Quantum Channels

Peter Shor
MIT
Cambridge, MA

Claude Shannon, 1948

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.

John Pierce, 1973

I think that I have never met a physicist who understood information theory. I wish that physicists would stop talking about reformulating information theory and 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.

Shannon

Shannon's 1948 paper "A mathematical theory of communication" founded the field of information theory. It contained two theorems that we will discuss the quantum analogs of today: Source Coding and Channel Coding.

Shannon's Source Coding Theorem

Asymptotically, n symbols from a source X can be compressed to length $nH(X) + O(\sqrt{n})$.

For a memoryless source, which emits i.i.d. signals where signal x_i has probability p_i, the entropy H is:

$$H(X) = \sum_i -p_i \log p_i$$

This lecture will deal only with memoryless sources and channels.

Shannon's Channel Coding Theorem

A noisy channel N has capacity

$$\max_{p(X)} \quad I(X; N(X)),$$

where $p(X)$ is maximized over all probability distributions on the channel input and $I(X; N(X))$ is the mutual information between the input and the output.

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X,Y). \end{aligned}$$

$$H(X) = \sum_i -p_i \log p_i$$

What is a quantum channel?

It is a map from density matrices to density matrices allowed by quantum mechanics.

There are three equivalent characterizations.

- A completely positive trace preserving linear map.
- The partial trace of a unitary transformation on a larger Hilbert space.
- A map $\rho \to \sum_k A_k \rho A_k^\dagger$, with the A_k being matrices with $\sum_k A_k^\dagger A_k = I$

Equivalence of Definitions of Quantum Channel

Suppose we have a channel defined by the partial trace of a unitary map.

This channel takes

$$\rho \to \rho \otimes \left| 0^k \right\rangle \left\langle 0^k \right| \to U \left(\rho \otimes \left| 0^k \right\rangle \left\langle 0^k \right| \right) U^\dagger \to \mathrm{Tr}_{\mathcal{S}} U \left(\rho \otimes \left| 0^k \right\rangle \left\langle 0^k \right| \right) U^\dagger$$

But $\mathrm{Tr}_{\mathcal{S}}\, \sigma = \sum_i \langle e_i | \sigma | e_i \rangle$ where $| e_i \rangle$ is a basis for $\mathcal{S}$, so

$$\rho \to \sum_i \langle e_i | U \left(\rho \otimes \left| 0^k \right\rangle \left\langle 0^k \right| \right) U^\dagger | e_i \rangle = \sum_i A_i \rho A_i^\dagger$$

where

$$A_i = \langle e_i | U \left| 0^k \right\rangle$$

Equivalence of Definitions of Quantum Channel II

Assume a trace preserving completely positive linear map $\mathcal{N}$.

Positive means that $\mathcal{N}$ takes positive matrices to positive matrices.

Completely positive means that $\mathcal{N} \otimes I$ takes positive matrices to positive matrices.

The map $\rho \to \rho^T$ is a positive map that is not completely positive.

We will show that $\mathcal{N}$ is completely defined by the action of $\mathcal{N} \otimes I$ on $|\psi\rangle = \frac{1}{\sqrt{d}} \sum_i |i\rangle |i\rangle$.

Recall

$$|\,\psi\rangle\langle\psi\,| = \frac{1}{d}\sum_{i=1}^{d} |\,i\rangle_A\,|\,i\rangle_B\,{}_A\langle i\,|\,{}_B\langle i\,|$$

.

$${}_B\langle v\,|\,\psi\rangle\,\langle\psi\,|\,v\rangle_B = |\,v^*\rangle_A\,{}_A\langle v^*\,|$$

By complete positivity, $\mathcal{N}\otimes I\,|\,\psi\rangle\langle\psi\,| = \sigma$ for some positive matrix σ. We can diagonalize $\sigma = \sum_j \lambda_j\,|\,f_j\rangle\langle f_j\,|$. We thus have

$$\begin{aligned}
\mathcal{N}(|\,v\rangle\langle v\,|) &= {}_B\langle v^*\,|\,(\mathcal{N}\otimes I)(\rho)\,|\,v^*\rangle_B \\
&= {}_B\langle v^*\,|\,\sigma\,|\,v^*\rangle_B \\
&= \sum_j \lambda_j\,{}_B\langle v^*\,|\,f_j\rangle\,\langle f_j\,|\,v^*\rangle_B
\end{aligned}$$

Each of these terms $\sqrt{\lambda_i}\,\langle v^*\,|\,f_j\rangle$ will give some linear map $A_j\,|\,v\rangle$, giving the Krauss decomposition.

