A Counter-example to Additivity:
Using Entanglement to Boost Communication Capacity

M. B. Hastings
T-13, Complex Systems Group
Los Alamos National Laboratory

Thanks: J. Yard, P. Hayden, A. Harrow

- Classical Information Theory Background
- Communicating over a Quantum Channel
- (Non)-Additivity

Communicating over a noisy channel:

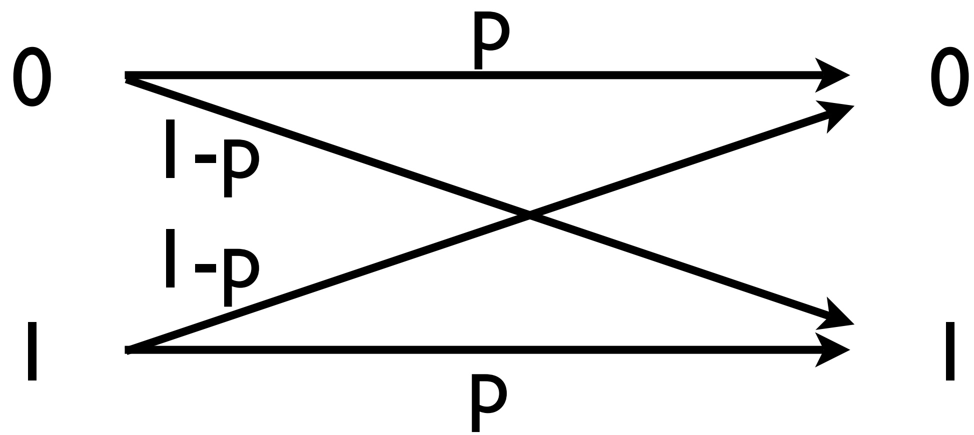
Alice ⟶ Bob

input X   output Y

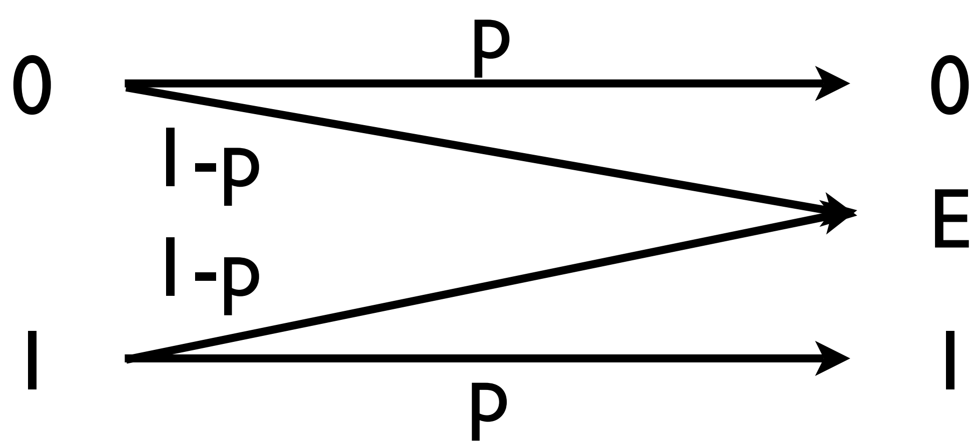
Channel defined by allowed inputs and outputs and by probability P(Y|X)

Communicating over a noisy channel (examples):

Binary symmetric channel:



Binary erasure channel:



Meaning of entropy (Shannon's noisy channel coding theorem):

Error correction:

k bits: 10....1 → [Encode] → N>k bits: 11.....1 → [Channel] → N bits (with errors): 01.....1

01.....1 → [Decode] → k bits: 10....1

A simple code (repetition code):

1 → 11.....1 → [Channel] → Pick majority

For p>1/2, error probability exponentially small in N, but we encode at rate k/N=1/N, so rate->0

Communicating over a noisy channel: the capacity

Choose inputs with probability $p(X)$

Output: $p(Y) = \sum_X P(Y|X)p(X)$

conditional information

The capacity: $\max_{p(X)} S(B) - S(B|A)$

$$S(B) = -\sum_Y p(Y)\log_2(p(Y))$$

$$S(B|A) = -\sum_X p(X)\Big[\sum_Y P(Y|X)\log_2(P(Y|X))\Big]$$

How much noise is output, minus how much noise is due to the channel, equals the information transmitted.

