

Mixing and measurement:

Recall that Shannon entropy is Schur concave — If the probability vector (p_x) is majorized by (q_y) (i.e. $(p_x) \prec (q_y)$), then $H(X) \geq H(Y)$.

Both relations reflect that "X is more random than Y." Thus a doubly stochastic channel

$$p_x = \sum_y D_{xy} q_y \quad (\text{where rows and columns of } D \text{ sum to } 1)$$

does not decrease entropy.

Von Neumann entropy $H(\rho)$ of density operator ρ is Shannon entropy of $\lambda(\rho)$ — the vector of eigenvalues of ρ . Recall that we have shown (as a consequence of the HJW Thm.) that ρ is realized as an ensemble of pure states

$$\{|\psi_x\rangle, p_x\} \rightarrow \rho = \sum_x p_x |\psi_x\rangle \langle \psi_x|$$

Then $p \prec \lambda(\rho)$; therefore $H(\rho) \leq H(X)$

The VN entropy of density operator is no larger than the Shannon entropy of the mixture. The entropies are equal iff the states $\{|\psi_x\rangle\}$ are mutually orthogonal. In general

$$\sqrt{p_x} |\psi_x\rangle = \sum_i V_{xi} \sqrt{\lambda_a} |a\rangle$$

unitary $\quad \quad \quad$ eigenvectors of ρ

$$\Rightarrow p_x = \sum_i |V_{xi}|^2 \lambda_a$$

$\quad \quad \quad$ doubly stochastic D_{xi}

As we'll soon see, when we perform a measurement on the system with density operator ρ , the amount of information we can gain about ρ

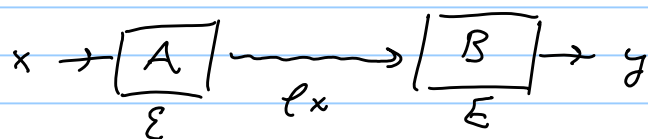
is no more than $H(\rho)$ bits. So mixing of nonorthogonal states is irreversible: that is, information about which state was prepared is irretrievably lost.

If we perform the orthogonal measurement that projects onto the basis $\{|y\rangle\}$, then outcome y occurs with probability

$$p_y = \langle y | \rho | y \rangle = \sum_a \lambda_a |\langle a | y \rangle|^2$$

where $\rho = \sum_a \lambda_a |a\rangle\langle a|$; since $|\langle a | y \rangle|^2$ is doubly stochastic, $(p_y) \prec \lambda(\rho)$ and hence $H(Y) \geq H(\rho)$, with equality iff the measurement is in basis in which ρ is diagonal. Any other orthogonal measurement produces an outcome that is "more random" than the measurement in the diagonal basis.

Let's ask a sharper question — how much information can we gain by making a measurement? Consider a game



Alice prepares a state by sampling from the ensemble $E = \{\rho_x, p(x)\}$. Bob performs POVM with Kraus operators $E = \{E_y\}$, $\sum E_y^\dagger E_y = I$. Then conditional prob of outcome y if state prepared is ρ_x is

$$p(y|x) = \text{tr}(E_y^\dagger E_y \rho_x)$$

and joint distribution is $p(x, y) = p(y|x)p(x)$

Averaged over Alice's prep. and Bob's outcome, Bob's information gain about Alice's state

is $I(X;Y)$. Bob's best strategy (his optimal measurement) maximizes his information gain. This maximum value of the mutual information is a property of the ensemble, which we call the "accessible information" of the ensemble:

$$Acc(\mathcal{E}) = \max_{\mathcal{E}} I(X;Y).$$

If the states $\{p_x\}$ are mutually orthogonal, we may choose $\{E_y\}$ to be projectors onto support of these states; then

$$p(y|x) = \delta_{y,x} \Rightarrow H(X|Y) = 0 \text{ and } I(X;Y) = H(X).$$

In this case, the optimal measurement determines Alice's preparation perfectly. But if ensemble is not orthogonal, then

$$H(X|Y) > 0 \text{ and } I(X;Y) < H(X);$$

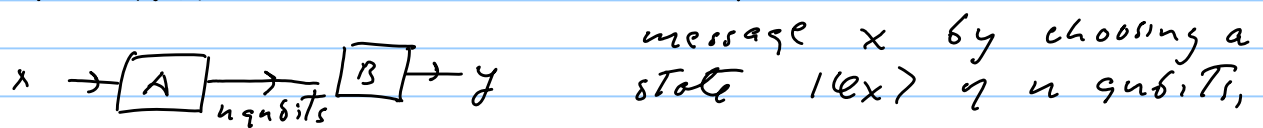
some info about what Alice prepared is inaccessible.

