16 Channel coding

A **binary symmetric channel** with bit flip probability f is a model of a noisy classical communication channel. With each use of the channel we get to input a binary value, and the channel output is, with probability $1-f$, the same value or, with probability f, the other value. Furthermore, each use of the channel behaves independently of the other uses. If we make n uses of the channel and the input is $X^{(n)} = (X_1, \ldots, X_n)$ and the output is $Y^{(n)} = (Y_1, \ldots, Y_n)$

$$P_{Y^{(n)}|X^{(n)}}((y_1, \ldots, y_n)|(x_1, \ldots, x_n)) = \Pi_{i=1}^{n} P_{Y|X}(y_i|x_i), \tag{16.1}$$

where

$$P_{Y|X}(y|x) = \begin{cases} 1-f \text{ if } y = x \\ f \text{ if } y \neq x. \end{cases} \tag{16.2}$$

A channel where each use behaves identically and independently, which is the property described by equation (16.1), is called a **memoryless channel**.

Suppose that Alice is connected to Bob by this channel, so that she makes inputs and Bob receives outputs. Suppose Alice wants to transmit a single bit M to Bob using the channel. Let us denote by $\hat{M}$ the decoding of the message that Bob makes based on the channel output. If only one use of the channel is made then it is not hard to see that, for $f \leq 1/2$, the worst-case error probability

$$\max_{m \in \{0,1\}} \Pr(\hat{M} \neq M | M = m)$$

will be at least f. Provided $f < 1/2$, it is possible to do better by making more uses of the channel: A "repetition code" of "blocklength" n, makes n uses of the channel to send one bit. If $M = 0$ then Alice sends a string of n zeros and if $M = 1$ then she sends a string of n ones. Bob decodes the output of the channel by "majority vote": If more than half of the output symbols are ones, then he decodes to one, and otherwise to zero. For $n = 3$, we have

$$\Pr(\hat{M} = 0 | M = 1) = \Pr(\hat{M} = 1 | M = 0) = f^3 + \binom{3}{1} f^2(1-f), \tag{16.3}$$

which is roughly 0.03 when $f = 0.1$. It is not hard to show that, for any $f < 1/2$, the two error probabilities go to zero as n increases, but so does the *rate* of the code which is $1/n$ bits per channel use. As we will see, in general it is possible to do much better.

16.1 Coding over quantum channels

Definition 1. A (k, ϵ)-code for an operation $\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}$ consists of

- an encoding function $\underline{\rho}$ from $\mathcal{A}_M = \{1, \ldots, k\}$ to states of A and
- a decoding POVM $E : \mathcal{A}_M \to \mathcal{L}(\mathcal{H}_\mathsf{B})$

such that

$$\forall m \in \mathcal{A}_M, \mathrm{Tr}(\mathbb{1} - E(m))\mathcal{N}\underline{\rho}(m) \leq \epsilon. \tag{16.4}$$

We call k the "size" of the code and define

$$k_\epsilon(\mathcal{N}) := \max\{k \in \mathbb{N} : \text{ There exists a } (k, \epsilon)\text{-code for } \mathcal{N}\}.$$

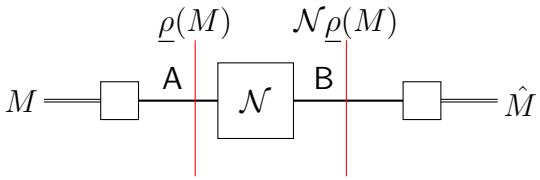


Figure 16.1: Alice uses $\underline{\rho}$ to encode a random variable M (taking values in $\mathcal{A}_M$) as a state $\underline{\rho}(M)$ of system A. The operation $\mathcal{N}$ is applied leaving B in the state $\mathcal{N}\underline{\rho}(M)$. Bob measures the POVM E on B producing a result $\hat{M}$ (also taking values in $\mathcal{A}_M$).