Meaning of entropy (Shannon's noisy channel coding theorem):

Shannon '48

We can encode k bits into N bits, such that the error probability goes to zero as N goes to infinity, with k/N asymptotically approaching C, the capacity of the channel.

This gives a meaning to the capacity of the channel.

Amazing things about channel coding:

- C is not zero!
- C is actually quite large. C for the erasure channel is equal to p.
- We can calculate C.
- We do the calculation by a **single-letter** formula, despite using correlations to correct errors.

Additivity of capacities.

- Classical Information Theory Background
- Communicating over a Quantum Channel
- (Non)-Additivity

Communicating over a noisy quantum channel:

Channel is a linear map on density matrices.

Alice $\longrightarrow$ Bob

input ρ output $\mathcal{E}(\rho)$

$$\mathcal{E}(\rho) = \sum_{s=1}^{D} A(s)\rho A^{\dagger}(s) \qquad \sum_{s=1}^{D} A^{\dagger}(s)A(s) = I$$

Communicating over a noisy quantum channel:

Quantum entropy: $H(\rho) = -\mathrm{tr}\Big(\rho \log_2(\rho)\Big)$

Signal words: input state ρ_i with probability p_i

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

recall S(B)

recall S(B|A)

Holevo capacity for sending classical information over a quantum channel:

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

"I wish that physicists would ... give us a general expression for the capacity of a channel with quantum effects taken into account rather than a number of special cases."

-J. R. Pierce, 1973, in a retrospective on Shannon's paper.

Communicating over a noisy quantum channel:

Why the Holevo capacity is hard to evaluate: should we entangle?

$$\chi_{\max}(\mathcal{E}^{\otimes n}) \stackrel{?}{=} n\chi_{\max}(\mathcal{E})$$

Additive: $\chi_{\max}(\mathcal{E}_1 \otimes \mathcal{E}_2) = \chi_{\max}(\mathcal{E}_1) + \chi_{\max}(\mathcal{E}_2)$

Non-Additive: $\chi_{\max}(\mathcal{E}_1 \otimes \mathcal{E}_2) > \chi_{\max}(\mathcal{E}_1) + \chi_{\max}(\mathcal{E}_2)$

The additivity conjecture: the first case is true for all quantum channels.

Equivalence of additivity conjectures (Shor, 2004):

- Additivity of Holevo capacity
- Additivity of minimum output entropy
- Additivity of entanglement of formation
- Strong super-additivity of entanglement of formation

Why Additivity Is Important:

- We can boost capacity using entangled inputs.
- If additivity fails, then we physicists have not answered Pierce's question. It is not practical to compute the capacity maximizing over arbitrary entangled inputs.
- Additivity in the classical case gives meaning to the capacity of a channel.

- Classical Information Theory Background
- Communicating over a Quantum Channel
- (Non)-Additivity

The minimum output entropy conjecture:

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

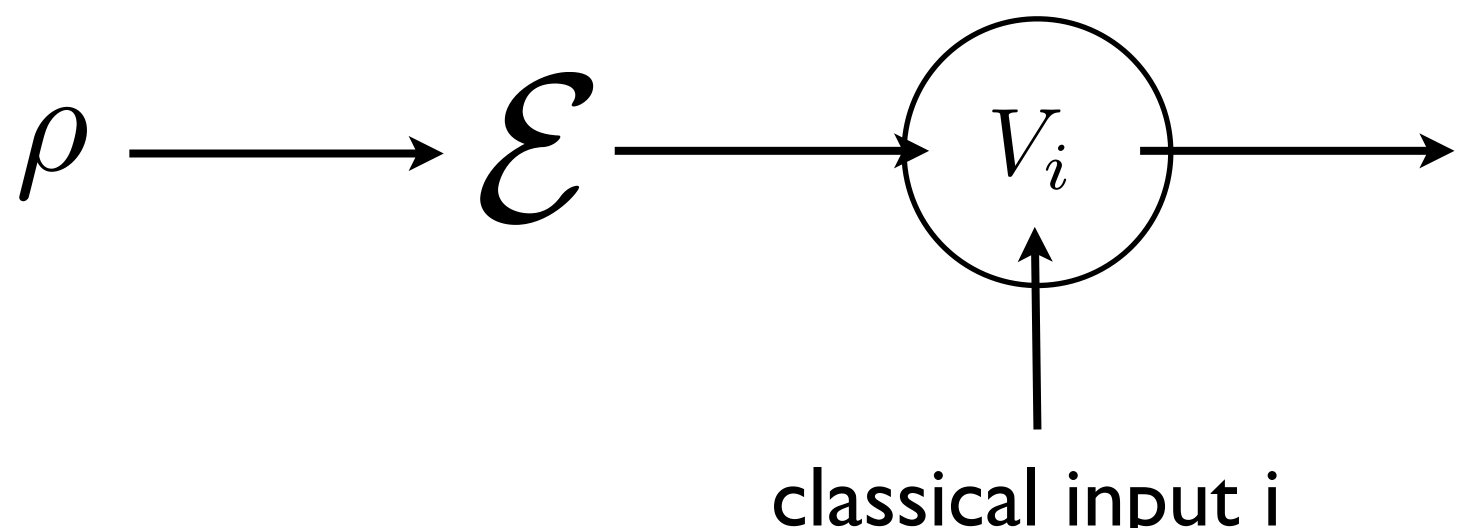
$$H^{\min}(\mathcal{E}_1 \otimes \mathcal{E}_2) \stackrel{?}{=} H^{\min}(\mathcal{E}_1) + H^{\min}(\mathcal{E}_2)$$

Relation to additivity of Holevo capacity: by reducing the output entropy for a given input state, we can communicate more effectively over the channel.

Violation of minimum output entropy conjecture implies violation of additivity of Holevo capacity

(for a different, but related channel, Shor 2004)

$$\mathcal{E}'(\rho, i) = V_i^\dagger \mathcal{E}(\rho) V_i$$



$$\chi_{\max}(\mathcal{E}') = \log(N) - H^{\min}(\mathcal{E})$$

A Counterexample to Additivity:

MBH '08
(see also counterexamples to p-norm multiplicativity by Winter and Hayden)

Two random channels, related by complex conjugation:

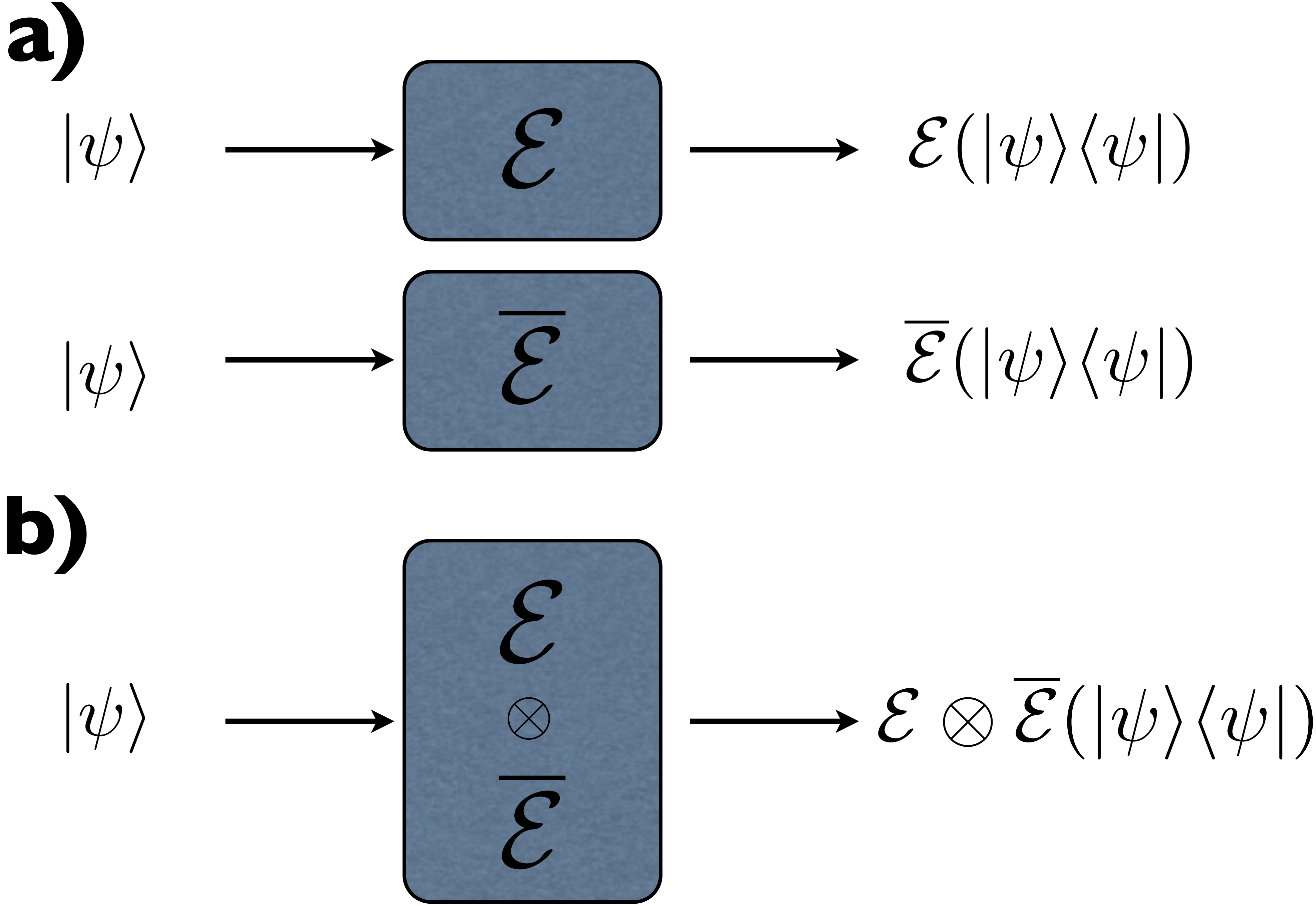
$$\mathcal{E}(\rho) = \sum_{i=1}^{D} p_i U_i^\dagger \rho U_i$$

