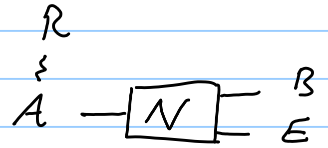
Ph/CS 219
Quantum Channel Capacities
1 June 2009

Note Title 5/22/2009

Last time we introduced the concept of "coherent information" and noted its relevance to sending quantum information through a noisy quantum channel. A channel $\mathcal{N}^{A\to B}$ has a dilation, the <u>isometry</u> $\mathcal{N}^{A\to BE}$

Suppose that input density operator ρ_A is purified by reference system R. Sending A through the channel prepares the tripartite pure state ϕ^{RBE}



The coherent information from R to B for channel $\mathcal{N}$ and input ρ_A is

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

it does not depend on the choice of purification, since a unitary on R does not change $H(B)$ or $H(E)$. It can also be expressed as

$$I_c(R\rangle B) = \frac{1}{2}\left[I(R;B) - I(R;E)\right],$$

since
$$I(R;B) = H(R) + H(B) - H(RB) = H(R) + H(B) - H(E)$$
$$I(R;E) = H(R) + H(E) - H(RE) = H(R) + H(E) - H(B),$$

and hence it quantifies how much stronger the correlation of R is with B than with E.

If the signal transmitted through the channel can be perfectly corrected, then Bob can apply a decoding map with dilation $\mathcal{D}^{B\to \hat{B}B'}$ such that

$$\phi^{RA} \xrightarrow{\mathcal{N}} \phi^{RBE} \xrightarrow{\mathcal{D}} \phi^{R\hat{B}} \otimes \psi^{B'E}$$

We argued that Bob can decode perfectly only if
$$H(R) = I_c(R\rangle B) \quad \text{or} \quad H(RE) = H(R) + H(E)$$

That is, for perfect correctability we

require that the state of RE is a product state — R and E are uncorrelated, or "decoupled."

By considering n uses of the channel, and choosing R to purify the maximally mixed state on the code space, we concluded that the regularized coherent information is an upper bound on the achievable rate for high-fidelity quantum communication:

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

Conversely, if R is maximally entangled with the code space, decoupling of RE suffices to ensure that any state in the code space can be perfectly decoded. If ϕ^{RBE} is the purification of RE density operator $\sigma^{RE} = \sigma^R \otimes \sigma^E$, then we can split B into two subsystems $B = \hat{B}B'$ such that $\hat{B}$ purifies σ^R and B' purifies σ^E; i.e.

$$\phi^{RBE} = \Phi^{R\hat{B}} \otimes \psi^{B'E}$$ where $\Phi^{R\hat{B}}$ is max. entangled,

and therefore Bob can construct a decoding map $\mathcal{D}^{B\to\hat{B}E'}$ that extracts Alice's logical state in the subsystem $\hat{B}$.

Furthermore, approximate decoupling of RE suffices for approximate correctability.

Recall that the fidelity of density operators is defined as

$$F(\rho,\sigma) = \left(\mathrm{tr}\sqrt{\rho^{\frac{1}{2}}\sigma\rho^{1/2}}\right)^2 = \|\sqrt{\rho}\sqrt{\sigma}\|_1^2$$

and is related to L^1 distance between ρ and σ by

$$F(\rho,\sigma) \geq 1 - \|\rho-\sigma\|_1 \quad \text{(see Appendix B)}$$

Also, if $|\psi_\rho\rangle$ is a purification of ρ, then

$$F(\rho,\sigma) = \max |\langle \psi_\sigma | \psi_\rho \rangle|^2 \quad (\text{"Uhlmann's Theorem"})$$

(where the max is over all possible purifications of σ). So, suppose ρ^{RE} is close to a product state:

$$\| \sigma^{RE} - \sigma^R_{max} \otimes \sigma^E \|_1 \le \varepsilon$$

(where σ^R_{max} is the maximally mixed state on R). Then σ^{RE} has a purification that has large overlap with the purification of $\sigma^R_{max} \otimes \sigma^E$:

$$|\langle \tilde{\phi}^{RBE} | \phi^{RBE} \rangle|^2 \geq 1 - \varepsilon$$

where $|\phi^{RBE}\rangle$ is the purification of σ^{RE} and $|\tilde{\phi}\rangle^{RBE} = |\Phi\rangle^{R\hat{B}} \otimes |\psi\rangle^{B'E}$ is the purification of $\sigma^R_{max} \otimes \sigma^E$

When we trace out a subsystem, fidelity is monotonic (states cannot become easier to distinguish), so applying the decoding map $\mathcal{D}^{B\to\hat{B}}$ to $|\phi^{RBE}\rangle$ yields $F(\Phi^{R\hat{B}}, \mathcal{D}^{B\to\hat{B}}(\sigma^{RB})) \geq 1-\varepsilon$

In other words, after decoding the density operator of R and Bob's decoded subsystem $\hat{B}$ is $\sigma^{R\hat{B}}$ where

$$\langle \Phi^{R\hat{B}} | \sigma^{R\hat{B}} | \Phi^{R\hat{B}} \rangle \geq 1 - \| \sigma^{RE} - \sigma^R_{max} \otimes \sigma^E \|_1 .$$

We conclude: approx. decoupling implies approx. correctability.

