Superadditivity of communication capacity using entangled inputs

M. B. Hastings[1]

[1] *Center for Nonlinear Studies and Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, 87545*

The design of error-correcting codes used in modern communications relies on information theory to quantify the capacity of a noisy channel to send information[1]. This capacity can be expressed using the mutual information between input and output for a single use of the channel: although correlations between subsequent input bits are used to correct errors, they cannot increase the capacity. For quantum channels, it has been an open question whether entangled input states can increase the capacity to send classical information[2]. The additivity conjecture[3, 4] states that entanglement does not help, making practical computations of the capacity possible. While additivity is widely believed to be true, there is no proof. Here we show that additivity is false, by constructing a random counter-example. Our results show that the most basic question of classical capacity of a quantum channel remains open, with further work needed to determine in which other situations entanglement can boost capacity.

In the classical setting, Shannon presented a formal definition of a noisy channel $\mathcal{E}$ as a probabilistic map from input states to output states. In the quantum setting, the channel becomes a linear, completely positive, trace-preserving map from density matrices to density matrices, modeling noise in the system due to interaction with an environment. Such a channel can be used to send either quantum or classical information. In the first case, a dramatic violation of *operational additivity* was recently shown, in that there exist two channels, each of which has zero capacity to send quantum information no matter how many times it is used, but which can be used in tandem to send quantum information[5].

Here we address the classical capacity of a quantum channel. To specify how information is encoded in the channel, we must pick a set of states ρ_i which we use as input signals with with probabilities p_i. Then the Holevo formula[2] for the capacity is:

$$\chi = H\Big(\sum_i p_i \mathcal{E}(\rho_i)\Big) - \sum_i p_i H\Big(\mathcal{E}(\rho_i)\Big), \qquad (1)$$

where $H(\rho) = -\mathrm{Tr}(\rho \ln(\rho))$ is the von Neumann entropy. The maximum capacity of a channel is the maximum over all input ensembles:

$$\chi_{\max}(\mathcal{E}) = \max_{\{p_i\},\{\rho_i\}} \chi(\mathcal{E}, \{p_i\}, \{\rho_i\}). \qquad (2)$$

Suppose we have two different channels, $\mathcal{E}_1, \mathcal{E}_2$. To compute this capacity, it seems necessary to consider entangled input states between the two channels. Similarly, when using the same channel multiple times, it may be useful to use input states which are entangled across multiple uses of the same channel. The additivity conjecture (see Figure 1) is the conjecture that this does not help and that instead

$$\chi_{\max}(\mathcal{E}_1 \otimes \mathcal{E}_2) = \chi_{\max}(\mathcal{E}_1) + \chi_{\max}(\mathcal{E}_2). \qquad (3)$$

The additivity conjecture makes it possible to compute the classical capacity of a quantum channel. Further, Shor[4] showed that several different additivity conjectures in quantum information theory are all equivalent. These are the additivity conjecture for the Holevo capacity, the additivity conjecture for entanglement of formation[6], strong superadditivity of entanglement of formation[7], and the additivity conjecture for minimum output entropy[3]. In this Letter, we show that all of these conjectures are false, by constructing a counterexample to the last of these conjectures. Given a channel $\mathcal{E}$, define the minimum output entropy $H^{\min}$ by

$$H^{\min}(\mathcal{E}) = \min_{|\psi\rangle} H(\mathcal{E}(|\psi\rangle\langle\psi|)). \qquad (4)$$

The minimum output entropy conjecture is that for all channels $\mathcal{E}_1$ and $\mathcal{E}_2$, we have

$$H^{\min}(\mathcal{E}_1 \otimes \mathcal{E}_2) = H^{\min}(\mathcal{E}_1) + H^{\min}(\mathcal{E}_2). \qquad (5)$$

A counterexample to this conjecture would be an entangled input state which has a lower output entropy, and hence is *more* resistant to noise, than any unentangled state (see Figure 2).

Our counterexample to the additivity of minimum output entropy is based on a random construction, similar to those Winter and Hayden used to show violation of the maximal p-norm multiplicativity conjecture for all $p > 1$[8, 9, 10]. For $p = 1$, this violation would imply violation of the minimum output entropy conjecture; however, the counterexample found in [9] requires a matrix size which diverges as $p \to 1$. We use different system and environment sizes (note that $D << N$ in our construction below) and make a different analysis of the probability of different output entropies. Other violations are known for p close to 0[11].

We define a pair of channels $\mathcal{E}$ and $\overline{\mathcal{E}}$ which are complex conjugates of each other. Each channel acts by randomly

choosing a unitary from a small set of unitaries U_i ($i = 1...D$) and applying that to ρ. This models a situation in which the unitary evolution of the system is determined by an unknown state of the environment. We define

$$\begin{aligned} \mathcal{E}(\rho) &= \sum_{i=1}^{D} P_i U_i^\dagger \rho U_i, \\ \overline{\mathcal{E}}(\rho) &= \sum_{i=1}^{D} P_i \overline{U}_i^\dagger \rho \overline{U}_i, \end{aligned} \tag{6}$$

where the U_i are N-by-N unitary matrices, chosen at random from the Haar measure, and the probabilities P_i are chosen randomly as described in the Supplemental Equations. The P_i are all roughly equal. We pick

$$1 << D << N. \tag{7}$$

We show in the Supplemental Equations that

Theorem 1. *For sufficiently large D, for sufficiently large N, there is a non-zero probability that a random choice of U_i from the Haar measure and of P_i (as described in Supplemental Equations) will give a channel $\mathcal{E}$ such that*

$$\begin{aligned} H^{\min}(\mathcal{E} \otimes \overline{\mathcal{E}}) &< H^{\min}(\mathcal{E}) + H^{\min}(\overline{\mathcal{E}}) \\ &= 2H^{\min}(\mathcal{E}). \end{aligned} \tag{8}$$

The size of N depends on D.

