Last time we introduced the concept of "coherent information" and noted its relevance to sending quantum information through a noisy quantum channel. A channel \( \mathcal{N} : A \rightarrow B \) has a dilation, i.e., isometry \( \mathcal{N} : A \rightarrow B \otimes E \).

Suppose that input density operator \( \rho_A \) is purified by reference system \( \mathcal{R} \). Sending \( A \) through the channel prepares the tripartite pure state \( \rho_{\mathcal{R}B E} \).

The coherent information from \( \mathcal{R} \) to \( B \) for channel \( \mathcal{N} \) and input \( \rho_A \) is

\[
I_c(\mathcal{R} \rightarrow B) = -H(\mathcal{R}|B) = H_B - H(E);
\]

it does not depend on the choice of purification, since a unitary on \( \mathcal{R} \) does not change \( H(\mathcal{R}|B) \) or \( H(E) \).

It can also be expressed as

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

since

\[
I(\mathcal{R}, B) = H(\mathcal{R}) + H(B) - H(\mathcal{R}|B) = H(\mathcal{R}) + H(B) - H(E),
\]

\[
I(\mathcal{R}; E) = H(\mathcal{R}) + H(E) - H(\mathcal{R}|E) = H(\mathcal{R}) + H(E) - H(B).
\]

Hence, it quantifies how much stronger the correlation of \( \mathcal{R} \) 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 \( \otimes B \rightarrow \hat{B} \) such that

\[
\rho_{\mathcal{R}B} \xrightarrow{\mathcal{N}} \rho_{\mathcal{R}B E} \xrightarrow{\hat{D}} \rho_{\hat{R}B \otimes \hat{B}' E}
\]

We argued that Bob can decode perfectly only if

\[
H(\mathcal{R}) = I_c(\mathcal{R} \rightarrow B) \quad \text{or} \quad H(E) = H(\mathcal{R}) + H(\mathcal{E})
\]

That is, for perfect correctness, we
require that the state \( \rho_{RE} \) 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:

\[
Q(n) \leq \lim_{n \to \infty} \max_{\mathcal{A}_n} \frac{1}{n} I_c \left( R^n \to B^n \right) .
\]

Conversely, if \( R \) is maximally entangled with the code space, decoupling \( \rho_{RE} \) suffices to ensure that any state in the code space can be perfectly decoded. If \( \rho_{RBE} \) is the purification of \( \rho_{RE} \), satisfying \( \rho_{RBE} = \rho_{R} \otimes \rho_{BE} \), then we can split \( B \) into two subsystems \( B = B' \wedge \hat{B} \) such that \( \hat{B} \) purifies \( \rho_{R} \) and \( B' \) purifies \( \rho_{BE} \), i.e.,

\[
\rho_{RBE} = \rho_{R} \hat{B} \otimes \rho_{BE} \quad \text{where } \rho_{R} \hat{B} \text{ is max. entangled,}
\]

and therefore Bob can construct a decoding map \( \hat{B} \to \hat{B} \) that extracts Alice's bosonic state in the subsystem \( \hat{B} \).

Furthermore, approximate decoupling of \( \rho_{RE} \) suffices for approximate connectability.

Recall that the fidelity of density operators is defined as

\[
F(\rho, \sigma) = \left( \text{tr} \sqrt{\sqrt{\rho} \sigma \sqrt{\rho}} \right)^2 = \| \sqrt{\rho} \sqrt{\sigma} \|_1^2
\]

and is related to \( L^1 \) distance between \( \rho \) and \( \sigma \) by

\[
F(\rho, \sigma) \geq 1 - \| \rho - \sigma \|_1 \quad \text{(See Appendix B)}.
\]
Also, if $|\psi_e\rangle$ is a purification of $\rho$, then

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

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

$$\| \rho^{RE} - \rho^{R} \otimes \rho^{E} \|_1 \leq \varepsilon
$$

(where $\rho^{max}$ is the maximally mixed state on $R$).

