16.5 HSW Theorem: Achievability part

Given a positive operator L with eigendecomposition $L = \sum_i \lambda_i |\alpha_i\rangle\langle\alpha_i|$, we define for $u \in \mathbb{R}$,

$$L^u = \sum_{i:\lambda_i \neq 0} \lambda_i^u |\alpha_i\rangle\langle\alpha_i|.$$

Note that for any $u, v \in \mathbb{R}$, $L^u L^v = L^{u+v}$ and that L^0 is the projector onto $\mathrm{supp}(L)$. The function $L \mapsto L^{1/2}$ is **operator monotone**, meaning if operators M and L satisfy $M \geq L \geq 0$ then $M^{1/2} \geq L^{1/2} \geq 0$.

Lemma 1 (Hayashi-Nagaoka)**.** For any real $c > 0$ and operators S, R (on the same Hilbert space) such that $0 \leq S \leq \mathbb{1}$ and $0 \leq R$,

$$\mathbb{1} - (S+R)^{-1/2} S (S+R)^{-1/2} \leq (1+c)(\mathbb{1} - S) + (2 + c + c^{-1})R \tag{16.1}$$

Proof. **(Not examinable.)** For any real c and operators A and B

$$(A - cB)^\dagger (A - cB) = A^\dagger A + c^2 B^\dagger B - c(A^\dagger B + B^\dagger A) \geq 0. \tag{16.2}$$

If $c > 0$ then

$$A^\dagger B + B^\dagger A \leq c^{-1} A^\dagger A + c B^\dagger B. \tag{16.3}$$

Setting $A = R^{1/2}$ and $B = R^{1/2}(X - \mathbb{1})$ for some hermitian X this becomes

$$R(X - \mathbb{1}) + (X - \mathbb{1})R \leq c^{-1} R + c(X - \mathbb{1})R(X - \mathbb{1}). \tag{16.4}$$

Equivalently,

$$\begin{aligned} XRX =& (X - \mathbb{1})R(X - \mathbb{1}) + R + R(X - \mathbb{1}) + (X - \mathbb{1})R & (16.5)\\ \leq& (1+c)(X - \mathbb{1})R(X - \mathbb{1}) + (1 + c^{-1})R & (16.6) \end{aligned}$$

and, since $S \geq 0$,

$$XRX \leq (1+c)(X - \mathbb{1})(S+R)(X - \mathbb{1}) + (1 + c^{-1})R. \tag{16.7}$$

Since square-root is operator monotone, and $R \geq 0$ and $0 \leq S \leq \mathbb{1}$ we have

$$(S+R)^{1/2} \geq S^{1/2} \geq S. \tag{16.8}$$

If we take $X = (S+R)^{-1/2}$ then $\Pi := X(S+R)X$ is the projector onto $\mathrm{supp}(S+R)$ and

$$(X - \mathbb{1})(S+R)(X - \mathbb{1}) = \Pi + S + R - 2(S+R)^{1/2} \leq \Pi + R - S, \tag{16.9}$$

where the inequality is (16.8). Combining this upper bound with (16.7) yields

$$XRX \leq (1+c)(\Pi - S) + (2 + c + c^{-1})R \tag{16.10}$$

and $XRX = \Pi - (S+R)^{-1/2} S (S+R)^{-1/2}$ so

$$\Pi - (S+R)^{-1/2} S (S+R)^{-1/2} \leq (1+c)(\Pi - S) + (2 + c + c^{-1})R. \tag{16.11}$$

Because $\mathbb{1} - \Pi \geq 0$ and $c > 0$, $(\mathbb{1} - \Pi) \leq (1+c)(\mathbb{1} - \Pi)$. Adding this to (16.11) yields (16.1). □

Lemma 2. If we have a code which can send a uniformly distributed message M from a set $\mathcal{A}_M$ of size k with average error probability $\Pr(\hat{M} \neq M) = \bar{\epsilon}$ then the same code can send a message from a subset of $\mathcal{A}_M$ of size $\lceil \frac{k}{2} \rceil$ with worst-case error probability no more than $2\bar{\epsilon}$.