For any pure state input, the output entropy of $\mathcal{E}$ is at most $\ln(D)$ and that of $\mathcal{E} \otimes \overline{\mathcal{E}}$ is at most $2\ln(D)$. To show theorem (1), we first exhibit an entangled state with a lower output entropy for the channel $\mathcal{E} \otimes \overline{\mathcal{E}}$. The entangled state we use is the maximally entangled state:

$$|\Psi_{\rm ME}\rangle = (1/\sqrt{N}) \sum_{\alpha=1}^{N} |\alpha\rangle \otimes |\alpha\rangle. \tag{9}$$

As shown in Lemma 1 in the Supplemental Equations, the output entropy for this state is bounded by

$$H\Big(\mathcal{E} \otimes \overline{\mathcal{E}}(|\Psi_{\rm ME}\rangle\langle\Psi_{\rm ME}|)\Big) \leq 2\ln(D) - \ln(D)/D. \tag{10}$$

We then use the random properties of the channel to show that no product state input can obtain such a low output entropy. Lemmas 2-5 in the Supplemental Equations show that, with non-zero probability, the entropy $H^{\min}(\mathcal{E})$ is at least $\ln(D) - \delta S^{\max}$, for

$$\delta S^{\max} = c_1/D + p_1(D)\mathcal{O}(\sqrt{\ln(N)/N})), \tag{11}$$

where c_1 is a constant and $p_1(D) = \text{poly}(D)$. Thus, since for large enough D, for large enough N we have $2\delta S^{\max} < \ln(D)/D$, the theorem follows.

The output entropy can be understood differently: for a given pure state input, can we determine from the output which of the unitaries $U_i^\dagger$ was applied? Recall that

$$U^\dagger \otimes \overline{U}^\dagger |\Psi_{\rm ME}\rangle = |\Psi_{\rm ME}\rangle. \tag{12}$$

for any unitary U. This means that, for the maximally entangled state, if a unitary $U_i^\dagger$ was applied to one subsystem, and $\overline{U}_i^\dagger$ was applied to the other subsystem, we cannot determine which unitary i was applied by looking at the output. This is the key idea behind Eq. (10).

Note that the minimum output entropy of $\mathcal{E}$ must be less than $\ln(D)$ by an amount at least of order $1/D$. Suppose U_1 and U_2 are the two unitaries with the largest l_i. Choose a state $|\psi\rangle$ which is an eigenvector of $U_1 U_2^\dagger$. For this state, we cannot distinguish between the states $U_1^\dagger|\psi\rangle$ and $U_2|\psi\rangle$, and so

$$H^{\min}(\mathcal{E}) \leq \ln(D) - (2/D)\ln(2). \tag{13}$$

Our randomized analysis bounds how much further the output entropy of the channel $\mathcal{E}$ can be lowered for a random choice of U_i.

Our work raises the question of how strong a violation of additivity is possible. The relative violation we have found is numerically small, but it may be possible to increase this, and to find new situations in which entangled inputs can be used to increase channel capacity, or novel situations in which entanglement can be used to protect against decoherence in practical devices. The map $\mathcal{E}$ is similar to that used[12] to construct random quantum expanders[13, 14], raising the possibility that deterministic expander constructions can provide stronger violations of additivity.

While we have used two different channels, it is also possible to find a single channel $\mathcal{E}$ such that $H^{\min}(\mathcal{E} \otimes \mathcal{E}) < 2H^{\min}(\mathcal{E})$, by choosing U_i from the orthogonal group. Alternately, we can add an extra classical input used to "switch" between $\mathcal{E}$ and $\overline{\mathcal{E}}$, as suggested to us by P. Hayden.

The equivalence of the different additivity conjecture[4] means that the violation of any one of the conjectures has profound impacts. The violation of additivity of the Holevo capacity means that the problem of channel capacity remains open, since if a channel is used many times, we must do an intractable optimization over all entangled inputs to find the maximum capacity. However, we conjecture that additivity holds for all channels of the form

$$\mathcal{E} = \mathcal{F} \otimes \overline{\mathcal{F}}. \tag{14}$$

Our intuition for this conjecture is that we believe that *multi-party* entanglement (between the inputs to three or more channels) is not useful, because it is very unlikely for all channels to apply the same unitary; note that the state $\Psi_{\rm ME}$ has a low minimum output entropy precisely

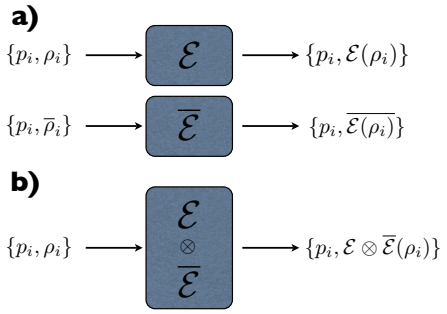

FIG. 1: Communicating classical information over a quantum channel. A set of states ρ_i are used with probabilities p_i as signal states on the channel. In (a), we use input states which are unentangled between channels $\mathcal{E}$ and $\overline{\mathcal{E}}$. In (b), we allow entanglement. The capacity of $\mathcal{E}$ is equal to $\overline{\mathcal{E}}$. The question addressed is whether entangling, as shown in (b), can increase this capacity.

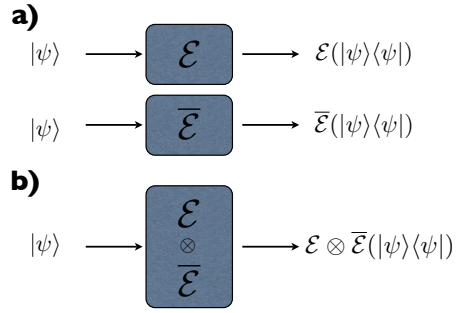

FIG. 2: Minimum output entropy of a quantum channel. A pure state $|\psi\rangle$ is input to the channel. While the input is a pure state, the output may be a mixed state. We attempt to minimize the entropy of the output state over all pure input states. The question addressed is whether an entangled input state, as shown in (b), can have a lower output entropy for channel $\mathcal{E} \otimes \overline{\mathcal{E}}$, than the sum of the minimum output entropies for the two channels.

because it is left unchanged as in Eq. (12) if both channels apply corresponding unitaries. This *two-letter* additivity conjecture would allow us to restrict our attention to considering input states with a bipartite entanglement structure, possibly opening the way to computing the capacity for arbitrary channels.

Acknowledgments— I thank J. Yard, P. Hayden, and A. Harrow. This work was supported by U. S. DOE Contract No. DE-AC52-06NA25396.