Aside: Proof of Uhlmann's Th$^{\underline{m}}$

Purification of ρ can be expressed as

$$\sum_a \sqrt{\lambda_a}\, |e_a\rangle \otimes |f_a\rangle = (\rho^{\frac{1}{2}} \otimes I)|\tilde{\Phi}\rangle \text{ where } |\tilde{\Phi}\rangle = \sum_a |e_a\rangle \otimes |f_a\rangle$$

and $\rho = \sum_a \lambda_a |e_a\rangle\langle e_a|$, and an arbitrary purification of σ is

$|\psi_\sigma\rangle = \sum_i \sqrt{\eta_i}\,|g_i\rangle\otimes|h_i\rangle = (\sigma^{\frac{1}{2}}\otimes I)|\tilde{\psi}\rangle$ (where $|\tilde{\psi}\rangle = \sum_i |g_i\rangle\otimes|h_i\rangle$)

$$= (\sigma^{\frac{1}{2}}\otimes I)(V\otimes W^T)|\tilde{\Phi}\rangle = \sigma^{\frac{1}{2}}(VW\otimes I)|\tilde{\Phi}\rangle$$

where $\sigma = \sum_i \eta_i\,|g_i\rangle\langle g_i|$ and V, W are unitary.

Thus $\langle\psi_\sigma|\psi_\rho\rangle = \langle\tilde{\Phi}|(U^\dagger\otimes I)\,\sigma^{\frac{1}{2}}\rho^{\frac{1}{2}}|\tilde{\Phi}\rangle$

(where $U = VW$) $= \mathrm{tr}\left(U^\dagger\sigma^{\frac{1}{2}}\rho^{\frac{1}{2}}\right)$

Using the polar decomp $A = U'\sqrt{A^\dagger A}$ applied to $A = \sigma^{\frac{1}{2}}\rho^{\frac{1}{2}}$

this is $\langle\psi_\sigma|\psi_\rho\rangle = \mathrm{tr}\left(U^\dagger U'\sqrt{\rho^{\frac{1}{2}}\sigma\rho^{\frac{1}{2}}}\right)$

whose modulus is maximized by choosing $U = U'$ so that

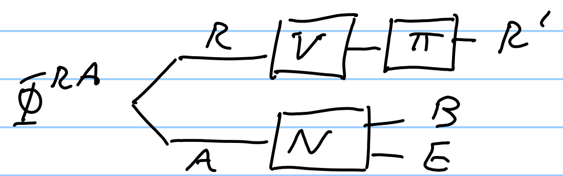
$$\max|\langle\psi_\sigma|\psi_\rho\rangle|^2 = \left(\mathrm{tr}\sqrt{\rho^{\frac{1}{2}}\sigma\rho^{\frac{1}{2}}}\right)^2$$ as claimed.

Monotonicity is a corollary: $F(\rho_{AB},\sigma_{AB}) \le F(\rho_A,\sigma_A)$, because any purifications of ρ_{AB} and σ_{AB} are also purifications of ρ_A and σ_A.

Achievability of Coherent Info

To show that coherent info is an achievable rate, we use a random quantum code. When using the channel n times, chose a random subspace of A^n as input to $(\mathcal{N}^{A\to B})^{\otimes n}$.

That is, consider



Π projects R to a fixed subspace R' and V is a unitary on R, so V determines what subspace is projected.

Φ^{RA} is a maximally entangled state of RA, so R' purifies the maximally mixed state on a code space determined by V.

Now we can average over V. One can show that for any state σ^{RE} on RE, if R' is random subspace of R determined by V, then

$$\left(\int dV \; \| \sigma^{R'E}(V) - \sigma^{R'}_{max} \otimes \sigma^{E} \|_1 \right)^2 \le |R'E| \, tr\left(\sigma^{RE}\right)^2$$

Here V is the normalized unitarily-invariant (Haar) measure on the unitary group acting on R.

In the case where we used the channel n times, the state on B^n is nearly maximally mixed on a typical subspace of dimension $|B^n| \tilde{=} 2^{nH(B)}$, the state on E^n is nearly maximally mixed on a typical subspace of dimension $|E^n| \tilde{=} 2^{nH(E)}$, and the state on R^n is nearly maximally mixed on a typical subspace of dimension $|R^n| \tilde{=} 2^{nH(R)}$. We apply the encoding unitary to this typical subspace R^n before projecting onto R' with $|R'| = 2^{n(\text{Rate})}$,

where "Rate" is the rate of the code in qubits per use of the channel.

Suppressing the small δ in the estimate of the dimension, we estimate $tr\left(\rho^{RE}\right)^n = tr\left(\rho^{B}\right)^n \simeq \dfrac{1}{|B^n|}$

and we conclude that, when we average over codes, the deviation of $R'E$ from a product state is suppressed for

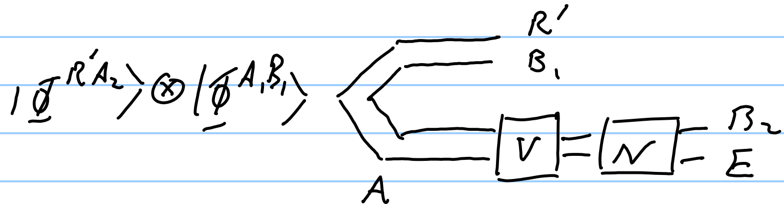
$$\frac{|R'||E^n|}{|B^n|} \approx \frac{2^{n(\text{Rate})} \, 2^{nH(E)}}{2^{nH(B)}} << 1$$

or $\quad$ Rate $< H(B) - H(E) = I_c(R \rangle B)$.