Then $\rho^{RE}$ has a purification that has large overlap with the purification of $\rho^{R} \otimes \rho^{E}$:

$$|\langle \rho^{R} \rho^{E} | \hat{\rho}^{R} \hat{\rho}^{E} \rangle|^2 \geq 1 - \varepsilon
$$

where $|\rho^{R} \rho^{E} \rangle$ is the purification of $\rho^{RE}$ and

$$|\langle \rho^{R} \rho^{E} | \hat{\rho}^{R} \hat{\rho}^{E} \rangle|^2 \geq 1 - \| \rho^{RE} - \rho^{R} \otimes \rho^{E} \|_1
$$

when we trace out a subsystem, fidelity is monotonic (states cannot become easier to distinguish), so applying the decoding map $\gamma_{R} : \hat{\rho}^{R} \rightarrow \hat{\rho}$ to $|\rho^{R} \rho^{E} \rangle$ yields

$$F(\rho^{R} \hat{\rho}^{R}, \gamma_{R} \hat{\rho} \rightarrow \hat{\rho} (\rho^{RB})) \geq 1 - \varepsilon
$$

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

$$|\langle \rho^{R} \hat{\rho}^{R} | \hat{\rho}^{R} \hat{\rho}^{R} \rangle|^2 \geq 1 - \| \rho^{RE} - \rho^{R} \otimes \rho^{E} \|_1.
$$

We conclude: approx. decoupling implies approx. correctability.

Aside: Proof of Uhlmann's Thm

Purification of $\rho$ can be expressed as

$$\sum_{a} \sqrt{\lambda_a} |e_a \rangle \otimes |f_a \rangle = (\rho^{\frac{1}{2}} \otimes I) |\hat{\rho}\rangle \quad \text{where} \quad |\hat{\rho}\rangle = \sum_{a} |e_a \rangle \otimes |f_a \rangle
$$

and $\rho = \sum_{a} \lambda_a |e_a \rangle \langle e_a|$, and an arbitrary purification of $\rho$ is
\[ |\psi_0\rangle = \sum_i \sqrt{\eta_i} |g_i\rangle |\psi_i\rangle = (\mathbf{f}^T \otimes I) |\tilde{\psi}\rangle \quad (\text{where } |\tilde{\psi}\rangle = \sum_i |g_i\rangle |\psi_i\rangle) \]

\[ = (\mathbf{f}^T \otimes I) (V \otimes W^T) |\tilde{\psi}\rangle = e^{-i/2} (V \otimes W^T) |\tilde{\psi}\rangle \]

where \( e = \sum_i \eta_i |g_i\rangle \langle g_i| 11 \) and \( V, W \) are unitary.

Thus \[ \langle \psi_0 | \psi_e \rangle = \langle \tilde{\psi} | (U^T \otimes I) e^{-i/2} |\tilde{\psi}\rangle \]

(\text{where } \ U = VW) = \operatorname{Tr} (U^T e^{-i/2})

Using the polar decomposes \( A = U \sqrt{A^+A} \) applied to \( A = \mathbf{f}^T e^{i/2} \),

this is \[ \langle \psi_0 | \psi_e \rangle = \operatorname{Tr} (U^T U \sqrt{e^{i/2} e^{-i/2}}) \]

whose modulus is maximized by choosing \( U = U' \) so that \[ \max | \langle \psi_0 | \psi_e \rangle | = (\operatorname{Tr} e^{i/2} e^{-i/2}) \text{ as claimed.} \]

Monotonicity is a corollary: \( F(\rho_{AB}, \sigma_{AB}) \leq F(\rho_A, \sigma_A) \),

because any purifications of \( \rho_{AB} \) and \( \sigma_{AB} \) are also purifications of \( \rho_A \) and \( \sigma_A \).

**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 subspose of \( A^n \) as input to \( (N \rightarrow B)^n \).

That is, consider

\[ \Phi^{RA} \xrightarrow{R} \Phi^{RA} \]

It projects \( R \) to a fixed subspose \( R' \) and \( V \) is a unitary on \( R \), so \( V \) determines what subspose is projected.