Proof. Label the elements of $\mathcal{A}_M$, $m_1, m_2, \ldots, m_k$, such that the probabilities

$$p_i := \Pr(\hat{M} \neq M | M = m_i)$$

are in increasing order $p_1 \leq p_2 \leq \cdots \leq p_k$.

$$\begin{aligned} \bar{\epsilon} =& \frac{1}{k} \left(\sum_{i=1}^{\lfloor k/2 \rfloor} p_i + \sum_{i=\lceil k/2 \rceil}^{k} p_i \right) && (16.12) \\ \geq& \frac{1}{k} \left(0 + (k + 1 - \lceil k/2 \rceil) p_{\lceil k/2 \rceil} \right) && (16.13) \\ \geq& \frac{1}{k} \frac{k+1}{2} p_{\lceil k/2 \rceil} \geq \frac{1}{2} p_{\lceil k/2 \rceil}. && (16.14) \end{aligned}$$

Therefore the subset $\{m_1, \ldots, m_{\lceil k/2 \rceil}\}$ has the required property. □

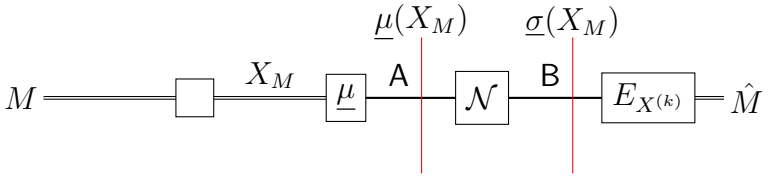

Figure 16.1: Depiction of the random codes used in the proof of Theorem 3.

Theorem 3. Suppose that $\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}$ is an operation, $k \geq 1$ is an integer, and M is a uniformly distributed RV taking values in $\{1, \ldots, k\}$. Given any finite set $\mathcal{A}_X$, distribution P_X on that set, and map $\underline{\mu} : \mathcal{A}_X \to \mathcal{L}(\mathcal{H}_\mathsf{A})$ from $\mathcal{A}_X$ to states of A let

$$\sigma_{\mathsf{XB}} := \sum_{x \in \mathcal{A}_X} P_X(x) |x\rangle\langle x|_\mathsf{X} \otimes \underline{\sigma}(x)_\mathsf{B}, \tag{16.15}$$

where $\underline{\sigma}(x)_\mathsf{B} := \mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}} \underline{\mu}(x)_\mathsf{A}$. Given k "codewords" $(x_1, x_2, \ldots, x_k) \in \mathcal{A}_X^k$ we encode M as the state $\underline{\mu}(x_M)_\mathsf{A}$; the operation $\mathcal{N}$ is applied; then we measure a decoding POVM $E : \{0, 1, \ldots, k\} \to \mathcal{L}(\mathcal{H}_\mathsf{B})$ producing result $\hat{M}$. We claim that there exists a set of k codewords and a decoding POVM E such that the average probability of error $\Pr(\hat{M} \neq M)$ is equal to $\bar{\epsilon}$, where $\bar{\epsilon}$ satisfies

$$k \geq \frac{\bar{\epsilon}}{16} \frac{1}{\beta_{\bar{\epsilon}/2}(\sigma_{\tilde{\mathsf{X}}\mathsf{B}} \| \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_\mathsf{B})} \tag{16.16}$$

and therefore[1]

$$k_\epsilon(\mathcal{N}^{\mathsf{B}\leftarrow\mathsf{A}}) \geq \frac{\epsilon}{64} \frac{1}{\beta_{\epsilon/4}(\sigma_{\tilde{\mathsf{X}}\mathsf{B}} \| \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_\mathsf{B})}. \tag{16.17}$$

[1] I missed out a factor of half in the lecture.

