Entanglement of random subspaces via the Hastings bound

Motohisa Fukuda
University of California, Davis CA 95616

Christopher King
Northeastern University, Boston MA 02115

July 30, 2009

Abstract

Recently Hastings [16] proved the existence of random unitary channels which violate the additivity conjecture. In this paper we use Hastings’ method to derive new bounds for the entanglement of random subspaces of bipartite systems. As an application we use these bounds to prove the existence of non-unital channels which violate additivity of minimal output entropy.

Contents

1 Introduction 2
1.1 Entanglement of subspaces 2
1.2 Violations of additivity 5

2 The Hastings bound 6
2.1 Notation 6
2.2 Random embeddings 6
2.3 Definition of the tube 7
2.4 Statement of the bound 7

3 Proof of Theorem 2 8

4 Proof of Theorem 3 11

5 Proof of the Hastings bound 13
5.1 Random states 13
5.2 Random embeddings yield random output states 14
5.3 Typical channels 14
5.4 The proof 15
1 Introduction

In his 2008 paper [16] M. Hastings proved the existence of channels which exhibit non-additivity of minimal output entropy. This result settled a long-standing open problem in quantum information theory. Hastings’ paper is interesting from many points of view, not least because it introduced some essentially new ideas into the field of random channels. To review the history a little, in earlier work Hayden, Leung and Winter [17] had derived bounds for the entanglement of random subspaces of bipartite spaces (these bounds are recalled below in Theorem 1). They used concentration of measure arguments to analyze the entropy of random states in high-dimensional spaces, together with the “ε-net” method to control the entropy of all states in a subspace. Their analysis led to the proofs by Hayden and Winter [18] of the existence of channels violating additivity of Renyi entropy for all \( p > 1 \). Further progress in this direction appeared in the recent work of Collins and Nechita [8, 9] on Renyi entropies of entangled states and subspaces. However the \( p = 1 \) case remained open until Hastings provided the new ingredients to complete this program.

Our goal in this paper is to apply these new methods from the paper [16] to the analysis of random subspaces of bipartite spaces. As an application we derive results about the entanglement of a generic high-dimensional subspace, and show that in some regimes this provides strictly tighter bounds than the Hayden, Leung and Winter estimates. We also use these bounds to deduce the existence of non-unital channels which violate additivity of minimal output entropy, and in fact show that such violation is generic for high-dimensional channels. In the process of deriving these results we formulate an abstract version of Hastings’ method, and we believe that this formulation will be useful for the study of other generic properties of random subspaces.

The idea of using Hastings’ method to study entanglement of random subspaces also appeared recently in the work of Brandao and Horodecki [5]. Their work is particularly interesting because it uses a combination of standard concentration of measure arguments together with some of the new ideas of Hastings. There is some overlap between their paper and ours, and in particular we re-derive their entanglement bound as a special case of our Theorem 2. However we also extend their results in several ways, both by considering different dimensions for input and output spaces, and by presenting explicit bounds for the size of the additivity violations.

The paper is organized as follows. In the rest of this Introduction we recall the entanglement bounds derived by Hayden, Leung and Winter, and state the new bounds derived using Hastings’ method. We then use these bounds to prove the existence of a new class of channels with non-additive minimal output entropy. Section 2 contains the main result of this paper, which is a general formulation of the Hastings bound. With an eye to possible future applications we state this as broadly as possible, namely as a condition which guarantees convergence to zero of the...
probabilities of a sequence of events in the output space. In Sections 3 and 4 we use the Hastings bound to derive the entanglement results in Section 1. In Section 5 we prove the Hastings bound, using methods similar to those in the paper [14]. Finally the Appendix contains some technical estimates needed for the derivations of the bounds.

1.1 Entanglement of subspaces

Consider a subspace $C$ of the bipartite system $A \otimes B$. The entanglement of $C \subset A \otimes B$ is defined to be

$$E(C) = \inf_{|\phi\rangle} S\left(\text{Tr}_B|\phi\rangle\langle\phi|\right)$$  \hspace{1cm} (1.1)

where the infimum runs over normalized states $|\phi\rangle$ in $C$, $S(\cdot)$ is the von Neumann entropy, and $\text{Tr}_B|\phi\rangle\langle\phi|$ is the reduced density matrix of the orthogonal projector onto $|\phi\rangle$. Note that $E(C) \geq 0$ with equality if and only if $C$ contains a product state.

In the search for counterexamples to additivity, one is interested in finding subspaces with large entanglement. Thus the quantity $\sup E(C)$ is of interest, where the supremum runs over all subspaces of a fixed dimension. This supremum depends only on the dimensions of $A, B, C$.

Let $d = \text{Dim } A, n = \text{Dim } B$ and $s = \text{Dim } C$, then the maximum entanglement of a $s$-dimensional subspace in $A \otimes B$ is

$$E_{\text{max}}(s,d,n) = \sup\{E(C) : C \subset A \otimes B, \quad \text{Dim } C = s\}$$ \hspace{1cm} (1.2)

Hayden, Leung and Winter [17] obtained the following lower bounds for $E_{\text{max}}$.

**Theorem 1 (Hayden, Leung and Winter)** Assume that $3 \leq d \leq n$. Then

$$E_{\text{max}}(s,d,n) \geq \log d - c_1 \frac{d}{n} - c_2 \left(\frac{s + 1}{dn}\right)^{2/5} \log d$$ \hspace{1cm} (1.3)

where $c_1 \simeq 1.44$ and $c_2 \simeq 19.84$.

The main results we present in this paper are new lower bounds for $E_{\text{max}}(s,d,n)$. The bounds are valid for sufficiently large dimensions $n$ and $s$, and for any dimension $d \geq 2$. Theorem 2 concerns the case where $s$ scales linearly with $n$, and Theorem 3 covers the case where $s/n \rightarrow 0$ as $n \rightarrow \infty$.

In order to state our first result we need to introduce the solution of the following optimization problem: for $x,y > 0$ define

$$h_d(x, y) = \inf_{0 < \gamma < 1} \inf_{z > 1} \left\{ \frac{z \log z + (d - z) \log \left(\frac{d + 1}{d - z}\right)}{x (\gamma + (1 - \gamma) \log(1 - \gamma))} : \frac{- \log z - (d - 1) \log \left(\frac{d + 1}{d - z}\right)}{- \log(1 - \gamma)} = y \right\}$$ \hspace{1cm} (1.4)

**Theorem 2** Let $d \geq 2, 0 < r_1 \leq r_2$, and $h > h_d(r_1, r_2)$. There is $n_0 < \infty$ such that for $n \geq n_0$, and all $s$ satisfying $r_1 \leq s/n \leq r_2$,

$$E_{\text{max}}(s,d,n) > \log d - h \left(\frac{s}{nd}\right)$$ \hspace{1cm} (1.5)
The above result is a generic property, meaning that with probability approaching one as $s, n \to \infty$, the right side of (1.5) is a lower bound for the entanglement $E(C)$ of a randomly selected subspace $C$. It is possible to analyze the function $h_d$ in detail but for our purposes here it is sufficient to note that it satisfies an upper bound which is uniform in $d$. As mentioned before, using related methods Brandao and Horodecki [5] proved Theorem 2 in the case $r_1 = r_2 = 1$.

Our second result concerns the case where $s/n \to 0$. Define

$$h_0 = \inf_{0<\gamma<1} \frac{-\log(1-\gamma)}{\gamma + (1-\gamma)\log(1-\gamma)} \simeq 3.351 \quad (1.6)$$

**Theorem 3** Let $d \geq 2$, and $h > h_0$. Consider sequences $\{s_k, n_k\}$ such that

$$\frac{s_k}{n_k} \to 0, \quad \frac{n_k \log s_k}{s_k^{3/2}} \to 0 \quad \text{as} \quad k \to \infty \quad (1.7)$$

There is $k_0 < \infty$ such that for $k \geq k_0$,

$$E_{\text{max}}(s_k, d, n_k) > \log d - h \left(\frac{s_k}{n_k d}\right) \quad (1.8)$$

Again we note that the lower bound in (1.8) is generic for random subspaces in high dimensions. The bounds (1.5), (1.8) and (1.3) can be compared for small values of the ratio $s/nd$. It can be seen that the right side of (1.5), (1.8) behaves like $\log d - c(s/nd)$, while the right side of (1.3) behaves like $\log d - c'(s/nd)^{2/5} \log d$ for some constants $c, c'$. Thus (1.5), (1.8) provide a sharper bound in the regime $s << nd$.

**Corollary 4** Let $d \geq 2$, $h > h_0$ and $0 < \epsilon < 1/2$. Then, there exists $s_0$ such that

$$E_{\text{max}}(s, d, sd) > \log d - h \left(\frac{1}{s^{1/2-\epsilon} d}\right) \quad (1.9)$$

for all $s \geq s_0$.

