PHYS 7895: Quantum Information Theory Fall 2015

Lecture 24 — November 16, 2015

Prof. Mark M. Wilde *Scribe: Mark M. Wilde*

1 Overview

In the last lecture, we discussed a strategy for classical communication over a quantum channel called sequential decoding.

In this lecture, we discuss an application to communication over optical channels. We then prove the Holevo-Schumacher-Westmoreland theorem, which gives a characterization of the classical capacity of a quantum channel. Finally, we show how this characterization simplifies for entanglement-breaking channels.

2 Sequential Decoding for Optical Communication

We now provide a physical realization of the sequential decoding strategy in the context of optical communications. In this setting, we suppose that a lossy bosonic channel, specified by the following Heisenberg relations, connects Alice to Bob:

$$\hat{b} = \sqrt{\eta}\hat{a} + \sqrt{1-\eta}\hat{e}, \tag{1}$$

where $\hat{a}$, $\hat{b}$, and $\hat{e}$ are the respective field operators for Alice's input mode, Bob's output mode, and an environmental input mode (assumed to be in its vacuum state). The transmissivity $\eta \in [0,1]$ is the fraction of Alice's input photons that make it to Bob on average. We assume that Alice is constrained to using mean photon number N_S per channel use. The strategy for achieving the classical capacity of this channel is for Alice to induce a classical-quantum channel, by selecting $\alpha \in \mathbb{C}$ and preparing a coherent state $|\alpha\rangle$ at the input of the channel in (1). A coherent state in quantum optics is defined as the following coherent superposition of photon number states:

$$|\alpha\rangle \equiv \exp\left\{\frac{-|\alpha|^2}{2}\right\} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle.$$

It is often described as being the ideal state of a single mode of the light field output from a laser. The most useful property of coherent states for classical communication over a pure-loss bosonic channel is that it retains its purity. That is, if Alice inputs the state $|\alpha\rangle$ to the pure-loss bosonic channel with transmissivity η, then the state output for Bob and Eve is

$$|\sqrt{\eta}\alpha\rangle \otimes |\sqrt{1-\eta}\alpha\rangle,$$

so that we recover a pure coherent state for Bob when tracing over the second mode. The resulting induced classical-quantum channel to Bob is of the following form:

$$\alpha \to |\sqrt{\eta}\alpha\rangle.$$

By choosing the distribution $p_X(x)$ in the achievability result for pure-state cq channels to be an isotropic, complex Gaussian with variance N_S:

$$p_{N_S}(\alpha) \equiv (1/\pi N_S) \exp\left\{ -|\alpha|^2/N_S \right\},$$

we have that $g(\eta N_S)$ is an achievable rate for classical communication, where

$$g(x) \equiv (x+1)\log(x+1) - x \log x.$$

The quantity $g(\eta N_S)$ is the entropy of the average state of the ensemble $\{p_{N_S}(\alpha),\ |\sqrt{\eta}\alpha\rangle\}$:

$$\int d^2\alpha\ p_{N_S}(\alpha) |\sqrt{\eta}\alpha\rangle\langle\sqrt{\eta}\alpha|,$$

which is a thermal state with mean photon number ηN_S. Each quantum codeword selected from the ensemble $\{p_{N_S}(\alpha),\ |\alpha\rangle\}$ has the following form:

$$|\alpha^n(m)\rangle \equiv |\alpha_1(m)\rangle \otimes \cdots \otimes |\alpha_n(m)\rangle.$$

We assume $\eta = 1$ above and for the rest of this section without loss of generality. Thus, the sequential decoder consists of measurements of the following form for all $m \in \mathcal{M}$:

$$\left\{ |\alpha^n(m)\rangle\langle\alpha^n(m)|,\ I^{\otimes n} - |\alpha^n(m)\rangle\langle\alpha^n(m)| \right\}. \tag{2}$$