Since decoupling is well satisfied when we average over the choice of the encoding unitary V, then RE decouples well for some particular V (and in fact for a typical V).

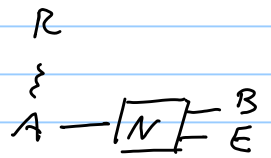
Father Protocol

It is also instructive to estimate the rate for entanglement-assisted quantum communication. Now the sender A and receiver B share a supply of entangled qubits that are consumed during the protocol.



In the iid version of the protocol (many uses of the channel) Alice and Bob share a maximally entangled state $\Phi^{A_1B_1}$ and Alice's input qubits A_2 are maximally entangled with reference system R' (the state Φ^{RA_2}). To encode Alice applies a typical unitary V that acts collectively on the input system A_2 and her half of the entangled qubits. Bob's decoding map can act collectively on his half of the shared entanglement and the output he receives through the (n uses of) the noisy channel. For Bob to be able to decode successfully, it suffices that $R'E$ decouple.

This protocol for entanglement-assisted quantum communication is called the "Father protocol" because it has a variety of interesting "children" that can be derived as consequences.



Recall again that for any input density operator ρ_A, we may consider its purification ϕ^{RA} and the pure state ϕ^{RBE} resulting from sending A through the channel with dilation $N^{A\to BE}$. The Father resource inequality expresses an achievable rate for the quantum communication in the Father protocol, and also the cost in Bell pairs for achieving that rate, in terms of properties of ϕ^{RBE}. Namely

$$\langle \mathcal{N}^{A\to B} : \rho_A \rangle + \frac{1}{2} I(R;E)\,[qq] \geq \frac{1}{2} I(R;B)\,[q\to q]$$

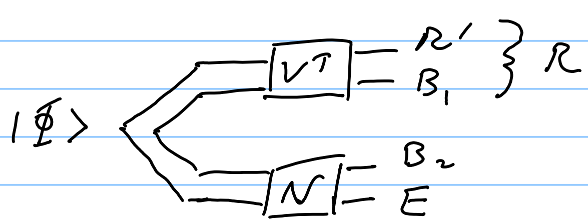
This means that, asymptotically, by using the noisy channel n times, $\frac{n}{2} I(R;B) - o(n)$ qubits can be sent from A to B with high fidelity while consuming $\frac{n}{2} I(R;E) + o(n)$ ebits of entanglement. (Here $o(n)$ means a quantity increasing more slowly than linearly in n.) The entropic quantities depend on the density operator ρ_A, and the resource inequality expresses a task that can be achieved for any ρ_A, so we are free to choose ρ_A that optimizes the rate.

To help you remember the father inequality, note that $I(R,E)$ quantifies something bad — the noise. The higher $I(R;E)$ is, the more entanglement we need to do something useful. On the other hand, $I(R;B)$ quantifies something good — the correlation that survives transmission through the noisy channel. The higher $I(R;B)$ is, the higher the rate of quantum communication. (But the factor ½ you will just need to remember.)

We can relate the Father protocol to an even more primitive task called the "mother protocol". Recalling that

$$(I \otimes V)|\Phi\rangle = (V^T \otimes I)|\Phi\rangle$$

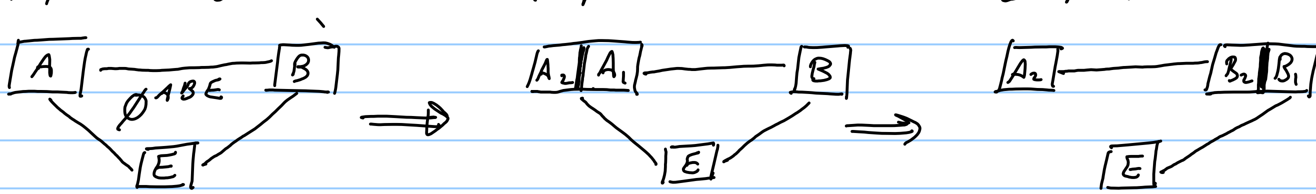
when $|\Phi\rangle$ is maximally entangled, the father transforms into



Now there is a tripartite state ϕ^{RB_2E} where Roy holds R

Roy divides R into subsystems $R = R'B_1$ (where the decomposition depends on V); he keeps R' and passes B_1 to Bob. If $|B_1|$ is large enough, then $R'E$ decouples, which means that the system maximally entangled with R' can be recovered by Bob after decoding. This means that the corresponding father protocol conveys $\log|R'|$ qubits from Alice to Bob while consuming $\log|B_1|$ ebits of entanglement.

Changing Roy's name to Alice, and relabeling the subsystems, the mother protocol can be described this way. Alice, Bob, and Eve share the tripartite pure state ϕ^{ABE}. Alice divides her system into subsystems, $A = A_1A_2$; she keeps A_2 and sends A_1 to Bob. Her goal is to send enough qubits to Bob so that what she holds is no longer correlated with Eve. At that point, the purification of Eve's state is entirely in Bob's hands and Bob also holds the purification of A_2; i.e. Bob's system at the end of the protocol has decomposition $A_1B = B_1B_2$ where B_1 purifies E and B_2 purifies A_2.



In the i.i.d. version of the mother A, B, E share many identical copies $(\phi^{ABE})^{\otimes n}$. Alice Schumacher compresses to a typical subspace of dimension $n(H(A)+o(n))$ and then sends a random subsystem A_1 to Bob. Bob decodes by dividing his system into B_1B_2.