To see this, let $n = \lceil s^{3/2-\epsilon} \rceil$ and then Theorem 3 implies that there exists $s_0$ such that

$$E_{\text{max}}(s, d, \lceil s^{3/2-\epsilon} \rceil) > \log d - h \left(\frac{s}{\lceil s^{3/2-\epsilon} \rceil d}\right) > \log d - h \left(\frac{1}{s^{1/2-\epsilon} d}\right) \quad (1.10)$$

for $s \geq s_0$. Without loss of generality, we can assume that $d < s_0^{1/2-\epsilon}$, which implies $sd \leq \lceil s^{3/2-\epsilon} \rceil$ for $s \geq s_0$. As described in the next section, there is a correspondence between subspaces of bipartite spaces and quantum channels. Thus a subspace which satisfies the bound in (1.10) corresponds to some quantum channel where the dimensions of input and output spaces and the number of Kraus operators are $s, d, \lceil s^{3/2-\epsilon} \rceil$ respectively. However, this channel can be rewritten by using at most $sd$ Kraus operators [34],[28]. Therefore this channel corresponds to some $s$-dimensional subspace, say $C$, of $\mathbb{C}^d \otimes \mathbb{C}^{sd}$, for which $E(C)$ satisfies the same lower bound (1.10).
1.2 Violations of additivity

The subspace $C \subset A \otimes B$ is defined by an embedding $W : \mathbb{C}^s \to \mathbb{C}^d \otimes \mathbb{C}^n$ satisfying $W^*W = I$, where $C$ is the image of $W$. This embedding defines two conjugate channels $\Phi_W$ and $\Phi_W^C$ via

$$\Phi_W(\rho) = \text{Tr}_{C^d} W \rho W^*, \quad \Phi_W^C(\rho) = \text{Tr}_{C^n} W \rho W^*$$

(1.11)

Letting $W^*$ denote the complex conjugate of the matrix $W$, the complex conjugate channels $\Phi_W^C$ and $\Phi_W^C$ are defined by

$$\Phi_W(\rho) = \text{Tr}_{C^d} W^* \rho W, \quad \Phi_W^C(\rho) = \text{Tr}_{C^n} W^* \rho W^*$$

(1.12)

It follows that

$$E(C) = S_{\min}(\Phi_W) = S_{\min}(\Phi_W^C)$$

(1.13)

Our next result gives a universal upper bound for the minimum entropy of any product channel of the form $\Phi \otimes \Phi$, depending only on the dimensions of the spaces (a similar bound was derived in [5] for the case $s = n$).

**Theorem 5** Let $p = s/dn$, and assume that $sd/n \geq 1$, then

$$S_{\min}(\Phi \otimes \Phi) \leq (1 - p) \log(d^2 - 1) - p \log p - (1 - p) \log(1 - p)$$

(1.14)

Theorem 5 will be proved in the Appendix. We will now use Theorems 2 and 5 to demonstrate the existence of channels of the form $\Phi \otimes \Phi$ violating additivity. For such a product channel the violation of additivity is given by

$$\Delta S(\Phi) = S_{\min}(\Phi) + S_{\min}(\Phi) - S_{\min}(\Phi \otimes \Phi) = 2S_{\min}(\Phi) - S_{\min}(\Phi \otimes \Phi)$$

(1.15)

Theorem 2 guarantees the existence of subspaces satisfying the bound (1.5), and hence also the existence of channels $\Phi$ for which $S_{\min}(\Phi)$ satisfies the same bound. Taking $s = n$ (so $r_1 = r_2 = 1$), this implies the existence of channels for which

$$S_{\min}(\Phi) > \log d - \frac{h_d(1, 1)}{d}$$

(1.16)

for sufficiently large $n$. Combining the bounds (1.16) and (1.14), and estimating $\log(d^2 - 1) \leq 2 \log d$ in (1.14), we obtain

$$\Delta S(\Phi) \geq p \log(pd^2) + (1 - p) \log(1 - p) - \frac{2}{d} h_d(1, 1)$$

(1.17)

where $p = s/nd = 1/d$. Using (1.17) and the inequality $(1 - p) \log(1 - p) \geq -p$, we get for sufficiently large $n$

$$\Delta S(\Phi) \geq \frac{1}{d} \left[ \log d - 2h_d(1, 1) - 1 \right]$$

(1.18)
For $d > \exp[2h_d(1, 1) + 1]$ the right side of (1.18) is positive and hence these channels violate additivity (recall that $h_d(1, 1)$ is upper bounded uniformly in $d$). Furthermore, the method of proof shows that this violation occurs with positive probability for a randomly selected subspace, and hence for a randomly selected channel. Since the unital channels have measure zero in the set of all channels of fixed dimensions $s, n, d$, this implies the existence of non-unital channels which violate additivity.

2 The Hastings bound

This section contains an ‘abstract’ version of the Hastings bound. Much of the notation was introduced previously in [14], and we will use several technical results from that paper.

2.1 Notation

$\mathcal{M}_n$ will denote the algebra of complex $n \times n$ matrices; the identity matrix will be written $I$; $\mathcal{U}(n)$ will denote the group of unitary matrices. The set of states in $\mathcal{M}_n$ is defined as

$$S_n = \{ \rho \in \mathcal{M}_n : \rho = \rho^*, \quad \text{Tr} \rho = 1 \} \quad (2.1)$$

The set of pure states in $\mathcal{M}_n$ is identified with the unit vectors in $\mathbb{C}^n$ and denoted

$$V_n = \{ \ket{\psi} \in \mathbb{C}^n : \bra{\psi} \psi \rangle = 1 \} \quad (2.2)$$

We write $\mathcal{R}(s, n, d)$ for the set of all embeddings $W : \mathbb{C}^s \to \mathbb{C}^d \otimes \mathbb{C}^n$, with $W^*W = I$. There is a one-to-one correspondence between such embeddings and pairs of complementary channels $\Phi_W : \mathcal{M}_s \to \mathcal{M}_n$, $\Phi_W^C : \mathcal{M}_s \to \mathcal{M}_d$ defined by

$$\Phi_W(\rho) = \text{Tr}_{C^d} W \rho W^*, \quad \Phi_W^C(\rho) = \text{Tr}_{C^n} W \rho W^* \quad (2.3)$$

Thus $\mathcal{R}(s, n, d)$ is also the set of all such pairs of conjugate channels. The image of the pure input states under the action of a channel $\Phi_W^C : \mathcal{M}_s \to \mathcal{M}_d$ will be denoted

$$\text{Im}(\Phi_W^C) = \{ \Phi_W^C(\ket{\phi}\bra{\phi}) \in S_d : \ket{\phi} \in V_s \} \quad (2.4)$$

2.2 Random embeddings

We define a probability measure $\mathcal{P}_{s,n,d}$ on the set of embeddings $\mathcal{R}(s, n, d)$ as follows. Let $W_0$ be a fixed embedding $W_0 : \mathbb{C}^s \to \mathbb{C}^d \otimes \mathbb{C}^n$ satisfying $W_0^*W_0 = I$. Then every embedding $W \in \mathcal{R}(s, n, d)$ can be written as

$$W = UW_0, \quad U^*U = I \quad (2.5)$$

for some unitary matrix $U \in \mathcal{U}(nd)$. Let $\text{Stab}(W_0)$ be the subgroup of unitary matrices which leave invariant every vector in the image of the embedding $W_0$. Then two unitary matrices $U_1, U_2$ define
the same embedding if \( U_1^{-1} U_2 \in \text{Stab}(W_0) \). Thus \( R(s, n, d) \) can be identified with the left cosets of the group of unitary matrices with respect to the subgroup \( \text{Stab}(W_0) \). Let \( \Pi \) be the projection from \( \mathcal{U}(nd) \) onto these cosets. Then the normalized Haar measure \( Haar \) on \( \mathcal{U}(nd) \) descends to a normalized measure \( \Pi^*(\text{Haar}) \) on this set of cosets, and this defines our probability measure \( P_{s,n,d} \).

2.3 Definition of the tube

We recall the notion of the ‘tube’ at a state \( \rho \), as defined in \([14]\). First, for any \( \rho \in \mathcal{S}_d \) and \( 0 < \gamma < 1 \) define \( L_\gamma(\rho) \) to be the following line segment pointing from \( \rho \) toward the maximally mixed state \( I/d \):

\[
L_\gamma(\rho) = \left\{ r\rho + (1-r)\frac{1}{d}I : \gamma \leq r \leq 1 \right\}
\] (2.6)

Then the tube at \( \rho \) is defined to be the set of states which lie within a small distance of the set \( L_\gamma(\rho) \). Also for any event \( C \subset \mathcal{S}_d \), the tube at \( C \) is the union of the tubes at all states in \( C \).

