

How Much Information is Contained in a Quantum State?

The Capacity of Quantum Channels

Peter Shor
AT&T Labs
Florham Park, NJ

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.

Claude E. Shannon, "A Mathematical Theory of Communication,"
The Bell System Technical Journal

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.

John R. Pierce, "The early days of information theory,"
IEEE Trans. Info. Theory.

Shannon's theorems

Definition of Entropy:

If a signal takes the value i with probability p_i , its entropy is

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

Source Coding

A source X can be compressed to length $H(X)$.

Channel Coding

A noisy channel N has capacity

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

where

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

Entropy of a quantum state

Classical Case

Given n photons, each in state $|\uparrow\rangle$ 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 $|\uparrow\rangle$ or $|\nearrow\rangle$, 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 (looking at heat, work, etc), von Neumann deduced that the entropy of a quantum system with density matrix ρ is

$$H_{\text{vN}}(\rho) = -\text{Tr}(\rho \log \rho)$$

Recall ρ was positive semidefinite, 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, $H_{\text{vN}}(\rho) = H_{\text{Shan}}(\lambda_i)$ so the von Neumann entropy is the Shannon entropy of the eigenvalues.

(Recall $\text{Tr} \rho = 1 = \sum_i \lambda_i$.)

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

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.

Answer (from classical information theory): Optimize over all possible measurements M on the signals (with outcomes M_1, M_2, \dots).

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

Example 1: Two states in ensemble

$$v_1 = \begin{array}{c} \updownarrow \\ \end{array} \quad v_2 = \begin{array}{c} \nearrow \\ \end{array}$$

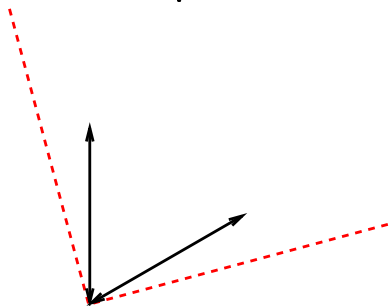
$$v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad v_2 = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}$$

Then

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

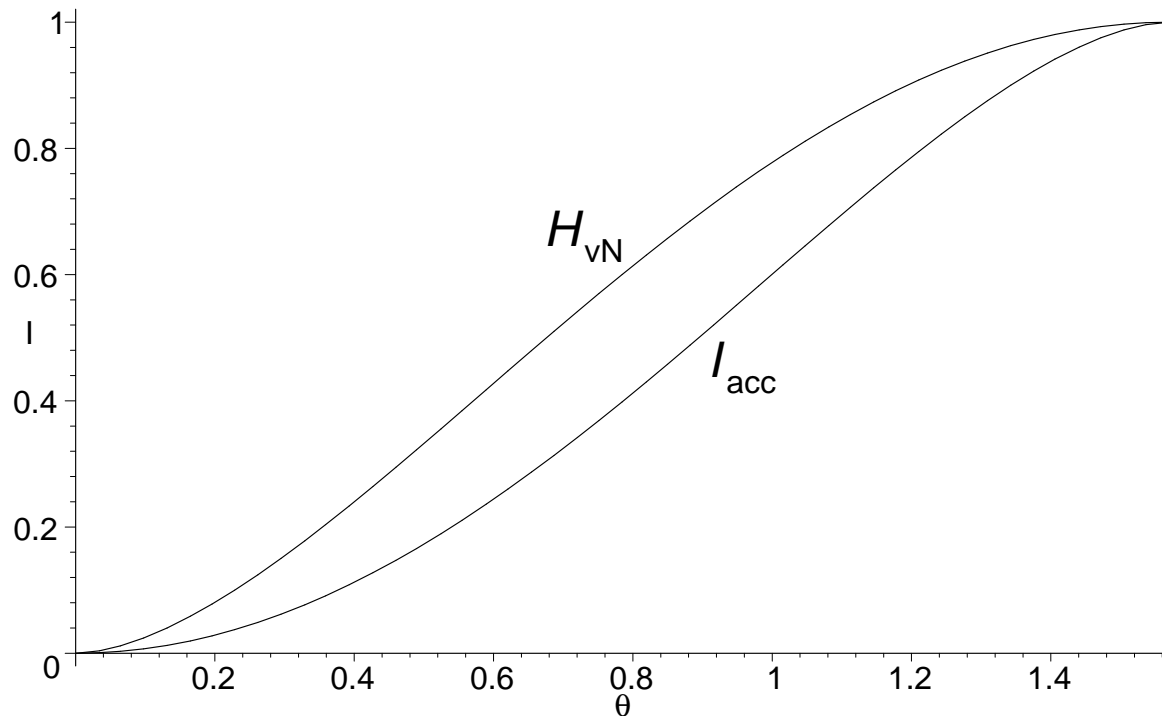
and $H_{\text{vN}} = H\left(\frac{1}{2} + \frac{\cos \theta}{2}\right)$.

