

Back to the Partial Trace

The partial trace over \mathcal{H}_A of a system in $\mathcal{H}_A \otimes \mathcal{H}_B$ can be written given an *orthonormal basis* $\{|e_i\rangle\}_i$ for \mathcal{H}_A as,

$$\begin{aligned}\mathrm{tr}_A (|x_1\rangle\langle x_2| \otimes |y_1\rangle\langle y_2|) &= \sum_i \langle e_i| (|x_1\rangle\langle x_2| \otimes |y_1\rangle\langle y_2|) |e_i\rangle \\ &\equiv \sum_i \langle e_i|x_1\rangle\langle x_2|e_i\rangle |y_1\rangle\langle y_2| \\ &= |y_1\rangle\langle y_2| \mathrm{tr} (|x_1\rangle\langle x_2|),\end{aligned}$$

which is the definition we have already seen. In general, for

$$U = \sum_i |x_i\rangle\langle x'_i| |y_i\rangle\langle y'_i|,$$

we write

$$\langle e_k|U|e_l\rangle = \sum_i \langle e_k|x_i\rangle\langle y_i|e_l\rangle |x'_i\rangle\langle y'_i|.$$

which is not necessarily a unitary operator!

Quantum Operations

Suppose system $\rho \in \mathcal{H}_T$ does not evolve in a closed system:

$$\rho \xrightarrow{\mathcal{E}} \text{tr}_E (U(\rho \otimes |e_0\rangle\langle e_0|)U^\dagger)$$

for some unitary U and state $|e_0\rangle$ for the *environment* \mathcal{H}_E . This mapping is called a *quantum operation* $\mathcal{E}(\rho)$. Let $\{|e_k\rangle\}_k$, be an orthonormal basis for \mathcal{H}_E ,

$$\begin{aligned}\mathcal{E}(\rho) &= \sum_k \langle e_k | U(\rho \otimes |e_0\rangle\langle e_0|)U^\dagger | e_k \rangle \\ &= \sum_k E_k \rho E_k^\dagger\end{aligned}$$

where,

$$E_k = \langle e_k | U | e_0 \rangle.$$

The set $\{E_k\}_k$ is called the *operator-sum* representation of $\mathcal{E}(\rho)$.

Properties of the Operator-Sum

- E_k is an operator acting only in the space \mathcal{H}_T containing ρ ,
- Since

$$\begin{aligned} 1 = \text{tr}(\rho) &= \text{tr}(\mathcal{E}(\rho)) = \text{tr}\left(\sum_k E_k \rho E_k^\dagger\right) \\ &= \text{tr}\left(\sum_k E_k^\dagger E_k \rho\right) \text{ for all } \rho, \\ &\Rightarrow \sum_k E_k^\dagger E_k = \mathbb{I}. \end{aligned}$$

- Composition has also an operator-sum representation,

$$\begin{aligned} \mathcal{G}(\rho) \equiv \mathcal{F} \odot \mathcal{E}(\rho) &= \sum_l F_l \left(\sum_k E_k \rho E_k^\dagger \right) F_l^\dagger \\ &= \sum_{l,k} (F_l E_k) \rho (F_l E_k)^\dagger = \sum_s G_s \rho G_s^\dagger. \end{aligned}$$

Physical Interpretation

- Let U be applied on **target-environment** in state $\rho \otimes |e_0\rangle\langle e_0|$, and
- A **complete projective measurement** $\{|e_k\rangle\langle e_k|\}_k$ is applied on E

\Rightarrow It does not change the state of T (if the outcome remains unknown),

\Rightarrow

$$\begin{aligned}\rho_k &\propto \text{tr}_E (|e_k\rangle\langle e_k| U(\rho \otimes |e_0\rangle\langle e_0|) U^\dagger |e_k\rangle\langle e_k|) \\ &= \langle e_k| U(\rho \otimes |e_0\rangle\langle e_0|) U^\dagger |e_k\rangle = E_k \rho E_k^\dagger.\end{aligned}$$

Normalizing $\rho_k = \frac{E_k \rho E_k^\dagger}{\text{tr}(E_k \rho E_k^\dagger)}$ gives that the probability $p(k)$ for outcome k is:

$$p(k) = \text{tr} (|e_k\rangle\langle e_k| U(\rho \otimes |e_0\rangle\langle e_0|) U^\dagger |e_k\rangle\langle e_k|) = \text{tr} (E_k \rho E_k^\dagger).$$