**Definition 6** Let \( \rho \in \mathcal{S}_d \), then the Tube at \( \rho \) is defined as

\[
\text{Tube}(\rho) = \left\{ \theta \in \mathcal{S}_d : \text{dist}(\theta, L_\gamma(\rho)) \leq 2 \sqrt{\frac{\log n}{n}} + 13 d \sqrt{\frac{\log d}{s}} \right\}
\] (2.7)

where \( \text{dist}(\theta, L(\rho)) = \inf_{\tau \in L(\rho)} ||\theta - \tau||_\infty \). For any output event \( C \subset \mathcal{S}_d \) the Tube at \( C \) is defined as

\[
\text{Tube}(C) = \bigcup_{\rho \in C} \text{Tube}(\rho)
\] (2.8)

2.4 Statement of the bound

Suppose that for each triplet \((s, n, d)\) there is given an event \( C(s, n, d) \subset \mathcal{S}_d \). We want to find conditions which will show that for sufficiently large dimensions \( s, n \) there is a nonzero probability that for a randomly selected embedding \( W \) the event \( C(s, n, d) \) will not contain any output states of the form \( \Phi_W^C(|\phi\rangle\langle\phi|) \). We analyze this by considering the sequence of complementary events, namely the events \( \{W : \text{Im}(\Phi_W^C) \cap C(s, n, d) \neq \emptyset\} \), and showing that their probabilities approach zero as \( s, n \to \infty \).

**Theorem 7 (The Hastings Bound)** Let \( \{C(s, n, d) \subset \mathcal{S}_d\} \) be a collection of output events defined for each triplet of dimensions \((s, n, d)\). Fix \( d \geq 2 \), consider sequences \( \{s_k, n_k\} \to \infty \), and define

\[
B_k = \{W \in R(s_k, n_k, d) : \text{Im}(\Phi_W^C) \cap C(s_k, n_k, d) \neq \emptyset\}
\] (2.9)

Suppose there is \( \gamma \in (0, 1) \) such that

\[
\lim_{k \to \infty} \left( d^2 \log n_k + n_k d \log d + (n_k - d)M(\gamma, k) - s_k \log(1 - \gamma) \right) = -\infty
\] (2.10)
where

\[ M(\gamma, k) = \sup \{ \text{Tr} \log \rho : \rho \in \text{Tube}(C(s_k, n_k, d)) \} \]  

(2.11)

Then

\[ \lim_{k \to \infty} \mathcal{P}_{s_k, n_k, d}(B_k) = 0 \]  

(2.12)

As a consequence of the Theorem, if the conditions are satisfied then for sufficiently large \( k \) we have \( \mathcal{P}_{s_k, n_k, d}(B_k) < 1 \), and hence there must exist embeddings \( W \) such that \( \Phi_{W}(|\phi \rangle \langle \phi|) \notin C(s_k, n_k, d) \) for any input state \( |\phi \rangle \). Theorem 7 will be proved later in Section 5. First we use it to deduce Theorems 2 and 3.

### 3 Proof of Theorem 2

In this section we apply the Hastings bound to prove Theorem 2. Fix dimension \( d \) and the parameter \( h \). In the following, we consider a sequence of integers \( n \) large enough so that we can choose \( s = s(n) \) satisfying \( r_1 \leq s/n \leq r_2 \) for each \( n \). We will prove the existence of an integer \( n_0 \) such that (1.5) holds for all \( n \geq n_0 \), where \( n_0 \) will not depend on the choice of \( s(n) \). Define

\[ C(s, n, d) = \left\{ \rho \in \mathcal{S}_d : S(\rho) \leq \log d - h \left( \frac{s}{nd} \right) \right\} \]  

(3.1)

Let \( \lambda_i \) be the eigenvalues of \( \rho \), then

\[ S(\rho) - \log d = -\frac{1}{d} \sum_{i=1}^{d} (\lambda_i d) \log(\lambda_i d) \]  

(3.2)

Define

\[ f(x) = x \log x - x + 1 \]  

(3.3)

then it follows that

\[ C(s, n, d) = \left\{ \rho \in \mathcal{S}_d : \sum_{i=1}^{d} f(\lambda_i d) \geq h \left( \frac{s}{n} \right) \right\} \]  

(3.4)

Next define

\[ F(x) = -\log x + x - 1 \]  

(3.5)

and note that for any \( \theta \in \mathcal{S}_d \)

\[ d \log d + \text{Tr} \log \theta = -\sum_{i=1}^{d} F(\theta_i d) \]  

(3.6)
where \( \{ \theta_i \} \) are the eigenvalues of \( \theta \). Thus recalling the definition (2.11) it follows that

\[
d \log d + M(\gamma, n) = - \inf \left\{ \sum_{i=1}^{d} F(\theta_i d) : \theta \in \text{Tube}(C(s, n, d)) \right\}
\]

(3.7)

Now from definitions (2.7), (2.8) it follows that if \( \theta \in \text{Tube}(C(s, n, d)) \) then for some \( r \in [\gamma, 1] \)

\[
\theta_i = z_i + \epsilon_i, \quad z_i = r \lambda_i + (1 - r) \frac{1}{d} \sum_{i=1}^{d} \epsilon_i = 0
\]

(3.8)

where the eigenvalues \( \lambda_i \) satisfy

\[
\sum_{i=1}^{d} f(\lambda_i d) \geq h \left( \frac{s}{n} \right)
\]

(3.9)

and where

\[
|\epsilon_i| \leq 2 \sqrt{\frac{\log n}{n}} + 13 d \sqrt{\frac{\log d}{s}}
\]

(3.10)

The Fannes inequality \cite{11, 4, 14} implies that

\[
\left| \sum_{i=1}^{d} f(\theta_i d) - \sum_{i=1}^{d} f(z_i d) \right| \leq \eta \equiv d \epsilon_m (\log d + \log \frac{1}{\epsilon_m})
\]

(3.11)

where

\[
\epsilon_m = \sum_{i=1}^{d} |\epsilon_i| \leq 2 d \sqrt{\frac{\log n}{n}} + 13 d^2 \sqrt{\frac{\log d}{s}}
\]

(3.12)

Note that \( \eta \to 0 \) as \( n, s \to \infty \).

We now apply Lemma 12 from \cite{14} which says that for all \( x > 0 \) and all \( r \in [\gamma, 1] \)

\[
f(x) \leq \frac{f(rx + 1 - r)}{f(1 - \gamma)}
\]

(3.13)

Applying this to (3.9), (3.11) we deduce that

\[
\sum_{i=1}^{d} f(\theta_i d) \geq h f(1 - \gamma) \left( \frac{s}{n} \right) - \eta
\]

(3.14)

Thus we finally arrive at the inequality (putting \( x_i = \theta_i d \))

\[
d \log d + M(\gamma, n) \leq - \inf \left\{ \sum_{i=1}^{d} F(x_i) : \sum_{i=1}^{d} f(x_i) \geq h f(1 - \gamma) \left( \frac{s}{n} \right) - \eta \right\}
\]

(3.15)
Define
\[
m_d(y) = \inf_{\sum_{i=1}^{d} x_i = d} \left\{ \sum_{i=1}^{d} F(x_i) : \sum_{i=1}^{d} f(x_i) \geq y \right\}
\] (3.16)
then (3.15) can be written
\[
d \log d + M(\gamma, n) \leq -m_d \left( h f(1 - \gamma) \left( \frac{s}{n} \right) - \eta \right)
\] (3.17)
In Section 5.7 of [14] the following identity was derived:
\[
m_d(y) = \inf_{z > 1} \left\{ -\log z - (d - 1) \log \frac{d - z}{d - 1} : z \log z + (d - z) \log \frac{d - z}{d - 1} = y \right\}
\] (3.18)
It was also shown in [14] that the function \( m_d \) is increasing and hence has an inverse \( m_d^{-1} \). Given \( y > 0 \) there is a unique \( z > 1 \) satisfying \( z \log z + (d - z) \log \frac{d - z}{d - 1} = y \), and this function also has an inverse. Thus
\[
m_d^{-1}(w) = z \log z + (d - z) \log \frac{d - z}{d - 1}
\] (3.19)
where \( z \) is the unique solution of \( -\log z - (d - 1) \log \frac{d - z}{d - 1} = w \). Since both of these functions are increasing this can be written as the minimization
\[
m_d^{-1}(w) = \inf_{z > 1} \left\{ z \log z + (d - z) \log \frac{d - z}{d - 1} : -\log z - (d - 1) \log \frac{d - z}{d - 1} = w \right\}
\] (3.20)
Thus recalling (1.4) we have
\[
h_d(x, y) = \inf_{0 < \gamma < 1} \frac{1}{x f(1 - \gamma)} m_d^{-1} \left( -\log(1 - \gamma) \, y \right)
\] (3.21)
Let \( \gamma_m \) be the value where the infimum is achieved, then
\[
m_d(x f(1 - \gamma_m) h_d(x, y)) = -\log(1 - \gamma_m) \, y
\] (3.22)
Returning to (3.17), and using the bound \( s/n \geq r_1 \),
\[
d \log d + M(\gamma, n) \leq -m_d \left( h f(1 - \gamma) \, r_1 - \eta \right)
\] (3.23)
By assumption \( h > h_d(r_1, r_2) \), and also \( \eta \to 0 \) as \( n \to \infty \), hence there is \( \delta > 0 \) such that for \( n \) sufficiently large
\[
h f(1 - \gamma) \, r_1 - \eta > h_d(r_1, r_2) f(1 - \gamma) \, r_1 + \delta
\] (3.24)
and thus
\[
d \log d + M(\gamma, n) \leq -m_d \left( h_d(r_1, r_2) f(1 - \gamma) \, r_1 + \delta \right)
\] (3.25)
Furthermore as was shown in [14]

$$m'_d(y) = \frac{d(1 - z^{-1})}{d \log z - y} > \frac{1}{z}$$

(3.26)

where $z$ is the unique solution of $z \log z + (d - z) \log \frac{d - z}{d - 1} = y$. The maximum value of $z$ is $d$ hence we obtain

$$m'_d(y) \geq \frac{1}{d}$$

(3.27)