The optimal measurement is



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

We see that $I_{\text{acc}} < H_{\text{vN}}(\rho)$.



A plot of H_{vN} and I_{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 $H_{\text{vN}} = H\left(\frac{1}{2} + \frac{\cos\theta}{2}\right)$ and the bottom the accessible information $I_{\text{acc}} = 1 - H\left(\frac{1}{2} + \frac{\sin\theta}{2}\right)$.

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, leaves other two equally likely

$$I_{\text{acc}} = \log 3 - 1 = .585 \text{ bits}$$

$$H_{\text{VN}} = 1$$

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

Holevo Bound χ

Suppose we have a source emitting ρ_i with probability p_i .

$$\chi = H_{\text{VN}}\left(\sum_i p_i \rho_i\right) - \sum_i p_i H_{\text{VN}}(\rho_i)$$

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 $I_{\text{acc}} < \chi$.

How can we use an ensemble of quantum states to send classical information?

Once we have chosen the measurement, we have essentially determined a classical channel. Shannon's classical coding theorem says that Alice can find a codebook using states from the ensemble such that she can asymptotically send Bob I_{acc} bits per state.

Example 2, Continued:

Suppose we use just two of the three signal states differing by 60° .

v_i :  (prob $\frac{1}{2}$)

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

Optimal Measurement for two vectors:

$$I_{\text{acc}} = 1 - H\left(\frac{1}{2} + \frac{\sqrt{3}}{2}\right) = .6454 \text{ bits}$$

This is larger than the accessible information for the ensemble containing all three states with equal probability, showing that the accessible information is not concave.

Let us go back to the situation where Alice is sending to Bob the states of the ensemble in Example 2 with equal probabilities.

Can Alice use this non-concavity of accessible information to let Bob extract more information from her ensemble?

She can give him hints. For the three-vector ensemble above, Alice can first narrow Bob's possibilities down to two vectors, and then he can use the optimal measurement to distinguish between these. This lets him extract more information from the reduced state.

This situation does not occur for classical probability distributions. Suppose Alice has three labels, each corresponding to a probability distribution on Bob's classically correlated information. If Alice sends Bob more information about the label, he can now necessarily extract less information from the reduced state.

Proof: Let the A be Alice's label, and B be Bob's state. Let C be the extra information (clue) sent from Alice to Bob.

$$\begin{aligned} I(A; B) &= I(C, A; B) \\ &= I(C; B) + I(A; B|C) \geq I(A; B|C) \end{aligned}$$

The accessible information for probability distributions over the three states at 120° angles is maximized when just two of them are used.

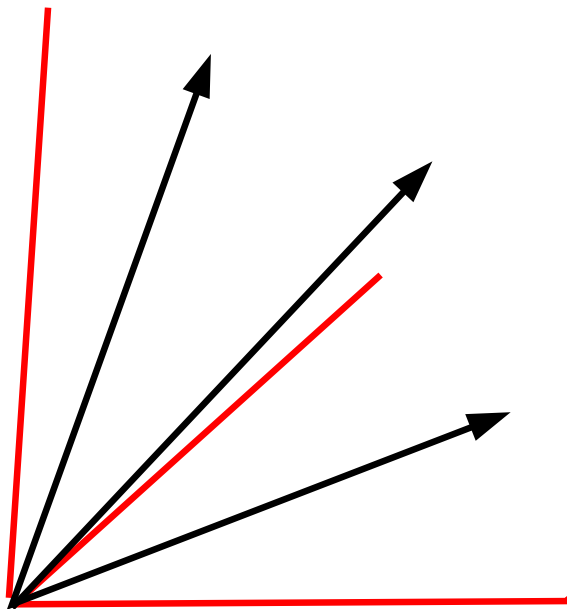
Is the most information we can send (channel capacity) using the three states of example 2?