The mother resource inequality expresses how many qubits of quantum communication from A to B suffice to decouple A_2 and E, and how many ebits of entanglement reside in A_2B_2 when the protocol ends:

$$\langle \phi^{ABE} \rangle + \frac{1}{2} I(A;E)\,[q\to q] \geq \frac{1}{2} I(A;B)\,[qq] + \langle \phi'^{B_1E} \rangle$$

That is, $\frac{n}{2}\left[I(A;E) + o(1)\right]$ qubits of communication decouple A_2 and E; meanwhile A and B harvest $\frac{n}{2}\left[I(A;B) - o(1)\right]$ ebits of entanglement. This mother protocol is "dual" to the father protocol – now quantum communication is consumed and quantum entanglement is achieved, rather than the other way around. $I(A;E)$ quantifies the noise in the entanglement that A+B share at the beginning of the protocol, and $I(A;B)$ quantifies the correlation between A+B at the beginning.

The mother can be viewed as a generalization of the entanglement concentration protocol discussed earlier, extended in 3 ways:

1. The initial state shared by A+B can be mixed rather than pure.
2. The communication from A to B is quantum rather than classical.
3. We quantify the amount of communication required.

In addition, as we have seen, the mother resource inequality implies the father, if we think of the communication from Alice to Bob in the mother as the offloading of part of R from Roy to Bob in the father, so that the amount of quantum communication in the mother is the quantum entanglement consumed by the father. Noting that

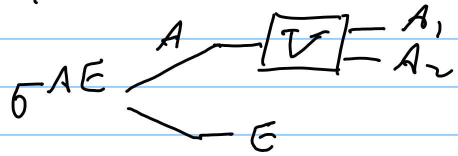
$$H(R) = \frac{1}{2} I(R;B) + \frac{1}{2} I(R;E),$$

we see that if Roy sends $\frac{n}{2} I(R;E) + o(n)$ qubits to Bob, he retains a reference system R'

with $\frac{n}{2} I(R;B) - o(n)$ qubits, which becomes the number of qubits in the code used in the father protocol, while the $\frac{n}{2} I(R;E) + o(n)$ qubits sent by Roy in the mother becomes the number of ebits consumed in the father.

Achievable rate in mother protocol

Consider an arbitrary mixed state σ^{AE} of AE. Consider a fixed decomposition into subsystems $A = A_1 A_2$. Apply a unitary V to A before discarding A_1 to obtain marginal state $\sigma^{A_2 E}(V)$.



The "decoupling inequality" expresses how close $A_2 E$ is to a product state when we average V over unitaries acting on A with respect to Haar measure:

$$\left(\int dV \, \| \sigma^{A_2 E}(V) - \sigma_{max}^{A_2} \otimes \sigma^{E} \|_1 \right)^2 \leq \frac{|A| \cdot |E|}{|A_1|^2} \, \mathrm{tr} \left(\sigma^{AE} \right)^2$$

(where $\sigma_{max}^{A_2}$ is the maximally entangled state on A_2).

This generalizes the result found in a homework exercise, which concerned the case where E is trivial and σ^A is pure; there you derived:

$$\left(\int dV \, \| \sigma^{A_2}(V) - \sigma_{max}^{A_2} \|_1 \right)^2 \leq \frac{|A_2|}{|A_1|} = \frac{|A|}{|A_1|^2} .$$

($\sigma^{A_2}(V)$ nearly maximally mixed for $|A_2| \ll |A_1|$.)

In the i.i.d. version of the mother A becomes the typical subspace of A^n, E the typical subspace of E^n, AE the typical subspace of $(AE)^n$. Since AE is nearly maximally mixed on space of dim $\simeq 2^{nH(AE)}$ we have $\mathrm{tr}(\sigma^{AE})^2 \simeq 2^{-nH(AE)}$. Therefore, when we average over V, the state on A_2E is nearly a product state provided

$$\frac{1}{|A_1|^2} 2^{nH(A)} 2^{nH(E)} 2^{-nH(AE)} << 1$$

or $\quad |A_1|^2 >> 2^{nI(A;E)}$.

There is also an $o(1)$ contribution to the RHS of the decoupling inequality from portion of $(\phi^{ABE})^{\otimes n}$ that lies outside typical subspace, which vanishes in limit $n\to\infty$.

It suffices then for Alice to send

$$\log|A_1| = \frac{n}{2} I(A;E) + o(n)$$

qubits to Bob. And since

$$H(A) = \tfrac{1}{2}[I(A;E) + I(A;B)] \qquad \text{(because } \phi^{ABE} \text{ is pure)}$$

Alice retains

$$\log|A_2| = \frac{n}{2} I(A;B) - o(n)$$

qubits. Since these are nearly maximally mixed and uncorrelated with E, Alice's retained qubits are nearly maximally entangled with a subsystem of Bob's qubits; Alice and Bob share $\frac{n}{2}I(A;B) - o(n)$ ebits. This proves the mother resource inequality. (It works when we average over V, and therefore for some particular V — in fact for typical V.)