The map

$$|\,v\rangle \to \sqrt{\lambda_j}\,\langle v^* \,|\, f_j\rangle$$

is linear. This is because $|\,v\rangle \to |\,v^*\rangle$ is antilinear, and $|\,w\rangle \to \langle w\,|\,f_j\rangle$ is antilinear. Any linear transformation can be represented as a matrix

$$|\,v\rangle \to A_j\,|\,v\rangle$$

If the input and output dimensions of $\mathcal{N}$ are d, then σ is a $d^2 \times d^2$ matrix, so its eigenvector decomposition $|\,f_i\rangle$ contains at most d^2 vectors, and we get at most d^2 terms in the Krauss representation

$$\mathcal{N}(\rho) = \sum_i A_i \rho A_i^\dagger.$$

With a little more work, we can show that $\langle f_j \,|\, f_i\rangle = 0$ for $i \neq j$ yeilds $\mathrm{Tr} A_i^\dagger A_j = 0$.

Entropy of a quantum state

Classical Case

Given n photons, each in state $|\updownarrow\rangle$ or $|\leftrightarrow\rangle$, with probability $\frac{1}{2}$. Any two of these states are completely distinguishable. The entropy is n bits.

Quantum Case

Given n photons, each in state $|\updownarrow\rangle$ or $|\swarrow\!\!\nearrow\rangle$, with probability $\frac{1}{2}$. If the angle between the polarizations is small, any two of these states are barely distinguishable. Intuitively, the entropy should be much less than n bits.

By thermodynamic arguments, von Neumann deduced the entropy of a quantum system with density matrix ρ is

$$S(\rho) = -\mathrm{Tr}(\rho \log \rho)$$

Recall ρ was positive semideifinite, so $\rho \log \rho$ is defined.

If ρ is diagonal with eigenvalues λ_i, then $\rho \log \rho$ is diagonal with eigenvalues $\lambda_i \log \lambda_i$.
Thus, $S(\rho) = H_{\mathrm{Shan}}(\lambda_i)$ so the von Neumann entropy is the Shannon entropy of the eigenvalues.
(Recall $\mathrm{Tr}\rho = 1 = \sum_i \lambda_i$.)

You can ask: is this the right definition for information theory?

Schumacher Compression
(Quantum source coding theorem)

Given a memoryless source producing pure states v_1, v_2, v_3, ...
with probabilities p_1, p_2, p_3, ...

We want to send them to a receiver using as few qubits as possible.

Theorem (Schumacher, 1994):
You can send n symbols using

$$nS(\rho) + o(n)$$

qubits, with fidelity approaching 1 as $n \to \infty$, where $\rho = \sum_i p_i v_i v_i^\dagger$ is the density matrix of the source.

Fidelity

Classical source coding works with high probability: the probability that the received sequeuce is exactly the signal goes to 1 as the block length n goes to ∞.

This is too strong a criterion for quantum source coding. We ask that the *fidelity* between the signal sent and the received state ρ goes to 1 as the block length n goes to ∞.

The fidelity between a pure state sent $|\, v\rangle$ and a received density matrix ρ is $\langle v \,|\, \rho \,|\, v\rangle = v^\dagger \rho v$.

If the fidelity goes to 1, any measurement on the received signal ρ will have almost the same probability distribution of outcomes as the same measurement on $vv^\dagger$, the state sent.

Proof of Classical Source Coding Theorem

Assume we have a source X emitting symbols s_1, s_2, ... with probabilities p_1, p_2, ... Consider a sequence of n symbols from this source.

Then a *typical sequence* has close to the right number (np_i) of each symbol s_i.

Theorem: Almost all the time, the source emits a typical sequence. There are $2^{nH_{\mathrm{Shan}}(X)+o(n)}$ typical sequences.

Typical Subspaces

Have states v_1, v_2, ..., v_k with probabilities p_1, p_2, ..., p_k.

Look at eigenvectors of density matrix ρ.
Assign to each of eigenvector a probability equal to the corresponding eigenvalue.
Any two eigenvectors are orthogonal.
Let the eigenvectors be $\hat{v}_1$, $\hat{v}_2$, ..., $\hat{v}_d$ with probabilities $\hat{p}_1$, $\hat{p}_2$, ..., $\hat{p}_d$.

Suppose we have n of these states.
The *typical subspace* $\mathcal{S}$ is the subspace generated by typical sequences of eigenvectors.

$\mathcal{S}$ has dimension $2^{S(\rho)n+o(n)}$.

How to do Schumacher compression.

Have states v_1, v_2, ..., v_k with probabilities p_1, p_2, ..., p_k. These give density matrix ρ. Let $\mathcal{S}$ be the typical subspace of $\rho^{\otimes n}$.

To compress:
Measure whether output of source lies in $\mathcal{S}$.
If *yes*, get the state projected onto $\mathcal{S}$. Can send using
$\log \dim \mathcal{S} \approx nS(\rho)$ qubits.
If *no*, this is a low probability event; send anything.

Why does Schumacher Compression work?

Recall that the density matrix determines the outcomes of any experiment.