Supplemental Equations:

To choose the P_i, we first choose a set of amplitudes l_i as follows. For $i = 1, ..., D$ pick $l_i \geq 0$ independently from a probability distribution with

$$P(l_i) \propto l_i^{2N-1} \exp(-NDl_i^2), \tag{15}$$

where the proportionality constant is chosen such that $\int_0^\infty P(l_i)\mathrm{d}l_i = 1$. This distribution is the same as that of the length of a random vector chosen from a Gaussian distribution in N complex dimensions. Then, define

$$L = \sqrt{\sum_i l_i^2}. \tag{16}$$

Then we set

$$P_i = l_i^2/L^2, \tag{17}$$

so that

$$\begin{aligned} \mathcal{E}(\rho) &= \sum_{i=1}^{D} \frac{l_i^2}{L^2} U_i^\dagger \rho U_i, \\ \overline{\mathcal{E}}(\rho) &= \sum_{i=1}^{D} \frac{l_i^2}{L^2} \overline{U}_i^\dagger \rho \overline{U}_i, \end{aligned} \tag{18}$$

The only reason in what follows for not choosing all the probabilities equal to $1/D$ is that the choice we made will allow us to appeal to certain exact results on random bipartite states later.

We also define the conjugate channel

$$\mathcal{E}^C(\rho) = \sum_{i=1}^{D}\sum_{j=1}^{D} \frac{l_i l_j}{L^2} \mathrm{Tr}\Big(U_i^\dagger \rho U_j\Big) |i\rangle\langle j|, \tag{19}$$

As shown in [15]

$$H^{\min}(\mathcal{E}) = H^{\min}(\mathcal{E}^C). \tag{20}$$

In the $\mathcal{O}(...)$ notation that follows, we will take

$$1 << D << N. \tag{21}$$

We use "computer science" big-O notation throughout, rather than "physics" big-O notation. That is, if we state that a quantity is $\mathcal{O}(N)$, it means that it is asymptotically bounded by a constant times N, and may in fact be much smaller. For example, $\sqrt{N}$ is $\mathcal{O}(N)$ in computer science notation but not in physics notation.

Theorem 1 follows from two lemmas below, 1 and 5, which give small corrections to the naive estimates of $2\ln(D)$ and $\ln(D)$ for the entropies. Lemma 1 upper bounds $H^{\min}(\mathcal{E} \otimes \overline{\mathcal{E}})$ by $2\ln(D) - \ln(D)/D$. Lemma 5 shows that for given D, for sufficiently large N, with non-zero probability, the entropy $H^{\min}(\mathcal{E})$ is at least $\ln(D) - \delta S^{\max}$, for

$$\delta S^{\max} = c_1/D + p_1(D)\mathcal{O}(\sqrt{\ln(N)/N})), \tag{22}$$

where c_1 is a constant and $p_1(D) = \mathrm{poly}(D)$. Thus, since for large enough D, for large enough N we have $2\delta S^{\max} < \ln(D)/D$, the theorem follows.

Lemma 1. *For any D and N, we have*

$$\begin{aligned} H^{\min}(\mathcal{E} \otimes \overline{\mathcal{E}}) &\leq \frac{1}{D}\ln(D) + \frac{D-1}{D}\ln(D^2) \\ &= 2\ln(D) - \frac{1}{D}\ln(D). \end{aligned} \tag{23}$$

Proof. Consider the maximally entangled state, $|\Psi_{\rm ME}\rangle = (1/\sqrt{N})\sum_{\alpha=1}^{N}|\alpha\rangle\otimes|\alpha\rangle$. Then,

$$\begin{aligned}\mathcal{E}\otimes\overline{\mathcal{E}}(|\Psi_{\rm ME}\rangle\langle\Psi_{\rm ME}|) &= \Big(\sum_i \frac{l_i^4}{L^4}\Big)|\Psi_{\rm ME}\rangle\langle\Psi_{\rm ME}| \\ &\quad+\sum_{i\neq j}\Big(\frac{l_i^2 l_j^2}{L^4}\Big)\Big(U_i^\dagger\otimes\overline{U}_j^\dagger\Big)|\Psi_{\rm ME}\rangle\langle\Psi_{\rm ME}|\Big(U_i\otimes\overline{U}_j\Big).\end{aligned} \tag{24}$$

Since the states $|\Psi_{\rm ME}\rangle\langle\Psi_{\rm ME}|$ and $(U_i^\dagger\otimes\overline{U}_j^\dagger)|\Psi_{\rm ME}\rangle\langle\Psi_{\rm ME}|(U_i\otimes\overline{U}_j)$ are pure states, the entropy of the state in (24) is bounded by

$$H(\mathcal{E}\otimes\overline{\mathcal{E}}(|\Psi_{\rm ME}\rangle\langle\Psi_{\rm ME}|)) \leq -\Big(\sum_i\frac{l_i^4}{L^4}\Big)\ln(\sum_i\frac{l_i^4}{L^4}) - \sum_{i\neq j}\Big(\frac{l_i^2l_j^2}{L^4}\Big)\ln(\frac{l_i^2l_j^2}{L^4}). \tag{25}$$

To show Eq. (25), let $\rho_{ij} = U_i^\dagger\otimes\overline{U}_j^\dagger|\Psi_{\rm ME}\rangle\langle\Psi_{\rm ME}|U_i\otimes\overline{U}_j$. Note that $\rho_{ii}=\rho_{jj}=|\Psi_{\rm ME}\rangle\langle\Psi_{\rm ME}|$ for all i,j. Then, the entropy is equal to

$$\begin{aligned}H(\mathcal{E}\otimes\overline{\mathcal{E}}(|\Psi_{\rm ME}\rangle\langle\Psi_{\rm ME}|)) &= -{\rm Tr}\Big[\Big(\sum_i\frac{l_i^4}{L^4}\Big)|\Psi_{\rm ME}\rangle\langle\Psi_{\rm ME}|\ln\Big(\mathcal{E}\otimes\overline{\mathcal{E}}(|\Psi_{\rm ME}\rangle\langle\Psi_{\rm ME}|)\Big)\Big] \\ &\quad-\sum_{i\neq j}{\rm Tr}\Big[\frac{l_i^2l_j^2}{L^4}\rho_{ij}\ln\Big(\mathcal{E}\otimes\overline{\mathcal{E}}(|\Psi_{\rm ME}\rangle\langle\Psi_{\rm ME}|)\Big)\Big].\end{aligned} \tag{26}$$

Using the fact that the logarithm is an operator monotone function[16], we find that $\ln\Big(\mathcal{E}\otimes\overline{\mathcal{E}}(|\Psi_{\rm ME}\rangle\langle\Psi_{\rm ME}|)\Big) \geq \ln(\sum_i\frac{l_i^4}{L^4}|\Psi_{\rm ME}\rangle\langle\Psi_{\rm ME}|)$, and also that $\ln\Big(\mathcal{E}\otimes\overline{\mathcal{E}}(|\Psi_{\rm ME}\rangle\langle\Psi_{\rm ME}|)\Big) \geq \ln(\frac{l_i^2l_j^2}{L^4}\rho_{ij})$ for all i,j. Inserting these inequalities into Eq. (26), we arrive at Eq. (25).