There is no useful general formula for $Acc(\mathcal{E})$, but we can derive a useful upper bound:

For $\mathcal{E} = \{|\psi_x\rangle, p(x)\}$ (an ensemble of pure states),

$$Acc(\mathcal{E}) \leq H(\rho), \text{ where } \rho = \sum p(x) |\psi_x\rangle \langle \psi_x|$$

this (a special case of) the "Holevo bound", and it is not tight in general; it can be generalized also to an alphabet of mixed states, as we'll discuss later. Note that if Alice tries to encode a

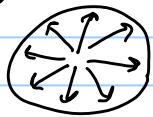


then

$$I(X;Y) \leq H(\rho^{(n)}) \leq n,$$

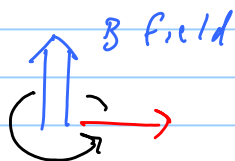
since $H(\rho) \leq \log_2 d$ and here $d = 2^n$

Thus, a qubit can convey at most one bit of classical info from Alice to Bob. Alice might try to encode more classical info by



choosing from a large alphabet of pure states — i.e. vectors on the Bloch sphere. But these states are imperfectly distinguishable — so Bob cannot extract more than a bit about what Alice prepared.

Why should we care about how well information can be encoded in nonorthogonal states? The players in the game might be, rather than "Alice" and "Bob," "Nature" and "experimenter."



For example, experimenter wants to measure magnetic field B , so he prepares spin in a known initial state, then in a specified time the spin precesses by an a priori unknown amount.

The experimenter then wants to collect info about the direction in which the spin is pointing — i.e. he wants to distinguish nonorthogonal states. If the spins decohere, then he will need to distinguish among an alphabet of mixed states.

To derive the Holevo bound (and other interesting consequences) we exploit properties of Von Neumann entropy. In some ways VN entropy is like Shannon entropy, but in some important ways it is different.

Subadditivity: $H(AB) \leq H(A) + H(B)$
(a HW exercise)

which implies

$$I(A; B) = H(A) + H(B) - H(AB) \geq 0;$$

quantum mutual information is nonnegative.

$I(A;B)$ expresses how strongly correlated are the systems A and B . It vanishes iff the state of AB is a product state

$$\rho_{AB} = \rho_A \otimes \rho_B = \sum_{a,b} |a,b\rangle \lambda_a \chi_b \langle a,b|$$

i.e. the vector of eigenvalues is a product distribution
 $\Rightarrow H(AB) = H(A) + H(B)$

An interesting difference between Shannon and VN entropy concerns conditional entropy. Classically

$$H(X|Y) \geq H(Y) \Rightarrow$$

$$H(X|Y) = H(XY) - H(Y) \geq 0$$

$$\text{Recall that } H(X|Y) = \langle -\log p(x|y) \rangle = \langle -\log \frac{p(x,y)}{p(y)} \rangle$$

quantifies our remaining uncertainty about X once Y is known.

But quantumly we can have $H(AB) < H(B)$

$$\Rightarrow H(A|B) = H(AB) - H(B) < 0$$

For example, if AB is pure, then $H(AB) = 0$ and $H(B) = H(A) = E$ is entanglement of A and B

$$\Rightarrow H(A|B) = -E < 0$$

It is as though our remaining uncertainty about A when B is known is negative — we are "more than certain." Actually, negative uncertainty has a sensible operational interpretation, which we'll come to later.

Strong Subadditivity

Classically, Shannon mutual info satisfies "strong subadditivity":

$$I(X;YZ) \geq I(X;Y)$$

"Obviously" the info you gain about X when you know Y and Z is no less than when you know only Y ! Strong subadditivity follows from the "chain rule" for mutual info (which holds classically or quantumly). Note that

$$I(X;Y) = H(X) - H(X|Y)$$

$$I(X;YZ) = H(X) - H(X|YZ)$$

$$I(X;Z|Y) = H(X|Y) - H(X|YZ)$$

which implies the identity

$$I(X;YZ) = I(X;Y) + I(X;Z|Y) \quad (\text{"chain rule"})$$

But classically it is also obvious that

$$I(X;Z|Y) = \sum_y p(y) I(X;Z|y) \geq 0 \quad (\text{a convex combination of nonnegative quantities})$$

From which follows: $I(X;YZ) \geq I(X;Y)$

Quantumly it is also true that $I(A;C|B) \geq 0$,

and therefore $I(A;BC) \geq I(A;B)$

But in the quantum case strong subadditivity is a rather deep theorem — there is no known elementary proof.