Answer: No!

We can use tensor products of the states as codewords.

How to do better: 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 $2 \cdot 0.6545 = 1.309$ bits.



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

$$\chi(\{\rho_i\}; \{p_i\}) = H_{\text{VN}}\left(\sum_i p_i \rho_i\right) - \sum_i p_i H_{\text{VN}}(\rho_i)$$

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.

Schumacher Compression (Quantum noiseless coding theorem)

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

Want to send them to a receiver using as few qubits as possible.

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

$$nH_{\text{vN}}(\rho) + o(n)$$

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

Typical Sequences

These are used extensively in classical coding theory.

Suppose that we have a source emitting symbol s_i with probability p_i

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

Here, “close to the right number” can be defined as being anywhere within $c\sqrt{n}$ and ϵn of it, depending on which definition is appropriate for the context.

With high probability, a sequence emitted by a source is *typical*.

Proof of Classical Source Coding Theorem

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

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

Typical Subspaces

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

Look at the eigenvectors of the density matrix ρ .

Assign to each of eigenvector a probability equal to the corresponding eigenvalue.

Let the eigenvectors be $\hat{v}_1, \hat{v}_2, \dots, \hat{v}_d$ with probabilities $\hat{p}_1, \hat{p}_2, \dots, \hat{p}_d$.

Any two eigenvectors are orthogonal, so these eigenvectors behave classically.

Suppose we have n of these states.

The *typical subspace* S is the subspace generated by typical sequences of eigenvectors.

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

How to do Schumacher compression.

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

To compress:

Measure whether output of source lies in S .

If *yes*, get the state projected onto S . Can send using $\log \dim S \approx nH_{\text{VN}}(\rho)$ qubits.

If *no*, this is a low probability event; send anything.

Why does this work?

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

Using the eigenvectors $\hat{v}_1, \hat{v}_2, \dots, \hat{v}_d$ with probabilities $\hat{p}_1, \hat{p}_2, \dots, \hat{p}_d$ gives same probability of the outcomes as using states v_1, v_2, \dots, v_k with probabilities p_1, p_2, \dots, p_k .

Now, we know from the classical theory of typical sequences that the probability of a *no* outcome is very small with \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 S . Sending the state projected into S gives the right outcomes with high fidelity.

Theorem (pure state capacity)

We are given pure quantum states v_1, v_2, \dots, 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 $\chi = H_{\text{VN}}(\rho)$

How do we prove this?

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

Square root 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 $1 - \tilde{u}_i^\dagger \phi^{-1/2} \tilde{u}_i$.

The overall probability of error is

$$1 - \frac{1}{N} \sum_i \tilde{u}_i^\dagger \phi^{-1/2} \tilde{u}_i.$$

This can be shown to be small for $N < \dim S$ and u_i a random code.

Random Coding

We choose codewords

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

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 square root measurement.

We now discuss communication over quantum channels.

Formula for arbitrary memoryless quantum channel \mathcal{N} .

\mathcal{N} must be trace-preserving completely positive operator.