We claim that the right-hand side of Eq. (25) is bounded by

$$-\Big(\sum_i\frac{l_i^4}{L^4}\Big)\ln(\sum_i\frac{l_i^4}{L^4}) - \sum_{i\neq j}\Big(\frac{l_i^2l_j^2}{L^4}\Big)\ln(\frac{l_i^2l_j^2}{L^4}) \leq \frac{1}{D}\ln(D) + \frac{D-1}{D}\ln(D^2). \tag{27}$$

To show Eq. (27), define $P_{same} = \sum_i l_i^4/L^4$. We claim that $P_{same}\geq 1/D$. To see this, consider the real vectors $(l_1^2/L^2,...,l_D^2/L^2)$ and $(1,...,1)$. The inner product of these vectors is equal to 1 since $\sum_i l_i^2/L^2 = 1$ while the norms of the vectors are $\sqrt{P_{same}}$ and $\sqrt{D}$, respectively. Applying the Cauchy-Schwarz inequality to this inner product, we find that $P_{same}\geq 1/D$ as claimed. Then the left-hand side of Eq. (27) is equal to

$$\begin{aligned}-\Big(\sum_i\frac{l_i^4}{L^4}\Big)\ln(\sum_i\frac{l_i^4}{L^4}) - \sum_{i\neq j}\Big(\frac{l_i^2l_j^2}{L^4}\Big)\ln(\frac{l_i^2l_j^2}{L^4}) &= -P_{same}\ln(P_{same}) - (1-P_{same})\ln(1-P_{same}) \\ &\quad-(1-P_{same})\sum_{i\neq j}\Big(\frac{l_i^2l_j^2}{(1-P_{same})L^4}\Big)\ln\Big(\frac{l_i^2l_j^2}{(1-P_{same})L^4}\Big) \\ &\leq -P_{same}\ln(P_{same}) - (1-P_{same})\ln(1-P_{same}) \\ &\quad+(1-P_{same})\ln(D^2-D).\end{aligned} \tag{28}$$

The last line of Eq. (28) is maximized at $P_{same}=1/D$, giving Eq. (27), which implies Eq. (23). □

Lemma 2. *Consider a random bipartite pure state $|\psi\rangle\langle\psi|$ on a bipartite system with subsystems B and E with dimensions N and D respectively. Let ρ_E be the reduced density matrix on E. Then, the probability density that ρ_E has a given set of eigenvalues, $p_1,...,p_D$, is bounded by*

$$\begin{aligned}&\tilde{P}(p_1,...,p_D)\prod_i {\rm d}p_i \\ &= \mathcal{O}(N)^{\mathcal{O}(D^2)}D^{(N-D)D}\delta(1-\sum_{i=1}^D p_i)\prod_{i=1}^D p_i^{N-D}{\rm d}p_i. \\ &= \mathcal{O}(N)^{\mathcal{O}(D^2)}\delta(1-\sum_{i=1}^D p_i)\prod_{i=1}^D F(p_i){\rm d}p_i,\end{aligned} \tag{29}$$

where we define

$$F(p) = D^{N-D} p^{N-D} \exp[-(N-D)D(p-1/D)] \tag{30}$$

Note that $F(p) \leq 1$ *for all* $0 \leq p \leq 1$.

Similarly, consider a random state pure state $\rho = |\chi\rangle\langle\chi|$ *on an* N *dimensional space, and a channel* $\mathcal{E}^C(...)$ *as defined in Eq. (18), with unitaries* U_i *chosen randomly from the Haar measure and the numbers* l_i *chosen as described in Eq. (15) and with* $N > D$. *Then, the probability density that the eigenvalues of* $\mathcal{E}^C(\rho)$ *assume given values* $p_1, ..., p_D$ *is bounded by the same function* $\tilde{P}(p_1, ..., p_D) \prod_i \mathrm{d}p_i$ *as above.*

Proof. As shown in[17, 18], the exact probability distribution of eigenvalues is

$$P(p_1, ..., p_D) \prod_i \mathrm{d}p_i \propto \delta(1 - \sum_{i=1}^{D} p_i) \prod_{1 \leq j < k \leq D} (p_j - p_k)^2 \prod_{i=1}^{D} p_i^{N-D} \mathrm{d}p_i, \tag{31}$$

where the constant of proportionality is given by the requirement that the probability distribution integrate to unity. The proportionality constant is $\mathcal{O}(N)^{\mathcal{O}(D^2)} D^{(N-D)D}$ as we show below, and for $0 \leq p_i \leq 1$

$$\prod_{1 \leq j < k \leq D} (p_j - p_k)^2 \prod_{i=1}^{D} p_i^{N-D} \leq \prod_{i=1}^{D} p_i^{N-D}, \tag{32}$$

so Eq. (29) follows. The second equality in (29) holds because $\sum_i (p_i - 1/D) = 0$.

Given a random pure state $|\chi\rangle\langle\chi|$, with U_i and l_i chosen as described above, then the state $\mathcal{E}^C(|\chi\rangle\langle\chi|)$ has the same eigenvalue distribution as the reduced density matrix of a random bipartite state, so the second result follows. To see that the eigenvalue distribution of a random bipartite state in DN dimensions is indeed the same as that of $\mathcal{E}^C(|\chi\rangle\langle\chi|)$, we consider the reduced density matrix on the N dimensional system of the random bipartite state and show that it has the same statistical properties as $\mathcal{E}(|\chi\rangle\langle\chi|)$. We choose the DN different amplitudes of the unnormalized bipartite state from a Gaussian distribution. Equivalently, for each $i = 1, ..., D$ corresponding to a given state in the environment, we choose an N dimensional vector $|v_i\rangle$ from a Gaussian distribution. Thus, before normalization, the reduced density matrix of the random bipartite state on the N dimensional system has the same statistics as the sum $\sum_{i=1}^{D} |v_i\rangle\langle v_i|$ where the $|v_i\rangle$ are states drawn from a Gaussian distribution. The state $\mathcal{E}(|\chi\rangle\langle\chi|)$ is the sum $\sum_{i=1}^{D} (l_i^2/L^2) U_i^\dagger |\chi\rangle\langle\chi| U_i$. The l_i^2 have the same statistics as $|v_i|^2$, while the directions of the vectors $U_i^\dagger |\chi\rangle$ are independent and uniformly distributed, as are the directions of the $|v_i\rangle$. The factor of L^2 takes into account the normalization, so that $\mathcal{E}(|\chi\rangle\langle\chi|)$ indeed has the same statistics as the normalized bipartite state as claimed.