Thus,

$$\mathcal{E}(\rho) = \sum_k p(k) \rho_k = \sum_k E_k \rho E_k^\dagger.$$

An Example

Suppose the environment applies a $U \equiv \text{CNOT}$:

$$\text{CNOT}|c\rangle|t\rangle \mapsto |c\rangle|c \oplus t\rangle, c, t \in \{0, 1\},$$

with the environment E starting in state $|0\rangle$. We determine the operator-sum representation:

$$\text{CNOT} = |00\rangle\langle 00| + |01\rangle\langle 01| + |10\rangle\langle 11| + |11\rangle\langle 10|.$$

Thus,

$$E_0 = \langle 0|\text{CNOT}|0\rangle = |0\rangle\langle 0|$$

$$E_1 = \langle 1|\text{CNOT}|0\rangle = |1\rangle\langle 1|,$$

and therefore for $P_0 = |0\rangle\langle 0|$ and $P_1 = |1\rangle\langle 1|$,

$$\mathcal{E}(\rho) = P_0\rho P_0 + P_1\rho P_1.$$

Freedom in Operator-Sum

Consider,

$$E_1 = \frac{\mathbb{I}}{\sqrt{2}} \quad \text{and} \quad E_2 = \frac{Z}{\sqrt{2}}$$
$$F_1 = |0\rangle\langle 0| \quad \text{and} \quad F_2 = |1\rangle\langle 1|.$$

Observe that,

$$F_1 = (E_1 + E_2)/\sqrt{2} \quad \text{and} \quad F_2 = (E_1 - E_2)/\sqrt{2}.$$

It follows that,

$$\begin{aligned} \mathcal{F}(\rho) &= \frac{1}{2} \left((E_1 + E_2)\rho(E_1^\dagger + E_2^\dagger) + (E_1 - E_2)\rho(E_1^\dagger - E_2^\dagger) \right) \\ &= E_1\rho E_1^\dagger + E_2\rho E_2^\dagger = \mathcal{E}(\rho). \end{aligned}$$

Theorem: *The operator-sum representations $\{E_i\}_i$ and $\{F_i\}_i$ define the same quantum operation *if and only if**

$$E_i = \sum_j u_{i,j} F_j \quad \text{with} \quad \{u_{i,j}\}_{i,j} \text{ unitary.}$$

Reminder: Freedom for Density Operators

Let $\{|\tilde{\psi}_x\rangle\}_x$ be an ensemble that produces $\frac{|\tilde{\psi}_x\rangle\langle\tilde{\psi}_x|}{\langle\tilde{\psi}_x|\tilde{\psi}_x\rangle}$ with probability $\langle\tilde{\psi}_x|\tilde{\psi}_x\rangle$. We need the following Theorem that we have seen last time:

Theorem 2.6: *Ensembles $\{|\tilde{\psi}_i\rangle\}_i$ and $\{|\tilde{\phi}_j\rangle\}_j$ generate the same density matrix if and only if*

$$|\tilde{\psi}_i\rangle = \sum_j u_{i,j} |\tilde{\phi}_j\rangle,$$

where $(u_{i,j})_{i,j}$ is unitary over the complex numbers for indices i , and j , and where we pad whichever set of vectors is smaller with additional vectors $\vec{0}$ so that the two sets have the same number of elements.

Proof of the Theorem (digression)

Assume the following *not-normalized* state shared among systems R and Q :

$$|\alpha\rangle = \sum_i |i_R\rangle |i_Q\rangle.$$

Now, assume we execute quantum operation \mathcal{E} acting only in Q :

$$\sigma = (\mathbb{I}_R \otimes \mathcal{E})(|\alpha\rangle\langle\alpha|).$$

We want to *recover* \mathcal{E} from σ . Let us associate to $|\psi\rangle \in Q$ the following $|\tilde{\psi}\rangle \in R$:

$$|\psi\rangle = \sum_j \psi_j |j_Q\rangle \Rightarrow |\tilde{\psi}\rangle = \sum_j \psi_j^* |j_R\rangle.$$

Proof of the Theorem (digression)