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

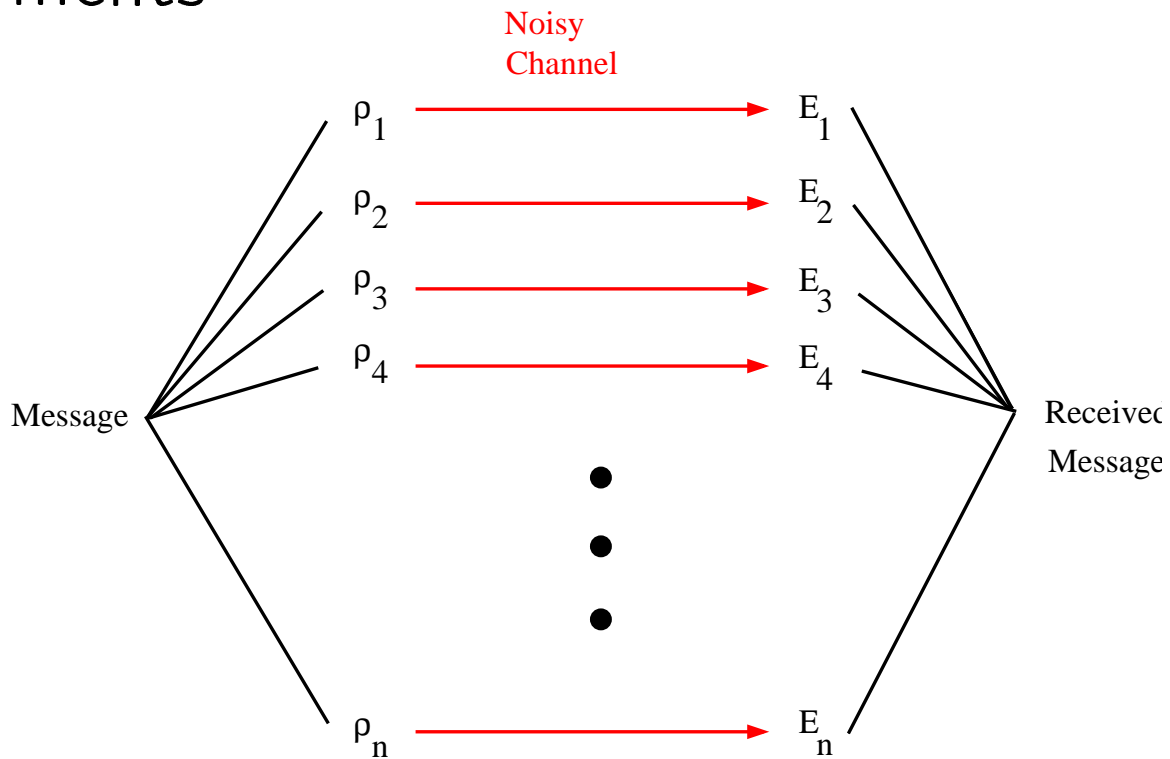
where

$$\sum_i A_i^\dagger A_i = I$$

Positive: takes positive semi-definite matrices to positive semi-definite matrices.

Completely positive: is positive even when tensored with the identity channel.

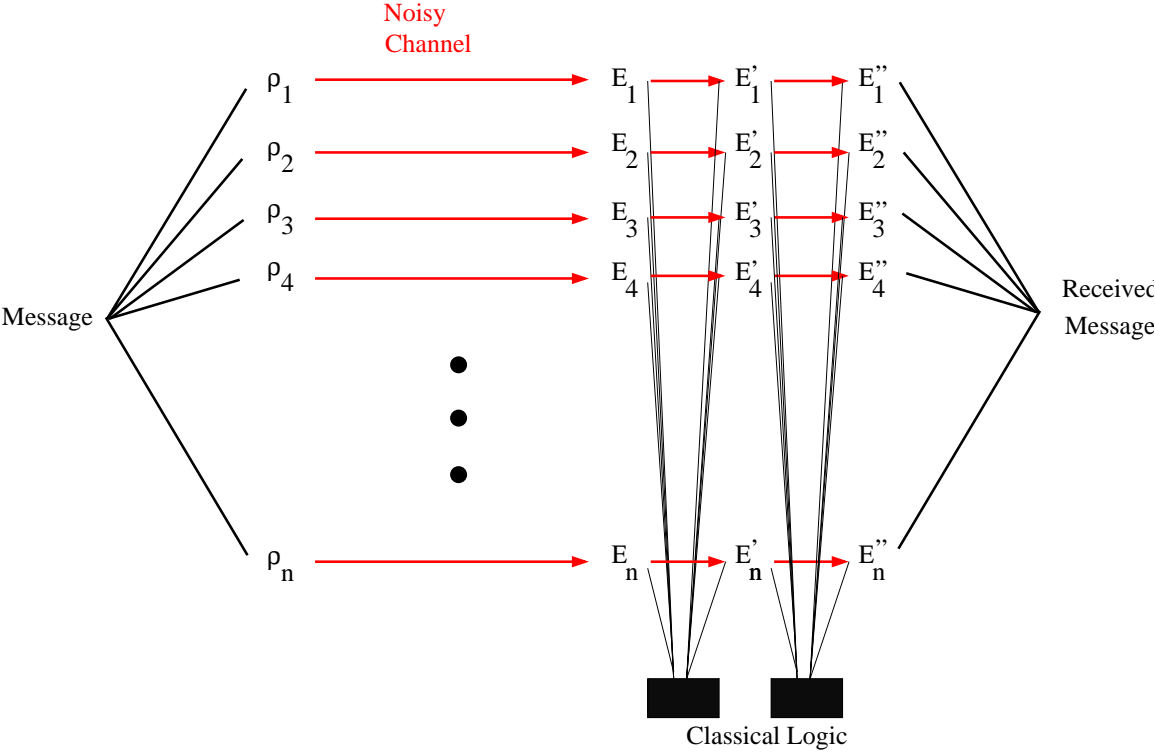
Unentangled Inputs, Separate Measurements