Using the eigenvectors $\hat{v}_1, \hat{v}_2, \ldots \hat{v}_d$ with probabilities $\hat{p}_1, \hat{p}_2, \ldots \hat{p}_d$ gives same probability of the outcomes as using states $v_1, v_2, \ldots v_k$ with probabilities $p_1, p_2, \ldots p_k$, since these two sources have the same density matrix.

We know from the classical theory of typical sequences that the probability of a *no* outcome is very small using $\hat{v}_i$ and $\hat{p}_i$. Thus, the probability of a *no* outcome is also very small with v_i and p_i.

This implies that the original state is almost surely very close to the typical subspace $\mathcal{S}$. Sending the state projected into $\mathcal{S}$ gives the right outcomes with high fidelity.

Accessible Information

Suppose that we have a source that outputs signal ρ_i with probability p_i. How much Shannon information can we extract about the sequence of i's?

Let X be the random variable telling which signal ρ_i was sent.

Optimize over all possible measurements M on the signals (with outcomes M_1, M_2, ...).

$$I_{\text{acc}} = \max_M I(X, M)$$

Example 1: Two states in ensemble

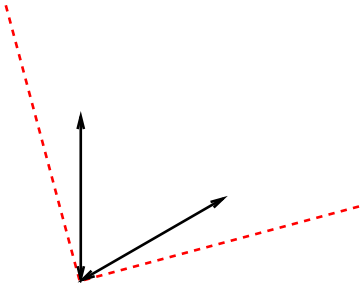
$v_1 =$ $v_2 =$ 

$$v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad v_2 = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}$$

Then

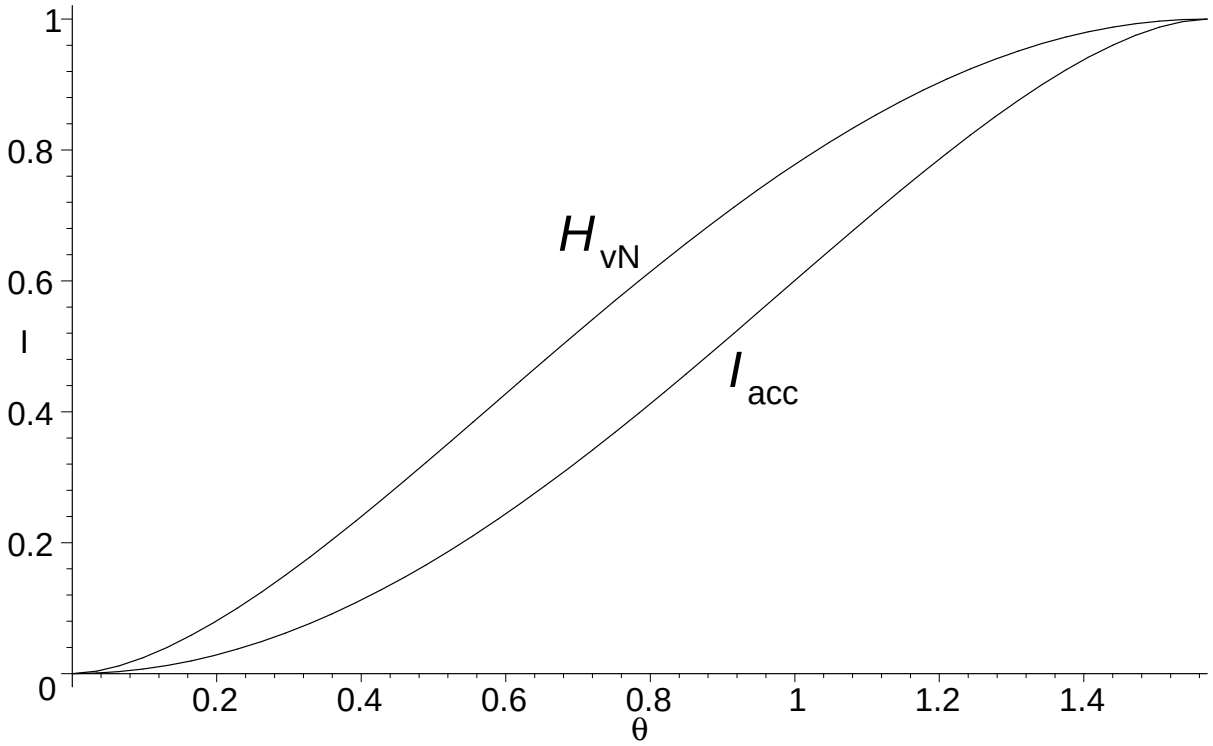
$$\rho = \frac{1}{2} \begin{pmatrix} 1 + \cos^2\theta & \sin\theta\cos\theta \\ \sin\theta\cos\theta & 1 - \cos^2\theta \end{pmatrix}$$

and $S = H(\frac{1}{2} + \frac{\cos\theta}{2})$. The optimal measurement is



and $I_{\text{acc}} = 1 - H(\frac{1}{2} + \frac{\sin\theta}{2})$.