Thus from (3.25) (applying the mean value theorem)

$$d \log d + M(\gamma, n) \leq -m_d(h_d(r_1, r_2) f(1 - \gamma) r_1) - \frac{1}{d} \delta$$

(3.28)

Setting $\gamma = \gamma_m$ we have

$$m_d(h_d(r_1, r_2) f(1 - \gamma_m) r_1) = -\log(1 - \gamma_m) r_2$$

(3.29)

Thus finally returning to (2.10) we have

$$d^2 \log n + nd \log d + (n - d)M(\gamma_m, n) - s \log(1 - \gamma_m)$$

$$= d^2(\log n + \log d) - \frac{sd}{n} \log(1 - \gamma_m) + (n - d) \left( d \log d + M(\gamma_m, n) - \frac{s}{n} \log(1 - \gamma_m) \right)$$

$$\leq d^2(\log n + \log d) - r_2 d \log(1 - \gamma_m) + (n - d) \left( \log(1 - \gamma_m) r_2 - \frac{1}{d} \delta - \frac{s}{n} \log(1 - \gamma_m) \right)$$

$$\leq d^2(\log n + \log d) - r_2 d \log(1 - \gamma_m) - (n - d) \frac{1}{d} \delta$$

(3.30)

where we used $s/n \leq r_2$. Since $\delta > 0$ the right side of (3.30) diverges to $-\infty$ as $n \to \infty$. Thus applying Theorem 7 yields the result.

4 Proof of Theorem 3

Following the steps of the proof of Theorem 2 leads to

$$d \log d + M(\gamma, k) \leq -m_d \left( h f(1 - \gamma) \left( \frac{s_k}{n_k} \right) - \eta \right)$$

(4.1)

where

$$\eta = d \epsilon_m (\log d + \log \frac{1}{\epsilon_m}), \quad \epsilon_m \leq 2d \sqrt{\frac{\log n_k}{n_k}} + 13 d^2 \sqrt{\frac{\log d}{s_k}}$$

(4.2)

The assumptions that $s_k \to \infty$ and $n_k \log s_k/s_k^{3/2} \to 0$ imply that

$$\frac{\eta}{s_k/n_k} \to 0 \quad \text{as} \quad k \to \infty$$

(4.3)

and hence the first term $h f(1 - \gamma) (s_k/n_k)$ on the right side of (4.1) dominates $\eta$. Since $s_k/n_k \to 0$ we must consider the behavior of $m_d(y)$ as $y \to 0$. 

11
Lemma 8 There is $y_0 > 0$ such that $y^{-1} m_d(y)$ is decreasing for all $0 < y \leq y_0$. Furthermore
\[
\lim_{y \to 0} \frac{m_d(y)}{y} = 1 \quad (4.4)
\]

Lemma 8 will be proved in the Appendix. We now use it to analyze (4.1). Since $s_k/n_k \to 0$, and using (4.3), it follows from (4.4) that for any $\epsilon > 0$ there is $k_0$ such that for $k > k_0$,
\[
m_d \left( h f(1 - \gamma) \left( \frac{s_k}{n_k} \right) - \eta \right) \geq \left( 1 - \epsilon \right) h f(1 - \gamma) \left( \frac{s_k}{n_k} \right) - \eta \quad (4.5)
\]

Hence from (4.1)
\[
d \log d + M(\gamma, k) \leq -(1 - \epsilon) h f(1 - \gamma) \left( \frac{s_k}{n_k} \right) + \eta \quad (4.6)
\]

Turning now to (2.10) we have for $k > k_0$
\[
d^2 \log n_k + n_k d \log d + (n_k - d) M(\gamma, k) - s_k \log(1 - \gamma) \leq d^2 \left( \log n_k + \log d \right) - \frac{s_k d}{n_k} \log(1 - \gamma) \\
\leq \left( 1 - \epsilon \right) h f(1 - \gamma) \left( \frac{s_k}{n_k} \right) - \eta + \frac{s_k}{n_k} \log(1 - \gamma) \quad (4.7)
\]

By assumption there is $\gamma_m$ such that
\[
h > -\frac{\log(1 - \gamma_m)}{f(1 - \gamma_m)} \quad (4.8)
\]

From (4.3) it follows that there is $\delta > 0$ such that for $k$ sufficiently large
\[
(1 - \epsilon) h f(1 - \gamma_m) - \eta \frac{n_k}{s_k} + \log(1 - \gamma_m) > \delta \quad (4.9)
\]

Thus from (4.7)
\[
d^2 \log n_k + n_k d \log d + (n_k - d) M(\gamma_m, k) - s_k \log(1 - \gamma_m) \leq d^2 \left( \log n_k + \log d \right) - \frac{s_k d}{n_k} \log(1 - \gamma_m) - \frac{s_k (n_k - d) \delta}{n_k} \quad (4.10)
\]

Since $s_k/\log n_k \to \infty$ the right side of (4.10) diverges to $-\infty$ as $k \to \infty$. QED
5 Proof of the Hastings bound

First we recall some of the ideas and notation from [14]. The set of eigenvalues of a state $\rho$ is denoted $\text{spec}(\rho)$. Also $\Delta_d$ denotes the simplex of $d$-dimensional probability distributions:

$$\Delta_d = \{(x_1, \ldots, x_d) \in \mathbb{R}^d : x_i \geq 0, \sum_{i=1}^d x_i = 1\}$$ (5.1)

5.1 Random states

The pure states $V_n$ can be identified with the unit sphere in $\mathbb{R}^{2n}$. This provides a probability measure on $V_n$, namely the normalized uniform measure which we denote $\sigma_n$. Thus saying that $|\psi\rangle \in V_n$ is a random vector means that $|\psi\rangle$ has the uniform distribution $\sigma_n$.

Let $|z\rangle = (z_1 \cdots z_{dn})^T$ be a unit vector in $V_{dn}$. Then $|z\rangle$ can be written as a $n \times d$ matrix $M$, with entries

$$M_{ij}(z) = z_{(i-1)d+j}, \quad i = 1, \ldots n, \quad j = 1, \ldots, d$$ (5.2)

satisfying $\text{Tr} M^* M = \sum_{ij} |z_{ij}|^2 = 1$. Define the map $G : V_{dn} \rightarrow \mathcal{M}_d$ by

$$G(z) = M(z)^* M(z)$$ (5.3)

The eigenvalues of $G(z)$ lie in $\Delta_d$. When $|z\rangle \in V_{dn}$ is a random vector, the probability density $\mu_{d,n}$ of these eigenvalues is known explicitly [25], [35]: for any event $A \subset \Delta_d$

$$\mu_{d,n}(A) = Z(n,d)^{-1} \int_A \prod_{1 \leq i < j \leq d} (w_i - w_j)^2 \prod_{i=1}^d w_i^{n-d} \delta\left(\sum_{i=1}^d w_i - 1\right) [dw]$$ (5.4)

where $Z(n,d)$ is a normalization factor. We recall the following bound which was derived in [14].