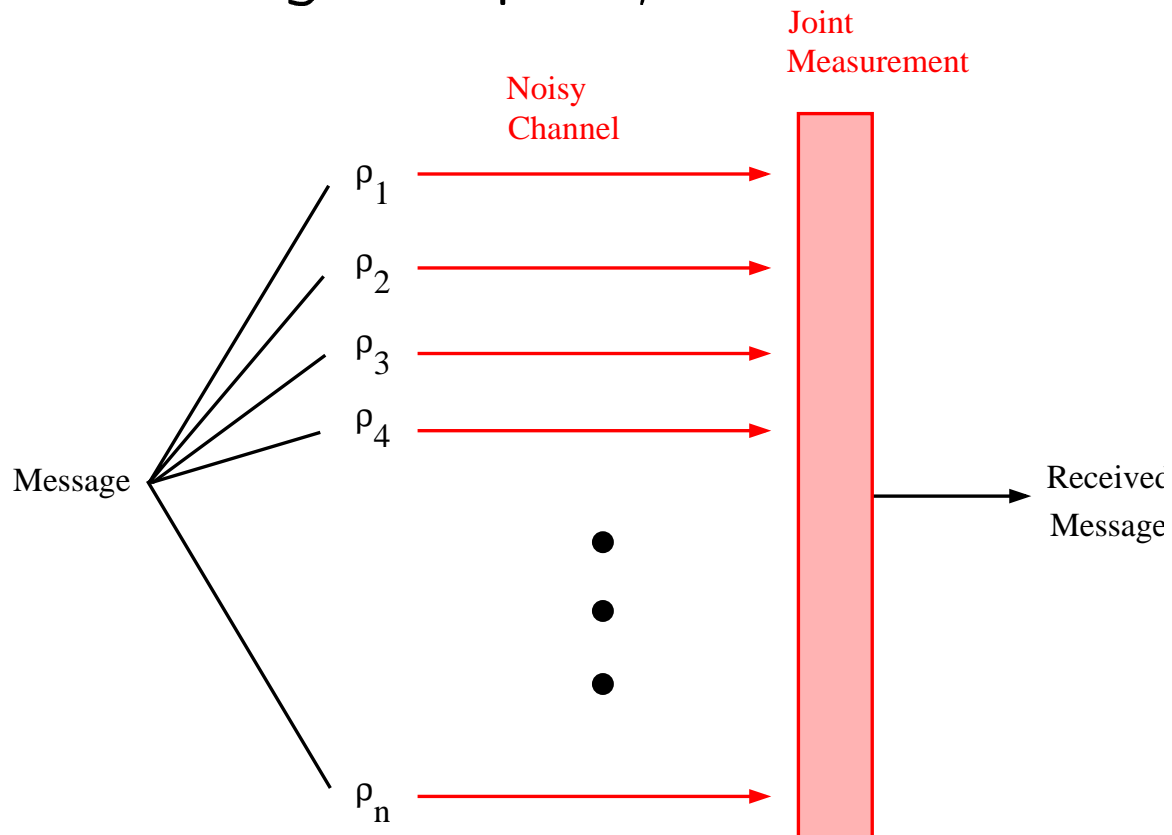
Maximize over probability distributions on inputs to the channel ρ_i, p_i :