\( \Phi^{RA} \) 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 \( \Phi^{RE} \) on \( RE \), if \( R' \) is random
subspace of $\mathcal{R}$ determined by $V$, then 

\[
(\frac{\| dV \|}{\| R' E(V) - R' \|} \leq \frac{\| R' \|}{\| E \|} \leq \text{tr}(R'E))^2
\]

Here $V$ is the normalized unitarily-invariant (Haar) measure on the unitary group acting on $\mathcal{R}$.

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

Suppressing the small $S$ in the estimate of the dimension, we estimate 

\[
\text{tr}(R'E)^n \approx \text{tr}(E^n) \approx \frac{1}{|\mathcal{B}^n|}
\]

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

\[
|\mathcal{R}'| |\mathcal{E}^n| \approx 2^{n(\text{Rate})} 2^{nH(\mathcal{E})}
\]

\[
|\mathcal{B}^n| \approx 2^{nH(\mathcal{B})} \ll 1
\]

or $\text{Rate} < H(\mathcal{B}) - H(\mathcal{E}) = I_\mathcal{C}(\mathcal{R} \to \mathcal{B})$.

Since decoupling is well satisfied when we average over the choice of the encoding unitary $V$, then $R'E$ 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 $i.i.d.$ version of the protocol, many uses of the channel Alice and Bob share a maximally entangled state $|\Phi_{AB}\rangle$, and Alice's input qubits $A$ are maximally entangled with reference system $R'$ (the state $|\Phi^{R'}_{RA}\rangle$). To encode Alice applies a typical unitary $V$ that acts collectively on the input system $A$ 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 (noisy, 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 $A = (N|E)$ density operator $\rho_A$, we may consider its purification $\psi^{RA}$ and the pure state $\rho^{RBE} = \rho^{RBE}$ resulting from sending $A$ through the channel with assistance $N^{A\rightarrow BE}$.

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 $\rho^{RBE}$. Namely
\[ \langle N^{A \rightarrow B} \mid \rho_A \rangle + \frac{1}{n} I(R; E) \leq \frac{1}{2} I(R; B) + o(n) \]

This means that, asymptotically, by using the noisy channel \( n \) times, \( \frac{1}{n} I(R; B) = o(n) \) qubits can be sent from \( A \) to \( B \) with high fidelity while consuming \( \frac{1}{n} I(R; E) = o(1) \) ebits of entanglement. (Here \( o(1) \) means a quantity increasing more slowly than linearly in \( n \).) The entropic quantities depend on the density operator \( \rho_A \), and the resource inequality expresses a task that can be achieved for any \( \rho_A \), so we are free to choose \( \rho_A \) 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 \( \frac{1}{n} \) you will just need to remember.)

We can relate the father protocol to an even more primitive task called the "mother protocol". Recalling that
\[
(I \otimes V) \mid \Phi^+ \rangle = (V^\dagger \otimes I) \mid \Phi^+ \rangle
\]
when \( \mid \Phi^+ \rangle \) is maximally entangled, the father transforms into
\[ \begin{array}{c}
| V^\dagger \rangle & \quad & B_1 \quad & \quad & R \\
\longrightarrow & & & & \\
| R \rangle & \quad & B_1 \quad & \quad & \emptyset_R B_2 E
\end{array} \]
Now there is a tripartite state, 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 bipartite pure state \( \varnothing^{A_B E} \). Alice divides her system into subsystems, \( A = A_1 A_2 \); 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 \( A_1 B = B_1 B_2 \) where \( B_1 \) purifies \( E \) and \( B_2 \) purifies \( A_2 \).

\[
\begin{array}{ccc}
A & \varnothing^{A_B E} & B \\
\updownarrow & \downarrow & \updownarrow \\
E & A_1 A_2 & B \\
\updownarrow & \downarrow & \updownarrow \\
E & A_2 & B_1 B_2 \\
\end{array}
\]

In the i.i.d. version of the mother \( A, B, E \) share many identical copies \( (\varnothing^{A_B E})^n \). 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 \( B_1 B_2 \).