$$\overline{\mathcal{E}}(\rho) = \sum_{i=1}^{D} p_i \overline{U}_i^\dagger \rho \overline{U}_i$$

U_i are randomly chosen unitaries.

$$1 << D << N$$

This channel models interaction with random environment.



a)
$|\psi\rangle$
$\mathcal{E}$
$\mathcal{E}(|\psi\rangle\langle\psi|)$
$|\psi\rangle$
$\overline{\mathcal{E}}$
$\overline{\mathcal{E}}(|\psi\rangle\langle\psi|)$
b)
$|\psi\rangle$
$\mathcal{E}$
$\otimes$
$\overline{\mathcal{E}}$
$\mathcal{E}\otimes\overline{\mathcal{E}}(|\psi\rangle\langle\psi|)$

Why additivity fails:

1) Upper bound $H^{\min}(\mathcal{E} \otimes \overline{\mathcal{E}})$

$$H^{\min}(\mathcal{E} \otimes \overline{\mathcal{E}}) \leq 2\log_2(D) - \log_2(D)/D$$

Proof based on an explicit low entropy input state for the combined channel:

$$\psi_{ME} = \frac{1}{\sqrt{N}} \sum_{\alpha=1}^{N} |\alpha\rangle \otimes |\alpha\rangle$$

---

2) Lower bound $H^{\min}(\mathcal{E})$

For most such channels,

$$H^{\min}(\mathcal{E}) \geq \log_2(D) - \text{const.}/D - O(\sqrt{\ln(N)/N})$$

(proof based on randomness)

Low entropy input state for the combined channel:

$$\psi_{ME} = \frac{1}{\sqrt{N}} \sum_{\alpha=1}^{N} |\alpha\rangle \otimes |\alpha\rangle$$

Note that: $U_i^\dagger \otimes \overline{U}_i^\dagger |\psi_{ME}\rangle = |\psi_{ME}\rangle$

So, of the D^2 possible outputs, D of them are the same, when we choose the same unitary for both channels.