Finally, we show how to upper bound the proportionality constant. One approach is to keep track of constant factors of N in the derivation of [17, 18]. Another approach, which we explain here, is to lower bound the integral $\int \delta(1 - \sum_{i=1}^{D} p_i) \prod_{1 \leq j < k \leq D} (p_j - p_k)^2 \prod_{i=1}^{D} p_i^{N-D} \mathrm{d}p_i$. As a lower bound on the integral, we restrict to a subregion of the integration domain: we assume that the i-th eigenvalue p_i falls into a narrow interval of width $1/N$, and we choose these intervals such that $|p_i - p_j| \geq 1/N$ for $i \neq j$ and such that $|p_i - 1/D| \leq \mathcal{O}(D)/N$. To do this, for example, we can require that the i-th eigenvalue p_i obey $1/D + (2i - D - 3/2)/N \leq p_i \leq 1/D + (2i - D - 1/2)/N$. Then, in this subregion, $\prod_{1 \leq j < k \leq D} (p_j - p_k)^2 \geq (1/N)^{D^2}$, and $\prod_{i=1}^{D} p_i^{N-D} \geq (1/D - \mathcal{O}(D/N))^{(N-D)D}$. The centers of the intervals were chosen such that if each eigenvalue is at the center, then $\sum_i p_i = 1$; we can then estimate the volume of the subregion as $\approx \sqrt{1/D}(1/N)^{D-1}$. Combining these estimates, we lower bound the integral as desired. □

Remark: In order to get some understanding of the probability of having a given fluctuation in the entropy, we consider a Taylor expansion about $p_i = 1/D$. The next three paragraphs are not intended to be rigourous and are not used in the later proof. Instead, they are intended to, first, give some rough idea of the probability of a given fluctuation in the entropy, and, second, explain why ϵ-nets do not suffice to give sufficiently tight bounds on the probability of having a given fluctuation in the entropy and hence why we turn to a slightly more complicated way of estimating this probability in lemmas 3-5.

If all the probabilities p_i are close to $1/D$, so that $p_i = 1/D + \delta p_i$ for small δp_i, we can Taylor expand the last line of (29), using $p_i^{N-D} = \exp[(N-D)\ln(p_i)]$, to get:

$$\tilde{P}(p_1, ..., p_D) \approx \mathcal{O}(N)^{\mathcal{O}(D^2)} \exp[-(N-D)D^2 \sum_i \delta p_i^2/2 + ...]. \tag{33}$$

Similarly, we can expand

$$S = -\sum_i p_i \ln(p_i) \approx \ln(D) - D\sum_i \delta p_i^2/2 + ... \tag{34}$$

Using Eq. (33,34), we find that the probability of having $S = \ln(D) - \delta S$ is roughly $\mathcal{O}(N)^{\mathcal{O}(D^2)} \exp[-(N-D)D\delta S]$.

Using ϵ-nets, these estimates (33,34) give some motivation for the construction we are using, but just fail to give a good enough bound on their own: define an ϵ-net with distance $d << 1$ between points on the net. There are then $\mathcal{O}(d^{-2N})$ points in the net. Then, the probability that, for a random U_i, l_i, at least one point on the net has a given δS is bounded by $\approx \exp[-ND\delta S + 2N\ln(1/d)]$. Thus, the probability of having a $\delta S = \ln(D)/2D$ is less than one for $d \geq D^{-1/4}$. However, in order to use ϵ-nets to show that it is unlikely to have any state $|\psi\rangle$ with given δS, we need to take a sufficiently dense ϵ-net. If there exists a state $|\psi^0\rangle$ with given δS^0, then any state within distance d will have, by Fannes inequality[19], a $\delta S \geq \delta S^0 - d^2 \ln(D/d^2)$, and therefore we will need to take a d of roughly $1/\sqrt{D}$ in order to use the bounds on δS for points on the net to get bounds on δS^0 with an accuracy $\mathcal{O}(1/D)$.

However, in fact this Fannes inequality estimate is usually an overestimate of the change in entropy. Given a state $|\psi^0\rangle$ with a large δS^0, random nearby states χ can be written as a linear combination of $|\psi^0\rangle$ with a random orthogonal vector $|\phi\rangle$. Since $\mathcal{E}^C(|\phi\rangle\langle\phi|)$ will typically by close to a maximally mixed state for random $|\phi\rangle$, and typically will also have almost vanishing trace with $\mathcal{E}^C(|\psi^0\rangle\langle\psi^0|)$, the state $\mathcal{E}^C(|\chi\rangle\langle\chi|)$ will typically be close to a mixture of $\mathcal{E}^C(|\psi^0\rangle\langle\psi^0|)$ with the maximally mixed state, and hence will also have a relatively large δS. This idea motivates what follows.

Definitions: We will say that a density matrix ρ is "close to maximally mixed" if the eigenvalues p_i of ρ all obey

$$|p_i - 1/D| \leq c_{MM}\sqrt{\ln(N)/(N-D)}, \tag{35}$$

where the constant c_{MM} will be chosen later. For any given channel $\mathcal{E}^C$, let $P_{\mathcal{E}^C}$ denote the probability that, for a randomly chosen $|\chi\rangle$, the density matrix $\mathcal{E}^C(|\chi\rangle\langle\chi|)$ is close to maximally mixed. Let Q denote the probability that a random choice of U_i from the Haar measure and a random choice of numbers l_i produces a channel $\mathcal{E}^C$ such that $P_{\mathcal{E}^C}$ is less than $1/2$. Note: we are defining Q to be the probability of a probability here. Then,

Lemma 3. *For an appropriate choice of c_{MM}, the probability Q can be made arbitrarily close to zero for all sufficiently large D and N/D.*