The proof of the decoupling inequality is in Appendix A. Note that a simple heuristic dimension counting argument shows that it is plausible, at least in the i.i.d. case that is relevant for the asymptotic achievability result. Suppose that the state on AE is maximally mixed on a subspace of dim $|B|$, i.e., a uniform mixture of $|B|$ mutually orthogonal pure states. Then we trace out A_1. But for $|A_1| << |A_2E|$, we expect that each of the $|B|$ states in the ensemble realizing ρ^{AE} is likely to be nearly maximally mixed on A_1; thus for each of these $|B|$ states, tracing out A_1 generates a density operator on A_2E which is a nearly uniform mixture of $|A_1|$ mutually orthogonal states. Furthermore, as long as $|A_1B| \ll |A_2E|$, all of the $|A_1B|$ states are likely to be nearly mutually orthogonal — tracing out A_1 produces a nearly uniform density operator with rank $\approx |A_1B|$. Once $|A_1|$ is large enough though, the rank $|A_1B|$ matches the dimension of A_2E, so that the state on A_2E is maximally mixed and in particular is a product state. This occurs for

$$|A_1|\cdot|B| \approx |A_2E| = \frac{|AE|}{|A_1|}$$

or $\quad |A_1|^2 \approx \dfrac{|AE|}{|B|}$, reproducing the conclusion we inferred from the decoupling inequality.

Children of the Father

We can derive a further consequence by combining the father resource inequality

$$\text{Father: } \langle \mathcal{N}^{A\to B} : \rho_A \rangle + \tfrac{1}{2} I(R;E)\,[qq] \geq \tfrac{1}{2} I(R;B)\,[q\to q]$$

with the superdense coding inequality

$$\text{SD:} \quad [q\to q] + [qq] \geq 2\,[c\to c]$$

(we use one qubit of quantum comm. and one ebit to achieve 2 bits of classical comm.)

Suppose we use the $\frac{1}{2} I(R;B)$ qubits of $[q\to q]$ and an additional $\frac{1}{2} I(R;B)$ ebits to achieve $I(R;B)$ bits of $[c\to c]$. Because

$$\tfrac{1}{2} I(R;E) + \tfrac{1}{2} I(R;B) = H(R),$$

we conclude

$$\langle N^{A\to B} : \rho_A \rangle + H(R)\,[qq] \geq I(R;B)\,[c \to c],$$

which establishes an achievable rate for entanglement-assisted classical communication.

We may define $C_E(N)$ as the supremum of achievable rates per use of the channel for sending classical info reliably over the noisy quantum channel, if entanglement can be consumed at zero cost. This "entanglement-assisted classical capacity" of the quantum channel thus satisfies

$$C_E(N) \geq \max_{\rho_A} I(R;B)$$

In this case, there is a matching upper bound, and thus the inequality is actually an equality. In this case, therefore, we have a single-letter formula and the cost of the task is fully understood. Furthermore, the resource inequality tells us how much entanglement consumption suffices to attain the capacity.

We can derive another consequence of the father by using some of the quantum communication generated by the father to repay the entanglement that was borrowed to activate (i.e. catalyze) the father protocol.

$$[q \to q] \geq [qq] \Rightarrow \tfrac{1}{2} I(R;E)\,[q \to q] \geq \tfrac{1}{2} I(R;E)\,[qq].$$

After replacing the entanglement consumed, the net amount of quantum communication achieved per use of the channel is

$$\tfrac{1}{2} I(R;B) - \tfrac{1}{2} I(R;E) = H(B) - H(E) = I_c(R \rangle B).$$

We have derived the achievability result

$$\langle \mathcal{N}^{A\to B} : \rho_A \rangle \geq I_c(R\rangle B)\,[q\to q],$$

at least in this catalyzed setting, and the same rate can also be achieved without any initial supply of entanglement. Together with the upper bound derived in the homework, we obtain a regularized formula for quantum capacity:

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

Unfortunately, though, since the coherent information can be superadditive, we don't know how to reduce this expression to a single-letter formula for the quantum capacity.

Children of the mother

We obtain a useful consequence of the mother resource inequality

$$\text{Mother: } \langle \phi^{ABE} \rangle + \tfrac{1}{2} I(A;E)\,[q\to q] \geq \tfrac{1}{2} I(A;B)\,[qq] + \langle \phi'^{B_1 E} \rangle$$

by combining with the teleportation resource inequality

$$\text{TP: } [qq] + 2\,[c\to c] \geq [q\to q]$$

(one qubit can be transmitted by consuming one ebit and sending two bits.)

We can replace the quantum communication in the mother by classical communication if we use $\frac{1}{2} I(A;E)$ ebits generated by the mother, together with $I(A;E)$ bits of classical communication to replace the quantum communication consumed by the mother. Then the net amount of entanglement generated is $\frac{1}{2}I(A;B) - \frac{1}{2}I(A;E) = I_c[A>B]$,

and we obtain the resource inequality

$$\langle \phi^{ABE} \rangle + I(A;E)[c \to c] \geq I_c(A>B)[qq] + \langle \phi'^{B_1 E} \rangle,$$

which is called the "Hashing inequality." It quantifies an achievable rate for distilling maximal entanglement from a state shared by A and B using one-way classical communication from A to B. Furthermore, the Hashing inequality tells us how much classical communication suffices.

In the case where the state on AB is pure, $I_c(A>B) = H(A) - H(AB) = H(A)$, and we recover our earlier conclusion concerning entanglement concentration for pure states: $\langle \phi^{AB} \rangle \geq H(A)[qq]$,

$H(A)$ ebits can be extracted asymptotically from n copies of ϕ^{AB}. In this case the resource inequality says that the sufficient amount of $[c \to c]$ is $I(A;E) = 0$ — no classical communication is required. But this result is a bit misleading; some classical communication is needed, but only $O(\sqrt{n})$ bits, or $O(n^{-\frac{1}{2}})$ per copy, which becomes negligible asymptotically.