Remember that,

$$|\psi\rangle\langle\psi| = \sum_{i,j} \psi_i \psi_j^* |i\rangle\langle j|.$$

Notice that,

$$\begin{aligned} \langle \tilde{\psi} | \sigma | \tilde{\psi} \rangle &= \langle \tilde{\psi} | \left(\sum_{i,j} |i_R\rangle\langle j_R| \otimes \mathcal{E}(|i_Q\rangle\langle j_Q|) \right) | \tilde{\psi} \rangle \\ &= \sum_{i,j} \psi_i \psi_j^* \mathcal{E}(|i_Q\rangle\langle j_Q|) = \mathcal{E}(|\psi\rangle\langle\psi|). \end{aligned}$$

Assume the *some decomposition* of σ :

$$\sigma = \sum_i |s_i\rangle\langle s_i|,$$

where $|s_i\rangle$ need not be normalized.

Proof of the Theorem (digression)

Define the following mapping:

$$E_i : |\psi\rangle \mapsto \langle \tilde{\psi} | s_i \rangle.$$

Observe that:

$$\begin{aligned} \sum_i E_i |\psi\rangle \langle \psi| E_i^\dagger &= \sum_i \langle \tilde{\psi} | s_i \rangle \langle s_i | \tilde{\psi} \rangle \\ &= \langle \tilde{\psi} | \sigma | \tilde{\psi} \rangle \\ &= \mathcal{E}(|\psi\rangle \langle \psi|). \end{aligned}$$

A single “state” σ allows for recovering \mathcal{E} and an operator-sum representation for it!

The Proof

Suppose $\{E_i\}_i$ and $\{F_j\}_j$ are two sets of operation elements such that $\sum_i E_i \rho E_i^\dagger = \sum_j F_j \rho F_j^\dagger$ for all ρ . We show that for unitary $\{u_{ij}\}_{i,j}$:

$$E_i = \sum_j u_{ij} F_j.$$

Let $|e_i\rangle \equiv \sum_k |k_R\rangle \otimes (E_i |k_Q\rangle)$ and $|f_j\rangle \equiv \sum_k |k_R\rangle \otimes (F_j |k_Q\rangle)$.

By definition of σ , it follows that $\sigma = \sum_i |e_i\rangle \langle e_i| = \sum_j |f_j\rangle \langle f_j|$.

From [Theorem 2.6](#), we get:

$$|e_i\rangle = \sum_j u_{ij} |f_j\rangle,$$

where $\{u_{ij}\}_{i,j}$ is unitary.

The proof(final)

For arbitrary $|\psi\rangle$ we have,

$$\begin{aligned} E_i |\psi\rangle &= \langle \tilde{\psi} | e_i \rangle \\ &= \sum_j u_{ij} \langle \tilde{\psi} | f_j \rangle \\ &= \sum_j u_{ij} F_j |\psi\rangle, \end{aligned}$$

it follows that:

$$E_i = \sum_j u_{ij} F_j.$$

- The other direction is easier...

Noisy Quantum Channels

Bit flip: $E_0 = \sqrt{1-p}\mathbb{I}$ and $E_1 = \sqrt{p}X$ is the operator-sum representation of a bit flip $|b\rangle \mapsto |1-b\rangle$ with probability p .

Phase flip: $E_0 = \sqrt{1-p}\mathbb{I}$ and $E_1 = \sqrt{p}Z$ flips the phase $|b\rangle \mapsto (-1)^b|b\rangle$ with probability p . When $p = 1/2$ we get,

$$\mathcal{E}(\rho) = P_0\rho P_0 + P_1\rho P_1,$$

this is equivalent to apply secretly measurement $\{P_0, P_1\}$.

Depolarizing channel: Applies the quantum operator

$$\mathcal{E}(\rho) = \frac{p\mathbb{I}}{2} + (1-p)\rho.$$

Using the fact that for all ρ , $\frac{\mathbb{I}}{2} = \frac{\rho + X\rho X + Y\rho Y + Z\rho Z}{4}$, the operator-sum representation is,

$$\mathcal{E}(\rho) = \left(1 - \frac{3p}{4}\right)\rho + \frac{p}{4}(X\rho X + Y\rho Y + Z\rho Z)$$

which has operation elements: $\left\{\sqrt{1 - \frac{3p}{4}}\mathbb{I}, \frac{\sqrt{p}}{2}X, \frac{\sqrt{p}}{2}Y, \frac{\sqrt{p}}{2}Z\right\}$.