There are a few other ways to express strong subadditivity that are sometimes useful:

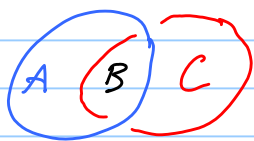
$$H(A) - H(A|BC) \geq H(A) - H(A|B) \Rightarrow$$

$$H(A|B) \geq H(A|BC) \quad (A \text{ becomes less uncertain when we know } C \text{ as well as } B)$$

Also:

$$H(A) + H(B) - H(AB) \leq H(A) + H(BC) - H(ABC)$$

$$\Rightarrow H(ABC) + H(B) \leq H(AB) + H(BC)$$



If AB and BC are two overlapping systems, then ABC is their union, and B is their intersection. (This reduces to subadditivity when intersection B is trivial.)

Monotonicity

One important consequence of strong subadditivity (SSA) is monotonicity of quantum mutual information.

This means that an operation applied to B cannot increase the mutual information of A and B .

To derive monotonicity, we recall that an operation $B \rightarrow B'$ can be realized by an isometry

$B \rightarrow B'E$, where E is a suitable "environment." Furthermore, the VN entropy is invariant under a unitary change of basis: $H(\rho) = H(U\rho U^\dagger)$. Therefore

$$H(AB) = H(AB'E) \text{ and } H(B) = H(B'E) \Rightarrow$$

$$I(A; B) = I(A; B'E)$$

Then from SSA we have $I(A; B'E) \geq I(A; B')$

$$\Rightarrow \boxed{I(A; B) \geq I(A; B')}$$

This makes sense: an operation on B cannot increase the correlation of B with A (though it can reduce the correlation).

Holevo Bound

one important consequence of SSA is the Holevo bound. In the accessible information game, we consider a three part system. Q is the quantum system that Alice prepares and Bob measures. Alice records the state that she prepares in register X and Bob records his measurement outcome in the register Y . We are interested in the mutual info $I(X; Y)$ of Alice's record and Bob's record. The joint state of XQY after Alice prepares is

$$\rho_{XQY} = \sum_x p(x) |x\rangle\langle x| \otimes (\rho_x \otimes |0\rangle\langle 0|),$$

which becomes after Bob measures:

$$\rho_{X'Q'Y'} = \sum_{x,y} p(x) |x\rangle\langle x| \otimes E_y \rho_x E_y^\dagger \otimes |y\rangle\langle y|$$

SSA says that, after the measurement

$$I(X'; Y') \leq I(X'; Q'Y')$$

Furthermore, since the measurement is an operation applied to QY , monotonicity of mutual information implies

$$I(X'; Q'Y') \leq I(X; QY) = I(X; Q)$$

So now we should compute $I(X; Q) = H(Q) - H(Q|X)$ (a property of the ensemble). We have:

$H(Q) = H(\rho)$ where $\rho = \sum_x p(x) \rho_x$ - the VN entropy of the density operator for ensemble.

$$H(Q|X) = \sum_x p(x) H(\rho_x)$$

- the entropy of the signal state conditioned on the preparation, averaged over the ensemble. To compute it explicitly:

$H(Q|X) = H(XQ) - H(X)$, where $H(X)$ is Shannon entropy of the ensemble, and

$$\rho_{XQ} = \sum_x p(x) (|x\rangle\langle x| \otimes \rho_x)$$

$$\begin{aligned} \Rightarrow H(XQ) &= - \sum_x \text{tr} [p(x) \rho_x \log p(x) \rho_x] \\ &= H(X) - \sum_x p(x) \text{tr} \rho_x \log \rho_x = H(X) + \sum_x p(x) H(\rho_x) \end{aligned}$$