Observing that

$$|\alpha^n(m)\rangle = D(\alpha_1(m)) \otimes \cdots \otimes D(\alpha_n(m)) |0\rangle^{\otimes n},$$

where $D(\alpha) \equiv \exp\left\{ \alpha \hat{a}^\dagger - \alpha^* \hat{a} \right\}$ is the unitary "displacement" operator from quantum optics and $|0\rangle^{\otimes n}$ is the n-fold tensor product vacuum state, we see that that the decoder can implement the measurement in (2) in three steps:

1. Displace the n-mode codeword state by

$$D(-\alpha_1(m)) \otimes \cdots \otimes D(-\alpha_n(m)).$$

2. Perform a "vacuum-or-not" measurement of the form

$$\left\{ |0\rangle\langle 0|^{\otimes n},\ \ I^{\otimes n} - |0\rangle\langle 0|^{\otimes n} \right\}.$$

   If the vacuum outcome occurs, decode as the m^{th} codeword. Otherwise, proceed.

3. Displace by $D(\alpha_1(m)) \otimes \cdots \otimes D(\alpha_n(m))$ with the same method as in Step 1.

The receiver just iterates this strategy for every codeword in the codebook. This strategy is in fact capacity-achieving, but we do not prove that here.

3 General Classical–Quantum Channels

Suppose now that Alice and Bob are connected by a mixed-state cq channel of the following form:

$$x \to \rho_B^x. \tag{3}$$

It is possible to show that the following rate is achievable for classical communication:

$$\max_{p_X(x)} I(X;B)_\omega, \tag{4}$$

where

$$\omega_{XB} \equiv \sum_x p_X(x)|x\rangle\langle x|_X \otimes \rho_B^x. \tag{5}$$

The main idea for doing this is to use the sequential decoding strategy, but with each projector (codeword test) set to be a conditionally typical projector. How is this defined? Suppose we have an ensemble of states $\{p_X(x), \rho_B^x\}$. Let $\rho_B^x = \sum_y p_{Y|X}(y|x)|y_x\rangle\langle y_x|$ be a spectral decompositino of ρ_B^x. For a sequence $x^n \equiv x_1 \cdots x_n$, we write

$$\rho_{B^n}^{x^n} \equiv \rho_{B_1}^{x_1} \otimes \cdots \otimes \rho_{B_n}^{x_n}. \tag{6}$$

The (weak) conditionally typical subspace corresponding to a sequence x^n is defined as

$$T_{B^n|x^n}^\delta \equiv \text{span}\left\{|y_{x^n}^n\rangle : \left|-\log p_{Y^n|X^n}(y^n|x^n) - H(B|X)_\omega\right| \leq \delta\right\}. \tag{7}$$

The conditionally typical projector onto this space is written as $\Pi_{B^n|x^n}^\delta$. Then one can show that

$$\mathbb{E}_{X^n}\left\{\text{Tr}\{\Pi_{B^n|X^n}^\delta \rho_{B^n}^{X^n}\}\right\} \geq 1 - \varepsilon \tag{8}$$

for all $\varepsilon \in (0,1)$, $\delta > 0$, and sufficiently large n. It also follows from the definition that

$$\Pi_{B^n|x^n}^\delta \rho_{B^n}^{x^n} \Pi_{B^n|x^n}^\delta \leq 2^{-n[H(B|X)_\omega - \delta]} \Pi_{B^n|x^n}^\delta. \tag{9}$$

So the strategy consists of picking codewords independently at random according to a distribution $p_X(x)$ which maximizes (4). The quantum codewords that Bob receives have the following form:

$$\{\rho_{B^n}^{x^n(m)}\}_{m \in \mathcal{M}}. \tag{10}$$

To test whether Bob received the mth message, he performs the following measurement:

$$\{\Pi_{B^n|x^n(m)}^\delta, \hat{\Pi}_{B^n|x^n(m)}^\delta\}, \tag{11}$$

where $\hat{\Pi}_{B^n|x^n(m)}^\delta \equiv I^{\otimes n} - \Pi_{B^n|x^n(m)}^\delta$The error probability when sending the mth message for this scheme is then

$$1 - \text{Tr}\{\Pi_{B^n|x^n(m)}^\delta \hat{\Pi}_{B^n|x^n(m-1)}^\delta \cdots \hat{\Pi}_{B^n|x^n(1)}^\delta \rho_{B^n}^{x^n(m)} \hat{\Pi}_{B^n|x^n(1)}^\delta \cdots \hat{\Pi}_{B^n|x^n(m-1)}^\delta \Pi_{B^n|x^n(m)}^\delta\}. \tag{12}$$