State Merging

The state-merging resource inequality answers the question: how much quantum communication is needed from A to B to transfer the purification of E's state shared by AB to a state held solely by B, assuming classical communication from A to B has zero cost. To derive state merging from the mother, we use all of the entanglement generated by the mother to teleport additional qubits from A to B. Adding

TP: $\frac{1}{2}I(A;B)[qq] + I(A;B)[c\to c] \geq \frac{1}{2}I(A;B)[q\to q]$

to the mother inequality, and noting that the net amount of quantum communication consumed is

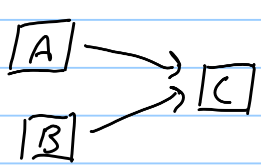
$$\frac{1}{2}I(A;E) - \frac{1}{2}I(A;B) = H(E) - H(B) = H(AB) - H(B) = H(A|B),$$

we obtain

State Merging: $\langle \phi^{ABE}\rangle + H(A|B)[q\to q] + I(A;B)[c\to c] \geq \langle \phi'^{B_1E}\rangle.$

state-merging is achieved with an amount of quantum communication given by the conditional entropy $H(A|B)$.

What is the classical version of state merging? If Alice and Bob have correlated classical bits, how many bits does Alice need to send to Bob so that Bob knows what Alice had? The answer is the conditional entropy $H(X|Y)$, which is achieved by what information theorists call "Slepian-Wolf coding". Alice sorts her messages into $2^{n(H(X|Y)+\delta)}$ bins and sends only the label of the bin. With high probability, Bob finds that only one message in that bin is jointly typical with his information.



Similarly, if A and B both send to C: Bob compresses the info from his source to $nH(Y) + o(n)$ letters. Then Alice need send only $nH(X|Y) + o(n)$ letters to C. Together, AB compress their shared information source to $nH(XY)$ letters, the same compression they would have been able to achieve if they were sending from the same location instead of two different locations. Therefore

Slepian-Wolf coding gives a precise operational interpretation to the informal statement that $H(X|Y)$ quantifies Bob's remaining ignorance about XY when he already knows Y.

In the same sense, state merging gives such an operational meaning to conditional entropy in the quantum setting: $H(A|B)$ is the number of qubits Bob needs to receive from Alice in order to possess the purification of system E (if classical communication is for free). The conditional entropy quantifies Bob's "ignorance" about this jointly held purification.

Classically, $H(X|Y)$ is nonnegative, and it is zero if Bob is already certain about XY. But quantumly, $H(A|B)$ can be negative. How can Bob have "negative uncertainty" about AB? If $H(A|B) < 0$ (equivalently $I(A;B) > I(A;E)$), then the mother produces more entanglement than the amount of quantum communication it consumes. In that case, the state merging inequality becomes the hashing inequality

Hashing: $\langle \phi^{ABE} \rangle + I(A;E)[c\to c] \geq -H(A|B)[qq] + \langle \phi'^{B_1E} \rangle$

Now the state merging has no quantum cost, and AB hold $-H(A|B)$ ebits at the end of the protocol. This shared $[qq]$ they have deposited in the bank can be used for teleportation in future rounds of state merging, reducing the quantum communication cost. The "negative uncertainty" Bob has today can reduce his uncertainty in the quantum communication tasks he will need to perform tomorrow.

Operational meaning of strong subadditivity

The observation that $H(A|B)$ is the quantum communication cost of state merging provides a simple "operational proof" of the strong subadditivity of quantum mutual information. SSA says

$$I(A;BC) = H(A) - H(A|BC) \geq I(A;B) = H(A) - H(A|B)$$

or equivalently: $$H(A|BC) \leq H(A|B)$$

If $H(A|B)$ is positive, this is the obvious statement that it is no harder to merge A with Bob's system if Bob holds C as well as B.

If $H(A|B)$ is negative, this is the obvious statement that Alice and Bob can distill no less entanglement with one-way classical communication if Bob holds C as well as B.

Appendix A: The decoupling inequality

We want to show

$$\left(\int dU \, \| \sigma^{A_2 E}(U) - \sigma_{max}^{A_2} \otimes \sigma^{E} \|_1 \right)^2 \leq \frac{|AE|}{|A_1|^2} \, tr\left(\sigma^{AE}\right)^2$$

where U acts on $A = A_1 A_2$.

We note that $\| \sigma^{A_2 E} - \sigma_{max}^{A_2} \otimes \sigma^{E} \|_2^2$

$$= tr\left(\sigma^{A_2 E}\right)^2 - \frac{1}{|A_2|} tr\left(\sigma^{E}\right)^2$$

(because $tr\left(\sigma_{max}^{A_2}\right)^2 = 1/|A_2|$).

Now evaluate

$$\int dU \; tr\left(\sigma^{A_2 E}(U)\right)^2$$

$$= \int dU \; tr_{A_1}\left(U \sigma^{AE} U^\dagger\right) \otimes tr_{A_1'}\left(U \sigma^{A'E'} U^\dagger\right) S^{A_2 A_2'} \otimes S^{EE'}$$

where $S^{AA'}$ denotes the swap operator on AA'.