The condition (16.4) means that, if Alice uses a (k, ϵ)-code to transmit a random variable M, as in the figure, then the worst-case error probability is no more than ϵ:

$$\forall m \in \mathcal{A}_M : \Pr(\hat{M} \neq M | M = m) = \mathrm{Tr}(\mathbb{1} - E(m))\mathcal{N}\underline{\rho}(m) \leq \epsilon. \tag{16.5}$$

16.1.1 Memoryless quantum channels

A **memoryless quantum channel** is completely specified by its behaviour for a single use, which is given by some operation $\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}$ where A is an input system for a single use of the channel and B the corresponding output system. Making n uses of the channel on n input systems $\mathsf{A}_1, \ldots, \mathsf{A}_n$ in an arbitrary state $\rho_{\mathsf{A}_1\cdots\mathsf{A}_n}$ results in output systems $\mathsf{B}_1, \ldots, \mathsf{B}_n$ (each of the same dimension as B) in the state

$$\mathcal{N}^{\otimes n}\rho_{\mathsf{A}_1\cdots\mathsf{A}_n} \text{ where } \mathcal{N}^{\otimes n} = \bigotimes_{i=1}^{n} \mathcal{N}^{\mathsf{B}_i\leftarrow\mathsf{A}_i}$$

and $\mathcal{N}^{\mathsf{B}_i\leftarrow\mathsf{A}_i} = \mathbf{id}^{\mathsf{B}_i\leftarrow\mathsf{B}}\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}\mathbf{id}^{\mathsf{A}\leftarrow\mathsf{A}_i}$. The **classical capacity** of the channel (in bits per channel use) is

Definition 2.

$$C(\mathcal{N}) := \lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\lfloor\log(k_\epsilon(\mathcal{N}^{\otimes n}))\rfloor. \tag{16.6}$$

16.2 The Holevo bound and the HSW theorem

Suppose that X is a random variable, $\underline{\sigma}$ is a function from $\mathcal{A}_X$ to states of a system B, and the system B is in the state $\underline{\sigma}(X)_\mathsf{B}$. If X is stored in system $\tilde{\mathsf{X}}$ then the state of $\tilde{\mathsf{X}}\mathsf{B}$ is $\sigma_{\tilde{\mathsf{X}}\mathsf{B}} = \sum_{x\in\mathcal{A}_X} P_X(x)|x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes \underline{\sigma}(x)_\mathsf{B}$.

Proposition 3. In the situation just described $H(\mathsf{B}|\tilde{\mathsf{X}})_\sigma = \sum_x P_X(x) S(\underline{\sigma}(x)_\mathsf{B})$.

Proof. For each x, let $\underline{\sigma}(x) = \sum_{j=1}^{d_\mathsf{B}} \lambda(j|x)|\alpha_j^{(x)}\rangle\langle\alpha_j^{(x)}|$ be an eigendecomposition. $\sum_{j=1}^{d_\mathsf{B}} \lambda(j|x) = 1$ and $\lambda(j|x) \geq 0$. Therefore, we can imagine that there is a random variable J distributed such that $P_{J|X=x}(j) = \lambda(j|x)$ and, therefore, $P_{JX} = \lambda(j|x)P_X$ and

$$\sigma_{\tilde{\mathsf{X}}\mathsf{B}} = \sum_x \sum_j P_{JX}(j,x)|x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes |\alpha_j^{(x)}\rangle\langle\alpha_j^{(x)}|_\mathsf{B}$$

is an eigendecomposition. Therefore, using the chain rule,

$$\begin{aligned} S(\sigma_{\tilde{\mathsf{X}}\mathsf{B}}) =& S(P_{XJ}) = H(X,J) = H(X) + H(J|X) && (16.7)\\ =& H(X) + \sum_x H(J|X=x)P_X(x) && (16.8)\\ =& S(P_X) + \sum_x S(P_{J|X=x})P_X(x) && (16.9)\\ =& S(\sigma_{\tilde{\mathsf{X}}}) + \sum_x S(\underline{\sigma}(x)_\mathsf{B})P_X(x), && (16.10) \end{aligned}$$

and, since $H(\mathsf{B}|\tilde{\mathsf{X}}) = S(\sigma_{\tilde{\mathsf{X}}\mathsf{B}}) - S(\sigma_{\tilde{\mathsf{X}}})$, we have the result. □

If a POVM E is measured on B with result Y, then the **Holevo bound** says that

$$I(X:Y) \leq I(\tilde{\mathsf{X}}:\mathsf{B})_\sigma = S\left(\sum_{x\in\mathcal{A}_X} P_X(x)\underline{\sigma}(x)_\mathsf{B}\right) - \sum_{x\in\mathcal{A}_X} P_X(x) S\left(\underline{\sigma}(x)_\mathsf{B}\right). \tag{16.11}$$