Using arguments similar to the pure-state channel, after an expectation over the messages and codebook, we can argue that this is $\approx$

$$\begin{aligned}&\text{Tr}\{\Pi_{B^n}^\delta \rho_{B^n}^{x^n(m)} \Pi_{B^n}^\delta\} \\ &- \text{Tr}\{\Pi_{B^n|x^n(m)}^\delta \hat{\Pi}_{B^n|x^n(m-1)}^\delta \cdots \hat{\Pi}_{B^n|x^n(1)}^\delta \Pi_{B^n}^\delta \rho_{B^n}^{x^n(m)} \Pi_{B^n}^\delta \hat{\Pi}_{B^n|x^n(1)}^\delta \cdots \hat{\Pi}_{B^n|x^n(m-1)}^\delta \Pi_{B^n|x^n(m)}^\delta\},\end{aligned} \tag{13}$$

where $\Pi^{\delta}_{B^n}$ is the unconditionally typical projector for $\sum_x p_X(x)\rho^x_B$. We then apply the non-commutative union bound and argue as before to get the following upper bound on the error probability:

$$2\sqrt{\varepsilon' + |\mathcal{M}|\, 2^{-n[I(X;B)-2\delta]}}, \tag{14}$$

so that this decays exponentially with n by picking $|\mathcal{M}| = 2^{n[I(X;B)-3\delta]}$. Combined with derandomization and expurgation, we can conclude that $I(X;B)$ is an achievable rate.

4 General Channels

To get a strategy for any channel, note that we can induce a cq channel from any channel $\mathcal{N}_{A\to B}$ by picking an input ensemble of the form $\{p_X(x), \rho^x_A\}$. Then the cq state representing the input-output correlations is as follows:

$$\sigma_{XB} \equiv \sum_x p_X(x)|x\rangle\langle x|_X \otimes \mathcal{N}_{A\to B}(\rho^x_A). \tag{15}$$

So using the above strategy, the quantum codewords for Bob are

$$\{\mathcal{N}^{\otimes n}_{A\to B}(\rho^{x^n(m)}_{A^n}) \equiv \mathcal{N}_{A\to B}(\rho^{x_1(m)}_A) \otimes \cdots \otimes \mathcal{N}_{A\to B}(\rho^{x_n(m)}_A)\}_{m\in\mathcal{M}}, \tag{16}$$

and the above demonstrates that an achievable rate is

$$I(X;B)_\sigma. \tag{17}$$

Optimizing over all possible input ensembles gives a quantity known as the Holevo information of the channel:

$$\chi(\mathcal{N}) \equiv \max_{\{p_X(x),\rho^x\}} I(X;B)_\sigma. \tag{18}$$

Of course, we could then form codes for the tensor-product channel $\mathcal{N}^{\otimes k}$, and by a limiting argument, we find that the following regularized Holevo information is achievable:

$$\lim_{k\to\infty} \frac{1}{k}\chi(\mathcal{N}^{\otimes k}). \tag{19}$$

This is the best known expression for an achievable rate for a quantum channel, but for some channels, it simplifies so that we can say that $\chi(\mathcal{N})$ is achievable.