The mother resource inequality expresses how many qubits of quantum communication from \( A \& B \) suffice to decouple \( A_2 \) and \( E \), and how many ebits of entanglement reside in \( A_1 B_2 \) when the protocol ends.
\[ \langle \Phi^{ABE} \rangle + \frac{1}{2} I(A;E) \{ q \rightarrow q \} > \frac{1}{2} I(A;B) \{ q \} + \langle \Phi^{BIE} \rangle \]

That is, \[ \frac{1}{2} [ I(A;E) + 0 \{ I \} ] \] quantum communication decouples A and E; meanwhile A and B harvest \[ \frac{1}{2} [ I(A;B) - 0 \{ I \} ] \] 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 to 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 \( \frac{1}{2} I(R;E) + O(n) \) qubits to Bob, he retains a reference system \( R' \)
with \( \frac{1}{2} I(R; B) - o(n) \) qubits, which becomes the number of qubits in the code used in the father protocol, while the \( \frac{1}{2} I(R; E) + o(n) \) qubits sent by Ray in the mother becomes the number of qubits consumed in the father.

**Achievable rate in mother protocol**

Consider an arbitrary mixed state \( \rho_{AE} \) of \( AE \). Consider a fixed decomposition in to subsystems \( A = A_1, A_2 \):

\[ \begin{array}{c}
\text{6-} \rho_{AE} \\
\text{6-} \rho_{AE} \\
\text{6-} \rho_{AE} \\
\text{6-} \rho_{AE} \\
\text{6-} \rho_{AE} \\
\text{6-} \rho_{AE}
\end{array} \]

Apply a unitary \( V \) to \( A \) before discarding \( A_2 \) to obtain marginal state \( \rho_{AE(V)} \).

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 the measure:

\[
(SdV \| 6A_2E(V) - 6A_2E \| / 1) \leq \frac{|A_2| - |E|}{|A_1|^2} \tr(6 \rho_{AE})^2
\]

(where \( 6_{max} \) is the maximally entangled state on \( A_1 \)).

This generalizes the result found in a homework exercise, which concerned the case where \( E \) is trivial and \( 6 \) \( A \) is pure; there you derived:

\[
(SdV \| 6A_2 \| / 1) \leq \frac{|A_2|}{|A_1|} = \frac{|A_1|}{|A_1|^2}.
\]

(6 \( A_2 \) \( V \) nearly maximally mixed for \( |A_2| \ll |A_1| \).)

In the i.i.d. version the mother \( A \) becomes the typical subspace \( \rho_n \), \( E \) the typical subspace \( \rho_n \), \( AE \) the typical subspace \( \rho_n \( AE \) \). Since \( AE \) is
nearly maximally mixed on space \( \gamma_{\text{dim}} \simeq 2^{-n \cdot H(AE)} \)
we have \( I_{\gamma\{\gamma\}} \simeq 2^{-n \cdot H(AE)} \). Therefore, when
we average over \( V \), the state on \( \gamma_{AE} \) is nearly
a product state provided
\[
\left| A_{11} \right| ^2 \simeq 2^{-n \cdot H(A)} 2^{-n \cdot H(E)} 2^{-n \cdot H(AE)} \ll 1
\]
or
\[
\left| A_{11} \right|^2 \gg 2^{-n \cdot I(A;E)} \]

It suffices then for Alice to send
\[
\log |A_{11}| = \frac{n}{2} I(A;E) + o(n)
\]
qubits to Bob. And since
\[
H(A) = \frac{1}{2} \left[ I(A;E) + I(A;B) \right] \quad \text{because } \emptyset_{ABE} \text{ is pure}
\]
Alice retains
\[
\log |A_{11}| = \frac{n}{2} I(A;B) - o(n)
\]
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; since Alice and Bob share
\[
\frac{n}{2} I(A;B) - o(n) \text{ qubits, this proves the mutual resource inequality. (It works when we average over } V \text{, and therefore for some particular } V \text{.)}
\]