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

Unentangled Inputs, Adaptive Separate Measurements



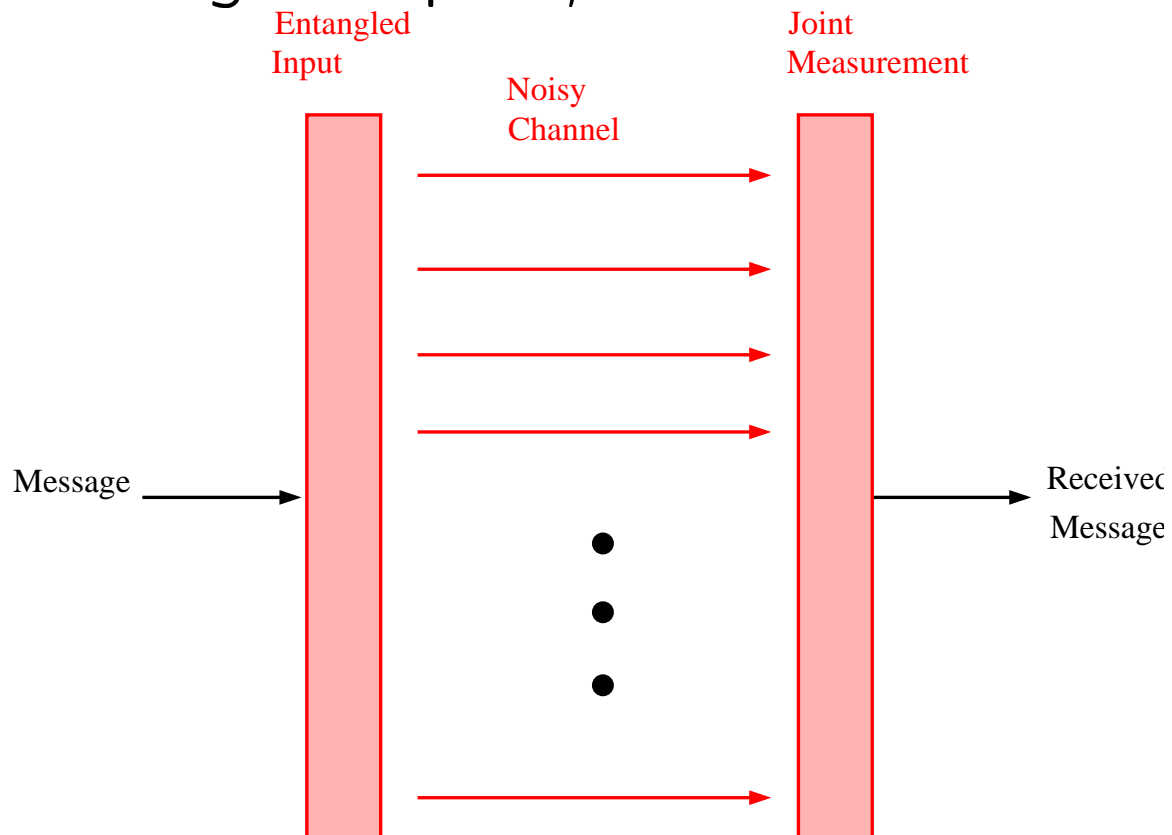
Unentangled Inputs, Joint Measurements



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:

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

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.

Progress on additivity conjecture.

Using ideas of Audenaert, Braunstein, Matsumoto, Shimono, and Winter, I can show that the following four questions are equivalent:

- Additivity of the minimum entropy output of a quantum channel.
- Additivity of the $C_{1,\infty}$ capacity of a quantum channel.
- Additivity of the entanglement of formation of a quantum state.
- Strong superadditivity of entanglement of formation of a quantum state.

First idea: Stinespring representation

A quantum channel

$$\mathcal{N} : \mathcal{H}_{\text{in}} \rightarrow \mathcal{H}_A$$

can be represented as a unitary transformation

$$\mathcal{U} : \mathcal{H}_{\text{in}} \rightarrow \mathcal{H}_A \otimes \mathcal{H}_B$$

followed by a partial trace

$$\text{Tr}_B : \mathcal{H}_A \otimes \mathcal{H}_B \rightarrow \mathcal{H}_A$$

Then if $\mathcal{U}(\rho) = \sigma$, we get

$$C_{1,\infty}(\mathcal{N}) = \max_{\rho} H(\mathcal{N}(\rho)) - E_F(\sigma)$$

$$E_F(\sigma) = \min_{\substack{p_i, |\phi_i\rangle \\ \sum_i p_i |\phi_i\rangle\langle\phi_i| = \sigma}} \sum_i p_i \text{Tr}_B |\phi_i\rangle\langle\phi_i|$$