Here the equality comes from $I(\tilde{\mathsf{X}}:\mathsf{B})_\sigma = H(\mathsf{B})_\sigma - H(\mathsf{B}|\tilde{\mathsf{X}})_\sigma$ and the previous proposition. The inequality follows almost immediately from the data processing inequality for QMI: Let $\mathcal{D}^{\tilde{\mathsf{Y}}\leftarrow\mathsf{B}}$ be the operation

$$\mathcal{D}^{\tilde{\mathsf{Y}}\leftarrow\mathsf{B}} : \tau_\mathsf{B} \mapsto \sum_{y\in\mathcal{A}_Y} |y\rangle\langle y|_{\tilde{\mathsf{Y}}} \mathrm{Tr} E(y)_\mathsf{B} \tau_\mathsf{B} \tag{16.12}$$

which measures the POVM E on B and store the result Y in system $\tilde{\mathsf{Y}}$ in the standard way. Then the DPI for QMI tells us that

$$I(X:Y) = I(\tilde{\mathsf{X}}:\tilde{\mathsf{Y}})_{\mathcal{D}^{\tilde{\mathsf{Y}}\leftarrow\mathsf{B}}\sigma_{\tilde{\mathsf{X}}\mathsf{B}}} \leq I(\tilde{\mathsf{X}}:\mathsf{B})_{\sigma_{\tilde{\mathsf{X}}\mathsf{B}}}. \tag{16.13}$$

16.2.1 The Holevo information of an operation

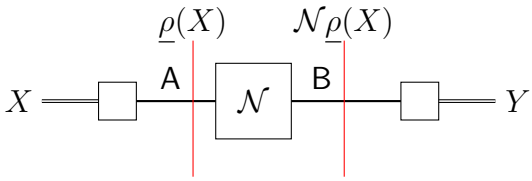
Suppose that, rather than the state $\underline{\sigma}(X)_{\mathsf{B}}$ being prepared directly, it is the result of first preparing the state $\underline{\rho}(X)_{\mathsf{A}}$ which is then which is then acted on by an operation $\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}$. If we maximise $I(X:Y)$ over all choices of the set $\mathcal{A}_X$, distribution P_X and function $\underline{\rho}$, as well as over all POVMs E, then we can regard the result as a measure of how large a classical correlation can be created via the operation. From the Holevo bound, it follows immediately that the maximal value of $I(X:Y)$ is bounded above by $\chi(\mathcal{N})$, the **Holevo information** of the operation $\mathcal{N}$, which is defined by

Definition 4.

$$\chi(\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}) := \max_{\mathcal{A}_X, P_X, \underline{\rho}} I(\tilde{\mathsf{X}}:\mathsf{B})_{\sigma_{\tilde{\mathsf{X}}\mathsf{B}}} = \max_{\mathcal{A}_X, P_X, \underline{\rho}} \left(S\left(\sum_x P_X(x)\mathcal{N}\underline{\rho}(x)\right) - \sum_x P_X(x) S\left(\mathcal{N}\underline{\rho}(x)\right)\right)$$

where $\sigma_{\tilde{\mathsf{X}}\mathsf{B}} = \sum_{x\in\mathcal{A}_X} P_X(x)|x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes (\mathcal{N}\underline{\rho}(x))_{\mathsf{B}}$.

Remark 5. It is possible to show that the maximum is achieved for some P_X and $\underline{\rho}$ when $|\mathcal{A}_X| = d_{\mathsf{A}}^2$.



Proposition 6. Suppose a system A is prepared in the state $\underline{\rho}(X)$, and then operation $\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}$ acts, leaving B in the state $\mathcal{N}\underline{\rho}(X)$, and then a POVM E is measured on B producing result Y. (The situation illustrated in the figure above.) For any $\mathcal{A}_X$, P_X, $\underline{\rho}$, $\mathcal{A}_Y$ and POVM E, the mutual information $I(X:Y)$ is no more than the Holevo information $\chi(\mathcal{N})$ of the operation $\mathcal{N}$.

♣♣The proof of the following proposition is an exercise for example class 4.