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 \( \gamma_{AE} \) is maximally mixed on a subspace of \( \text{dim } 181 \), i.e., a uniform mixture of 181 mutually orthogonal pure states. Then we trace out \( A_1 \). But for \( |A_{1i}| < |A_{11}| \), we expect that each of the 181 states in the ensemble realizing \( \gamma_{AE} \) is likely to be nearly maximally mixed.
on $A_1$, thus for each of these 181 states, tracing out $A_1$ generates a density operator on $A_2E$ which is a nearly uniform mixture of $1A_1$ mutually orthogonal states. Furthermore, as long as $|A_1B| < |A_2E|$, all of the $1A_1B$ states are likely to be nearly mutually orthogonal — tracing out $A_1$ produces a nearly uniform density operator with rank $\approx |A_1B|$. Once $|A_1B|$ 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 $|A_1B| \approx |A_2E| = \frac{|A_2E|}{|A_1B|}$

or $|A_1B| \approx \frac{|A_2E|}{|B|}$, 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

Father: $\langle NA^{A\rightarrow B}: \frac{1}{n} I(R;E) [q_{q\rightarrow q}] > \frac{1}{n} I(R;B) [q_{q\rightarrow q}]$ with the superdense coding inequality

SD: $[q_{q\rightarrow q}] + [E_{q\rightarrow q}] \geq 2 [C_{C\rightarrow C}]$

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

Suppose we use the $\frac{1}{n} I(R;B)$ qubits of $[q_{q\rightarrow q}]$ and an additional $\frac{1}{n} I(R;B)$ ebits to achieve $I(R;B)$ bits of $[C_{C\rightarrow C}]$. Because
\[ \frac{1}{2} I(R;E) + \frac{1}{2} I(R;B) = H(R), \]

we conclude

\[ \langle N^{A \rightarrow B} : P_A \rangle + H(R) [q_q] \geq I(R;B) [l \rightarrow c], \]

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 was entanglement-assisted classical capacity of the quantum channel thus satisfies

\[ C^E(N) \geq \max_{P_A} I(R;B) \]

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 latter 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.

\[ q \rightarrow q \geq [q_q] \Rightarrow \frac{1}{2} I(R;E) [q \rightarrow q] \geq \frac{1}{2} I(R;E) [q_q]. \]

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

\[ \frac{1}{2} I(R;B) - \frac{1}{2} I(R;E) = H(B) - H(E) = I_c(R;B). \]
We have derived the achievability result
\[
\left< N^{A\rightarrow B} : \rho_A \right> \geq I_c (R \rightarrow B) \left[ q \rightarrow q \right],
\]
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:
\[
Q(N) = \lim_{n \rightarrow \infty} \max_{\{\rho_n\}} \frac{1}{n} \sum_{i=1}^{n} I_c (R^n \rightarrow B^n).
\]

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:
\[
\text{Mother: } \left< \phi ^{AEB} \right> + \frac{1}{2} I(A;E) \left[ q \rightarrow q \right] \geq \frac{1}{2} I(A,B) \left[ q q \right] + \left< \phi ^{B,E} \right>
\]
by combining with the teleportation resource inequality:
\[
\text{TP: } \left[ q q \right] + 2 \left( c \rightarrow c \right) \geq \left[ q \rightarrow q \right]
\]
(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 \(\frac{1}{2} I(A;E)\) ebits 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
\( \langle \psi^{ABE} \rangle + I(A;E) \) \( \langle c \rightarrow c \rangle \geq I_c(A > B) \) \( \langle 99 \rangle \) + \( \langle \psi^{ABiE} \rangle \),
which is called the "Hashing inequality." It quantifies an achievable rate for distilling maximum 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 in 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: \( \langle \psi^{AB} \rangle \geq H(A) \) \( \langle 99 \rangle \).

\( H(A) \) bits can be extracted asymptotically from n copies of \( \psi^{AB} \). In this case the resource inequality says that the sufficient amount of \( \langle c \rightarrow c \rangle \) is \( I(A;E) = 0 \) — no classical communication is required. But this result is a bit misleading; some classical communication is needed, but only \( O(\sqrt{n}) \) 6 bits, or \( O(n^{-\frac{1}{2}}) \) per copy, which becomes negligible asymptotically.