Second idea: linear programming duality

Let

$$Y = \min_{\substack{p_i, |v_i\rangle \\ \sum_i p_i |v_i\rangle\langle v_i| = \rho}} \sum p_i H(\mathcal{N}(|v_i\rangle\langle v_i|))$$

This is the second term in the $C_{1,\infty}$ capacity formula. By linear programming duality, we also have

$$Y = \max_{\tau} \text{Tr } \tau \rho : \\ \tau \text{ such that } \langle v | \tau | v \rangle \leq H(\mathcal{N}(|v\rangle\langle v|)) \quad \forall |v\rangle$$

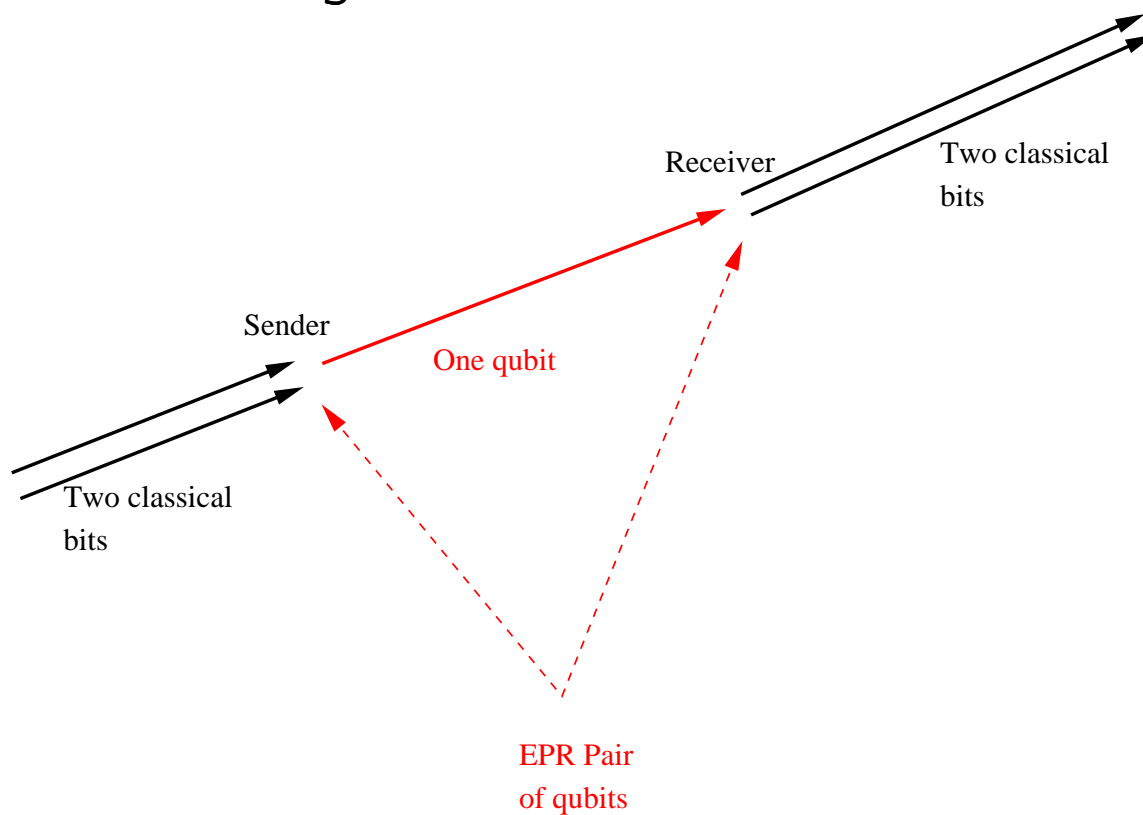
What things might increase the capacity of a quantum channel which don't affect the capacity of a classical channel?

- a) Entanglement between different channel uses?
Unknown.

- b) A classical feedback channel from the receiver to the sender? Not without a).