Consider simplest case (other cases have lower entropy):

$$p_i = 1/D$$

$\langle\psi_{ME}|U_j \otimes \bar{U}_k|U_l^\dagger \otimes \bar{U}_m^\dagger|\psi_{ME}\rangle = 0$ unless j=k, l=m OR j=l, k=m

Then, output eigenvalues are: 1/D and 1/D^2 (with multiplicity D^2-D)

$$\begin{aligned} H(\mathcal{E} \otimes \overline{\mathcal{E}}(\phi)) &\leq \frac{1}{D}\log_2(D) + (1 - 1/D)\log_2(D^2) \\ &= 2\log_2(D) - \log_2(D)/D \end{aligned}$$

Intuition for the $\mathrm{const.}/D$

For all such channels, $H^{\min}(\mathcal{E}) \leq \log_2(D) - 2/D$

Proof:

Assume, without loss of generality, $p_1 \geq p_2 \geq p_3 \geq \ldots$

Pick Ψ to be an eigenvector of $U_1 U_2^\dagger$

Then, $U_2^\dagger|\Psi\rangle = zU_1^\dagger|\Psi\rangle$ for some phase z

So, at most D-1 different outcomes: the first two unitaries cannot be distinguished!

$$
\begin{aligned}
H(\mathcal{E}(|\psi\rangle\langle\psi|)) &\leq \frac{2}{D}\log_2(D/2) + (1-2/D)\log_2(D) \\
&= \log_2(D) - \frac{2}{D}\log_2(2)
\end{aligned}
$$

Intuition for the $\text{const.}/D$

Suppose we can find a simultaneous eigenvector Ψ of: $U_1^\dagger, U_2^\dagger, U_3^\dagger, \ldots$

$$U_i^\dagger|\Psi\rangle = z_i|\Psi\rangle$$

Then, $H^{\min}(\mathcal{E}) = 0$

Suppose we can find a simultaneous eigenvector Ψ of: $U_1U_2^\dagger, U_3U_4^\dagger, U_5U_6^\dagger, \ldots$

Then, $H^{\min}(\mathcal{E}) \leq \log_2(D/2) = \log_2(D) - 1$

So, we choose random unitaries to avoid such simultaneous eigenvectors.

Outline of proof

- Choice of P_i
- Statistical properties of output density matrix for random input and random channel. Usually eigenvalues are close to 1/D (state is usually close to maximally mixed)
- We need to show that no state has low entropy. Epsilon-nets: multiply the number of different input states, by probability that a given one is low entropy. Fails! Too many possible inputs.
- Less pessimistic approach works...see below.

Choice of P_i

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

Length of a random vector chosen from Gaussian distribution in N complex dimensions

$$L = \sqrt{\sum_i l_i^2} \qquad P_i = l_i^2 / L^2$$

This is done so that, for a random input state, the output density matrix has the same statistics as the reduced density matrix of a random bipartite state with system dimension N, environment dimension D

Conjugate channel:

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

This outputs the other half of the bipartite state (the environment instead of the system). All non-zero eigenvalues the same.

Output probability density:

(using known results on bipartite states by Page, Lloyd, Pagels)

$$\begin{aligned} & P(p_1, ..., p_D) \\ \leq \quad & \mathcal{O}(N)^{\mathcal{O}(D^2)} \prod_{i=1}^{D} \Big(D^{N-D} p_i^{N-D} \exp[-(N-D)D(p_i - 1/D)] \Big) \end{aligned}$$

Relatively unimportant normalization factor

Sharply peaked near 1/D

Low output entropy is unlikely:

Taylor series (just to get oriented, not used in rigourous proof):

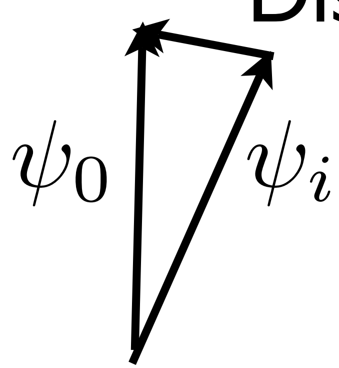
$$P(p_1, ..., p_D) \approx \mathcal{O}(N)^{\mathcal{O}(D^2)} \exp[-(N-D)D^2 \sum_i \delta p_i^2/2 + ...].$$

$$\begin{aligned} S = -\sum_i p_i \ln(p_i) &\approx \ln(D) - D\sum_i \delta p_i^2/2 + ... \\ &= \ln(D) - \delta S \end{aligned}$$

$$P \approx \mathcal{O}(N)^{\mathcal{O}(D^2)} \exp[-(N-D)D\delta S]$$

Epsilon-net estimates

Create a "net" of states ψ_i separated by a small distance d, so that the minimum output entropy state ψ_0 will be close to a state in the net.