Proposition 7. For any two operations $\mathcal{M}$ and $\mathcal{N}$, $\chi(\mathcal{N}\otimes\mathcal{M}) \geq \chi(\mathcal{N}) + \chi(\mathcal{M})$.

16.2.2 The Holevo-Schumacher-Westmoreland (HSW) theorem

It turns out that the classical capacity of an operation $\mathcal{N}$ is equal to its "regularised" Holevo information:

Theorem 8. For any operation $\mathcal{N}$, $C(\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n}\chi(\mathcal{N}^{\otimes n})$.

This result is known as the Holevo-Schumacher-Westmoreland (HSW) theorem, and we will prove it, starting with the "converse" which shows that the right-hand-side is an upper bound on the left-hand-side. (That the limit on the RHS exists follows from Proposition 7 and Fekete's lemma - a basic result in analysis.)

16.3 HSW theorem: Converse part

Proposition 9. For any $\epsilon \in [0, 1)$ and operation $\mathcal{N}$

$$\log(k_\epsilon(\mathcal{N})) \leq \frac{\chi(\mathcal{N}) + 1}{1 - \epsilon}. \tag{16.14}$$

Proof. Suppose that there is a (k, ϵ)-code for $\mathcal{N}$ which we use to transmit a uniformly distributed message M (taking values in $\{1, \ldots, k\}$). Let $\hat{M}$ be the result of measuring the decoding POVM. The bound on the worst-case error probability implies that

$$P(\hat{M} \neq M) \leq \epsilon. \tag{16.15}$$

Using the fact that M is uniformly distributed, Fano's inequality, and $h(x) \leq 1$ for all $x \in [0, 1]$, and (16.15), we have

$$\begin{aligned} I(M : \hat{M}) =& H(M) - H(M|\hat{M}) && (16.16)\\ =& \log(k) - H(M|\hat{M}) && (16.17)\\ \geq& \log(k) - \Big(\Pr(\hat{M} \neq M) \log(k - 1) + 1\Big) && (16.18)\\ \geq& (1 - \epsilon) \log(k) - 1. && (16.19) \end{aligned}$$

Rearranging this and using Proposition 6 we have

$$\log(k) \leq \frac{I(M : \hat{M}) + 1}{1 - \epsilon} \leq \frac{\chi(\mathcal{N}) + 1}{1 - \epsilon}, \tag{16.20}$$

as claimed. □

Corollary 10 (HSW theorem: Converse part)**.**

$$C(\mathcal{N}) \leq \lim_{n \to \infty} \frac{1}{n} \chi(\mathcal{N}^{\otimes n}). \tag{16.21}$$

Proof.

$$C(\mathcal{N}) \leq \lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \frac{\chi(\mathcal{N}^{\otimes n}) + 1}{1 - \epsilon} = \lim_{\epsilon \to 0} (1 - \epsilon)^{-1} \lim_{n \to \infty} \frac{1}{n} \chi(\mathcal{N}^{\otimes n}).$$

□

We will return to the achievability part later, but before that, we will show how the formula given for the capacity simplifies for certain kinds of channel.

16.4 Capacity of entanglement breaking channels

Definition 11. An operation $\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}$ is **entanglement breaking** if, for all systems R and states ρ_{RA}, $\sigma_{\mathsf{RB}} = \mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}\rho_{\mathsf{RA}}$ is separable.

Theorem 12. If $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}$ and $\mathcal{N}^{\mathsf{D}\leftarrow\mathsf{C}}$ are operations, and $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}$ is entanglement-breaking, then $\chi(\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}} \otimes \mathcal{N}^{\mathsf{D}\leftarrow\mathsf{C}}) = \chi(\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}) + \chi(\mathcal{N}^{\mathsf{D}\leftarrow\mathsf{C}})$.

Proof. We already know that $\chi(\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}} \otimes \mathcal{N}^{\mathsf{D}\leftarrow\mathsf{C}}) \geq \chi(\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}) + \chi(\mathcal{N}^{\mathsf{D}\leftarrow\mathsf{C}})$ so we just need to show the opposite inequality.