4.1 The Converse Theorem

The second part of the classical capacity theorem is the converse theorem, and we provide a simple proof of it in this section. Suppose that Alice and Bob are trying to accomplish randomness distribution rather than classical communication—the capacity for such a task can only be larger than that for classical communication. Recall that in such a task, Alice first prepares a maximally correlated state $\overline{\Phi}_{MM'}$ so that the rate $C - \delta$ of randomness distribution is equal to $\frac{1}{n}\log_2 |M|$. Alice and Bob share a given state after encoding, channel transmission, and decoding. We now show that the regularized Holevo information bounds the rate of randomness distribution for any protocol that has vanishing error in the asymptotic limit. As a result, the regularized Holevo

information also upper bounds the capacity for classical communication. Consider the following chain of inequalities:

$$\begin{aligned}
\log|M| &= I\left(M;M'\right)_{\overline{\Phi}} && (20)\\
&\leq I\left(M;M'\right)_{\omega} + f(n,\varepsilon) && (21)\\
&\leq I\left(M;B^n\right)_{\omega} + f(n,\varepsilon) && (22)\\
&\leq \chi(\mathcal{N}^{\otimes n}) + f(n,\varepsilon). && (23)
\end{aligned}$$

The first equality follows because the mutual information of the common randomness state $\overline{\Phi}_{MM'}$ is equal to $n\left(C-\delta\right)$ bits. The first inequality follows from the error criterion and by applying the AFW inequality for quantum mutual information with $f(n,\varepsilon) \equiv 6\varepsilon\log|M| + 4h_2(\varepsilon)$. This function has the property that $\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}f(n,\varepsilon) = 0$. The second inequality results from the quantum data-processing inequality for quantum mutual information—recall that Bob processes the B^n system with a quantum instrument to get the classical system M'. Also, the quantum mutual information is evaluated on a classical–quantum state. The final inequality follows because this classical–quantum state has a particular distribution and choice of states, and this choice always leads to a value of the quantum mutual information that cannot be greater than the Holevo information of the tensor product channel $\mathcal{N}^{\otimes n}$. Putting everything together, we find that

$$C - \delta \leq \frac{1}{n}\chi(\mathcal{N}^{\otimes n}) + \frac{1}{n}f(n,\varepsilon). \tag{24}$$

Taking the limit as $n\to\infty$ and as $\varepsilon,\delta\to 0$ then establishes that an achievable rate C necessarily satisfies $C \leq \chi_{\text{reg}}(\mathcal{N})$, where $\chi_{\text{reg}}(\mathcal{N})$ is the regularized Holevo formula.

5 Additivity

Observe that the final upper bound in (23) on the rate C is the multi-letter Holevo information of the channel. It would be more desirable to have $\chi(\mathcal{N})$ as the upper bound on C rather than $\frac{1}{n}\chi(\mathcal{N}^{\otimes n})$ because the former is simpler, but the optimization problem set out in the latter quantity is simply impossible to compute with finite computational resources. However, the upper bound in (23) is the best known upper bound if we do not know anything else about the structure of the channel, and for this reason, the best known characterization of the classical capacity is the regularized Holevo information.

If we know that the Holevo information of the tensor product of a certain channel with itself is additive, then there is no need for the regularization $\chi_{\text{reg}}(\mathcal{N})$, and the HSW characterization reduces to a very good one: the Holevo information $\chi(\mathcal{N})$. There are many examples of channels for which the classical capacity reduces to the Holevo information of the channel, and we detail three such classes of examples in this section: the cq channels, the quantum Hadamard channels, and the quantum depolarizing channels. The proof that demonstrates additivity of the Holevo information for each of these channels depends explicitly on structural properties of each one, and there is unfortunately not much to learn from these proofs in order to say anything about additivity of the Holevo information of general quantum channels. Nevertheless, it is good to have some natural channels for which we can compute their classical capacity, and it is instructive to examine these proofs in detail to understand what it is about each channel that makes their Holevo information additive.