Lemma 9 For all $d$, for $n$ sufficiently large, and for any event $A \subset \Delta_d$

$$\mu_{d,n}(A) \leq \frac{1}{(d-1)!} \exp \left[d^2 \log n + (n-d) d \log d + (n-d) \sup_{w \in A} \sum_{i=1}^d \log w_i\right]$$ (5.5)

Now let $C \subset \mathcal{S}_d$ be any set which is invariant under conjugation by every unitary matrix in $\mathcal{U}(d)$. Then the event $\{|\psi\rangle : \text{Tr}_2 |\psi\rangle\langle\psi| \in C\}$ depends only on the eigenvalues of $\text{Tr}_2 |\psi\rangle\langle\psi|$, and thus its probability is determined by $\mu_{d,n}$. For an arbitrary set $C \subset \mathcal{S}_d$ we define

$$\bar{C} = \{V \rho V^* : \rho \in C, \quad V \in \mathcal{U}(d)\}, \quad \text{spec}(C) = \bigcup_{\rho \in C} \text{spec}(\rho)$$ (5.6)

13
Then \( \tilde{C} \) is invariant under conjugation by an arbitrary unitary matrix, and \( \text{spec}(C) = \text{spec}(\tilde{C}) \). Hence from (5.5) we deduce the bound
\[
G^*(\sigma_{nd})(C) \leq G^*(\sigma_{nd})(\tilde{C}) = \mu_{d,n}(\text{spec}(C)) \leq \frac{1}{(d-1)!} \exp \left[ d^2 \log n + (n-d) d \log d + (n-d) \sup_{w \in \text{spec}(C)} \sum_{i=1}^d \log w_i \right] \leq \frac{1}{(d-1)!} \exp \left[ d^2 \log n + (n-d) d \log d + (n-d) \sup_{\rho \in C} \text{Tr} \log \rho \right] \]
(5.7)

5.2 Random embeddings yield random output states

Let \( W \in \mathcal{R}(s, n, d) \) be a random embedding. Then for any pure state \( |\phi\rangle \in \mathcal{V}_s \), the vector \( |\psi\rangle = UW_0|\phi\rangle \) is a uniform random vector in \( \mathbb{C}^{nd} \). Thus \( \Phi_W^C(|\phi\rangle) = T_2|\psi\rangle\langle\psi| \) is the reduced density matrix of a random vector. This remains true if \( |\phi\rangle \) is a random input state. To formalize this relation we define the map
\[
H : \mathcal{R}(s, n, d) \times \mathcal{V}_s \to \mathcal{M}_d, \quad (W, |\phi\rangle) \mapsto \Phi_W^C(|\phi\rangle\langle\phi|) \]
(5.8)

**Lemma 10**
\[
H^*(\mathcal{P}_{s,n,d} \times \sigma_s) = G^*(\sigma_{dn}) \]
(5.9)

The proof of Lemma 10 is very similar to the proof of Lemma 7 in [14] and so we omit the details here. Lemma 10 implies that if \( W \) is chosen randomly according to the measure \( \mathcal{P}_{s,n,d} \) and \( |\phi\rangle \) is chosen randomly and uniformly in \( \mathcal{V}_s \), then the eigenvalues of the matrix \( \Phi_W^C(|\phi\rangle\langle\phi|) \) will have the distribution \( \mu_{d,n} \), as defined above in (5.4).

5.3 Typical channels

For a random embedding \( W \) ‘most’ output states of the channel \( \Phi_W^C \) are close to the maximally mixed state. More precisely, an embedding \( W \) will be called typical if \( \Phi_W^C \) maps at least one half of input states into a small ball centered at the maximally mixed output state \( I/d \). The ball is defined as follows:
\[
B_d(n) = \left\{ \rho \in \mathcal{S}_d : \left\| \rho - \frac{1}{d} I \right\|_\infty \leq 2 \sqrt{\frac{\log n}{n}} \right\} \]
(5.10)

**Definition 11** An embedding \( W \) is called typical if with probability at least \( 1/2 \) a randomly chosen input state is mapped by \( \Phi_W^C \) into the set \( B_d(n) \). The set of typical embeddings is denoted \( T \):
\[
T = \left\{ W : \sigma_s(|\phi\rangle : \Phi_W^C(|\phi\rangle\langle\phi|) \in B_d(n)) \geq 1/2 \right\} \]
(5.11)
As the next result shows, for large $n$ most embeddings are typical. This result was proved in [14] and we just quote the result here (note that in [14] the definition of $B_d(n)$ contained a free parameter $b$ which was required to be at least $\sqrt{3}$ – here we have set $b = 2$).

**Lemma 12** For each $d \geq 2$ taking $n$ sufficiently large, and for all $s$,

$$\mathcal{P}_{s,n,d}(T^c) \leq \frac{2d}{(d-1)!} \exp[-\alpha d^2 \log n], \quad \alpha = \frac{4(n-d)}{3n} - 1 \quad (5.12)$$

Thus as $n \to \infty$ with high probability a randomly chosen embedding will lie in the set $T$. In particular $\mathcal{P}_{s,n,d}(T^c) < 1$ for $n$ sufficiently large.

The next result says that for a typical embedding $W$ there is a fixed fraction of input states which are mapped by $\Phi_W^C$ into the tube at any output state $\rho$. This result is crucial for the proof and differs in some significant ways from the related proof in [14], thus we include full details in the Appendix.

**Lemma 13** Let $d, s \geq 2$, then for $n$ sufficiently large, for all $W \in T$ and $\rho \in \text{Im}(\Phi_W^C)$

$$\sigma_s(\langle \phi \rangle : \Phi_W^C(\langle \phi \rangle) \in \text{Tube}(\rho)) \geq \frac{1}{4} \left(1 - \gamma\right)^{s-1} \quad (5.13)$$

### 5.4 The proof

Define

$$E_k = \{(W, |\phi\rangle) : W \in B_k, \, \Phi_W^C(\langle \phi \rangle) \in \text{Tube}(C(s_k, n_k, d))\} \quad (5.14)$$

The proof will proceed by proving upper and lower bounds for the probability of $E_k$, that is $(\mathcal{P}_{s,n,d} \times \sigma_s)(E_k)$.

For the upper bound, note that by Lemma 10,

$$(\mathcal{P}_{s,n,d} \times \sigma_s)(E_k) \leq (\mathcal{P}_{s,n,d} \times \sigma_s)\{(W, |\phi\rangle) : \Phi_W^C(\langle \phi \rangle) \in \text{Tube}(C(s_k, n_k, d))\}$$

$$= (\mathcal{P}_{s,n,d} \times \sigma_s)(H^{-1}(\text{Tube}(C(s_k, n_k, d))))$$

$$= H^*(\mathcal{P}_{s,n,d} \times \sigma_s)(\text{Tube}(C(s_k, n_k, d)))$$

$$= G^*(\sigma_{da})(\text{Tube}(C(s_k, n_k, d))) \quad (5.15)$$

Recall the definition of $M(\gamma, k)$ in (2.11). Using the bound (5.7) we deduce

$$(\mathcal{P}_{s,n,d} \times \sigma_s)(E_k) \leq \frac{1}{(d-1)!} \exp \left[d^2 \log n_k + (n_k - d)d \log d + (n_k - d)M(\gamma, k)\right]$$

$$= \alpha(d) \exp \left[d^2 \log n_k + n_k d \log d + (n_k - d)M(\gamma, k)\right] \quad (5.16)$$

where $\alpha(d) = \exp[-d^2 \log d] / (d-1)!$.  


The derivation of the lower bound is very similar to that in [14], however we include it here for completeness. First we write

\[(s_n,d) \times \sigma_s(E_k) = \mathbb{E}_W[1_{B_k} \sigma_s(|\phi\rangle) : \Phi_W^C(|\phi\rangle\langle\phi|) \in \text{Tube}(C(s_k,n_k,d))] \geq \mathbb{E}_W[1_{B_k \cap T} \sigma_s(|\phi\rangle) \in \text{Tube}(C(s_k,n_k,d))] \]

where \(\mathbb{E}_W\) denotes expectation over \(\mathcal{R}(s_k,n_k,d)\) with respect to the measure \(\mathcal{P}_{s,n,d}\), and \(1_{B_k \cap T}\) is the characteristic function of the event \(B_k \cap T\). Given that \(W \in B_k\) there is a state \(|\theta\rangle \in \mathbb{C}^s\) such that

\[\Phi_W^C(|\theta\rangle\langle\theta|) \in C(s_k,n_k,d)\]

Since \(\text{Tube}(\Phi_W^C(|\theta\rangle\langle\theta|)) \subset \text{Tube}(C(s_k,n_k,d))\) it follows that

\[(s_n,d) \times \sigma_s(E_k) \geq \mathbb{E}_W[1_{B_k \cap T} \sigma_s(|\phi\rangle) \in \text{Tube}(\Phi_W^C(|\theta\rangle\langle\theta|))] \]

Applying Lemma [13] to (5.18) gives

\[\begin{align*}
(s_n,d) \times \sigma_s(E_k) & \geq \frac{1}{4} \left(1 - \gamma\right)^{s_k-1} \mathbb{E}_W[1_{B_k \cap T}] \\
& = \frac{1}{4} \left(1 - \gamma\right)^{s_k-1} \mathcal{P}_{s,n,d}(B_k \cap T) \\
& \geq \frac{1}{4} \left(1 - \gamma\right)^{s_k-1} \left(\mathcal{P}_{s,n,d}(B_k) - \mathcal{P}_{s,n,d}(T^c)\right)
\end{align*}\]