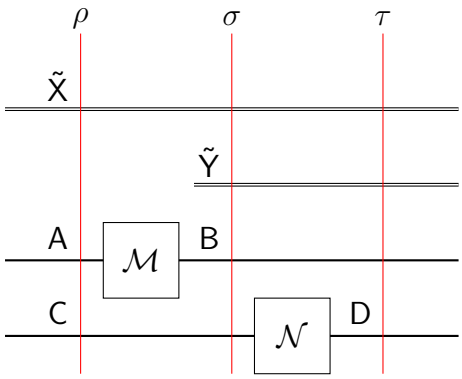


Figure 16.2: The systems, operations and states referred to in the proof.

Given some $\mathcal{A}_X$, P_X, and $\underline{\rho}$ let

1. $\rho_{\tilde{\mathsf{X}}\mathsf{AC}} = \sum_x P_X(x)|x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes \underline{\rho}(x)_{\mathsf{AC}}$.
2. $\sigma_{\tilde{\mathsf{X}}\mathsf{BC}} := \mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}\rho_{\tilde{\mathsf{X}}\mathsf{AC}} = \sum_x P_X(x)|x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes (\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}\underline{\rho}(x)_{\mathsf{AC}})$. Because $\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}$ is entanglement breaking, we know that, for all $x \in \mathcal{A}_X$,
$$\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}\underline{\rho}(x)_{\mathsf{AC}} = \sum_r q(r|x)\underline{\beta}(x,r)_{\mathsf{B}} \otimes \underline{\gamma}(x,r)_{\mathsf{C}}$$
where, for all r, $\underline{\beta}(x,r)_{\mathsf{B}}$ and $\underline{\gamma}(x,r)_{\mathsf{C}}$ are states, $q(r|x) \geq 0$ and $\sum_r q(r|x) = 1$.
3. $\sigma_{\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{BC}} := \sum_x \sum_r q(r|x)P_X(x)|x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes |r\rangle\langle r|_{\tilde{\mathsf{R}}}\underline{\beta}(x,r)_{\mathsf{B}} \otimes \underline{\gamma}(x,r)_{\mathsf{C}}$ defines an extension of $\sigma_{\tilde{\mathsf{X}}\mathsf{BC}}$.
4. $\tau_{\tilde{\mathsf{X}}\mathsf{BD}} := \mathcal{N}^{\mathsf{D}\leftarrow\mathsf{C}}\sigma_{\tilde{\mathsf{X}}\mathsf{BC}} = \mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}} \otimes \mathcal{N}^{\mathsf{D}\leftarrow\mathsf{C}}\rho_{\tilde{\mathsf{X}}\mathsf{AC}}$. So, $\tau_{\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{BD}} := \mathcal{N}^{\mathsf{D}\leftarrow\mathsf{C}}\sigma_{\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{BC}}$ defines an extension of $\tau_{\tilde{\mathsf{X}}\mathsf{BD}}$.

Let $\mathcal{R}^{\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{B}\leftarrow\tilde{\mathsf{X}}\tilde{\mathsf{R}}}$ be the operation which measures the computational basis of $\tilde{\mathsf{X}}\tilde{\mathsf{R}}$ (measures the PVM) and adds system B prepared in the state $\underline{\beta}(x,r)_{\mathsf{B}}$, that is

$$\mathcal{R}^{\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{B}\leftarrow\tilde{\mathsf{X}}\tilde{\mathsf{R}}} : \eta_{\tilde{\mathsf{X}}\tilde{\mathsf{R}}} \mapsto \sum_{x,r}(|x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes |r\rangle\langle r|_{\tilde{\mathsf{R}}}\eta_{\tilde{\mathsf{X}}\tilde{\mathsf{R}}}|x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes |r\rangle\langle r|_{\tilde{\mathsf{R}}}) \otimes \underline{\beta}(x,r)_{\mathsf{B}}.$$

Since $\tau_{\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{BD}} = \mathcal{R}^{\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{B}\leftarrow\tilde{\mathsf{X}}\tilde{\mathsf{R}}}\tau_{\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{D}}$ and $\tau_{\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{D}} = \mathrm{Tr}_{\mathsf{B}}\tau_{\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{BD}}$, the DPI for QMI tells us that

$$I(\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{B} : \mathsf{D})_\tau = I(\tilde{\mathsf{X}}\tilde{\mathsf{R}} : \mathsf{D})_\tau \tag{16.22}$$