Proof. **Random codes to specific codes:** Suppose that the k codewords are chosen at random from $\mathcal{A}_X$. We represent these as random variables $X_1, \ldots, X_k$ taking values in $\mathcal{A}_{\mathsf{X}}$ and write $X^{(k)} = (X_1, \ldots, X_k)$. We assume that both Alice and Bob know $X^{(k)}$, so Bob's decoding POVM can depend on $X^{(k)}$. We will use the following reasoning, which holds for any way of choosing the decoding POVM and any distribution of the k codewords:

$$\bar{\epsilon} = \min_{\underline{x} \in \mathcal{A}_X^k} \Pr(\hat{M} \neq M | X^{(k)} = \underline{x}) \tag{16.18}$$

$$\leq \Pr(\hat{M} \neq M) = \sum_{\underline{x} \in \mathcal{A}_X^k} P_{X^{(k)}}(\underline{x}) \Pr(\hat{M} \neq M | X^{(k)} = \underline{x}). \tag{16.19}$$

The expression (16.18) is the minimum average error probability attained by some particular choice of codewords $\underline{x} = (x_1, \ldots, x_k)$, so we can set this equal to the $\bar{\epsilon}$ of the theorem. This is a lower-bound on the average error probability (16.19) which is achieved when the codewords are chosen at random (with the decoding POVM chosen depending on the codewords).

The decoding POVM: Recall that

$$\beta_{\epsilon'}(\sigma_{\tilde{\mathsf{X}}\mathsf{B}} \| \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}}) = \min\{\beta(T_{\tilde{\mathsf{X}}\mathsf{B}}, \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}}) : \alpha(T_{\tilde{\mathsf{X}}\mathsf{B}}, \sigma_{\tilde{\mathsf{X}}\mathsf{B}}) \leq \epsilon, 0 \leq T_{\tilde{\mathsf{X}}\mathsf{B}} \leq \mathbb{1}\} \tag{16.20}$$

where $\alpha(T_{\tilde{\mathsf{X}}\mathsf{B}}, \sigma_{\tilde{\mathsf{X}}\mathsf{B}}) = 1 - \mathrm{Tr} T_{\tilde{\mathsf{X}}\mathsf{B}} \sigma_{\tilde{\mathsf{X}}\mathsf{B}}$ and $\beta(T_{\tilde{\mathsf{X}}\mathsf{B}}, \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}}) = \mathrm{Tr} T_{\tilde{\mathsf{X}}\mathsf{B}} \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}}$. In this particular case, defining

$$L(x)_{\mathsf{B}} := \mathrm{Tr}_{\tilde{\mathsf{X}}} |x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes \mathbb{1}_{\mathsf{B}} T_{\tilde{\mathsf{X}}\mathsf{B}} \tag{16.21}$$

we find that

$$\alpha(T_{\tilde{\mathsf{X}}\mathsf{B}}, \sigma_{\tilde{\mathsf{X}}\mathsf{B}}) = 1 - \sum_{x \in \mathcal{A}_X} \mathrm{Tr} L(x)_{\mathsf{B}} \underline{\sigma}(x)_{\mathsf{B}} \text{ and} \tag{16.22}$$

$$\beta(T_{\tilde{\mathsf{X}}\mathsf{B}}, \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}}) = \sum_{x, x' \in \mathcal{A}_X} P_X(x) P_X(x') \mathrm{Tr} L(x)_{\mathsf{B}} \underline{\sigma}(x')_{\mathsf{B}}. \tag{16.23}$$

By partial trace cyclicity, $L(x)_{\mathsf{B}} = \mathrm{Tr}_{\tilde{\mathsf{X}}} |x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes \mathbb{1}_{\mathsf{B}} T_{\tilde{\mathsf{X}}\mathsf{B}} |x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes \mathbb{1}_{\mathsf{B}}$, so $L(x)_{\mathsf{B}} \geq 0$ and, because $T_{\tilde{\mathsf{X}}\mathsf{B}} \leq \mathbb{1}_{\tilde{\mathsf{X}}\mathsf{B}}$,

$$|x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes \mathbb{1}_{\mathsf{B}} T_{\tilde{\mathsf{X}}\mathsf{B}} |x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes \mathbb{1}_{\mathsf{B}} \leq |x\rangle\langle x|_{\tilde{\mathsf{X}}} \otimes \mathbb{1}_{\mathsf{B}}$$

and, therefore, $L(x)_{\mathsf{B}} \leq \mathbb{1}_{\mathsf{B}}$.