Therefore,

$$\int dU \; tr\left(\sigma^{A_2 E}(U)\right)^2$$

$$= tr\left[\sigma^{AE} \otimes \sigma^{A'E'} \left(\int dU \left(U^\dagger \otimes U^\dagger\right) I^{A_1 A_1'} \otimes S^{A_2 A_2'} \left(U \otimes U\right)\right) \otimes S^{EE'}\right]$$

By the Lemma below, the integral is

$$\int dU \left(U^\dagger \otimes U^\dagger\right) I^{A_1 A_1'} \otimes S^{A_2 A_2'} \left(U \otimes U\right)$$

$$= c_I \, I^{AA'} + c_S \, S^{AA'} \quad \text{where}$$

$$c_I = \frac{1}{|A_2|}\left(\frac{1 - 1/|A_1|^2}{1 - 1/|A|^2}\right) \leq \frac{1}{|A_2|}, \quad c_S = \frac{1}{|A_1|}\left(\frac{1 - 1/|A_2|^2}{1 - 1/|A|^2}\right) \leq \frac{1}{|A_1|}.$$

Plugging the value of the integral into the trace:

$$\int dU \, \mathrm{tr}\left(\sigma^{A_2E}(U)\right)^2 \leq \frac{1}{|A_2|} \mathrm{tr}\left(\sigma^{E}\right)^2 + \frac{1}{|A_1|} \mathrm{tr}\left(\sigma^{AE}\right)^2$$

and we conclude

$$\int dU \, \| \sigma^{A_2E}(U) - \sigma^{A_2}_{max} \otimes \sigma^{E} \|_2^2 \leq \frac{1}{|A_1|} \mathrm{tr}\left(\sigma^{AE}\right)^2.$$

From the Cauchy-Schwarz inequality

$$\| M \|_1^2 \leq d \, \| M \|_2^2 \quad \text{and} \quad \langle \sqrt{f} \rangle^2 \leq \langle f \rangle,$$

we find

$$\left(\int dU \, \| \sigma^{A_2E}(U) - \sigma^{A_2}_{max} \otimes \sigma^{E} \|_1 \right)^2 \leq \frac{|A_2 E|}{|A_1|} \mathrm{tr}\left(\sigma^{AE}\right)^2$$

— this is the decoupling inequality.

It remains to prove:

Lemma: $\int dU \, (U^\dagger \otimes U^\dagger) \, I^{A_1A_1'} \otimes S^{A_2A_2'} \, (U \otimes U)$

$$= C_I \, I^{AA'} + C_S \, S^{AA'}$$

Proof: The integral commutes with $V \otimes V$, and therefore by Schur's Lemma is a weighted sum of projectors onto irreducible representations. The irreps are the symmetric and antisymmetric tensors, so that

$$\int dU \, (U^\dagger \otimes U^\dagger) \, I^{A_1A_1'} \otimes S^{A_2A_2'} \, (U \otimes U) = C_{sym} \, \Pi_{sym}^{AA'} + C_{anti} \, \Pi_{anti}^{AA'}$$

where $\Pi_{sym}^{AA'}$ projects onto the subspace symmetric under $A \leftrightarrow A'$ and $\Pi_{anti}^{AA'}$ projects onto the antisymmetric subspace. To compute c_{sym}, evaluate $\mathrm{tr}(\Pi_{sym}^{AA'} \, \cdot)$ of both sides. Using $\Pi_{sym}^{AA'} = \frac{1}{2}\left(I^{AA'} + S^{AA'}\right)$, we obtain

$$\frac{1}{2}\,\mathrm{tr}\left(I^{A_1A_1'}\otimes S^{A_2A_2'}\right)\left(I^{A_1A_1'}\otimes I^{A_2A_2'}+S^{A_1A_1'}\otimes S^{A_2A_2'}\right)$$

$$=\frac{1}{2}\left[\mathrm{tr}\left(I^{A_1A_1'}\otimes S^{A_2A_2'}\right)+\mathrm{tr}\left(S^{A_1A_1'}\otimes I^{A_2A_2'}\right)\right]$$

$$=\frac{1}{2}\left(|A_1|^2|A_2|+|A_1||A_2|^2\right)=C_{sym}\,\mathrm{tr}\,\Pi_{sym}^{AA'}=$$

$$=C_{sym}\frac{1}{2}|A|(|A|+1)$$

$$\Rightarrow\quad C_{sym}=\frac{|A_1|+|A_2|}{|A|+1}$$

(Here we used $\mathrm{tr}\,S^{AA'}=\mathrm{tr}\left(2\Pi_{sym}^{AA'}-I^{AA'}\right)=|A|(|A|+1)-|A|^2=|A|.$)

Similarly, $\Pi_{anti}^{AA'}=\frac{1}{2}\left(I^{AA'}-S^{AA'}\right)$ and $\mathrm{tr}\,\Pi_{anti}^{AA'}=\frac{1}{2}|A|(|A|-1)$

$$\Rightarrow\quad C_{anti}=\frac{|A_1|-|A_2|}{|A|-1}.$$

Then noting that
$$C_I=\frac{1}{2}\left(C_{sym}+C_{anti}\right)$$
$$C_S=\frac{1}{2}\left(C_{sym}-C_{anti}\right)$$

we obtain
$$C_I=\frac{|A|\cdot|A_1|-|A_2|}{|A|^2-1}=\frac{1}{|A_2|}\,\frac{|A_2|\left(|A_1|^2-1\right)}{|A_2|\left(|A_1|^2-1/|A_2|^2\right)},$$
$$C_S=\frac{|A|\cdot|A_2|-|A_1|}{|A|^2-1}=\frac{1}{|A_1|}\,\frac{|A_1|\left(|A_2|^2-1\right)}{|A_1|\left(|A_2|^2-1/|A_1|^2\right)},$$

which proves the lemma.