Proof. The probability Q is less than or equal to 2 times the probability that for a random U_i, random l_i, and random $|\chi\rangle$, the density matrix $\mathcal{E}^C(|\chi\rangle\langle\chi|)$ is not close to maximally mixed. From (29), and as we will explain further in the next paragraph, this probability is bounded by the maximum over p such that $|p - 1/D| > c_{MM}\sqrt{\ln(N)/(N-D)}$ of

$$\begin{aligned} & \mathcal{O}(N^2)^{\mathcal{O}(D^2)} F(p) \\ = \; & \mathcal{O}(N^2)^{\mathcal{O}(D^2)} D^{N-D} p^{N-D} \exp[-(N-D)D(p-1/D)] \\ \approx \; & \exp[\mathcal{O}(D^2)\ln(N) - (N-D)D^2 c_{MM}^2 (\ln(N)/(N-D))/2 + ...]. \end{aligned} \tag{36}$$

By picking c_{MM} large enough, we can make this probability

$$\max_{p, |p-1/D| > c_{MM}\sqrt{\ln(N)/(N-D)}} \Big(\mathcal{O}(N^2)^{\mathcal{O}(D^2)} F(p)\Big) \tag{37}$$

arbitrarily small for sufficiently large D and N/D.

The fact that $F(p) \leq 1$ for all $0 \leq p \leq 1$ is important in the claim that (37) indeed is a bound on the given probability. To compute the probability density for a given set of eigenvalues, p_i, such that for some j we have $|p_j - 1/D| > c_{MM}\sqrt{\ln(N)/(N-D)}$, we can use the bound $F(p) \leq 1$ to show that $\mathcal{O}(N^2)^{\mathcal{O}(D^2)} \prod_{i=1}^D F(p_i) \mathrm{d}p_i$ is bounded by $\mathcal{O}(N^2)^{\mathcal{O}(D^2)} F(p_j) \prod_{i=1}^D \mathrm{d}p_i$. Therefore, Eq. (37) gives a bound on the probability *density* under the assumption that for some j we have $|p_j - 1/D| > c_{MM}\sqrt{\ln(N)/(N-D)}$.

To turns this bound on the probability density into a bound on the probability, note that the total integration volume $\int \delta(1 - \sum_{i=1}^D p_i) \prod_{i=1}^D \mathrm{d}p_i$ is bounded by unity, and the set of p_i such that for some j we have $|p_j - 1/D| > c_{MM}\sqrt{\ln(N)/(N-D)}$ is a subset of the set of all p_i.

Finally, note that the maximum of Eq. (37) is achieved at $|p - 1/D| = c_{MM}\sqrt{\ln(N)/(N-D)}$ and it is straightforward to control the higher terms in the Taylor expansion of (36) in that case. □

The next lemma is the crucial step.

Lemma 4. *Consider a given choice of U_i and l_i which give a $\mathcal{E}^C$ such that $P_{\mathcal{E}^C} \geq 1/2$. Suppose there exists a state $|\psi^0\rangle$, such that $\mathcal{E}^C(|\psi^0\rangle\langle\psi^0|)$ has given eigenvalues $p_1, ..., p_D$. Let P_{near} denote the probability that, for a randomly chosen state $|\chi\rangle$, the density matrix $\mathcal{E}^C(|\chi\rangle\langle\chi|)$ has eigenvalues $q_1, ..., q_D$ which obey*

$$|q_i - (yp_i + (1-y)(1/D))| \leq \text{poly}(D)\mathcal{O}(\sqrt{\ln(N)/(N-D)}) \tag{38}$$

for some $y \geq 1/2$. Then,

$$P_{near} \geq \exp(-\mathcal{O}(N))(1/2 - 1/\text{poly}(D)), \tag{39}$$

where the power of D in the polynomial in (39) can be made arbitrarily large by an appropriate choice of the polynomial in (38).

Proof. Consider a random state χ. We can write $|\chi\rangle$ as a linear combination of $|\psi^0\rangle$ and a state $|\phi\rangle$ which is orthogonal to $\psi^0\rangle$ as follows:

$$|\chi\rangle = z\sqrt{1-x^2}|\psi^0\rangle + x|\phi\rangle, \tag{40}$$

where z is a phase: $|z| = 1$.

For random χ, the probability that $x^2 \leq 1/2$ is $\exp(-\mathcal{O}(N))$. We can also calculate this probability exactly. Let S_n be the surface area of a unit hypersphere in n dimensions. Then, the probability that $x^2 \leq 1/2$ is equal to

$$\begin{aligned} & S_{2N}^{-1} \int_0^{\pi/4} 2\pi \cos(\theta)\sin(\theta)^{2N-3} S_{2N-2}\, \mathrm{d}\theta \\ = \; & (1/\sqrt{2})^{2N-2} \\ = \; & \exp[-\ln(2)(N-1)] \\ = \; & \exp(-\mathcal{O}(N)). \end{aligned} \tag{41}$$

Since χ is random, the probability distribution of $|\phi\rangle$ is that of a random state with $\langle\phi|\psi^0\rangle = 0$. One way to generate such a random state $|\phi\rangle$ with this property is to choose a random state $|\theta\rangle$ and set

$$|\phi\rangle = \frac{1}{\sqrt{1-|\langle\psi^0|\theta\rangle|^2}}\Big(1 - |\psi^0\rangle\langle\psi^0|\Big)|\theta\rangle. \tag{42}$$

If we choose a random state $|\theta\rangle$, then with probability at least $1/2$, the state $\mathcal{E}^C(|\theta\rangle\langle\theta|)$ is close to maximally mixed. Further, for any given i, j, the probability that $|\langle\psi^0|U_i U_j^\dagger|\theta\rangle|$ is greater than $\mathcal{O}(\sqrt{\ln(D)/N})$ is $1/\text{poly}(D)$, and the polynomial $\text{poly}(D)$ can be chosen to be any given power of D by appropriate choice of the constant hidden in the $\mathcal{O}$ notation for $\mathcal{O}(\sqrt{\ln(D)/N})$. Therefore,

$$\Pr\Big[\text{Tr}\Big(|\mathcal{E}^C(|\theta\rangle\langle\psi^0|)|\Big) \geq \text{poly}(D)\sqrt{\ln(D)/N}\Big] \leq 1/\text{poly}(D), \tag{43}$$

with any desired power of D in the polynomial on the right-hand side (the notation $\text{Tr}(|...|)$ is used to denote the trace norm here).