We see that $I_{\rm acc} < S(\rho)$.



A plot of S and $I_{\rm acc}$ for the ensemble of two pure quantum states with equal probabilities that differ by an angle of θ, $0 \leq \theta \leq \pi/2$.

The top curve is the von Neumann entropy $S = H(\frac{1}{2} + \frac{\cos\theta}{2})$ and the bottom the accessible information $I_{\rm acc} = 1 - H(\frac{1}{2} + \frac{\sin\theta}{2})$.

POVM Measurements

(Positive Operator Valued Measurements).

We are given a set of positive semidefinite matrices E_i satisfying $\sum_i E_i = I$.

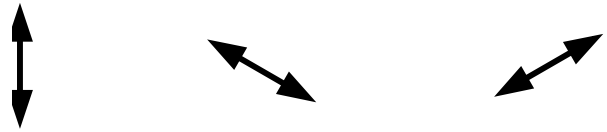
The probability of the i'th outcome is

$$p_i = \mathrm{Tr}(E_i \rho)$$

For von Neumann measurements, $E_i = \Pi_{S_i}$

To obtain the maximum information, we can assume that E_i's are pure states. Then $E_i = v_i v_i^\dagger$ for some vector v_i.

Example 2: Three signal states differing by 60°.

v_i:  (prob $\frac{1}{3}$)

$$v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad v_2 = \begin{pmatrix} -1/2 \\ \sqrt{3}/2 \end{pmatrix} \quad v_3 = \begin{pmatrix} -1/2 \\ -\sqrt{3}/2 \end{pmatrix}$$

Optimal Measurement:

POVM corresponding to vectors $w_i \perp v_i$.

$E_i = \frac{2}{3} w_i w_i^\dagger$

w_i:  (prob $\frac{1}{3}$)

Each outcome rules out one state, leaving the other two equally likely

$$I_{\text{acc}} = \log 3 - 1 = .585; \qquad S = 1$$

Again, we have $I_{\text{acc}} \leq S$.

Holevo Bound χ

Suppose we have a source emitting ρ_i with probability p_i.

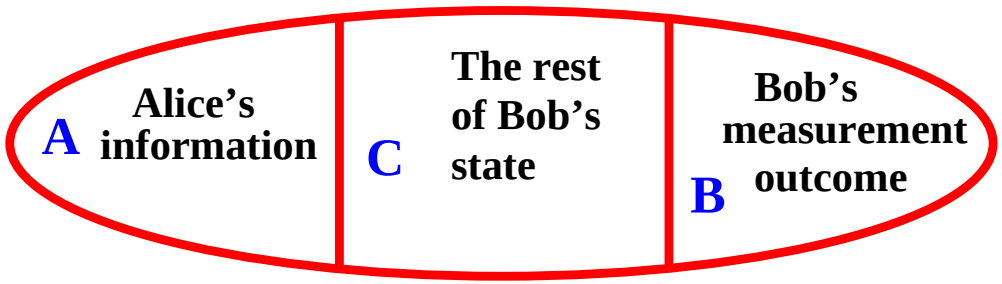
$$\chi = S(\sum_i p_i \rho_i) - \sum_i p_i S(\rho_i)$$

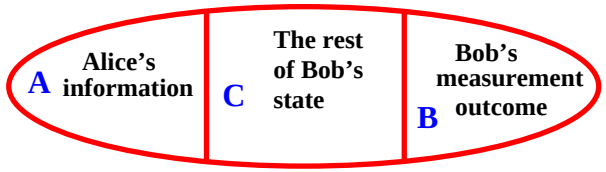
Theorem (Holevo, 1973)
$I_{\text{acc}} \leq \chi$

If all the ρ_i commute, the situation is essentially classical, and we get $I_{\text{acc}} = \chi$. Otherwise $I_{\text{acc}} < \chi$.

Proof of Holevo theorem.

Suppose Alice sends state ρ_i to Bob, and Bob makes a POVM to obtain a classical state. Recall that we can represent a POVM as a unitary transformation followed by a partial trace. If Alice keeps a record of which ρ_i she sent Bob in a quantum register, we have the following tripartite state.





The accessible information is

$$I(A:B) = S(\rho_A) + S(\rho_B) - S(\rho_{AB}).$$

The entropy of the average density matrix Bob receives is

$$S(\sum_i \rho_i) = S(\rho_{BC})$$

The average entropy of the density matrix Bob receives is

$$\sum_i S(\rho_i) = S(\rho_{ABC}) - S(\rho_A)$$

by the chain rule for entropy. We want to show

$$\begin{aligned} I(A:B) &\leq \chi \\ S(\rho_A) + S(\rho_B) - S(\rho_{AB}) &\leq S(\rho_{BC}) - [S(\rho_{ABC}) - S(\rho_A)] \end{aligned}$$

This follows from strong subadditivity.