Putting together the upper and lower bounds for \((s_n,d) \times \sigma_s(E_k)\) produces the following bound: for all \(d \geq 2\), for all \(0 < \gamma < 1\), and for \(n\) sufficiently large

\[\mathcal{P}_{s,n,d}(B_k) - \mathcal{P}_{s,n,d}(T^c) \leq 4 \left(\frac{1}{1 - \gamma}\right)^{s_k-1} \begin{align*}
(s_n,d) \times \sigma_s(E_k) & \geq 4 \alpha(d) \left(\frac{1}{1 - \gamma}\right)^{s_k-1} \exp \left[d^2 \log n_k + n_k d \log d + (n_k - d)M(\gamma,k)\right] \\
& = 4 \alpha(d) (1 - \gamma) \exp \left[d^2 \log n_k + n_k d \log d + (n_k - d)M(\gamma,k) - s_k \log(1 - \gamma)\right]
\end{align*}\]

Note that Lemma [12] implies \(\mathcal{P}_{s,n,d}(T^c) \rightarrow 0\) as \(k \rightarrow \infty\). Also, by assumption there is \(\gamma\) such that

\[d^2 \log n_k + n_k d \log d + (n_k - d)M(\gamma,k) - s_k \log(1 - \gamma) \rightarrow -\infty\]

as \(k \rightarrow \infty\). By choosing this value for \(\gamma\) we deduce that \(\mathcal{P}_{s,n,d}(B_k) \rightarrow 0\) as required. QED

Acknowledgments: This collaboration began at “the University of Arizona FRG Workshop” in June 2009 (the grant number: DMS-0757581), and a lot of progress was made and the results were presented during the July 2009 workshop “Thematic Program on Mathematics in Quantum Information” at the Fields Institute. The authors are grateful to the organizers of these workshops.
References

[1] G.G. Amosov, “Remark on the additivity conjecture for the quantum depolarizing channel”, Probl. Inf. Transm. 42 no.2, 69-76 (2006).

[2] G.G. Amosov, “The strong superadditivity conjecture holds for the quantum depolarizing channel in any dimension”, Phys. Rev. A 75 no. 6, P. 060304 (2007).

[3] G. G. Amosov, A. S. Holevo, and R. F. Werner, “On Some Additivity Problems in Quantum Information Theory”, Problems in Information Transmission, 36, 305 – 313 (2000).

[4] K.M.R. Audenaert, “A Sharp Fannes-type Inequality for the von Neumann Entropy”, J. Phys. A 40, 8127–8136 (2007).

[5] F.G.S.L. Brandao and M. Horodecki, “On Hastings counterexamples to the minimum output entropy additivity conjecture”, arXiv:0907.3210[quant-ph] (2009).

[6] B. V. Bronk, “Exponential Ensemble for Random Matrices”, J. Math. Phys. 6, 228 (1965).

[7] D. Bruss, L. Faoro, C. Macchiavello and M. Palma, “Quantum entanglement and classical communication through a depolarising channel”, J. Mod. Opt. 47 325 (2000).

[8] B. Collins and I. Nechita, “Random quantum channels I: graphical calculus and the Bell state phenomenon”, arXiv:0905.2313[quant-ph] (2009).

[9] B. Collins and I. Nechita, “Random quantum channels II: Entanglement of random subspaces, Renyi entropy estimates and additivity problems ”, arXiv:0906.1877[quant-ph] (2009).

[10] T. Cubitt, A. W. Harrow, D. Leung, A. Montanaro, A. Winter, “Counterexamples to additivity of minimum output p-Renyi entropy for p close to 0”, Comm. Math. Phys., 284:281-290 (2008).

[11] M. Fannes, “A continuity property of the entropy density for spin lattice systems”, Comm. Math. Phys. 31:291-294, (1973).

[12] A. Fujiwara and T. Hashizume, “Additivity of the capacity of depolarizing channels”, Phys. Lett. A 299, 469-475 (2002).

[13] M. Fukuda, “Simplification of additivity conjecture in quantum information theory”, Quant. Info. Proc., 6, 179–186, (2007); arXiv:quant-ph/0608010.

[14] M. Fukuda, C. King and D. Moser, “Comments on Hastings’ Additivity Counterexamples”, arXiv:0905.3697[quant-ph] (2009).

[15] M. Fukuda, M. M. Wolf, “Simplifying additivity problems using direct sum constructions”, arXiv:0704.1092[quant-ph], accepted by J. Math. Phys. 48 072101, (2007).
[16] M. B. Hastings, “A Counterexample to Additivity of Minimum Output Entropy”, Nature Physics 5, 255 - 257 (2009), arXiv:0809.3972v3 [quant-ph].

[17] P. Hayden, D. W. Leung and A. Winter, “Aspects of generic entanglement”, Comm. Math. Phys. 265(1):95-117 (2006).

[18] P. Hayden and A. Winter, “Counterexamples to the maximal p-norm multiplicativity conjecture for all p > 1”, Comm. Math. Phys. 284(1):263-280 (2008).

[19] A. S. Holevo, “The capacity of the quantum channel with general signal states”, IEEE Trans. Inf. Theory, 44(1):269-273 (1998).

[20] A. S. Holevo, “On complementary channels and the additivity problem”, Probab. Theory and Appl., 51, 133-143, (2005).

[21] C. King, “Additivity for unital qubit channels”, J. Math. Phys. 43 4641-4653 (2002).

[22] C. King, “The capacity of the quantum depolarizing channel”, IEEE Trans. Info. Theory 49, 221-229 (2003).

[23] C. King, M. Nathanson and M. B. Ruskai, “Multiplicativity properties of entrywise positive maps”, Lin. Alg. Appl. 404, 367-379 (2005).

[24] C. King and M. B. Ruskai, “Minimal Entropy of States Emerging from Noisy Quantum Channels”, IEEE Trans. Info. Theory, 47, 192-209 (2001).

[25] S. Lloyd S and H. Pagels, “Complexity as Thermodynamic Depth”, Annals of Physics 188, 186-213, (1988).

[26] C. King, K. Matsumoto, M. Nathanson, M. B. Ruskai, “Properties of Conjugate Channels with Applications to Additivity and Multiplicativity”, Markov Processes and Related Fields, volume 13, no. 2, 391 – 423 (2007).

[27] D. N. Page, “Average entropy of a subsystem”, Phys. Rev. Lett., 71:1291 (1993).

[28] M. B. Ruskai, A comment at the July 2009 workshop “Thematic Program on Mathematics in Quantum Information” at the Fields Institute.

[29] J. Sanchez-Ruiz, “Simple proof of Page’s conjecture on the average entropy of a subsystem”, Phys. Rev. E, 52:5653 (1995).

[30] B. Schumacher and M. D. Westmoreland, “Sending classical information via noisy quantum channels”, Phys. Rev. A 56(1):131-138 (1997).

[31] S. Sen, “Average entropy of a quantum subsystem”, Phys. Rev. Lett., 77(1):13 (1996).

[32] P. W. Shor, “Equivalence of additivity questions in quantum information theory”, Comm. Math. Phys. 246(3):453-472 (2004).
[33] R.F. Werner and A.S. Holevo, “Counterexample to an additivity conjecture for output purity of quantum channels”, J. Math. Phys. 43(9):4353-4357 (2002).

[34] K. Zyczkowski, I. Bengtsson “On duality between quantum maps and quantum states”, Open Syst. Inf. Dyn. 11, 3-42 (2004).

[35] K. Zyczkowski and H.-J. Sommers, “Induced measures in the space of mixed quantum states”, J. Phys. A 34, 7111-7125 (2001).
A Proof of Theorem 5

The bound for $S_{\min}(\Phi \otimes \Phi)$ is obtained using a Kraus representation $\Phi(\rho) = \sum_{i=1}^{d} A_i \rho A_i^*$ and the maximally entangled state. Let $|\psi\rangle$ and $|\phi\rangle$ be the maximally entangled states on $\mathbb{C}^s \otimes \mathbb{C}^s$ and $\mathbb{C}^n \otimes \mathbb{C}^n$ respectively. Then

\[
(\Phi \otimes \Phi)(|\psi\rangle\langle \psi|) = \sum_{i,j=1}^{d} (A_i \otimes A_j)|\psi\rangle\langle \psi| (A_i^* \otimes A_j^T) \tag{A.1}
\]

\[
= \frac{n}{s} \sum_{i=1}^{d} (A_i A_i^* \otimes I)|\phi\rangle\langle \phi| (A_i A_i^* \otimes I) + \sum_{i \neq j} (A_i \otimes A_j)|\psi\rangle\langle \psi| (A_i^* \otimes A_j^T) \tag{A.2}
\]

where we used the identity $(A \otimes A)|\psi\rangle = \sqrt{n/s} (AA^* \otimes I)|\phi\rangle$. Note that $\sum_i A_i A_i^* = \Phi(I_s)$, therefore

\[
\sum_{i=1}^{d} \text{Tr} A_i A_i^* = \text{Tr} \Phi(I_s) = \text{Tr} I_s = s
\]