State Merging

The state-merging resource inequality allows 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

Adding

\[ \frac{1}{2} I(A;B) \{ q \rightarrow q \} + I(A;B) \{ c \rightarrow c \} > \frac{1}{2} I(A;B) \{ q \rightarrow q \} \]

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

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

we obtain

\[ \text{State: } \langle \psi^{ABE} \rangle + H(A|B) \{ q \rightarrow q \} + I(A;B) \{ c \rightarrow c \} > \langle 0^{+} B; E \rangle. \]

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 an information theoretic
call "Slepian-Wolf coding." Alice sorts her
messages into \( 2^{-(H(X|Y)+\delta)} \) 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.

\[
\begin{array}{c}
A \\
B
\end{array} \rightarrow \begin{array}{c}
C
\end{array}
\]

Similarly, if \( A \) and \( B \) both send to \( C \):

Bob compresses the info from his source
to \( n H(Y) + o(n) \) letters. Then Alice
need send only \( n H(X|Y) + o(n) \) letters
to \( C \). Together, \( AB \) compress their
shared information source to \( n (H(XY)) \) 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:

\[
\text{Hashing: } \langle \psi^{AAE} \rangle + I(A;E) \langle c \rightarrow c \rangle > H(A|B)[qq] + \langle \psi^{BIE} \rangle
\]

Now the state merging has no quantum cost, and $AB$ hold $-H(A|B)$ qubits 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

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

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
\[
(SdU \ || \ \delta A E(U) - \delta_{\text{max}} \otimes \delta E \parallel_2) \leq \frac{|1_A E|}{|1_A v|} \text{Tr} (\delta A E)^2
\]
where \( U \) acts on \( A = A, A' \).

We note that
\[
|| \delta A E - \delta_{\text{max}} \otimes \delta E \parallel_2^2 = \text{Tr} (\delta A E)^2 - \frac{1}{|1_A v|} \text{Tr} (\delta E)^2
\]
(because \( \text{Tr} (\delta_{\text{max}})^2 = \frac{1}{|1_A v|} \)).

Now evaluate
\[
SdU \ \text{Tr} (\delta A E(U))^2
\]
\[
= SdU \ \text{Tr}_{A_1} \left( U \delta A E(U^+) \otimes \text{Tr}_{A_1} (U \delta A E(U^+)) S AA' \otimes S E E' \right)
\]
where \( S AA' \) denotes the swap operator on \( AA' \).

Therefore,
\[
(SdU \ \text{Tr} (\delta A E(U))^2)
\]
\[
= \text{Tr} \left[ \delta A E \otimes \delta A E' \left( SdU (U^+ U^+) I AA' \otimes S AA' (U^+ U) \right) S E E' \right]
\]

By the Lemma below, the integral is
\[
SdU (U^+ U^+) I AA' \otimes S AA' (U^+ U)
\]
\[
= C_I I AA' + C_S S AA'
\]
where
\[
C_I = \frac{1}{|1_A v|} \left( \frac{1 - \frac{1}{|1_A v|^2}}{2 - \frac{1}{|1_A v|^2}} \right) \leq \frac{1}{|1_A v|}, \quad C_S = \frac{1}{|1_A|} \left( \frac{1 - \frac{1}{|1_A v|^2}}{1 - \frac{1}{|1_A|^2}} \right) \leq \frac{1}{|1_A|}.
\]
Plugging the value of the integral into the trace:

\[
\left( \text{d}u \text{ tr}(E^{A_1 E}(U)) \right)^2 \leq \frac{1}{1A_1} \text{ tr}(E^2) + \frac{1}{1A_1} \text{ tr}(E A_1 E)^2
\]

and we conclude

\[
\text{d}u \| E^{A_1 E}(U) - E_{\max} \otimes E \|_2^2 \leq \frac{1}{1A_1} \text{ tr}(E A_1 E)^2.
\]

From the Cauchy–Schwarz inequality,

\[
\|M\|_1 \leq \text{d} \|M\|_2 \quad \text{and} \quad \langle \text{tr}^2 \rangle \leq \langle \text{tr} \rangle^2,
\]

we find

\[
\left( \text{d}u \| E^{A_2 E}(U) - E_{\max} \otimes E \|_2^2 \right) \leq \frac{1A_2 E}{1A_1} \text{ tr}(E A_1 E)^2
\]

— This is the decoupling inequality.