Therefore, $H(Q|X) = \sum_x p(x) H(\rho_x)$ and

$$I(X'; Y') \leq I(X, Q) = H(Q) - H(Q|X)$$

$$= H(\mathcal{E}) - \langle H(\rho_x) \rangle$$

VN entropy
of ensemble

average VN entropy
of the alphabet

The quantity $\chi(\mathcal{E}) = H(\mathcal{E}) - \langle H(\rho_x) \rangle$

is called the "Holevo Chi" (or Holevo information) of the ensemble $\mathcal{E} = \{p(x), \rho_x\}$ and we have

derived
$$A_{CL}(\mathcal{E}) = \max_E I(X; Y) \leq \chi(\mathcal{E})$$

For an ensemble of pure states, $H(\rho_x) = 0$ for all x , and

$$A_{CL}(\mathcal{E}) \leq H(\mathcal{E})$$

To summarize the argument succinctly,

$$I(X; Y)_{\text{after}}^{\text{SSA}} \leq I(X; QY)_{\text{after}}^{\text{Mono}} \leq I(X; Q)_{\text{before}} = \chi(\mathcal{E}),$$

where "after" refers to after Bob's measurement and before refers to before Bob's measurement. The first inequality follows from strong subadditivity, and the second from monotonicity of mutual information (itself a consequence of SSA).

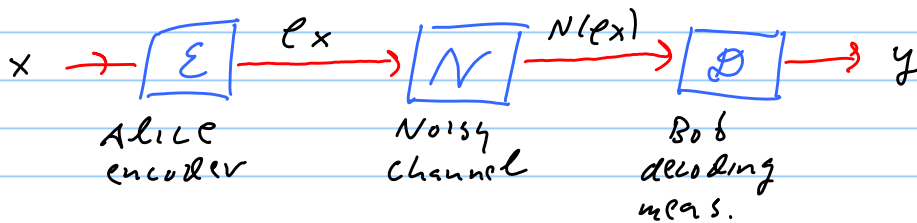
Since Holevo Chi is itself an instance of quantum mutual information, we can apply monotonicity to X . If a channel N maps Q to Q' , then

$$X' = I(X; Q') \leq I(X, Q) = X;$$

the channel cannot increase Holevo Chi of an ensemble

Classical Capacity of a Quantum Channel

Accessible information, and hence also Holevo Chi, are relevant for sending classical information through a noisy quantum channel



Alice encodes classical message by preparing quantum state ρ_x and Bob receives noisy version $N(\rho_x)$ of the state. Bob measures and tries to infer x . For prob distribution $p(x)$ on the messages, Bob's optimal measurement achieves the accessible information of the ensemble

$$\mathcal{E} = \{N(\rho_x), p(x)\}$$

which is bounded above by Holevo Chi of the ensemble

$$X(\mathcal{E}) = I(X; B)$$

- the quantum mutual information of the state

$$\rho_{XB} = \sum_x p(x) |x\rangle\langle x| \otimes N(\rho_x)$$

In a single shot, the optimal meas. does not achieve Holevo Chi, in general.

But if we use the channel many times, there is a capacity theorem saying that we can achieve X bits per letter asymptotically.

Before discussing achievability, consider the converse of the coding theorem, the upper bound on capacity. It is useful to recall the classical case, where the channel is the conditional prob. distribution $p(y|x)$. Suppose we use the channel n times, where there are 2^{nR} codewords (hence R bits conveyed per use of the channel). Let $\tilde{X}^n$ denote the uniform distribution on these codewords, which induces joint distribution $(\tilde{X}^n, \tilde{Y}^n)$ on channel inputs and outputs. The mutual information for this distribution is

$$I(\tilde{X}^n; \tilde{Y}^n) = H(\tilde{X}^n) - H(\tilde{X}^n | \tilde{Y}^n).$$

But $H(\tilde{X}^n) = nR$ for the uniform distribution on codewords, and $H(\tilde{X}^n | \tilde{Y}^n) \rightarrow 0$ as $n \rightarrow \infty$ if the output can be decoded with negligible error probability; thus

$$R = \frac{1}{n} [I(\tilde{X}^n; \tilde{Y}^n) + \varepsilon]$$

The achievable rate, then, is bounded above by

$$R = \max_{X^n} \frac{1}{n} I(X^n; Y^n)$$

(since the uniform distribution on codewords is a special case of a distribution X^n)

We can express

$$I(X^n; Y^n) = H(Y^n) - H(Y^n | X^n),$$

and the channel acts independently on each letter, so that

$$H(Y^n|X^n) = \sum_i \langle -\log p(y_i|x_i) \rangle = \sum_{i=1}^n H(Y_i|X_i)$$