In the random code, Bob will measure $E_{X^{(k)}}$ where, for each possible value $\underline{x} = (x_1, \ldots, x_k) \in \mathcal{A}_X^k$ of the random codewords $X^{(k)}$, we build a decoding POVM $E_{\underline{x}}$ based on the operators $L(x)$, as follows: For $m \in \{1, \ldots, k\}$ we set

$$E_{\underline{x}}(m) := \left(\sum_{i=1}^k L(x_i) \right)^{-1/2} L(x_m) \left(\sum_{i=1}^k L(x_i) \right)^{-1/2}. \tag{16.24}$$

Clearly these are positive operators, and

$$\sum_{m=1}^k E_{\underline{x}}(m) = \left(\sum_{m=1}^k L(x_m) \right)^0$$

which is the projector onto the support of $\sum_{m=1}^k L(x_m)$. Therefore, setting

$$E_{\underline{x}}(0) := \mathbb{1} - \sum_{i=1}^m E_{\underline{x}}(m) \tag{16.25}$$

we have $E_{\underline{x}}(0) \geq 0$ and $\sum_{m=0}^{k} E_{\underline{x}}(m) = \mathbb{1}$ so $E_{\underline{x}}$ is indeed a POVM.

Bounding $\Pr(\hat{M} \neq M)$ for the random code: Since the message M is uniformly distributed, we have

$$\Pr(\hat{M} \neq M) = \frac{1}{k} \sum_{m=1}^{k} \Pr(\hat{M} \neq M | M = m) \tag{16.26}$$

and because $X^{(k)}$ is independent of M,

$$\Pr(\hat{M} \neq M | M = m) = \sum_{\underline{x}} \Pr(\hat{M} \neq M | M = m, X^{(k)} = \underline{x}) \Pr(X^{(k)} = \underline{x} | M = m) \tag{16.27}$$

$$= \sum_{\underline{x}} \Pr(\hat{M} \neq M | M = m, X^{(k)} = \underline{x}) \Pr(X^{(k)} = \underline{x}) = \mathbb{E} p_m \tag{16.28}$$

where $\mathbb{E} p_m$ is the expectation of the random variable

$$p_m := \mathrm{Tr}(\mathbb{1} - E_{X^{(k)}}(m)) \underline{\sigma}(X_m). \tag{16.29}$$

For the decoding POVM we are using, this is

$$p_m = \mathrm{Tr}\left(\mathbb{1} - (S+R)^{-1/2} S (S+R)^{-1/2}\right) \underline{\sigma}(X_m) \tag{16.30}$$

where $S = L(X_m)$ and $R = \sum_{i \neq m} L(X_i)$. We already showed that $0 \leq S \leq \mathbb{1}$, and clearly $R \geq 0$, so the Hayashi-Nagaoka operator inequality tells us that

$$\forall c \geq 0,\ \mathbb{1} - (S+R)^{-1/2} S (S+R)^{-1/2} \leq (1+c)(\mathbb{1} - S) + (2 + c + c^{-1}) R. \tag{16.31}$$

Since $\underline{\sigma}(X_m) \geq 0$, for all m we have

$$\begin{aligned} p_m \leq & \mathrm{Tr}((1+c)(\mathbb{1} - S) + (2 + c + c^{-1}) R) \underline{\sigma}(X_m) \\ = & (1+c)\left(1 - \mathrm{Tr} L(X_m) \underline{\sigma}(X_m)\right) + (2 + c + c^{-1}) \sum_{i \neq m} \mathrm{Tr} L(X_i) \underline{\sigma}(X_m) \end{aligned} \tag{16.32}$$

where we used $\mathrm{Tr} \underline{\sigma}(X_m) = 1$. Now, using the fact that each X_m is distributed according to P_X, we have

$$\forall m\ \mathbb{E}\left(1 - \mathrm{Tr} L(X_m) \underline{\sigma}(X_m)\right) = 1 - \sum_{x \in \mathcal{A}_X} P_X(x) \mathrm{Tr}_{\mathsf{B}} L(x) \underline{\sigma}(x) = \alpha(T_{\tilde{\mathsf{X}}\mathsf{B}}, \sigma_{\tilde{\mathsf{X}}\mathsf{B}}) \tag{16.33}$$