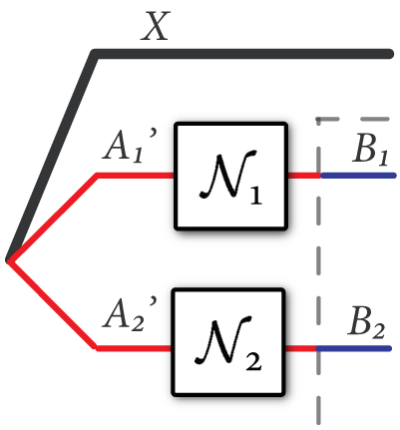

Figure 1: This figure displays the scenario for determining whether the Holevo information of two quantum channels $\mathcal{N}_1$ and $\mathcal{N}_2$ is additive. The question of additivity is equivalent to the possibility of quantum correlations being able to enhance the Holevo information of two quantum channels. The result proved in Theorem 1 is that the Holevo information is additive for the tensor product of an entanglement-breaking channel and any other quantum channel, so that quantum correlations cannot enhance the Holevo information in this case. This is perhaps intuitive because an entanglement-breaking channel destroys quantum correlations in the form of quantum entanglement.

The Holevo information of a quantum channel is generally not additive (by no means is this obvious!). The question of additivity for this case is *not* whether classical correlations can enhance the Holevo information, but it is *rather* whether quantum correlations can enhance it. That is, Alice can choose an ensemble of the form $\{p_X(x), \rho^x_{A_1A_2}\}$ for input to two uses of the quantum channel. The conditional density operators $\rho^x_{A_1A_2}$ can be entangled and these quantum correlations can potentially increase the Holevo information.

The question of additivity of the Holevo information of a quantum channel was a longstanding open conjecture in quantum information theory—many researchers thought that quantum correlations would not enhance it and that additivity would hold. But recent research has demonstrated a counterexample to the additivity conjecture, and perhaps unsurprisingly in hindsight, this counterexample exploits maximally entangled states to demonstrate superadditivity (see Section **??**). Figure 1 displays the scenario corresponding to the question of additivity of the Holevo information.

Additivity of Holevo information may not hold for all quantum channels, but it is possible to prove its additivity for certain classes of quantum channels. One such class for which additivity holds is the class of entanglement-breaking channels, and the proof of additivity is perhaps the simplest for this case.

Theorem 1 (Additivity for Entanglement-Breaking Channels)**.** *Suppose that a quantum channel $\mathcal{N}^{\text{EB}}$ is entanglement breaking and another channel $\mathcal{M}$ is arbitrary. Then the Holevo information $\chi(\mathcal{N}^{\text{EB}} \otimes \mathcal{M})$ of the tensor-product channel $\mathcal{N}^{\text{EB}} \otimes \mathcal{M}$ is the sum of the individual Holevo informations $\chi(\mathcal{N}^{\text{EB}})$ and $\chi(\mathcal{M})$:*

$$\chi(\mathcal{N}^{\text{EB}} \otimes \mathcal{M}) = \chi(\mathcal{N}^{\text{EB}}) + \chi(\mathcal{M}). \tag{25}$$

Proof. The trivial inequality $\chi(\mathcal{N}^{\text{EB}} \otimes \mathcal{M}) \geq \chi(\mathcal{N}^{\text{EB}}) + \chi(\mathcal{M})$ holds for any two quantum channels

$\mathcal{N}^{\text{EB}}$ and $\mathcal{M}$ because we can choose the input ensemble on the left-hand side to be a tensor product of the ones that individually maximize the terms on the right-hand side.

We now prove the non-trivial inequality $\chi(\mathcal{N}^{\text{EB}} \otimes \mathcal{M}) \leq \chi(\mathcal{N}^{\text{EB}}) + \chi(\mathcal{M})$ that holds when $\mathcal{N}^{\text{EB}}$ is entanglement breaking. Let $\rho_{XB_1B_2}$ be a state that maximizes the Holevo information $\chi(\mathcal{N}^{\text{EB}} \otimes \mathcal{M})$, where

$$\rho_{XB_1B_2} \equiv (\mathcal{N}^{\text{EB}}_{A_1 \to B_1} \otimes \mathcal{M})(\rho_{XA_1A_2}), \tag{26}$$

$$\rho_{XA_1A_2} \equiv \sum_x p_X(x)|x\rangle\langle x|_X \otimes \rho^x_{A_1A_2}. \tag{27}$$