Digression: Shannon Entropy

- $\mathcal{Z} = \{((x_i, y_j), p_{i,j})\}_{i,j}$ is a joint probability distribution for random variables (X, Y) ,
- $\mathcal{X} = \{(x_i, p_i)\}_i$ and $\mathcal{Y} = \{(y_j, q_j)\}_j$ are the marginal distributions for X and Y resp.

Basics:

$$H(X) = - \sum_i p_i \log p_i \text{ and } H(Y) = - \sum_j q_j \log q_j,$$

$$H(X, Y) = - \sum_{i,j} p_{i,j} \log p_{i,j}.$$

Relative Entropy: $H(X | Y) = - \sum_{i,j} p_{i,j} \log \Pr(X = x_i | Y = y_j)$.

Chain Rule:

$$H(X, Y) = H(X) + H(Y | X) = H(Y) + H(X | Y),$$

$$H(X, Y | Z) = H(X | Z) + H(Y | X, Z).$$

Shannon Mutual Information

The *mutual information* provided by Y about X is,

$$I(X; Y) = \sum_{i,j} p_{i,j} \log \left(\frac{p_{i,j}}{p_i q_j} \right) = H(X) - H(X | Y)$$

Mutual information and entropy satisfy the following identities:

1. $I(X; Y) = H(X) - H(X | Y)$,
2. $I(X; Y) = H(Y) - H(Y | X)$,
3. $I(X; Y) = H(X) + H(Y) - H(X, Y)$,
4. $I(X; Y) = I(Y; X)$,
5. $H(X | Y) \leq H(X)$,
6. $H(X, Y) \leq H(X) + H(Y)$.

Von Neumann Entropy

Let ρ be any mixed state. From the **spectral decomposition** theorem we can write,

$$\rho = \sum_i \lambda_i |e_i\rangle\langle e_i|$$

where $\{|e_i\rangle\}_i$ is an *orthonormal basis* and $\{\lambda_i\}_i$ is the set of *eigenvalues* (i.e. $\lambda_i \geq 0$). This is *equivalent* to the classical probability distribution $\{(e_i, \lambda_i)\}_i$.

This suggests to define the *entropy* associated to ρ as,

$$S(\rho) = -\text{tr}(\rho \log \rho) = -\sum_i \lambda_i \log \lambda_i.$$

Notice that,

- Any pure state $|\psi\rangle\langle\psi|$ is such that $S(|\psi\rangle\langle\psi|) = 0$ since it has an eigenvalue $\lambda = 1$.

Basic Properties of $S(\rho)$

1. For any ρ , $S(\rho) \geq 0$,
2. Any $\rho = \frac{1}{n} \sum_i |e_i\rangle\langle e_i| = \frac{\mathbb{I}_n}{n}$ is such that $S(\rho) = \log n$,
3. Let $\{(p_i, \rho_i)\}_i$ be a probability distribution over mixtures with support in **orthonormal subspaces** then,

$$S\left(\sum_i p_i \rho_i\right) = H(p_1, \dots, p_n) + \sum_i p_i S(\rho_i),$$

4. Let $\{p_i\}_i$ be a probability distribution, let $\{|e_i\rangle\}_i$ be an orthonormal basis, and let $\{\rho_i\}_i$ be any set of mixed states,

$$S\left(\sum_i p_i |e_i\rangle\langle e_i| \otimes \rho_i\right) = H(p_1, \dots, p_n) + \sum_i p_i S(\rho_i),$$

5. For any ρ and σ , $S(\rho \otimes \sigma) = S(\rho) + S(\sigma)$.

Quantum Mutual Information

Let $\rho^{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ be any composite system. We define,

$$\begin{aligned} S(A, B) &= -\text{tr}(\rho^{AB} \log \rho^{AB}) \\ S(A) &= S(\rho^A) \text{ where } \rho^A = \text{tr}_B(\rho^{AB}) \\ S(B) &= S(\rho^B) \text{ where } \rho^B = \text{tr}_A(\rho^{AB}). \end{aligned}$$

As for classical information:

conditional entropy: $S(A | B) = S(A, B) - S(B)$,

mutual information:

$$\begin{aligned} S(A; B) &\equiv S(A) + S(B) - S(A, B) = S(A) - S(A | B) \\ &= S(A, B) - S(B | A) - S(A | B) \\ &= S(A | B) + S(B) - S(B | A) - S(A | B) \\ &= S(B) - S(B | A) = S(B; A). \end{aligned}$$

Interpretation of $S(A | B)$

Classically, $H(X | Y)$ is the amount of (bit of) information missing to re-construct X from Y . What about $S(A | B)$?