Using, in addition, the independence of X_m and X_i for all $i \neq m$,

$$\forall i \neq m,\ \mathbb{E} \mathrm{Tr} L(X_i) \underline{\sigma}(X_m) = \sum_{x, x'} P_X(x) P_X(x') \mathrm{Tr} L(x) \underline{\sigma}(x') = \beta(T_{\tilde{\mathsf{X}}\mathsf{B}}, \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}}). \tag{16.34}$$

Taking the expectation of both sides of (16.32) and using (16.33) and (16.34) we obtain

$$\forall m,\ \Pr(\hat{M} \neq M | M = m) = \mathbb{E} p_m \leq (1+c)\alpha(T_{\tilde{\mathsf{X}}\mathsf{B}}, \sigma_{\tilde{\mathsf{X}}\mathsf{B}}) + (2 + c + c^{-1})(k-1)\beta(T_{\tilde{\mathsf{X}}\mathsf{B}} \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}}). \tag{16.35}$$

and therefore

$$\bar{\epsilon} \leq \Pr(\hat{M} \neq M) \leq (1+c)\alpha(T_{\tilde{\mathsf{X}}\mathsf{B}}, \sigma_{\tilde{\mathsf{X}}\mathsf{B}}) + (2 + c + c^{-1})(k-1)\beta(T_{\tilde{\mathsf{X}}\mathsf{B}} \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}}).$$

Minimising the right-hand-side over all $T_{\tilde{\mathsf{X}}\mathsf{B}}$ such that $\alpha(T_{\tilde{\mathsf{X}}\mathsf{B}}, \sigma_{\tilde{\mathsf{X}}\mathsf{B}}) \leq \epsilon'$ and $0 \leq T_{\tilde{\mathsf{X}}\mathsf{B}} \leq \mathbb{1}_{\tilde{\mathsf{X}}\mathsf{B}}$ we obtain

$$\bar{\epsilon} \leq (1+c)\epsilon' + (2 + c + c^{-1})(k-1)\beta_{\epsilon'}(\sigma_{\tilde{\mathsf{X}}\mathsf{B}} \| \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}}).$$

We have shown that, for any $c \geq 0$, there exists a code with k codewords and average error probability $\bar{\epsilon}$ where

$$k - 1 \geq \frac{\bar{\epsilon} - (1+c)\epsilon'}{2 + c + c^{-1}} \frac{1}{\beta_{\epsilon'}(\sigma_{\tilde{\mathsf{X}}\mathsf{B}} \| \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}})} \tag{16.36}$$

Taking $\epsilon' = \bar{\epsilon}/2$ and (the optimal value) $c = 1/3$

$$k \geq \frac{\bar{\epsilon}}{16} \frac{1}{\beta_{\bar{\epsilon}/2}(\sigma_{\tilde{\mathsf{X}}\mathsf{B}} \| \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}})}. \tag{16.37}$$

Finally, since there exists a set of codewords of k' and average probability of error $\epsilon/2$ such that

$$k' \geq \frac{\epsilon}{32} \frac{1}{\beta_{\epsilon/4}(\sigma_{\tilde{\mathsf{X}}\mathsf{B}} \| \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}})},$$

we know from Lemma 2 that there is set of codewords of size $\lceil k'/2 \rceil$ with *worst-case* error probability ϵ. □

Remark 4. We can interpret the operators $L(x)$ in the following way: For each $x \in \mathcal{A}_X$ let $F_x : \{0,1\} \to \mathcal{L}(\mathcal{H}_{\mathsf{B}})$ by a POVM with

$$F_x(0)_{\mathsf{B}} = L(x)_{\mathsf{B}}, F_x(1)_{\mathsf{B}} = \mathbb{1}_{\mathsf{B}} - L(x)_{\mathsf{B}}.$$

Suppose we measure $\tilde{\mathsf{X}}$ in the computational basis, obtaining an outcome X, and then measure the POVM F_X on B, and use the outcome of that measurement as our guess of which hypothesis is true. This procedure has exactly the α and β given above. We can interpret $L(x)$ as specifying a test which tries to decide between the particular state $\underline{\sigma}(x)_{\mathsf{B}}$ and the average $\sigma_{\mathsf{B}} = \sum_{x \in \mathcal{A}_X} P_X(x) \underline{\sigma}(x)_{\mathsf{B}}$.