Is this the most information we can send using the three states of example 2?

Answer: No!

Use just two of the states, each with probabilities 1/2

$$v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad v_2 = \begin{pmatrix} -1/2 \\ \sqrt{3}/2 \end{pmatrix}$$

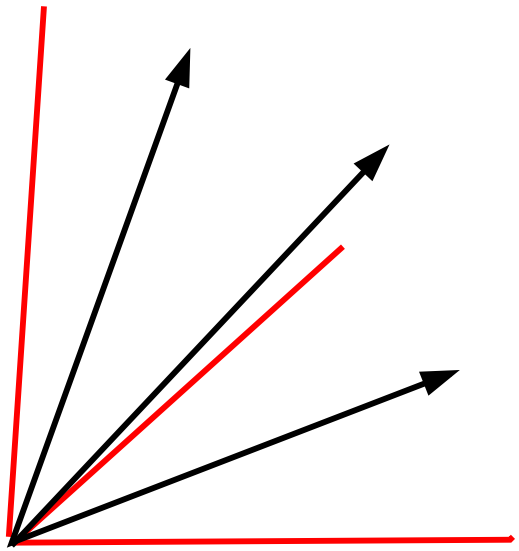
The measurement from example 1 now gives .6454 bits of information about the random variable identifying the state which was sent.

Is *this* the most information we can send using the three states of example 2?

Answer: No!

Use three codewords v_1v_1, v_2v_2, v_3v_3.

The optimal measurement for these three states gives 1.369 bits, which is larger than $2 \cdot .6454 = 1.298$.



What about still longer codewords?

Theorem (Holevo, Schumacher-Westmoreland)

The classical-information capacity obtainable using codewords composed of signal states ρ_i, where ρ_i has marginal probability p_i, is

$$\chi(\{\rho_i\};\{p_i\}) = S(\sum_i p_i \rho_i) - \sum_i p_i S(\rho_i)$$

We will give sketch of the proof of this formula in the special case of pure states ρ_i.

Does this give the capacity of a quantum channel $\mathcal{N}$?

Possible capacity formula:
Maximize $\chi(\{\mathcal{N}(\rho_i)\};\{p_i\})$ over all output states $\mathcal{N}(\rho)$ of the channel.

Theorem (pure state capacity)

We are given pure quantum states v_1, v_2, ..., v_k for use as signals. Let $\rho = \sum_i p_i v_i v_i^\dagger$. There are codes such that we send state v_i with probability p_i having asymptotic capacity $\chi = S(\rho)$

How do we prove this?

- random coding
- typical subspace
- "pretty good measurement"
  also called square root measurement

Random Coding

We choose codewords

$$u_i = v_{i_1} \otimes v_{i_2} \otimes \ldots \otimes v_{i_n}$$

where v_i is picked with probability p_i for each signal.

Then u_i will be close to the typical subspace of $\rho^{\otimes n}$.

To decode, we

- project into the typical subspace
- apply the "pretty good measurement"

Pretty good measurement

We have N vectors $\tilde{u}_i \in S$, which occur with equal probability $\frac{1}{N}$.
Given one of these, we want to distinguish between them.
Let $\phi = \sum_i \tilde{u}_i \tilde{u}_i^\dagger$