Also

\[
\langle \phi| (A_i A_i^* \otimes I)|\phi\rangle = \frac{1}{n} \text{Tr} [A_i A_i^*]
\]

hence

\[
\langle \phi| \sum_{i=1}^{d} (A_i A_i^* \otimes I)|\phi\rangle\langle \phi| (A_i A_i^* \otimes I)|\phi\rangle = \frac{1}{n^2} \sum_{i=1}^{d} (\text{Tr} [A_i A_i^*])^2 \geq \frac{s^2}{dn^2}
\]

Hence

\[
\langle \phi| (\Phi \otimes \Phi)(|\psi\rangle\langle \psi|)|\phi\rangle \geq \frac{s}{dn} \tag{A.3}
\]

This shows that one of the eigenvalues of $(\Phi \otimes \Phi)(|\psi\rangle\langle \psi|)$ is larger than or equal to $p = s/(dn)$. The rank of this matrix is $d^2$, hence it has at most $d^2 - 1$ other nonzero eigenvalues. Given that $sd \geq n$, the entropy is maximized when these other eigenvalues are equal to $(1 - p)/(d^2 - 1)$. This implies that the entropy cannot be larger than

\[
S \left( (\Phi \otimes \Phi)(|\psi\rangle\langle \psi|) \right) \leq g(p) = -p \log p - (1 - p) \log \left( \frac{1 - p}{d^2 - 1} \right)
\]

B Proof of Lemma 13

The result is very similar to the proof of Lemma 11 in [14], but with important differences in detail. Let $|\psi\rangle$ be a fixed state in $\mathcal{V}_s$, and let $|\theta\rangle$ be a random pure state in $\mathcal{V}_s$, with probability distribution $\sigma_s$. We write $x = \langle \psi|\theta\rangle$, and let $|\phi\rangle$ be the state orthogonal to $|\psi\rangle$ such that

\[
|\theta\rangle = x |\psi\rangle + \sqrt{1 - |x|^2} |\phi\rangle \tag{B.1}
\]
Thus $|\phi\rangle$ is also a random state, defined by its relation to the uniformly random state $|\theta\rangle$ in (B.1). The following result was proved in [14].

**Proposition 14** $x$ and $|\phi\rangle$ are independent. $|\phi\rangle$ is a random vector in $\mathcal{V}_{s-1}$ with distribution $\sigma_{s-1}$.

For all $0 \leq t \leq 1$

$$\sigma_s\{|\theta\rangle : |\langle\psi|\theta\rangle| = |x| > t\} = (1 - t^2)^{s-1}$$  \hfill (B.2)

Proposition 14 implies that as $s \to \infty$ the overlap $x = \langle\psi|\theta\rangle$ becomes concentrated around zero. In other words, with high probability a randomly chosen state will be almost orthogonal to any given fixed state. As a consequence, from (B.1) it follows that $|\phi\rangle$ will be almost equal to $|\theta\rangle$. This statement is made precise by noting that

$$\|\langle\theta| - |\phi\rangle\|_2 \leq \sqrt{2} |\langle\psi|\theta\rangle|$$  \hfill (B.3)

Then (B.2) immediately implies that

$$\sigma_s(|\theta\rangle : \|\langle\theta| - |\phi\rangle\|_2 > t) \leq \left(1 - \frac{t^2}{2}\right)^{s-1}$$  \hfill (B.4)

The second property relies on the form of the conjugate channel $\Phi^C$. If the Kraus decomposition for $\Phi$ is $\Phi(\rho) = \sum_{i=1}^d A_i \rho A_i^*$ then the Kraus decomposition for $\Phi^C$ is

$$\Phi^C(\rho) = \sum_{k,l=1}^d \text{Tr}(A_k \rho A_l^*) |k\rangle \langle l|$$  \hfill (B.5)

For any fixed channel $\Phi$ and random state $|\theta\rangle$, with high probability the norm of the matrix $\Phi^C(|\theta\rangle \langle\psi|)$ is small, and approaches zero as $n \to \infty$. We will prove the following bound: for any $\Phi \in \mathcal{R}(s,n,d)$, and for all $0 \leq t \leq 1$,

$$\sigma_s\{|\theta\rangle : \|\Phi^C(|\theta\rangle \langle\psi|)\|_2 > t\} \leq d^2 \left(1 - \left(\frac{t}{d}\right)^2\right)^{s-1}$$  \hfill (B.6)

As a first step toward deriving (B.6), note that for any vectors $|u\rangle$ and $|v\rangle$,

$$\|\Phi^C(|u\rangle \langle v|)\|_2 = \left(\sum_{k,l=1}^d |\langle v|A_l^* A_k|u\rangle|^2\right)^{\frac{1}{2}} \leq d \max_{k,l} |\langle v|A_l^* A_k|u\rangle|.$$  \hfill (B.7)

Since $\sum_{i=1}^d A_i^* A_i = I$ it follows that $\|A_i\|_{\infty} \leq 1$ for all $i = 1, \ldots, d$, which implies that

$$\|\Phi^C(|u\rangle \langle v|)\|_2 \leq d \||u\rangle\|_2 \||v\rangle\|_2$$  \hfill (B.8)
To derive (B.6) we apply (B.7) with \( u = \theta \) and \( v = \psi \) and deduce that
\[
\sigma_s(\{\theta\} : \| \Phi^C(\{\theta\}\langle \psi \}) \|_2 > t) \leq \sigma_s(\{\theta\} : \max_{k,l} |\langle \psi | A_k^* A_l | \theta \rangle | > \frac{t}{d})
\]
\[
\leq d^2 \sigma_s(\{\theta\} : |\langle \psi | A_k^* A_l | \theta \rangle | > \frac{t}{d})
\]
\[
\leq d^2 \left( 1 - \left( \frac{t}{d} \right)^2 \right)^{s-1}
\]
(B.9)
where the last equality follows from (B.2). Note that for each \( k,l \), the above \( |A_k^* A_l \psi \rangle \) is a fixed vector with norm less than 1.