Then, since

$$\Pr\Big[\Big||\phi\rangle - |\theta\rangle\Big| \geq \mathcal{O}(\sqrt{\ln(D)/N})\Big] \leq 1/\text{poly}(D), \tag{44}$$

we find that

$$\begin{aligned} & \Pr\Big[\text{Tr}\Big(|\mathcal{E}^C(|\phi\rangle\langle\psi^0|)|\Big) \geq \text{poly}(D)\sqrt{\ln(D)/N}\Big] \\ \leq \; & 1/\text{poly}(D), \end{aligned} \tag{45}$$

with again any desired power in the polynomial.

The probability that $\mathcal{E}^C(|\theta\rangle\langle\theta|)$ is close to maximally mixed is at least $1/2$, and so by (19,44) the probability that the eigenvalues $r_1, ..., r_D$ of $\mathcal{E}^C(|\phi\rangle\langle\phi|)$ obey

$$\begin{aligned} |r_i - 1/D| \; & \leq \; c_{MM}\sqrt{\ln(N)/(N-D)} + \text{poly}(D)(\ln(D)/N) \\ & \leq \; \text{poly}(D)\mathcal{O}(\sqrt{\ln(N)/N}) \end{aligned} \tag{46}$$

is at least $1/2 - 1/\mathrm{poly}(D)$. Let

$$y = 1 - x^2. \tag{47}$$

Thus, since

$$\begin{aligned} \mathcal{E}^C(|\chi\rangle\langle\chi|) &= (1-x^2)\mathcal{E}^C(|\psi^0\rangle\langle\psi^0|) + x^2\mathcal{E}^C(|\phi\rangle\langle\phi|) \\ &\quad + \Big(\overline{z}x\sqrt{1-x^2}\mathcal{E}^C(|\phi\rangle\langle\psi^0|) + h.c.\Big), \end{aligned} \tag{48}$$

using Eq. (45) we find that for given x, the probability that a randomly chosen $|\phi\rangle$ gives a state with eigenvalues $q_1, ..., q_D$ such that

$$|q_i - (yp_i + (1-y)(1/D))| \leq \mathrm{poly}(D)\mathcal{O}(\sqrt{\ln(N)/N}) \tag{49}$$

is $1/2 - 1/\mathrm{poly}(D)$. Combining this result with the $\exp(-\mathcal{O}(N))$ probability of $x^2 \leq 1/2$, the claim of the lemma follows. □

We now give the last lemma which shows a lower bound, with non-zero probability, on $H^{min}(\mathcal{E}^C)$. The basic idea of the proof is to estimate the probability that a random state input into a random channel $\mathcal{E}^C$ gives an output state with moderately low output entropy (defined slightly differently below in terms of properties of the eigenvalues of the output density matrix). We estimate this probability in two different ways. First, we estimate the probability of such an output state conditioned on $\mathcal{E}^C$ being chosen such that there exists some input state with an output entropy less than $\ln(D) - \delta S^{\max}$. Next, we estimate the probability of such an output state, without any conditioning on $\mathcal{E}^C$. By comparing these estimates, we are able to bound the probability of $\mathcal{E}^C$ having an input state which gives an output entropy less than $\ln(D) - \delta S^{\max}$.

Lemma 5. *If the unitary matrices U_i are chosen at random from the Haar measure, and the l_i are chosen randomly as described above, then the probability that $H^{min}(\mathcal{E}^C)$ is less than $\ln(D) - \delta S^{\max}$ is less than one for sufficiently large N, for appropriate choice of c_1 and p_1. The N required depends on D.*

Proof. Let P_{bad} denote the probability that $H^{min}(\mathcal{E}^C) < \ln(D) - \delta S^{\max}$. Then, with probability at least $P_{bad} - Q$, for random U_i and l_i, the channel $\mathcal{E}^C$ has $P_{\mathcal{E}^C} \geq 1/2$ and has $H^{min}(\mathcal{E}^C) < \ln(D) - \delta S^{\max}$.

Let $|\psi^0\rangle$ be a state which minimizes the output entropy of channel $\mathcal{E}^C$. By lemma 4, for such a channel, for a random state $|\chi\rangle$, the density matrix $\mathcal{E}^C(|\chi\rangle\langle\chi|)$ has eigenvalues $q_1, ..., q_D$ which obey

$$|q_i - (yp_i + (1-y)(1/D))| \leq \mathrm{poly}(D)\mathcal{O}(\sqrt{\ln(N)/N}) \tag{50}$$

for $y \geq 1/2$ with probability at least

$$\exp(-\mathcal{O}(N))(1/2 - 1/\mathrm{poly}(D)). \tag{51}$$

Therefore, for a random choice of U_i, l_i, χ, the state $\mathcal{E}^C(|\chi\rangle\langle\chi|)$ has eigenvalues q_i which obey Eq. (50) with probability at least

$$(P_{bad} - Q)\exp(-\mathcal{O}(N))(1/2 - 1/\mathrm{poly}(D)). \tag{52}$$

However, by Eq. (29), the probability of having such eigenvalues q_i is bounded by the maximum of the probability density $\tilde{P}(q_1, ..., q_D)$ over q_i which obey Eq. (50). Given the assumptions that $-\sum_i p_i \ln(p_i) \leq \ln(D) - \delta S^{\max}$, $y \geq 1/2$, and the constraint that $\sum_i p_i = 1$, the quantity $\tilde{P}(q_1, ..., q_D) \leq \mathcal{O}(N)^{\mathcal{O}(D^2)} \exp[-c_2(N-D)]$, where c_2 can be made arbitrarily large by choosing c_1 large (the proof of this statement is given in the next paragraph). We pick c_{MM} so that $Q < 1$ and then if $P_{bad} = 1$, we can pick c_1 and p_1 such that for sufficiently large N this quantity $\tilde{P}(q_1, ..., q_D)$ is less than that in (52), giving a contradiction. Comparing to the discussion below Eq. (40), we see that we need $c_2 > \ln(2)$ to get this contradiction. Therefore, $P_{bad} < 1$. In fact, since Q can be made arbitrarily close to zero, P_{bad} can be made arbitrarily close to zero for sufficiently large D, N.