Proposition 5 (HSW Theorem: Achievability)**.**

$$C(\mathcal{N}) = \lim_{\ell \to \infty} \frac{1}{\ell} C(\mathcal{N}^{\otimes \ell}) \geq \frac{1}{\ell} \chi(\mathcal{N}^{\otimes \ell}). \tag{16.38}$$

Proof. We consider codes for the operation $\mathcal{N}^{\otimes n}$. For a given $\mathcal{A}_X$, distribution P_X and map $\underline{\mu}$ from $\mathcal{A}_X$ to states of A, we can use $\mathcal{A}_X^n$ and P_X^n as the set and distribution in the premises of Theorem 3 and for the map use

$$\underline{\mu}^{\otimes n} : \mathcal{A}_X^n \to \mathcal{L}(\mathcal{H}_{\mathsf{A}}^{\otimes n}) : (x_1, \ldots, x_n) \mapsto \underline{\mu}(x_1)_{\mathsf{A}_1} \otimes \ldots \otimes \underline{\mu}(x_n)_{\mathsf{A}_i} \text{ where } \underline{\mu}(x)_{\mathsf{A}_i} = \mathbf{id}^{\mathsf{A}_i \leftarrow \mathsf{A}} \underline{\mu}(x)_{\mathsf{A}}.$$

The "null hypothesis" state which appears in Theorem 3 is

$$\sigma_{\tilde{\mathsf{X}}_1\mathsf{B}_1 \cdots \tilde{\mathsf{X}}_n\mathsf{B}_n} = \bigotimes_{i=1}^{n} \sigma_{\tilde{\mathsf{X}}_i\mathsf{B}_i} \text{ where } \sigma_{\tilde{\mathsf{X}}_i\mathsf{B}_i} = \sum_x P_X(x) |x\rangle\langle x|_{\tilde{\mathsf{X}}_i} \otimes \underline{\sigma}(x)_{\mathsf{B}_i}$$

where $\underline{\sigma}(x)_{\mathsf{B}_i} = \mathcal{N}^{\mathsf{B}_i \leftarrow \mathsf{A}_i} \underline{\mu}(x)_{\mathsf{A}_i}$ while the "alternative hypothesis" state is $\bigotimes_{i=1}^n \sigma_{\tilde{\mathsf{X}}_i} \otimes \sigma_{\mathsf{B}_i}$. Using Theorem 3 and the Quantum Stein's lemma, we have, for any $\epsilon > 0$

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} \lfloor \log\left(k_\epsilon(\mathcal{N}^{\otimes n})\right) \rfloor &\geq \lim_{n\to\infty} \frac{1}{n} \left(\log \frac{\epsilon}{32} - \log \beta_{\epsilon/4} \left(\bigotimes_{i=1}^n \sigma_{\tilde{\mathsf{X}}_i\mathsf{B}_i} \middle\| \bigotimes_{i=1}^n \sigma_{\tilde{\mathsf{X}}_i} \otimes \sigma_{\mathsf{B}_i} \right) \right) \\ &= D\left(\sigma_{\tilde{\mathsf{X}}\mathsf{B}} \| \sigma_{\tilde{\mathsf{X}}} \otimes \sigma_{\mathsf{B}}\right) = I(\tilde{\mathsf{X}} : \mathsf{B})_\sigma,\end{aligned}$$

