujiwara A and Hashizumé T 2002 Additivity of the capacity of depolarizing channels *Phys. Lett.* A **299** 469–75

[9] Fujiwara A and Nagaoka H 1988 Operational capacity and pseudoclassicality of a quantum channel *IEEE Trans. Inf. Theor.* **44** 1071–86

[10] Fukuda M 2005 Extending additivity from symmetric to asymmetric channels *J. Phys. A: Math. Gen.* **38** L753–8 (*Preprint* quant-ph/0505022)

[11] Hayashi M, Imai H, Matsumoto K, Ruskai M B and Shimono T 2005 Qubit channels which require four inputs to achieve capacity: implications for additivity conjectures *Quantum Inf. Comput.* **5** 13–31 (*Preprint* quant-ph/0403176)

[12] Giovannetti V, Lloyd S and Ruskai M B 2005 Conditions for multiplicativity of maximal l_p-norms of channels for fixed integer p *J. Math. Phys.* **46** 042105 (*Preprint* quant-ph/0408103)

[13] Holevo A S 2002 Remarks on the classical capacity of quantum channel *Preprint* quant-ph/0212025

[14] Horn R A and Johnson C R 1985 *Matrix Analysis* (Cambridge: Cambridge University Press)

[15] Horodecki M, Shor P and Ruskai M B 2003 Entanglement breaking channels *Rev. Math. Phys.* **15** 629–41 (*Preprint* quant-ph/030203)

[16] King C 2002 Additivity for unital qubit channels *J. Math. Phys.* **43** 4641–53

[17] King C 2003 The capacity of the quantum depolarizing channel *IEEE Trans. Inf. Theor.* **49** 221–9

[18] King C 2003 Maximal p-norms of entanglement breaking channels *Quantum Inf. Comput.* **3** 186–90

[19] King C 2003 An application of the Lieb–Thirring inequality in quantum information theory *Proc. ICMP* in press

[20] King C, Nathanson M and Ruskai M B 2002 Qubit channels can require more than two inputs to achieve capacity *Phys. Rev. Lett.* **88** 057901 (*Preprint* quant-ph/0109079)

[21] King C, Nathanson M and Ruskai M B 2005 Multiplicativity results for entrywise positive maps *Linear Alg. Appl.* **404** 367–9 (*Preprint* quant-ph/0409181)

[22] King C and Ruskai M B 2001 Minimal entropy of states emerging from noisy quantum channels *IEEE Trans. Info. Theor.* **47** 192–209

[23] Marshall A W and Olkin I 1979 *Inequalities: Theory of Majorization and its Applications* (New York: Academic)

[24] Ohya M, Petz D and Watanabe N 1997 On capacities of quantum channels *Prob. Math. Stat.* **17** 170–96

[25] Ohya M and Petz D 1993 *Quantum Entropy and Its Use* (Berlin: Springer)

[26] Ruskai M B 2003 Qubit entanglement breaking channels *Rev. Math. Phys.* **15** 643–62

[27] Ruskai M B, Szarek S and Werner E 2002 An analysis of completely positive trace-preserving maps M_2 *Linear Alg. Appl.* **347** 159

[28] Schumacher B and Westmoreland M D 2001 Optimal signal ensembles *Phys. Rev.* A **63** 022308 (*Preprint* quant-ph/9912122)

[29] Shirokov M E 2004 On the structure of optimal sets for tensor product channel *Preprint* quant-ph/0402178

[30] Shor P 2002 Additivity of the classical capacity of entanglement-breaking quantum channels *J. Math. Phys.* **43** 4334–40

[31] Shor P W 2004 Equivalence of additivity questions in quantum information theory *Commun. Math. Phys.* **246** 453–72

[32] Werner R F and Holevo A S 2002 Counterexample to an additivity conjecture for output purity of quantum channels *J. Math. Phys.* **43** 4353–7

[33] Wolf M M and Eisert J 2005 Classical information capacity of a class of quantum channels *New J. Phys.* **7** 93 (*Preprint* quant-ph/0412133)