Finally, we briefly show how c_2 can be made arbitrarily large by choosing c_1 sufficiently large. The natural way to do this is by treating this problem as a constrained maximization problem: maximize the probability $\tilde{P}(q_1, ..., q_D)$ subject to a contraint on the entropy of the p_i. This maximization can be done with Lagrange multipliers, and the final result is obtained after a direct, but slightly lengthy, calculation. We now show a slightly different way to obtain the same result. First, we claim that we can find constants x, y with $0 < x < 1 < y$, such that the probability that an

eigenvalue q_i falls outside the interval $(x/D, y/D)$ is bounded by $\mathcal{O}(N)^{\mathcal{O}(D^2)} \exp[-c_2(N-D)]$ for any desired c_2. To show this claim, we use the fact that this probability is bounded by $\mathcal{O}(N)^{\mathcal{O}(D^2)} \max(F(x/D), F(y/D))$. The function $F(x/D) = \exp[-(N-D)(x-1) + (N-D)\ln(x)] = \exp[(N-D)(-x+1+\ln(x))]$. We choose x sufficiently small that $-x+1+\ln(x) \leq c_2$ and similarly we choose y sufficiently large that $-y+1+\ln(y) \leq c_2$, and then we have bounded the probability of any eigenvalue q_i lying outside this interval. Thus, we can assume that the eigenvalues lie inside this interval.

Next, for any set of eigenvalues $\{q_i\}$ which all lie in this interval, we have

$$\prod_i F(q_i) \leq \exp[-(N-D)D^2 \sum_i (q_i - 1/D)^2/2y^2]. \tag{53}$$

Comparing Eq. (53) to Eq. (33), we have worsened by a constant in the exponent ($1/2y^2$ instead of $1/2$), but the inequality is now valid for all q_i in the interval $(x/D, y/D)$, not just as a Taylor expansion. We now also give a bound on the entropy. For any set of eigenvalues of the density matrix p_i, we have

$$\begin{aligned} S(\{p_i\}) &= -\sum_i p_i \ln(p_i) \\ &\equiv \ln(D) - \delta S \\ &\geq \ln(D) - D\sum_i (p_i - 1/D)^2. \end{aligned} \tag{54}$$

To derive Eq. (54), note that $\sum_i -p_i \ln(p_i) = \ln(D) + \sum_i [(1/D)\ln(1/D) - p_i \ln(p_i) + (p_i - 1/D)(\ln(1/D)+1)]$, because $\sum_i (p_i - 1/D) = 0$. Then, $\delta S = -\sum_i [(1/D)\ln(1/D) - p_i \ln(p_i) + (p_i - 1/D)(\ln(1/D)+1)]$. For $p_i = 1/D$, $-[(1/D)\ln(1/D) - p_i \ln(p_i) + (p_i - 1/D)(\ln(1/D)+1)] = 0$, while for $p_i = 0$, it is equal to $1/D$. The function $D(p_i - 1/D)^2$ is a quadratic function chosen to fit these two points (0 at $p_i = 1/D$ and $1/D$ at 0), and both $D(p_i - 1/D)^2$ and $(1/D)\ln(1/D) - p_i \ln(p_i) + (p_i - 1/D)(\ln(1/D)+1)$ have vanishing derivative at $p_i = 1/D$; it was to make the derivative vanish that we subtracted off that linear term. By checking the sign of the third derivative of $-p_i \ln(p_i)$ one may verify the inequality (54).

Comparing Eq. (54) to Eq. (34), we have lost the factor of $1/2$ in (54), but the result is now an inequality valid for all p_i, not just a Taylor expansion. Comparing Eq. (53) and Eq. (54), and using Eq. (50), we find

$$\prod_i F(q_i) \leq \exp\Big\{-\frac{(N-D)D[\delta S - \mathrm{poly}(D)\mathcal{O}(\sqrt{\ln(N)/N})]}{8y^2}\Big\}, \tag{55}$$

and so we can make c_2 arbitrarily large by choosing sufficiently large c_1.

□

This completes the proof of the theorem.

[1] Shannon, C. E., A Mathematical Theory of Communication, Bell Syst. Tech. Jour. **27**, 379-423 (1948).
[2] A. S. Holevo, Bounds for the Quantity of Information Transmitted by a Quantum Communication Channel, Probl. Info. Transm. (USS), **9**, 177-183 (1973).
[3] C. King and M. B. Ruskai, Minimal Entropy of States Emerging from Noisy Quantum Channels, IEEE Trans. Inf. Thy. **47**, 192-209 (2001).
[4] P. W. Shor, Equivalence of Additivity Question in Quantum Information Theory, Comm. Math. Phys. **246**, 453-472 (2004).
[5] G. Smith and J. Yard, Quantum Communication with Zero-Capacity Channels, Science **321**, 1812-1815 (2008).
[6] Bennett, C. H., DiVincenzo, D. P., Smolin, J. A., and Wootters, W. K., Mixed-state entanglement and quantum error correction, Phys. Rev. A **54**, 3824-3851 (1996).
[7] Bennatti, F. and Narnhofer, H., Additivity of the entanglement of formation, Phys. Rev. A **63**, 042306 (2001).
[8] Winter, A., The maximum output p-norm of quantum channels is not multiplicative for any $p > 2$, arXiv.org:0707.0402.
[9] Hayden, P., The maximal p-norm multiplicativity conjecture is false, arXiv.org:0707.3291.
[10] Hayden, P. and Winter, A., Counterexample to the maximal p-norm multiplicativity conjecture for all $p > 1$, arXiv.org:0807.4753.
[11] Cubitt, T., Harrow, A. W., Leung, D., Montanero, A., and Winter, A., Counterexamples to additivity of minimum output p-Renyi entropy for p close to 0, arXiv.org:0712.3628.
[12] Hastings, M. B., Random Unitaries Give Quantum Expanders, Phys. Rev. A **76**, 032315.

[13] Ben-Aroya, A. and Ta-Shma, A., Quantum Expanders and the Quantum Entropy Difference Problem, arXiv:quant-ph/0702129.
[14] Hastings, M. B. Entropy and Entanglement in Quantum Ground States, Phys. Rev. B **76**, 035114 (2007).
[15] King, C., Matsumoto, K., Nathanson, M., and Ruskai, M. B., Properties of Conjugate Channels with Applications to Additivity and Multiplicativity, arXiv:quant-ph/0509126, Markov Proc. Relat. Fields. **13**, 391-423 (2007).
[16] Bhatia, R., *Matrix Analysis*, Springer 1997.
[17] Page, D. N., Average Entropy of a Subsystem, Phys. Rev. Lett. **71**, 1291-1294 (1993).
[18] Lloyd, S. and Pagels, H., Complexity as Thermodynamics Depth, Ann. Phys. **188**, 186-213 (1988).
[19] Fannes, M., A Continuity Property of the Entropy Density for Spin Lattice Systems, Commun. Math. Phys. **31**, 291-294 (1973).