Entropy is subadditive $\Rightarrow$

$$H(Y^n) \leq \sum_i H(Y_i) \quad , \text{ and therefore}$$

$$\frac{1}{n} I(X^n; Y^n) \leq \frac{1}{n} \left(\sum_i I(X_i; Y_i) \right) \leq \max_X I(X; Y),$$

which implies

$$\lim_{n \rightarrow \infty} \max_{X^n} \frac{1}{n} I(X^n; Y^n) = \max_X I(X; Y) = C$$

The left-hand side is a valid expression for capacity (it is achievable), but not very useful, since it involves an arbitrary number of channel uses; we say it is a "regularized" expression. But because $I(X^n; Y^n)$ is subadditive, it reduces to a "single-letter formula" involving just one use of the channel

Now consider sending classical information over a quantum channel, where the quantum channel is used n times. For a code with rate R , consider the uniform distribution over the 2^{nR} codewords. If Bob can decode with negligible probability of error, then his optimal measurement can identify the codeword sent, and his information gain is

$$I(\tilde{X}^n; \tilde{Y}^n) = nR - \epsilon \quad \text{where } \epsilon \rightarrow 0 \text{ as } n \rightarrow \infty$$

By the Holevo bound, then, the rate satisfies

$$R = \frac{1}{n} [I(\tilde{X}^n; \tilde{Y}^n) + \epsilon] \leq \frac{1}{n} \left[\max_{X^n} I(X^n; B^n) + \epsilon \right]$$

If we define the classical capacity C of the quantum channel as the maximum rate that can be achieved with error prob. $\rightarrow 0$ as $n \rightarrow \infty$, then

$$C \leq \max_{X^n} \left[\frac{1}{n} I(X^n; B^n) \right]$$

In fact, this expression really is achievable using random coding, so it is a regularized expression for the classical capacity. Can it be reduced to a single-letter formula?

$$I(X^n; B^n) = H(B^n) - H(B^n | X^n)$$

subadditivity of VN entropy implies

$$H(B^n) \leq \sum_{i=1}^n H(B_i),$$

where B_i is the marginal density operator for the i th letter that Bob receives. But what can we say about $H(B^n | X^n)$?

If Alice uses the channel n times, then in principle she could choose her quantum signals to be entangled states of the n letters she sends, and in that case the signals that Bob receives could also be entangled. But suppose that Alice decides to send codewords that are product states (though of course the signals sent in channel uses $i=1, 2, \dots, n$ could be classically correlated). Bob will still want to use the optimal POVM, which might act collectively on the n letters he receives, but $I(X^n; B^n)$ is still an upper bound on the information Bob can gain. In general, then, Bob receives states chosen from some ensemble of product states

$$\mathcal{E} = \left\{ \rho_{x_1} \otimes \rho_{x_2} \otimes \dots \otimes \rho_{x_n}, p(x_1, x_2, \dots, x_n) \right\}$$

(where $p(x)$ may be correlated among $x_1, x_2, \dots, x_n$).

$$\text{Then } H(B^n | X^n) = \sum_{x_1, \dots, x_n} p(x_1, \dots, x_n) H(\rho_1 \otimes \dots \otimes \rho_n)$$

since the entropy of a product state is additive,

$$H(\mathcal{P}_{x_1} \otimes \dots \otimes \mathcal{P}_{x_n}) = \sum_i H(\mathcal{P}_{x_i}), \text{ we have}$$

$$\begin{aligned} H(B^n | X^n) &= \sum_{x_1} p_1(x_1) H(\mathcal{P}_{x_1}) + \dots + \sum_{x_n} p_n(x_n) H(\mathcal{P}_{x_n}) \\ &= \sum_{i=1}^n H(B_i | X_i) \end{aligned}$$

(where $p_i(x_i)$ is the marginal prob. distribution for the i th letter. So for the special case of product codewords, the mutual information is subadditive

$$I(X^n; B^n) \Big|_{\substack{\text{product} \\ \text{states}}} \leq \sum_i I(X_i; B_i) \leq n \max_X I(X; B)$$

So our upper bound on the capacity (which is achievable with product state codewords) becomes a single-letter formula:

$$C_1 = \max_X I(X; B) \equiv \chi(N)$$

(which is a property of the channel N). C_1 is the "product state classical capacity" of the channel N .