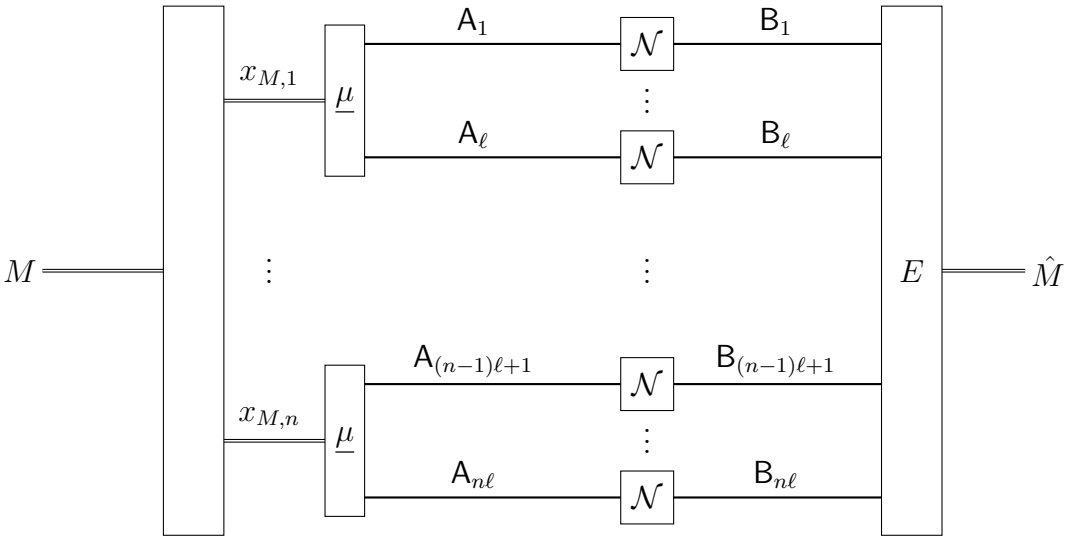

Figure 16.2: There exist codes of this form which, in the large n limit, achieve the rate $\frac{1}{\ell}\chi(\mathcal{N}^{\otimes\ell})$.

and therefore

$$C(\mathcal{N}) \geq I(\tilde{\mathsf{X}} : \mathsf{B})_\sigma.$$

Maximising this lower bound over choice of $\mathcal{A}_X$, P_X, and $\underline{\mu}$ we obtain

$$C(\mathcal{N}) \geq \chi(\mathcal{N}).$$

Now, $C(\mathcal{N}) = \frac{1}{\ell}C(\mathcal{N}^{\otimes\ell})$ so we also have, for all $\ell \in \mathbb{N}$,

$$C(\mathcal{N}) = \frac{1}{\ell}C(\mathcal{N}^{\otimes\ell}) \geq \frac{1}{\ell}\chi(\mathcal{N}^{\otimes\ell}).$$

Since the Holevo information is superadditive, we obtain the highest lower bound by taking the large ℓ limit, and this proves the "achievability" part of the HSW theorem. □

Note that, to achieve the lower bound $\frac{1}{\ell}\chi(\mathcal{N}^{\otimes\ell})$ on capacity we optimise over maps $\underline{\mu}$ which take some set $\mathcal{A}_X$ to states of ℓ input system, and then consider random codes for $\mathcal{N}^{\otimes\ell n}$ whose codewords X_i are strings $(X_{i,1}, \ldots X_{i,n})$ taking values in $\mathcal{A}_X^n$ with distribution P_X^n. This is the same as saying that each *symbol* $X_{i,j}$ in each string X_i is chosen i.i.d. with distribution P_X. The inputs to the channel which are states of the form

$$\underline{\mu}(x_{m,1})_{\mathsf{A}_1\cdots\mathsf{A}_\ell} \otimes \underline{\mu}(x_{m,2})_{\mathsf{A}_{\ell+1}\cdots\mathsf{A}_{2\ell}} \otimes \cdots \underline{\mu}(x_{m,n})_{\mathsf{A}_{(n-1)\ell+1}\cdots\mathsf{A}_{n\ell}}.$$

The possibility of having entanglement within the blocks of ℓ input systems does, for certain $\mathcal{N}$, allow us to do better as we make ℓ larger.