Suppose A is *classical*:

$$\rho_{AB} = \sum_{x \in \{0,1\}^n} P_X(x) |x\rangle\langle x| \otimes \rho_x.$$

Then,

$S(A | B) = S(X | B) =$ The amount of additional classical information sufficient in order to re-construct X from quantum system B [Winter99(Ph.D. thesis), Devetak-Winter03].

Notice: We have:

$$S(X | B) \geq H(X) - S(B),$$

with equality iff ρ_x is pure for all x and where $\rho^B = \sum_x P_X(x) \rho_x$.

In the Twilight Zone

Consider the maximally entangled state:

$$|\Psi\rangle^{AB} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle).$$

We know that,

$$\rho^A = \frac{1}{2}\mathbb{I}_2 = \rho^B$$

This means that

$$S(A, B) = 0 \text{ and } S(A) = S(B) = 1,$$

- This shows that

$$S(A, B) < S(A) \text{ is possible but } H(X) \leq H(X, Y),$$

- It also shows that,

$$S(B | A) = S(A, B) - S(A) = -1 \text{ is possible but } H(X | Y) \geq 0.$$

More Properties

Theorem: Let A, B, C be a composite quantum system. Then,

Conditioning reduces entropy: $S(A | B, C) \leq S(A | B)$.

Discarding never increases information: $S(A; B) \leq S(A; B, C)$.

Quantum operation never increases information: Let \mathcal{E} be a quantum operation acting on B alone. Then,

$$S(A'; B') \leq S(A; B).$$

Proof of 3: We have seen that $\mathcal{E}(\rho^{AB}) = \text{tr}_C(U(\rho \otimes |0\rangle\langle 0|)U^\dagger)$ for U acting only upon B and C .

$$S(A; B) = S(A; B, C) \stackrel{U}{=} S(A'; B', C') \geq S(A'; B'),$$

where the second equality comes from the fact that unitary transforms do not change the eigenvalues. The last inequality comes after discarding C .

Holevo Bound

The **Holevo bound** gives the amount of *classical information* that can be extracted from a quantum state.

Theorem: Suppose Alice prepares ρ_x with probability p_x . Let X be a r.v. for this choice. Suppose the receiver Bob performs a generalized measurement with measurement operators $\{M_y\}_y$ providing classical outcome Y . Then,

$$I(X; Y) \leq S(\rho) - \sum_x p_x S(\rho_x)$$

where $\rho = \sum_x p_x \rho_x$.

The expression

$$S(\rho) - \sum_x p_x S(\rho_x)$$

is called the *accessible classical information* in ρ .

Proof of the Holevo Bound

Consider the following tri-partite state:

$$\rho^{PQM} = \sum_x p_x |x\rangle\langle x|^P \otimes \rho_x^Q \otimes |0\rangle\langle 0|^M. \quad (1)$$

Let $\{E_y\}_y = \{M_y^\dagger M_y\}_y$ be the POVM with operators $\{M_y\}_y$ implemented as,

$$\mathcal{E}(\sigma \otimes |0\rangle\langle 0|) = \sum_y M_y \sigma M_y^\dagger \otimes |y\rangle\langle y|$$

The measurement operators applies the following transformation:

$$\rho^{P'Q'M'} = \sum_{x,y} p_x |x\rangle\langle x|^P \otimes M_y \rho_x^Q M_y^\dagger \otimes |y\rangle\langle y|^M.$$

$$\begin{aligned} S(P; Q) &= S(P; Q, M) \text{ since } M \text{ is initially in pure state,} \\ &\geq S(P'; Q', M') \text{ since quantum operations don't increase inf.,} \\ &\geq S(P'; M') \text{ since discarding don't increase inf.} \\ \Rightarrow S(P; Q) &\geq S(P'; M'). \end{aligned}$$

Proof of the Holevo Bound

$S(P; Q)$ is the Accessible Information

$$\rho^{PQM} = \sum_x p_x |x\rangle\langle x|^P \otimes \rho_x^Q \otimes |0\rangle\langle 0|^M. \quad (2)$$

Observe that,

- $S(P; Q) = S(P) + S(Q) - S(P, Q)$ by definition,
- $S(P) = H(p_0, \dots, p_n)$ by a property of $S(\cdot)$.
- $S(Q) = S(\rho)$ by construction, and
- $S(P, Q) = H(p_0, \dots, p_n) + \sum_x p_x S(\rho_x)$ by a property of $S(\cdot)$.