Appendix B: Fidelity and L^1 distance

We wish to show:

$$\sqrt{F(\rho,\sigma)} \equiv \| \sqrt{\rho}\sqrt{\sigma} \|_1 = \mathrm{tr}\sqrt{\rho^{\frac{1}{2}}\sigma\rho^{\frac{1}{2}}} \geq 1 - \tfrac{1}{2}\|\rho-\sigma\|_1 .$$

From the polar decomposition of M we obtain

$$\mathrm{tr}\sqrt{M^\dagger M} \geq \mathrm{tr}\, M \Rightarrow \sqrt{F(\rho,\sigma)} \geq \mathrm{tr}(\sqrt{\rho}\sqrt{\sigma}) .$$

And

$$\| \sqrt{\rho} - \sqrt{\sigma} \|_2^2 = \mathrm{tr}(\sqrt{\rho}-\sqrt{\sigma})^2 = 2 - 2\,\mathrm{Tr}(\sqrt{\rho}\sqrt{\sigma}) \geq 2 - 2\sqrt{F(\rho,\sigma)}$$

$$\Rightarrow \quad \sqrt{F(\rho,\sigma)} \geq 1 - \tfrac{1}{2}\|\sqrt{\rho}-\sqrt{\sigma}\|_2^2$$

Therefore, it suffices to show $\|\rho-\sigma\|_1 \geq \|\sqrt{\rho}-\sqrt{\sigma}\|_2^2$.

Note that $\rho - \sigma = \tfrac{1}{2}(\sqrt{\rho}-\sqrt{\sigma})(\sqrt{\rho}+\sqrt{\sigma}) + \tfrac{1}{2}(\sqrt{\rho}+\sqrt{\sigma})(\sqrt{\rho}-\sqrt{\sigma})$,

and we may write $\sqrt{\rho} - \sqrt{\sigma} = \sum_i \lambda_i |i\rangle\langle i| \Rightarrow$

$$|\sqrt{\rho}-\sqrt{\sigma}| = \sum_i |\lambda_i|\, |i\rangle\langle i| = U(\sqrt{\rho}-\sqrt{\sigma}) = (\sqrt{\rho}-\sqrt{\sigma})U$$

where $\{|i\rangle\}$ is the ON basis that diagonalizes $\sqrt{\rho}-\sqrt{\sigma}$

and U is the unitary transformation $U = \sum_i \mathrm{sign}(\lambda_i)\, |i\rangle\langle i|$.

Now, $\mathrm{tr}\,|\rho-\sigma| \geq \mathrm{tr}(\rho-\sigma)U$ (true for any unitary U)

$$= \mathrm{tr}\,|\sqrt{\rho}-\sqrt{\sigma}|(\sqrt{\rho}+\sqrt{\sigma}) = \sum_i |\lambda_i| \langle i|\sqrt{\rho}+\sqrt{\sigma}|i\rangle$$

$$\geq \mathrm{tr}\sum_i |\lambda_i| \langle i|\sqrt{\rho}-\sqrt{\sigma}|i\rangle = \sum_i |\lambda_i|^2 = \|\sqrt{\rho}-\sqrt{\sigma}\|_2^2$$

Thus $\|\rho-\sigma\|_1 \geq \|\sqrt{\rho}-\sqrt{\sigma}\|_2^2$, as we wanted to show.

By the way, it is sometimes convenient to have an upper bound on $F(\rho,\sigma)$ expressed in terms of the L^1 distance $\|\rho-\sigma\|_1$; for example,

$$F(\rho,\sigma) \le 1-\frac{1}{4}\|\rho-\sigma\|_1^2 .$$

① First show this is an equality for pure states

$$|\psi\rangle = \begin{pmatrix}\cos\alpha\\ \sin\alpha\end{pmatrix} \qquad |\phi\rangle = \begin{pmatrix}\sin\alpha\\ \cos\alpha\end{pmatrix} \Rightarrow |\langle\phi|\psi\rangle|^2 = \sin^2 2\alpha = F(\phi,\psi)$$

$$|\psi\rangle\langle\psi| - |\phi\rangle\langle\phi| = \begin{pmatrix}\cos 2\alpha & 0\\ 0 & -\cos 2\alpha\end{pmatrix} \Rightarrow$$

$$\| |\psi\rangle\langle\psi| - |\phi\rangle\langle\phi| \|_1^2 = 4\cos^2 2\alpha = 4\left[1-F(\phi,\psi)\right]$$

② Next note that L^1 distance is monotonic:

$$\|\rho_{AB}-\sigma_{AB}\|_1 \ge \|\rho_A-\sigma_A\|_1 .$$

This is true because L_1 distance is optimal distance between prob. distributions for POVM outcomes, and we can perform a POVM on AB that acts nontrivially only on A.

③ Finally, by Uhlmann's theorem,

$$F(\rho_A,\sigma_A) = F(\rho_{AB},\sigma_{AB})$$

where ρ_{AB},σ_{AB} are purifications with maximal fidelity.

$$= 1-\frac{1}{4}\|\rho_{AB}-\sigma_{AB}\|_1^2$$

$$\le 1-\frac{1}{4}\|\rho_A-\sigma_A\|_1^2 ,$$

where the last step uses monotonicity of L^1 distance.