It remains to prove:

\[
\text{Lemma:} \quad \text{d}u \ (U \otimes U) \ I^{A_1 A'} \otimes S^{A_1 A'} (U \otimes U)
\]

\[
= C_1 \ I^{A A'} + C_5 \ S^{A A'}
\]

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

\[
\text{d}u \ (U \otimes U) \ I^{A_1 A'} \otimes S^{A_1 A'} (U \otimes U) = C_{\text{sym}} \ I^{AA'} + C_{\text{anti}} \ I^{AA'}
\]

where \( I^{AA'} \) projects onto the subspace symmetric under \( A \leftrightarrow A' \) and \( I^{AA'} \) projects onto the antisymmetric subspace. To compute \( C_{\text{sym}} \), evaluate \( \text{tr}(I^{AA'} \otimes S^{AA'}) \) on both sides. Using \( I^{AA'} = \frac{1}{2}(I^{AA'} + S^{AA'}) \), we obtain...
\[
\frac{1}{2} \text{tr} \left( I^{A_{1}A_{1}'} \otimes S^{A_{2}A_{2}'} \right) \left( I^{A_{1}A_{1}'} \otimes I^{A_{2}A_{2}'} + S^{A_{1}A_{1}'} \otimes S^{A_{2}A_{2}'} \right)
\]
\[
= \frac{1}{2} \left[ \text{tr} \left( I^{A_{1}A_{1}'} \otimes S^{A_{2}A_{2}'} \right) + \text{tr} \left( S^{A_{1}A_{1}'} \otimes I^{A_{2}A_{2}'} \right) \right]
\]
\[
= \frac{1}{2} \left( 1A_{1}1 A_{1}1 + 1A_{1}1 1A_{1}1 \right) = C_{\text{sym}} \frac{\text{tr} \Pi_{\text{sym}}^{AA'}}{2} = C_{\text{sym}} \frac{1}{2} (1A_{1}1 (1A_{1}+1))
\]
\[
\Rightarrow \quad C_{\text{sym}} = \frac{1A_{1}1 + 1A_{1}1}{1A_{1}+1}
\]

(Here we used \( \text{tr} S^{AA'} = \text{tr} (2 \Pi_{\text{sym}}^{AA'} - I^{AA'}) = 1A_{1}(1A_{1}+1) - 1A_{1}1 = 1A_{1}1 \).

Similarly, \( \Pi_{\text{anti}}^{AA'} = \frac{1}{2} \left( I^{AA'} - S^{AA'} \right) \) and \( \text{tr} \Pi_{\text{anti}}^{AA'} = \frac{1}{2} (1A_{1}1 (1A_{1}1 - 1)) \),

\[
\Rightarrow \quad C_{\text{anti}} = \frac{1A_{1}1 - 1A_{1}1}{1A_{1}-1}
\]

Then noting that \( C_{\text{sym}} + C_{\text{anti}} = \frac{1}{2} \left( C_{\text{sym}} + C_{\text{anti}} \right) \),

\[
C_{s} = \frac{1}{2} \left( C_{\text{sym}} - C_{\text{anti}} \right)
\]

we obtain

\[
C_{s} = \frac{1A_{1}1 \cdot 1A_{1}1 - 1A_{1}1}{1A_{1}1^2 - 1} = \frac{1}{1A_{1}1} \left( \frac{1A_{1}1^2 - 1}{1A_{1}1(1A_{1}1^2 - 1/1A_{1}1^2)} \right)
\]

which proves the lemma.
Appendix B: Fidelity and $L^1$ distance

We wish to show:

$$\sqrt{F(\rho, \sigma)} \equiv \| \sqrt{\rho} \sqrt{\sigma} \|_1 = \text{Tr} \sqrt{\rho} \sqrt{\sigma} \sqrt{\rho} \sqrt{\sigma} \geq 1 - \frac{1}{2} \| \rho - \sigma \|_1.$$  