Measure using the POVM with elements

$$E_i = \phi^{-1/2} \tilde{u}_i \tilde{u}_i^\dagger \phi^{-1/2}$$

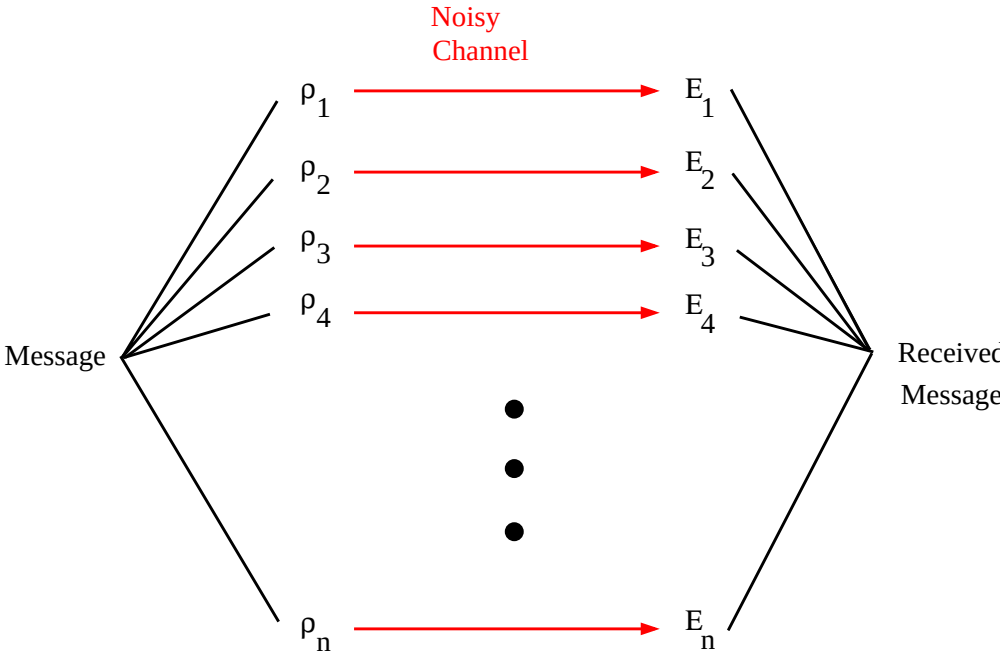
This is a POVM since

$$\sum_i E_i = \sum_i \phi^{-1/2}\, \tilde{u}_i \tilde{u}_i^\dagger\, \phi^{-1/2} = I$$

The probability of error if the state u_i is sent is $1 - (\tilde{u}_i^\dagger \phi^{-1/2} \tilde{u}_i)^2$.

This can be shown to be small for most u_i from a random code if $N < \dim S - o(\dim S)$.

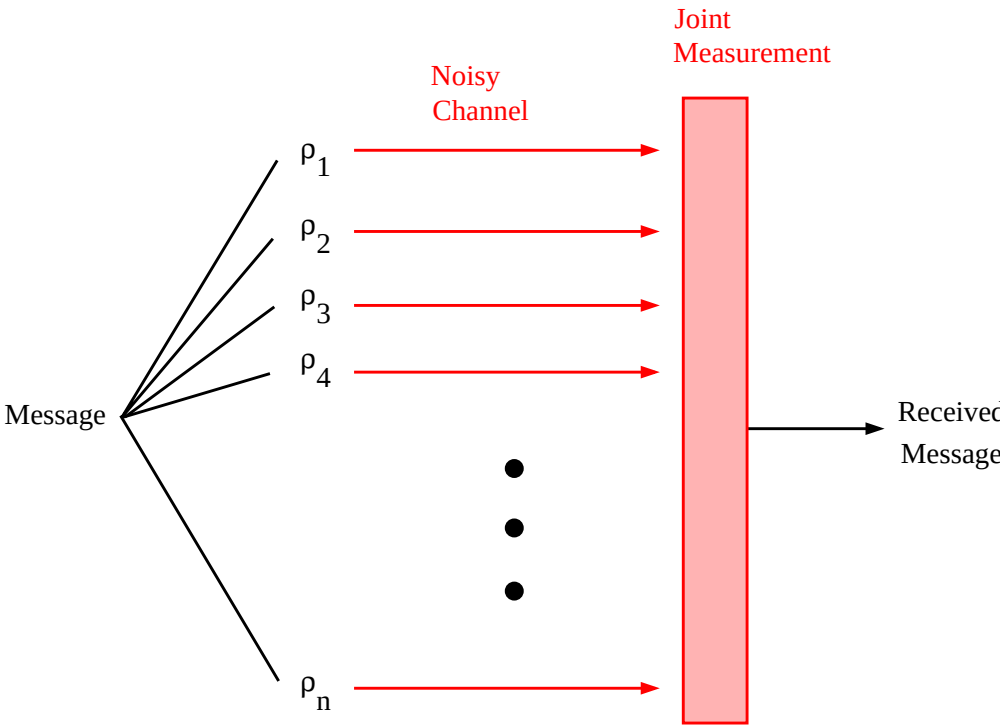
Unentangled Inputs, Separate Measurements



Maximize over probability distributions on inputs to the channel ρ_i, p_i:

$$I_{\mathrm{acc}}(\{\mathcal{N}(\rho_i)\};\{p_i\})$$

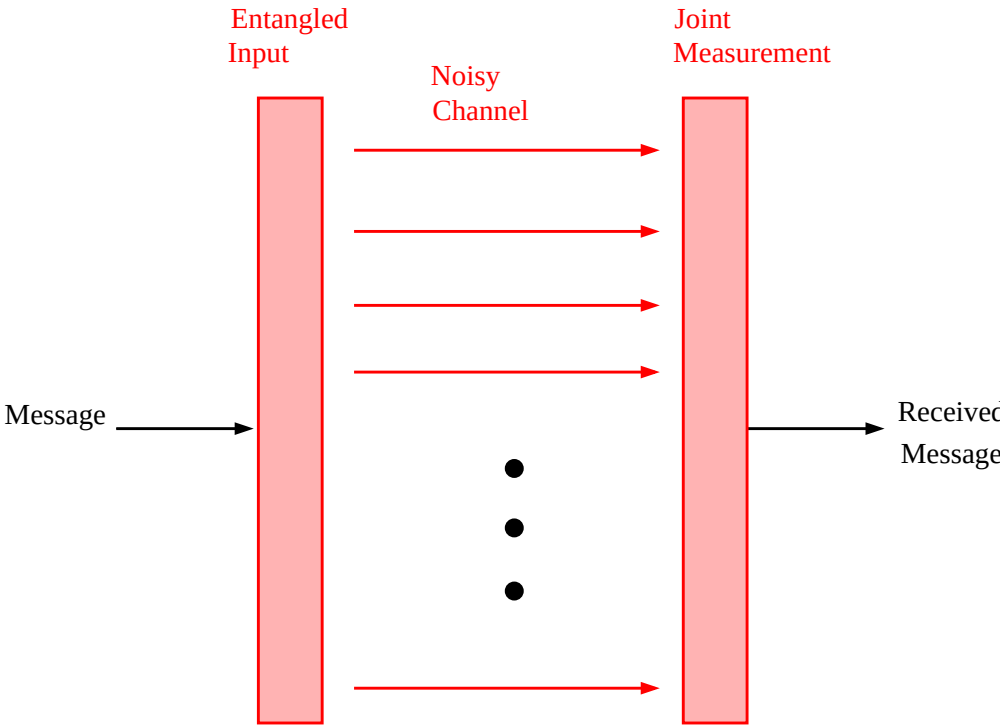
Unentangled Inputs, Joint Measurements



Maximize over probability distributions on inputs to the channel ρ_i, p_i:

$$\chi(\{\mathcal{N}(\rho_i)\}; \{p_i\})$$

Entangled Inputs, Joint Measurements



Maximize over probability distributions on inputs to the channel ρ_i, p_i where ρ_i is in the tensor product space of n inputs:

$$\lim_{n\to\infty} \frac{1}{n} \chi(\{\mathcal{N}^{\otimes n}(\rho_i)\}; \{p_i\})$$

Open Question

Is channel capacity additive?

Is $\max\ \chi(\mathcal{N}_1 \otimes \mathcal{N}_2) = \max\ \chi(\mathcal{N}_1) + \max\ \chi(\mathcal{N}_2)$?

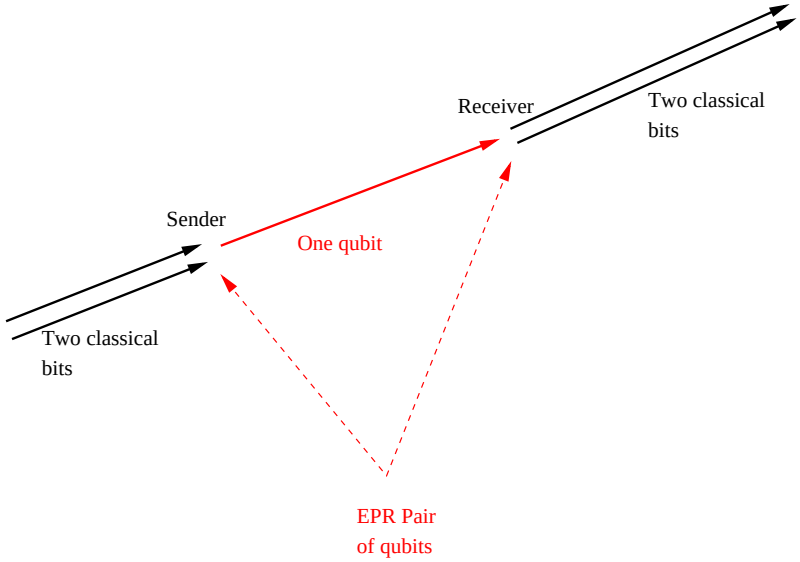
If it is, then χ gives the classical-information capacity of a quantum channel.

This turns out to be the same question as additivity of entanglement of formation considered in the previous lecture.

What things might increase the capacity of a quantum channel which don't affect the capacity of a classical channel?

- Entanglement between different channel uses? Unknown. This is the big open additivity question.
- A classical back channel from the receiver to the sender? This helps, but seems to make exact calculation of the capacity extremely difficult.
- Prior entanglement shared between the sender and the receiver. This helps and makes the formulas really nice.

Recall superdense coding lets you send two bits per qubit over a noiseless quantum channel if the sender and receiver share entanglement.



By Holevo's theorem, the bound without prior shared entanglement is one bit per qubit. Thus, for the noiseless quantum channel (the simplest case possible) entanglement between sender and receiver helps.

Suppose that we have a quantum channel $\mathcal{N}$. From superdense coding, if $\mathcal{N}$ is a noiseless quantum channel, the sender could communicate twice as much classical information to a receiver if they share EPR pairs than if they don't. How does this generalize to noisy channels? We call this quantity the entanglement-assisted capacity and denote it by C_E.