To show that C_1 is really an achievable rate, we use random coding. For a given input ensemble $\{\mathcal{P}_x, p(x)\}$ we generate n -letter codewords by sampling from the ensemble n times; when $\mathcal{P}_x$ is sent, Bob receives $N(\mathcal{P}_x)$. In Shannon's

case, we could say that with high prob Bob receives one of at least $2^{n(H(Y) - \delta)}$ typical messages, and that

for each message sent by Alice, Bob receives one of $2^{n(H(Y|X) + \delta)}$ typical messages. If Alice sends one of 2^{nR} messages, the prob of a decoding error, because the message received is in more than one decoding sphere is

$$\text{errn prob} \leq \frac{2^{nR} 2^{n(H(Y|X) + \delta)}}{2^{n(H(Y) - \delta)}}$$

$$\leq 2^{n(R - I(X; Y) + 2\delta)}$$

which approaches zero as $n \rightarrow \infty$ for $R < I - 2\delta$

Quantumly, Bob receives quantum message in a typical subspace of dimension $\text{last } 2^{n(H(B) - \delta)}$ and for each message sent by Alice, the signal is in a typical subspace of dimension $2^{n(H(B|X) + \delta)}$

To give an honest argument we should specify Bob's decoding PVM and estimate its error probability. But that is rather technical, so suffice it to say that the probability that signal received is accidentally in the decoding subspace of another signal is

$$\text{errn prob} < \frac{2^{nR} 2^{n(H(B|X) + \delta)}}{2^{n(H(B) - \delta)}}$$

$$= 2^{n(R - I(X; B) + 2\delta)}$$

so we can achieve the rate $R = I(X; B)$ asymptotically.

But is the capacity C really the same as the product state capacity C_1 ? It would be if

we could show $I(X^n; B^n) \leq \sum_i I(X_i; B_i)$

in general, for entangled signals as well as product signals. However, a recent discovery is that the

Holovo chi is not subadditive — there are channels

such that $\chi(N_1 \otimes N_2) > \chi(N_1) + \chi(N_2)$.

As a result, our understanding of the classical capacities of quantum channels is far from complete. We have only a regularized formula, not a single-letter formula.

Quantum Channel Capacity

The formula $C = \max_X I(X; Y)$ for capacity of a

quantum channel is quite robust. For example, the capacity does not increase if we allow sender and receiver to share randomness, or if we allow feedback from the receiver to the sender. But quantumly the situation is more complicated. For example, shared entanglement boosts the quantum capacity, as does classical communication from receiver to sender. So there are a variety of different natural notions of quantum capacity, all with different capacity formulas.

Perhaps the most natural quantum capacity (unassisted by entanglement, and with one-way quantum communication) is this:



Alice encodes pure state $|\psi\rangle$ in Hilbert space $\mathcal{H}^{(n)}$ (with $\log \dim \mathcal{H}^{(n)} = nR$) in an n -letter codeword which is sent to Bob with n uses of the noisy channel N . Bob applies a decoding map to the n -letter signal he receives obtaining ρ , which has fidelity with the input state $\langle \psi | \rho | \psi \rangle \geq 1 - \epsilon$.

We say rate R is achievable if there is a sequence of codes with rate at least R such that $\epsilon \rightarrow 0$ as $n \rightarrow \infty$. The quantum capacity $Q(N)$ is the supremum of achievable rates.

There is a regularized formula for $Q(N)$.

The channel $N^{A \rightarrow B}$ can be realized by an isometry $N^{A \rightarrow BE}$ where E is an environment.

For any density operator ρ_A on A consider its purification ψ_{RA} , where R is a reference system.

This is mapped by the channel to a pure state ϕ_{RBE} of reference system R , environment E , and Bob's system B .

We define the "one-shot" quantum capacity as

$$Q_1(N) = \max_{\rho_A} (-H(R|B))$$

The quantity $-H(R|B) = H(B) - H(RB)$
 can also be expressed as $-H(R|B) = H(B) - H(E)$
 since the state ϕ_{RBE} is pure. It is important enough
 to have its own name:

$$I_c(R \rangle B) = -H(R|B) = H(B) - H(E)$$

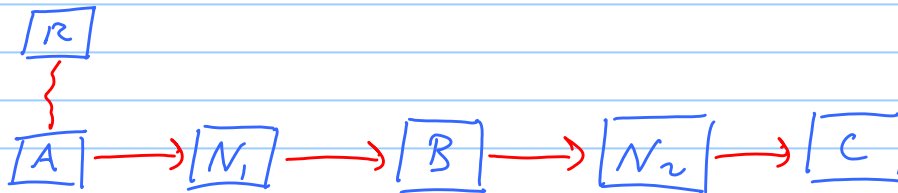
is the "coherent information" from R to B . For
 a classical channel $H(R|B)$ is always nonnegative
 and coherent info is never positive. In the quantum
 setting $I_c(R \rangle B) > 0$ means that the ref.
 system R has a stronger quantum correlation with
 B than with the environment E .