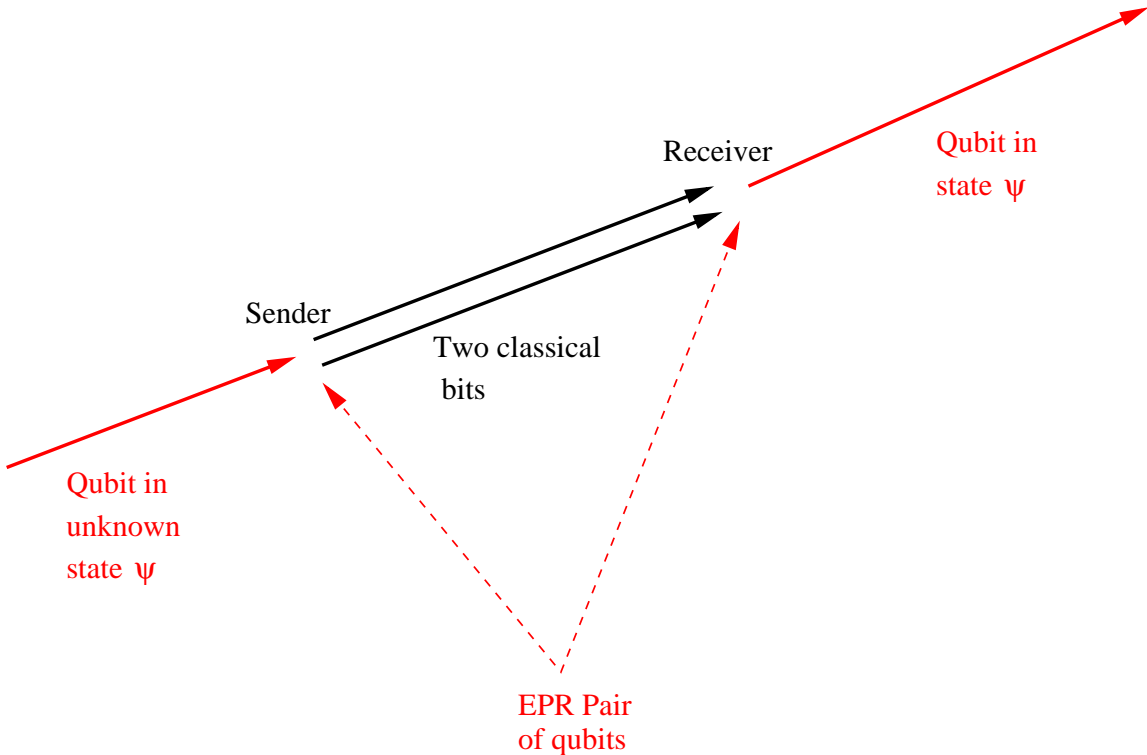
- c) Prior entanglement shared between the sender and the receiver. This helps!

The phenomenon called 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.

Teleportation is the converse to superdense coding. This lets a sender communicate a quantum bit to a receiver using two classical bits if the sender and receiver share entanglement.



This seems paradoxical, since there are an infinite number of quantum states a qubit can be in. But only one bit of information is extractable from a qubit, so it's not really a paradox.

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 .

Formula for entanglement-assisted capacity

Theorem (Bennett, Shor, Smolin, Thapliyal)

$$C_E = \max_{\Phi} H(\text{Tr}_B (\mathcal{N} \otimes \mathcal{I})\Phi) + H(\text{Tr}_A (\mathcal{N} \otimes \mathcal{I})\Phi) - H((\mathcal{N} \otimes \mathcal{I})\Phi)$$

The sender is A; the receiver B; Φ is a pure state on the tensor product of the input space of the channel and a quantum space that the sender keeps. 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.

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.

Quantum Information Capacity

We can also look at how much *quantum information* a channel can transmit. To say we have transmitted d bits of quantum information reliably, the sender needs to be able to take any state he is given in a 2^d -dimensional Hilbert space, encode it in quantum states, and send it over the quantum channel to the receiver, who then must decode it with high fidelity.

A classical feedback channel from the receiver to the sender, or a classical two-way side channel, will increase the quantum channel capacity.

This gives at least three capacities,

$$Q_1 \leq Q_{FB} \leq Q_2$$

Coherent Information

$$I_C(\mathcal{N}) = \max_{\rho \in \mathcal{H}_{\text{in}}} H_{\text{vN}}(\mathcal{N}(\rho)) - H_{\text{vN}}((\mathcal{N} \otimes \mathcal{I})\Phi_\rho)$$

where Φ_ρ is a purification of ρ , so $\text{Tr}_2 \Phi_\rho = \rho$.

$$Q_1(\mathcal{N}) = \lim_{n \rightarrow \infty} \frac{1}{n} I_C(\mathcal{N}^{\otimes n})$$