The action of $\mathcal{N}^{\text{EB}}_{A_1 \to B_1}$ is to break entanglement. Let $\rho_{XB_1A_2}$ be the state after only the entanglement-breaking channel $\mathcal{N}^{\text{EB}}_{A_1 \to B_1}$ acts. We can write this state as follows:

$$\rho_{XB_1A_2} \equiv \mathcal{N}^{\text{EB}}_{A_1 \to B_1}(\rho_{XA_1A_2}) \tag{28}$$

$$= \sum_x p_X(x)|x\rangle\langle x|_X \otimes \mathcal{N}^{\text{EB}}_{A_1 \to B_1}(\rho^x_{A_1A_2}) \tag{29}$$

$$= \sum_x p_X(x)|x\rangle\langle x|_X \otimes \sum_y p_{Y|X}(y|x)\ \sigma^{x,y}_{B_1} \otimes \theta^{x,y}_{A_2} \tag{30}$$

$$= \sum_{x,y} p_{Y|X}(y|x)p_X(x)|x\rangle\langle x|_X \otimes \sigma^{x,y}_{B_1} \otimes \theta^{x,y}_{A_2}. \tag{31}$$

The third equality follows because the channel $\mathcal{N}^{\text{EB}}$ breaks any entanglement in the state $\rho^x_{A_1A_2}$, leaving behind a separable state $\sum_y p_{Y|X}(y|x)\ \sigma^{x,y}_{B_1} \otimes \theta^{x,y}_{A_2}$. Then the state $\rho_{XB_1B_2}$ has the form

$$\rho_{XB_1B_2} = \sum_{x,y} p_{Y|X}(y|x)p_X(x)|x\rangle\langle x|_X \otimes \sigma^{x,y}_{B_1} \otimes \mathcal{M}(\theta^{x,y}_{A_2}). \tag{32}$$

Let $\omega_{XYB_1B_2}$ be an extension of $\rho_{XB_1B_2}$ where

$$\omega_{XYB_1B_2} \equiv \sum_{x,y} p_{Y|X}(y|x)p_X(x)|x\rangle\langle x|_X \otimes |y\rangle\langle y|_Y \otimes \sigma^{x,y}_{B_1} \otimes \mathcal{M}(\theta^{x,y}_{A_2}), \tag{33}$$

and $\text{Tr}_Y\{\omega_{XYB_1B_2}\} = \rho_{XB_1B_2}$. Then the following chain of inequalities holds

$$\chi(\mathcal{N}^{\text{EB}} \otimes \mathcal{M}) = I(X; B_1B_2)_\rho \tag{34}$$

$$= I(X; B_1)_\rho + I(X; B_2|B_1)_\rho \tag{35}$$

$$\leq \chi(\mathcal{N}^{\text{EB}}) + I(X; B_2|B_1)_\rho \tag{36}$$

The first equality follows because we took $\rho_{XB_1B_2}$ to be a state that maximizes the Holevo information $\chi(\mathcal{N}^{\text{EB}} \otimes \mathcal{M})$ of the tensor-product channel $\mathcal{N}^{\text{EB}} \otimes \mathcal{M}$. The second equality is an application of the chain rule for conditional mutual information (Property **??**). The inequality follows because the Holevo information $I(X; B_1)_\rho$ is with respect to the following state:

$$\rho_{XB_1} \equiv \sum_x p_X(x)|x\rangle\langle x|_X \otimes \mathcal{N}^{\text{EB}}_{A_1 \to B_1}(\rho^x_{A_1}), \tag{37}$$

whereas the Holevo information of the channel $\mathcal{N}^{\text{EB}}_{A_1 \to B_1}$ is defined to be the maximal Holevo information with respect to all input ensembles. Now let us focus on the term $I(X; B_2|B_1)_\rho$. Consider

that

$$\begin{aligned} I(X;B_2|B_1)_\rho &= I(X;B_2|B_1)_\omega && (38)\\ &\leq I(XB_1;B_2)_\omega && (39)\\ &\leq I(XYB_1;B_2)_\omega && (40)\\ &= I(XY;B_2)_\omega + I(B_1;B_2|XY)_\omega && (41)\\ &= I(XY;B_2)_\omega && (42)\\ &\leq \chi(\mathcal{M}). && (43) \end{aligned}$$