With these ingredients in place the proof of Lemma 13 can proceed. By assumption \( \Phi \) is a channel belonging to the typical set \( T \), and \( \rho = \Phi^C(\{\psi\}\langle \psi \}) \) is some state in \( \text{Im}(\Phi^C) \). Let \( |\theta\rangle \) be a random input state, then as in (B.1) we write
\[
|\theta\rangle = x |\psi\rangle + \sqrt{1 - |x|^2} |\phi\rangle
\]
It follows that
\[
|\theta\rangle\langle \theta | = |x|^2 |\psi\rangle\langle \psi | + (1 - |x|^2) |\phi\rangle\langle \phi | + \sqrt{1 - |x|^2} (x |\psi\rangle\langle \phi | + \bar{x} |\phi\rangle\langle \psi |)
\]
(B.10)
Write \( r = |x|^2 \), then (B.10) yields
\[
\Phi^C(\{\theta\}\langle \theta |) - \left( r \Phi^C(\{\psi\}\langle \psi |) + (1 - r) \frac{1}{d} I \right)
\]
\[
= (1 - r) \left( \Phi^C(\phi\rangle\langle \phi |) - \frac{1}{d} I \right) + \sqrt{r(1 - r)} \Phi^C(e^{i\xi} |\psi\rangle\langle \phi | + e^{-i\xi} |\phi\rangle\langle \psi |)
\]
(B.11)
where \( \xi \) is the phase of \( x \). Since \( r \leq 1 \) this implies
\[
\left\| \Phi^C(\{\theta\}\langle \theta |) - \left( r \Phi^C(\{\psi\}\langle \psi |) + (1 - r) \frac{1}{d} I \right) \right\|_\infty \leq \left\| \Phi^C(\phi\rangle\langle \phi |) - \frac{1}{d} I \right\|_\infty + \left\| \Phi^C(\psi\rangle\langle \phi |) \right\|_\infty
\]
(B.12)
Referring to the definition (2.7) of Tube(\( \rho \)), recall that \( \Phi^C(\{\theta\}\langle \theta |) \) belongs to Tube(\( \rho \)) if and only if for some \( r \) satisfying \( \gamma \leq r \leq 1 \),
\[
\left\| \Phi^C(\{\theta\}\langle \theta |) - \left( r \Phi^C(\{\psi\}\langle \psi |) + (1 - r) \frac{1}{d} I \right) \right\|_\infty \leq 2 \sqrt{\frac{\log n}{n} + 13d} \sqrt{\frac{\log d}{s}}
\]
(B.13)
Define the following three events in \( \mathcal{V}_s \):
\[
A_1 = \{ |\theta\rangle : r = |\langle \psi |\theta \rangle|^2 \geq \gamma \}
\]
(B.14)
\[
A_2 = \left\{ |\theta\rangle : \left\| \Phi^C(\phi\rangle\langle \phi |) - \frac{1}{d} I \right\|_\infty \leq \sqrt{\frac{48d^2 \log d}{s} + 2 \sqrt{\frac{\log n}{n}}} \right\}
\]
(B.15)
\[
A_3 = \left\{ |\theta\rangle : \left\| \Phi^C(\psi\rangle\langle \phi |) \right\|_\infty \leq \sqrt{\frac{6d^2 \log d}{s} + \sqrt{\frac{12d^2 \log d}{s}}} \right\}
\]
(B.16)
It follows from (B.12) and (B.13) that
\[ A_1 \cap A_2 \cap A_3 \subset \{ |\theta\rangle : \Phi^C(|\theta\rangle|\theta\rangle) \in \text{Tube}(\rho) \} \] (B.17)
Furthermore by Proposition 14, \( A_1 \) is independent of \( A_2 \) and \( A_3 \), hence
\[ \sigma_s(\Phi^C(|\theta\rangle|\theta\rangle) \in \text{Tube}(\rho)) \geq \sigma_s(A_1 \cap A_2 \cap A_3) = \sigma_s(A_1) \sigma_s(A_2 \cap A_3) \] (B.18)
Proposition 14 immediately yields
\[ \sigma_s(A_1) = (1 - \gamma)^{s-1} \] (B.19)
From (B.18) this gives
\[ \sigma_s(\Phi^C(|\theta\rangle|\theta\rangle) \in \text{Tube}(\rho)) \geq (1 - \gamma)^{s-1}(1 - \sigma_s(A_2^c) - \sigma_s(A_3^c)) \] (B.20)
In order to bound \( \sigma_s(A_3^c) \) we first use (B.8) to deduce
\[ \|\Phi^C(|\psi\rangle|\phi\rangle)\|_2 \leq \|\Phi^C(|\psi\rangle|\phi\rangle)\|_2 + d \|\theta\rangle - |\phi\rangle\|_2 \] (B.21)
Thus
\[
\sigma_s(A_3^c) = \sigma_s\left\{ |\theta\rangle : \|\Phi^C(|\psi\rangle|\phi\rangle)\|_2 > \sqrt{\frac{6d^2 \log d}{s}} + \sqrt{\frac{12d^2 \log d}{s}} \right\} \\
\leq \sigma_s\left\{ |\theta\rangle : \|\Phi^C(|\psi\rangle|\theta\rangle)\|_2 + d \|\theta\rangle - |\phi\rangle\|_2 > \sqrt{\frac{6d^2 \log d}{s}} + \sqrt{\frac{12d^2 \log d}{s}} \right\} \\
\leq \sigma_s\left\{ |\theta\rangle : \|\Phi^C(|\psi\rangle|\theta\rangle)\|_2 > \frac{6d^2 \log d}{s} \right\} + \sigma_s\left\{ |\theta\rangle : \|\theta\rangle - |\phi\rangle\|_2 > \sqrt{\frac{12d^2 \log d}{s}} \right\} \\
\leq (d^2 + 1) \left( 1 - \frac{6 \log d}{s} \right)^{s-1} \] (B.22)
where the last inequality follows from (B.9) and (B.4).
Turning now to \( \sigma_s(A_2^c) \), note first that
\[
\|\Phi^C(|\phi\rangle|\phi\rangle) - \frac{1}{d} I\|_\infty \leq \|\Phi^C(|\phi\rangle|\phi\rangle) - \Phi^C(|\theta\rangle|\theta\rangle)\|_\infty + \|\Phi^C(|\theta\rangle|\theta\rangle) - \frac{1}{d} I\|_\infty \\
\leq \left\| \Phi^C(|\phi\rangle|\phi\rangle) - \Phi^C(|\theta\rangle|\theta\rangle) \right\|_2 + \left\| \Phi^C(|\theta\rangle|\theta\rangle) - \frac{1}{d} I \right\|_\infty \\
\leq 2d \|\theta\rangle - |\phi\rangle\|_2 + \left\| \Phi^C(|\theta\rangle|\theta\rangle) - \frac{1}{d} I \right\|_\infty \] (B.24)
where we used (B.8) for the last inequality. As in (B.22) this gives

\[
\sigma_s(A_2^c) = \sigma_s\left\{ |\theta\rangle : \left\| \Phi^C(|\theta\rangle\langle\theta|) - \frac{1}{d} I \right\|_\infty > \sqrt{\frac{48d^2 \log d}{s}} + 2 \sqrt{\frac{\log n}{n}} \right\}
\]

\[
\leq \sigma_s\left\{ |\theta\rangle : \| |\theta\rangle - |\phi\rangle \|_2 > \sqrt{\frac{12 \log d}{s}} \right\}
\]

\[
+ \sigma_s\left\{ |\theta\rangle : \left\| \Phi^C(|\theta\rangle\langle\theta|) - \frac{1}{d} I \right\|_\infty > 2 \sqrt{\frac{\log n}{n}} \right\}
\]

\[
\leq \left( 1 - \frac{6 \log d}{s} \right)^{s-1} + \sigma_s\left\{ |\theta\rangle : \left\| \Phi^C(|\theta\rangle\langle\theta|) - \frac{1}{d} I \right\|_\infty > 2 \sqrt{\frac{\log n}{n}} \right\}
\quad \text{(B.25)}
\]

where we used (B.4) for the last inequality. By assumption \( \Phi \in T \), and therefore there is a set of input states \( L \) with \( \sigma_s(L) \geq 1/2 \) such that

\[
|\theta\rangle \in L \Rightarrow \left\| \Phi^C(|\theta\rangle\langle\theta|) - \frac{1}{d} I \right\|_\infty \leq 2 \sqrt{\frac{\log n}{n}} \quad \text{(B.26)}
\]

Thus

\[
\sigma_s\left\{ |\theta\rangle : \left\| \Phi^C(|\theta\rangle\langle\theta|) - \frac{1}{d} I \right\|_\infty > 2 \sqrt{\frac{\log n}{n}} \right\} \leq \sigma_s(L^c) \leq \frac{1}{2} \quad \text{(B.27)}
\]

Putting together the bounds (B.20), (B.22), (B.25) and (B.27) we get

\[
\sigma_s(\Phi^C(|\theta\rangle\langle\theta|) \in \text{Tube}(\rho)) \geq (1 - \gamma)^{s-1} \left( 1 - \sigma_s(A_2^c) - \sigma_s(A_3^c) \right)
\]

\[
\geq (1 - \gamma)^{s-1} \left( 1 - \left( 1 - \frac{6 \log d}{s} \right)^{s-1} - \frac{1}{2} - (d^2 + 1) \left( 1 - \frac{6 \log d}{s} \right)^{s-1} \right)
\]

\[
= (1 - \gamma)^{s-1} \left( \frac{1}{2} - (d^2 + 2) \left( 1 - \frac{6 \log d}{s} \right)^{s-1} \right) \quad \text{(B.28)}
\]

The proof now follows by noting that for all \( d, s \geq 2 \)

\[
(d^2 + 2) \left( 1 - \frac{6 \log d}{s} \right)^{s-1} \leq \frac{1}{4} \quad \text{(B.29)}
\]

and hence

\[
\sigma_s(\Phi^C(|\theta\rangle\langle\theta|) \in \text{Tube}(\rho)) \geq \frac{1}{4} (1 - \gamma)^{s-1} \quad \text{(B.30)}
\]
C Proof of Lemma 8

We use properties of the function $m_d$ derived in Section 5.7 of [14]. We have

$$m_d(y) = g(z), \quad g(z) = -\log z - (d - 1) \log \frac{d - z}{d - 1} \quad (C.1)$$

where $z = h^{-1}(y)$ and

$$h(z) = z \log z + (d - z) \log \frac{d - z}{d - 1} \quad (C.2)$$

As was shown in [14], both functions $g, h$ are increasing, and $h(1) = 0$. Since the functions are analytic, their behavior near $z = 1$ is determined by their power series expansions at $z = 1$. To leading order these are

$$g(1 + t) = \left( \frac{d}{2(d - 1)} \right) t^2 - \left( \frac{d^2 - 2d}{3(d - 1)^2} \right) t^3 + O(t^4)$$
$$h(1 + t) = \left( \frac{d}{2(d - 1)} \right) t^2 - \left( \frac{d^2 - 2d}{6(d - 1)^2} \right) t^3 + O(t^4) \quad (C.3)$$

Setting $y = h(1 + t)$, solving the second series for $t$ in terms of $y$, and substituting into the first series gives

$$m_d(y) = y - \left( \frac{d^2 - 2d}{3(d - 1)^2} \right) \left( \frac{2(d - 1)}{d} \right)^{3/2} y^{3/2} + \ldots \quad (C.4)$$

Thus for sufficiently small $y$ we have $m_d(y)/y \sim 1 - k\sqrt{y}$, which implies both statements in Lemma 8.