Distance between states is at most d.
Difference is some state ϕ

By Fannes inequality,

$$\delta S_i \geq \delta S^0 - d^2 \ln(D/d^2)$$

We need $d \sim 1/\sqrt{D}$ to get good bounds on δS^0

Probability that, for a given state in the net,

$\delta S^i = \ln(D)/2D$ is bounded by roughly

$$\exp[-ND\delta S] = \exp[-ND\ln(D)/2D] = D^{-N/2}$$

Number of points in net is roughly: d^{-2N}

So, we can take $d \sim D^{-1/4}$

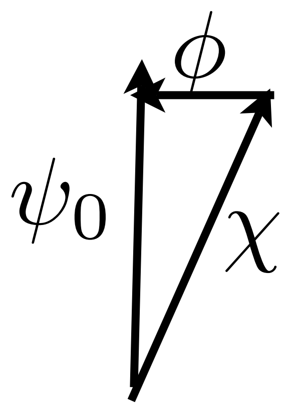
and with high probability no state in the net has the given $\delta S^i = \ln(D)/2D$

But, we need $d \sim 1/\sqrt{D}$ to get good bounds on δS^0

DOES NOT WORK!

Fannes inequality was too pessimistic here!

Pick random state χ



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

Typically, $\mathcal{E}^C(\phi)$ is close to the maximally mixed state.

So, if $\mathcal{E}^C(\psi)$ has eigenvalues p_i then

$\mathcal{E}^C(\chi)$ has eigenvalues q_i with

$$q_i \approx (1-x^2)p_i + x^2/D$$

With probability $\exp(-\mathcal{O}(N)), x^2 \leq 1/2$

So, conditioned on there being a state ψ_0 with low output entropy, then the probability that a random state χ has low entropy entropy is fairly high. But one can show it isn't.

The rest of the proof is just estimates.

Experimental relevance:

- Currently, it is too difficult to manipulate entangled states to expect any practical boost in capacity for any channel.
- However, we may be able to check that certain entangled states decohere less than unentangled states.
- Check simpler claim: that entangled state is more likely to remain unchanged after interacting with environment.
- Need to create large number of entangled pairs (N>>1), and interact in a non-linear way with environment.

The future of additivity?

- Pierce's channel capacity problem remains open.
- I conjecture additivity for channels of the form $\mathcal{E} = \mathcal{F} \otimes \overline{\mathcal{F}}$, giving a **two-letter** formula to solve the capacity problem. (This is actually 4 different conjectures, are they all equivalent again?)
- Can we use these ideas to protect states from decoherence?

Conjectured two-letter formula

A consequence if this conjecture is true

(consequence due to P. Hayden)

Note that: $\chi_{\max}(\mathcal{E}^{\otimes n}) \leq \frac{1}{2}\chi_{\max}(\mathcal{E}^{\otimes n} \otimes \overline{\mathcal{E}}^{\otimes n})$

Either: for all $\mathcal{E}$,
$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\chi_{\max}(\mathcal{E}^{\otimes n}) &= \frac{1}{2}\lim_{n\to\infty}\frac{1}{n}\chi_{\max}(\mathcal{E}^{\otimes n} \otimes \overline{\mathcal{E}}^{\otimes n}) \\ &= \frac{1}{2}\chi_{\max}(\mathcal{E} \otimes \overline{\mathcal{E}})\end{aligned}$$

Solution of classical capacity problem for all channels!

Or: for some $\mathcal{E}$, $\lim_{n\to\infty}\frac{1}{n}\chi_{\max}(\mathcal{E}^{\otimes n}) < \frac{1}{2}\lim_{n\to\infty}\frac{1}{n}\chi_{\max}(\mathcal{E}^{\otimes n} \otimes \overline{\mathcal{E}}^{\otimes n})$

Operational non-additivity!

I conjecture this case, for the same random channels as before

Conclusion

Entangled states improve communication capacity

Is there a two-letter formula?

Which channels are additive?

Can we use these ideas to protect against decoherence in other settings?