From the polar decomposition of $M$ we obtain

$$\text{Tr} \sqrt{M^* M} \geq \text{Tr} M \Rightarrow \sqrt{F(\rho, \sigma)} \geq \text{Tr}(\sqrt{\rho} \sqrt{\sigma}).$$

And

$$\| \sqrt{\rho} - \sqrt{\sigma} \|_2^2 = \text{Tr} (\sqrt{\rho} - \sqrt{\sigma})^2 = 2 - 2 \text{Tr} (\sqrt{\rho} \sqrt{\sigma}) \geq 2 - 2 \sqrt{F(\rho, \sigma)} \Rightarrow \sqrt{F(\rho, \sigma)} \geq 1 - \frac{1}{2} \| \sqrt{\rho} - \sqrt{\sigma} \|_2^2$$

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

Note that  

$$\rho - \sigma = \frac{1}{2} (\sqrt{\rho} - \sqrt{\sigma})(\sqrt{\rho} + \sqrt{\sigma}) + \frac{1}{2} (\sqrt{\rho} + \sqrt{\sigma})(\sqrt{\rho} - \sqrt{\sigma}),$$

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

$$\| \sqrt{\rho} - \sqrt{\sigma} \|_2^2 = \sum_i |\langle i | (\sqrt{\rho} - \sqrt{\sigma}) |i \rangle|^2 = \text{Tr} (\sqrt{\rho} - \sqrt{\sigma}) = (\sqrt{\rho} - \sqrt{\sigma}) \otimes$$

where $|i \rangle$ is the ON basis that diagonalizes $\sqrt{\rho} - \sqrt{\sigma}$ and $\otimes$ is the unitary transformation $U = \sum_i \text{diag}(\{i\}) |i \rangle \langle i |$.

Now,  

$$\text{Tr} | \rho - \sigma | \geq \text{Tr} (\rho - \sigma) U \text{ (true for any unitary $U$)}$$

$$= \text{Tr} | \sqrt{\rho} - \sqrt{\sigma} | (\sqrt{\rho} + \sqrt{\sigma}) = \sum_i |\langle i | \sqrt{\rho} + \sqrt{\sigma} |i \rangle|^2 \Rightarrow \text{Tr} \sum_i |\langle i | \sqrt{\rho} - \sqrt{\sigma} |i \rangle|^2 = \sum_i |\langle i | \sqrt{\rho} - \sqrt{\sigma} |i \rangle|^2 = \| \sqrt{\rho} - \sqrt{\sigma} \|_2^2$$

Thus $\| \rho - \sigma \|_2 \geq \| \sqrt{\rho} - \sqrt{\sigma} \|_2^2$, as we wanted to show.
By the way, it is sometimes convenient to have an upper bound on \( F(\phi, \psi) \) expressed in terms of the \( L^2 \) distance \( \| \phi - \psi \|_2 \), for example,
\[
F(\phi, \psi) \leq 1 - \frac{1}{4} \| \phi - \psi \|_2^2.
\]

0. First show this is an equality for pure states

\[
\langle \psi | \psi \rangle = (\cos^2 \alpha) \quad \langle \phi | \phi \rangle = (\sin^2 \alpha) \quad \Rightarrow \quad \langle \phi | \psi \rangle = \sin \alpha \cos \alpha \quad \Rightarrow
\]

\[
\| \phi - \psi \|_2^2 = 4 \cos^2 \alpha = 4 \left(1 - F(\phi, \psi)\right).
\]

2. Next note that \( L^2 \) distance is monotonic:

\[
\| \phi_{AB} - \psi_{AB} \|_2 \geq \| \phi_A - \psi_A \|_2.
\]

This is true because \( L^2 \) distance is optimal distance between prob. distributions for POVM on times, and we can perform a POVM on \( AB \) that acts nontrivially only on \( A \).

3. Finally, by Uhlmann’s Lemma,

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

where \( \rho_{AB}, \sigma_{AB} \) are purifications with maximum fidelity,

\[
= 1 - \frac{1}{4} \| \rho_{AB} - \sigma_{AB} \|_2^2
\]

\[
\leq 1 - \frac{1}{4} \| \rho_A - \sigma_A \|_2^2,
\]

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