By making use of these facts and the following instance of strong subadditivity

$$H(\mathsf{BD}|\tilde{\mathsf{X}})_\tau - H(\mathsf{B}|\tilde{\mathsf{X}})_\tau - H(\mathsf{D}|\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{B})_\tau = H(\mathsf{D}|\tilde{\mathsf{X}}\mathsf{B})_\tau - H(\mathsf{D}|\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{B})_\tau = I(\mathsf{D} : \tilde{\mathsf{R}}|\tilde{\mathsf{X}}\mathsf{B})_\tau \geq 0 \quad (16.23)$$

we find that

$$\begin{aligned} I(\tilde{\mathsf{X}} : \mathsf{BD})_\tau &\overset{(a)}{=} H(\mathsf{BD})_\tau - H(\mathsf{BD}|\tilde{\mathsf{X}})_\tau && (16.24)\\ &\overset{(b)}{\leq} H(\mathsf{B})_\tau + H(\mathsf{D})_\tau - H(\mathsf{BD}|\tilde{\mathsf{X}})_\tau && (16.25)\\ &\overset{(c)}{\leq} H(\mathsf{B})_\tau + H(\mathsf{D})_\tau - H(\mathsf{B}|\tilde{\mathsf{X}})_\tau - H(\mathsf{D}|\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{B})_\tau && (16.26)\\ &\overset{(d)}{=} I(\tilde{\mathsf{X}} : \mathsf{B})_\tau + I(\tilde{\mathsf{X}}\tilde{\mathsf{R}}\mathsf{B} : \mathsf{D})_\tau && (16.27)\\ &\overset{(e)}{=} I(\tilde{\mathsf{X}} : \mathsf{B})_\tau + I(\tilde{\mathsf{X}}\tilde{\mathsf{R}} : \mathsf{D})_\tau && (16.28)\\ &\overset{(f)}{\leq} \chi(\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}) + \chi(\mathcal{N}^{\mathsf{D}\leftarrow\mathsf{C}}). && (16.29) \end{aligned}$$

where (a) is by definition; (b) is $I(\mathsf{B} : \mathsf{D}) \geq 0$; (c) is (16.23); (d) is by definition; (e) is (16.22); and (f) is by definition of the Holevo information. Since this bound is true for *any* $\mathcal{A}_X$, P_X and $\underline{\rho}$, we have

$$\chi(\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}} \otimes \mathcal{N}^{\mathsf{D}\leftarrow\mathsf{C}}) = \max_{\mathcal{A}_X, P_X, \underline{\rho}} I(\tilde{\mathsf{X}} : \mathsf{BD})_\tau \leq \chi(\mathcal{M}^{\mathsf{B}\leftarrow\mathsf{A}}) + \chi(\mathcal{N}^{\mathsf{D}\leftarrow\mathsf{C}})$$

as required.

□

It is easily verified that the tensor product of two (or more) entanglement breaking operations is entanglement breaking. Therefore, if $\mathcal{N}$ is entanglement breaking, we have $\chi(\mathcal{N}^{\otimes n}) = \chi(\mathcal{N}^{\otimes(n-1)}) + \chi(\mathcal{N}) = n\chi(\mathcal{N})$ and, by the HSW theorem,

Corollary 13. If $\mathcal{N}$ is any entanglement-breaking operation $C(\mathcal{N}) = \chi(\mathcal{N})$.

Proposition 14. Given a conditional probability distribution $N_{Y|X}$ we can define an associated operation $\mathcal{N}^{\tilde{\mathsf{Y}}\leftarrow\tilde{\mathsf{X}}}$ by

$$\mathcal{N}^{\tilde{\mathsf{Y}}\leftarrow\tilde{\mathsf{X}}} : \rho_{\tilde{\mathsf{X}}} \mapsto \sum_{y,x} N_{Y|X}(y|x)|y\rangle\langle x|\rho_{\tilde{\mathsf{X}}}|x\rangle\langle y|.$$

$$C(\mathcal{N}^{\tilde{\mathsf{Y}}\leftarrow\tilde{\mathsf{X}}}) = \max_{P_X} I(X : Y) \text{ where } P_{XY}(x,y) = N_{Y|X}(y|x)P_X(x)$$

and the maximisation is over probability distributions P_X on $\mathcal{A}_X$.

♣♣Proving this is an exercise for example class 4.