The quantum capacity is the regularized
 quantity

$$Q(N) = \lim_{n \rightarrow \infty} \max_{A^n} \frac{1}{n} I_c(R^n \rangle B^n)$$

Here, too, unfortunately, coherent info is superadditive,
 and we don't know how to express $Q(N)$ as a
 single-letter formula.

We can understand, though, why coherent info
 provides an upper bound on quantum capacity



Consider composing two channels N_1 and N_2 .
 Monotonicity of mutual information implies

$$I(R; A) \geq I(R; B) \geq I(R; C)$$

$$\text{or } H(R) - H(R|A) \geq H(R) - H(R|B) \geq H(R) - H(R|C).$$

since $H(R)$ is unchanged by the channels, this becomes

$$I_c(R \rangle A) \geq I_c(R \rangle B) \geq I_c(R \rangle C)$$

This "quantum data-processing inequality" identifies coherent info as a quantity that cannot be increased by a quantum channel.

But now suppose that the first channel is the noisy channel $\mathcal{N}^{A \rightarrow B}$ while the second channel is Bob's decoding map $\mathcal{D}^{B \rightarrow C}$. Suppose that

ρ_A is maximally mixed on Alice's codespace, so that

$$H(\tilde{A}^n) = H(\tilde{R}^n) = nR \text{ where } R \text{ is the code rate}$$

$$\text{and } I(\tilde{R}^n | \tilde{A}^n) = H(\tilde{A}^n) - H(\tilde{A}^n \tilde{R}^n) = H(\tilde{A}^n) = H(\tilde{R}^n)$$

(since the state of $\tilde{A}^n \tilde{R}^n$ is pure). If Bob's recovery is perfect then the state of

$\tilde{C}^n \tilde{R}^n$ is also maximally entangled, and

$$I_c(\tilde{R}^n \rangle \tilde{C}^n) = H(\tilde{R}^n) = I_c(\tilde{R}^n \rangle \tilde{A}^n)$$

But therefore the data processing inequality implies

$$I_c(\tilde{R}^n \rangle \tilde{B}^n) = I_c(\tilde{R}^n \rangle \tilde{A}^n) = H(\tilde{R}^n)$$

Since the coherent information is

$$I_c(\tilde{R}^n \rangle \tilde{B}^n) = H(\tilde{B}^n) - H(\tilde{E}^n)$$

and $H(\tilde{B}^n) = H(\tilde{R}^n \tilde{E}^n)$ because the state of $\tilde{R}^n \tilde{B}^n \tilde{E}^n$ is pure, this condition becomes

$$I_c(|\tilde{R}^n\rangle|\tilde{B}^n\rangle) = H(\tilde{R}^n) \Rightarrow$$

$$H(\tilde{R}^n \tilde{E}^n) = H(\tilde{R}^n) + H(\tilde{E}^n)$$

- correctability means that there is no correlation between R and E (the environment "knows nothing" about which codeword was sent).

Since perfect decoding fidelity means

$$R = \frac{1}{n} I_c(|\tilde{R}^n\rangle|\tilde{B}^n\rangle)$$

we have the bound on achievable rates (and hence capacity)

$$Q(N) \leq \lim_{n \rightarrow \infty} \max_{A^n} \frac{1}{n} I_c(|R^n\rangle|B^n\rangle)$$

as we claimed.

Now we would like to argue that $I_c(R|B) = H(B) - H(E)$ is an achievable rate, by using "random quantum codes." We need to explain

- ① What is a "random quantum code?"
- ② That $H(RE) \approx H(R) + H(E)$ is sufficient as well as necessary for (approximate) correctability
- ③ That $H(RE) \approx H(R) + H(E)$ is true for random codes with rate less than coherent information (except for negligible corrections).