The first equality follows because the reduced state of $\omega_{XYB_1B_2}$ on systems X, B_1, and B_2 is equal to $\rho_{XB_1B_2}$. The first inequality follows from the chain rule: $I(X;B_2|B_1) = I(XB_1;B_2) - I(B_1;B_2) \leq I(XB_1;B_2)$. The second inequality follows from the quantum data-processing inequality. The second equality is again from the chain rule for conditional mutual information. The third equality is the crucial one that exploits the entanglement-breaking property. It follows by examining (33) and observing that the state $\omega_{XYB_1B_2}$ on systems B_1 and B_2 is product when conditioned on classical variables X and Y, so that the conditional mutual information between systems B_1 and B_2 given both X and Y is equal to zero. The final inequality follows because ω_{XYB_2} is a particular state of the form needed in the maximization of $\chi(\mathcal{M})$. □

Corollary 2. *The regularized Holevo information of an entanglement-breaking quantum channel $\mathcal{N}^{\text{EB}}$ is equal to its Holevo information:*

$$\chi_{\text{reg}}(\mathcal{N}^{\text{EB}}) = \chi(\mathcal{N}^{\text{EB}}). \tag{44}$$

Proof. The proof of this property uses the same induction argument as in Corollary **??** and exploits the additivity property in Theorem 1 above. □

5.1 Optimizing the Holevo Information

5.1.1 Pure States are Sufficient

The following theorem allows us to simplify the optimization problem given by the Holevo information of a channel—we show that it is sufficient to consider ensembles of pure states at the input.

Theorem 3. *It is sufficient to maximize the Holevo information over only pure states:*

$$\chi(\mathcal{N}) = \max_{\rho_{XA}} I(X;B)_\rho = \max_{\tau_{XA}} I(X;B)_\tau, \tag{45}$$

where

$$\tau_{XA} \equiv \sum_x p_X(x)|x\rangle\langle x|_X \otimes |\phi_x\rangle\langle\phi_x|_A, \tag{46}$$

and ρ_{XB} and τ_{XB} are the states that results from sending the A system of ρ_{XA} and τ_{XA} through the quantum channel $\mathcal{N}_{A\to B}$.

Proof. Suppose that ρ_{XA} is any cq state input to the channel. Consider a spectral decomposition of the states ρ_A^x:

$$\rho_A^x = \sum_y p_{Y|X}(y|x)\psi_A^{x,y}, \tag{47}$$

where the states $\psi_A^{x,y}$ are pure. Then let σ_{XYA} denote the following state:

$$\sigma_{XYA} \equiv \sum_x p_{Y|X}(y|x)p_X(x)|x\rangle\langle x|_X \otimes |y\rangle\langle y|_Y \otimes \psi_A^{x,y}, \tag{48}$$

so that $\mathrm{Tr}_Y\{\sigma_{XYA}\} = \rho_{XA}$. Also, observe that σ_{XYA} is a state of the form τ_{XA} with XY as the classical system. Let σ_{XYB} denote the state that results from sending the A system through the quantum channel $\mathcal{N}_{A\to B}$. Then the following relations hold:

$$I(X;B)_\rho = I(X;B)_\sigma \leq I(XY;B)_\sigma. \tag{49}$$

The equality follows because $\mathrm{Tr}_Y\{\sigma_{XYB}\} = \rho_{XB}$ and the inequality follows from the quantum data-processing inequality. It then suffices to consider ensembles with only pure states because the state σ_{XYB} is a state of the form τ_{XB} with the combined system XY acting as the classical system. □