By superdense coding and teleportation, the entanglement-assisted quantum capacity is exactly half of the entanglement-assisted classical capacity.

$$Q_E = \frac{1}{2} C_E$$

Formula for entanglement-assisted capacity

Theorem (Bennett, Shor, Smolin, Thapliyal)

$$\max_{\rho} S(\mathcal{N}(\rho)) + S(\rho) - S((\mathcal{N} \otimes \mathcal{I}))(\Phi_\rho))$$

Φ_ρ is a pure state on the tensor product of the input space of the channel and a quantum space that the sender keeps, with

$$\mathrm{Tr}_B \Phi_\rho = \rho.$$

When the channel is classical, this formula turns into the entropy of the input plus the entropy of the output less the entropy of the joint system, or the second expression for classical mutual information.

Generalization

Suppose that the sender and the receiver have a limited amount of entanglement (E ebits) they share. How much can capacity can they obtain from a quantum channel?

If the sender is not allowed to use entanglement between different channel uses, the answer is:

$$\max_{\rho_i : \bar{H}(\rho_i) \leq E} \bar{H}(\rho_i) + H(\mathcal{N}(\bar{\rho}_i)) - \bar{H}((\mathcal{N} \otimes \mathcal{I})\Phi_{\rho_i})$$

Here $\bar{H}$ means average over the entropy, and $\bar{\rho}_i$ means average over the state; Φ_{ρ_i} is the pure entangled state (shared between sender and receiver) whose partial traces are ρ_i. This formula interpolates between the Holevo-Schumacher-Westmoreland capacity and the entanglement-assisted capacity.

How to prove the formula for C_E
(the lower bound)

$$\max_{\rho} S(\mathcal{N}(\rho)) + S(\rho) - S((\mathcal{N} \otimes \mathcal{I}))(\Phi_\rho))$$

Suppose ρ is $\frac{1}{d}\,\mathrm{Id}$, a multiple of the identity. Then we do the same operations as for standard superdense coding, with the generalizations of the Pauli matrices.

Use Holevo formula for χ:

$$\chi(\{\rho_i\}; \{p_i\}) = S(\sum_i p_i \rho_i) - \sum_i p_i S(\rho_i)$$

The first term of χ gives the first two terms of C_E; the second term of χ gives the last term of C_E.

The Holevo formula for χ:

$$\chi(\{\rho_i\};\{p_i\}) = S(\sum_i p_i \rho_i) - \sum_i p_i S(\rho_i)$$

Recall in superdense coding Alice and Bob share a maximally entangled state. Alice applies a random Pauli matrix to her half and sends it through the channel.

The first term is the entropy of the average state of both halves of the maximally entangled state after this operation. But applying a random Pauli matrix turns, on average, both halves into maximally mixed states, turning the first term of χ into

$$S(\mathcal{N}(\rho)) + S(\rho)$$

where ρ is $\frac{1}{d}I$.

Proof sketch if $\rho \neq \mathrm{Id}$.

If ρ is a projection matrix, things work the same way as in the identity case.

If ρ is not a projection matrix, then take tensor product of n uses of the channel, $\mathcal{N}^{\otimes n}$, and the projection matrix $\rho = \pi_T$, where T is a typical subspace for $\rho^{\otimes n}$.

It turns out we need to show that

$$\lim_{n\to\infty} \frac{1}{n} S(\mathcal{N}^{\otimes n}(\pi_T)) = S(\mathcal{N}(\rho)).$$

This is intuitively correct, but not that easy to prove rigorously.

What things might increase the entanglement-assisted capacity of a quantum channel which don't affect the capacity of a classical channel?

- Entanglement between different channel uses? Does not help!
- A classical back channel from the receiver to the sender? Does not help!
- Both of the above simultaneously? Does not help!

Proofs via quantum reverse Shannon theorem (next slide).

Quantum Reverse Shannon Theorem:

In the presense of entanglement, a noiseless qubit channel can simulate n uses of any quantum channel with entanglement-assisted capacity C_E by sending $nC + o(n)$ qubits.

This conjecture would show that asymptotically, in the presense of free entanglement, quantum channels are characterized by one parameter, C_E. The analogous theorem is true for classical channels in the presense of a correlated source of random bits.

With Charlie Bennett, Igor Devetak, and Andreas Winter, we have a proof of this theorem for channels (1) transmitting signals generated by some stochastic source, or (2) transmitting tensor product states.

Does not appear to be quite true for general inputs, unless we allow for other forms of shared entanglement than EPR pairs.

Quantum capacity

The quantum capacity is defined as $\lim_{n\to\infty} \log d/n$, where n is the number of channel uses in a protocol, and the d is the dimension of the largest Hilbert space which can be transmitted through the channel such that the fidelity of transmission of the (average/lowest fidelity) state in it is $1-\epsilon$, for some fixed ϵ.

The quantum capacity of a channel can be shown to be

$$\lim_{n\to\infty} \frac{1}{n} \max_{\rho} H(\mathcal{N}^{\otimes n}(\rho)) - H((\mathcal{N}^{\otimes n} \otimes I))(\Phi_\rho))$$

where Φ_ρ is a state whose reduced state on the input space is ρ.