In other words, we have shown:

$$S(P; Q) = S(\rho) - \sum_x p_x S(\rho_x).$$

$$S(P'; M') = I(X; Y)$$

$$\begin{aligned} \rho^{P'M'} &= \text{tr}_Q \left(\rho^{P'Q'M'} \right) \\ &= \sum_{x,y} p_x \text{tr} (E_y \rho_x) |x\rangle\langle x|^{P'} \otimes |y\rangle\langle y|^{M'} \\ &= \sum_{x,y} p_x \Pr(Y = y | X = x) |x\rangle\langle x|^{P'} \otimes |y\rangle\langle y|^{M'}. \end{aligned}$$

By definition, we have that $S(P'; M') = S(P') + S(M') - S(P', M')$.

Moreover,

- $S(P') = S(\sum_x p_x |x\rangle\langle x|) = H(X)$,
- $S(M') = S(\sum_y p_y |y\rangle\langle y|) = H(Y)$,
- $S(P', M') = H(X, Y)$, since
 $p_x \Pr(Y = y | X = x) = \Pr(X = x \wedge Y = y)$.

Which results in $S(P', M') = H(X) + H(Y) - H(Y, X) = I(X; Y)$.

A Corollary

Corollary: Suppose P and Q are sharing a quantum state $|\psi\rangle^{PQ}$. Suppose P does a measurement defined by the operators $\{M_x\}_x$ (acting only on his part of the system). Q wants to determine as much information as possible about P 's random variable X containing the outcome of the measurement. In order to do that, Q applies a measurement $\{M'_y\}_y$ on her part of the system. Then,

$$I(X; Y) \leq S(\rho^P) = S(\rho^Q) \quad (3)$$

where $\rho^P = \text{tr}_Q(|\psi\rangle\langle\psi|^{PQ})$ and $\rho^Q = \text{tr}_P(|\psi\rangle\langle\psi|^{PQ})$.

Application to QKD: Suppose Alice and Bob share a state ρ^{AB} such that,

$$\rho^{AB} \approx \bigotimes_{i=1}^m |\Psi_{EPR}\rangle\langle\Psi_{EPR}| \Rightarrow \langle\Psi_{EPR}|^{\otimes m} \rho^{AB} |\Psi_{EPR}\rangle^{\otimes m} \geq 1 - 2^{-\alpha n}.$$

- Eve can only get negligible information about measurement on ρ^{AB} ,
- Alice and Bob share a random m -bit string by measuring each pair of ρ^{AB} with measurement $\{\underline{P_0 \otimes P_0}, P_0 \otimes P_1, P_1 \otimes P_0, \underline{P_1 \otimes P_1}\}$.

Proof of the Corollary

We show that: $I(X; Y) \leq S(\rho)$ where $\rho = \text{tr}_P(|\psi\rangle\langle\psi|)$ and (X, Y) correspond to the outcome of P 's and Q 's measurements.

P 's measurement as a trace-preserving quantum operation:

$$\mathcal{E}(\sigma \otimes |0\rangle\langle 0|) = \sum_x M_x \sigma M_x^\dagger \otimes |x\rangle\langle x|.$$

Consider for $p_x = \text{tr}((M_x \otimes \mathbb{I})(M_x \otimes \mathbb{I})^\dagger |\psi\rangle\langle\psi|)$,

$$\begin{aligned} \sigma \mathcal{E} &= \sum_x p_x |x\rangle\langle x| \otimes \text{tr}_P \left(\frac{(M_x \otimes \mathbb{I})|\psi\rangle\langle\psi|(M_x \otimes \mathbb{I})^\dagger}{p_x} \right) \\ &= \sum_x p_x |x\rangle\langle x| \otimes \rho_x \text{ where } \rho = \sum_x p_x \rho_x \text{ by construction.} \end{aligned}$$

This corresponds to the mixture Q must distinguish in order to get information about X . Using the Holevo bound we get,

$$I(X; Y) \leq S(\rho) - \sum_x p_x S(\rho_x) \leq S(\rho).$$