Lec 12–13 Oct 25, 27

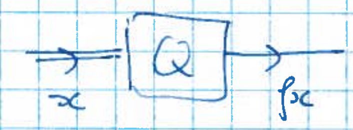
CO781 / QIC 890 Lec 12, Oct 25, 2016.

Goal: Classical capacity of quantum channels.

Stepping stone: classical capacity of Q-boxes (aka CQ channels)

A "Q-box" is specified by $\{\rho_x\}_{x \in \Omega}$.

If Alice inputs x, then Bob gets ρ_x



Obs 1: unlike an ensemble, no distribution is imposed on x.

Obs 2: If Bob applies a measurement $\mathcal{B}$,

then $B \circ Q$ reduces to a classical channel,

with unique $p(y|x)$ $\forall x$.

From last topic, asymptotic capacity of $B \circ Q$

$$= \max_{p(x)} I(X:Y)$$

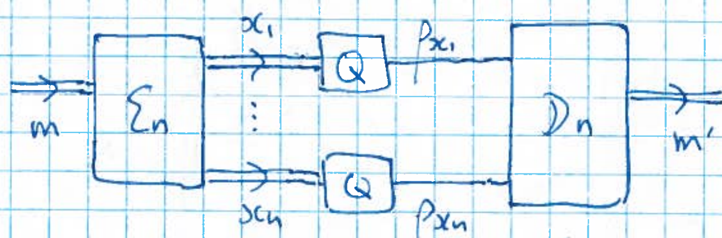
So, given n uses of the Q box, for large n, the rate of communication (in bits) per use of Q is at least

$$\underbrace{\max_{p(x)} \quad I_{acc}(\{p_x, \rho_x\})}$$

Composing two maximization

Can we do better?

Most general communication protocol using Q-boxes:



Alice' classical inputs to the Q boxes.

* correlations between x_i's may help ...

Lesson 1

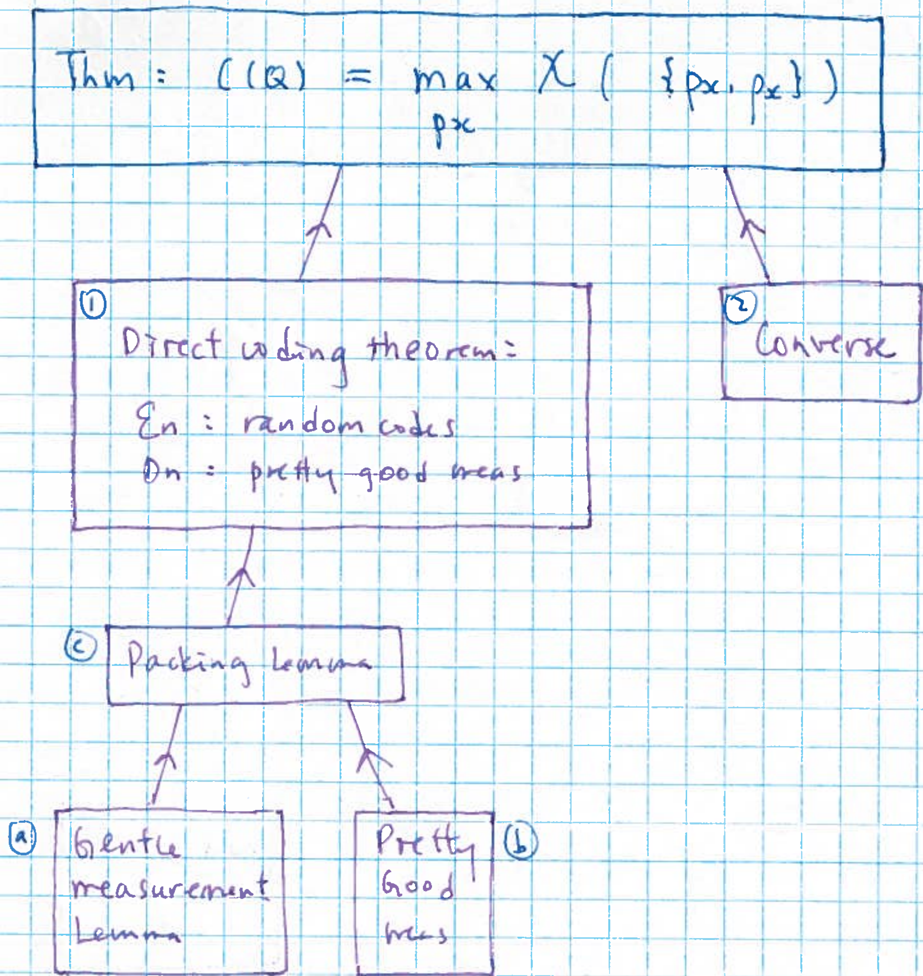
* Bob has to estimate m. Best strategy is a measurement on ALL n systems jointly!

Lesson 2

The notions of (M, n) code $\mathcal{C}_n$, $P_e(m)$, $P_e(\mathcal{C}_n)$, $\mathbb{E}\, P_e(\mathcal{C}_n)$

\# messages \# uses of Q

achievable rate, and capacity $C(Q)$ are similar to the classical setting



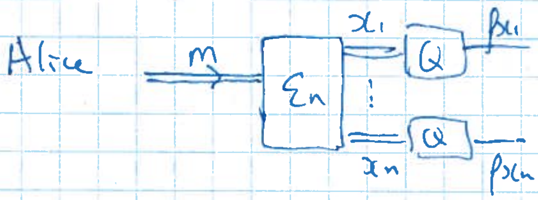
NB: $\max_{p_x} I_{acc}(\{p_x, \rho_x\}) \leq \max_{p_x} \chi(\{p_x, \rho_x\})$

- For the uniform distribution of the locking states, $I_{acc} \ll \chi$ but max over p_x removes the difference
- consider the trine states. $\max_{p_x} \chi(\{p_x, \rho_x\}) = 1$, attained by the uniform distribution. This is quite a bit higher than $I_{acc} \simeq 0.58$.

① Direct coding theorem

Given any Q box described by $\{\rho_x\}$, any dist^n $p(x)$

consider the (M, n) code :



- where the encoder is generated as follows (& told to Alice & Bob)

$$C_1 = x_{11}\, x_{12} \cdots x_{1n}$$
$$C_2 = x_{21}\, x_{22} \quad\; x_{2n}$$
$$\vdots$$
$$C_M = x_{M1}\, x_{M2} \cdots x_{Mn}$$

and each x_{ij} is drawn iid $\sim p(x)$, and rejected if C_i is not strongly typical.

- to transmit message i, Alice inputs $x_{i1}\, x_{i2} \cdots x_{in}$ into $Q^{\otimes n}$.
- Bob receives $\gamma_i = \rho_{x_{i1}} \otimes \rho_{x_{i2}} \otimes \cdots \otimes \rho_{x_{in}}$

Qn : for what M can Bob decode with error $\epsilon_n \to 0$?

Strongly typical sequence:

Let X be a r.v with sample space Ω & dist^n $p(x)$.

$C = x_1 x_2 \ldots x_n$ is δ-strongly typical ($C \in \$_{n,\delta}$)

if the empirical distribution in C, $q(x)$, is δ-close to $p(x)$ in the total variation distance, ie $\sum_x |p(x) - q(x)| \leq \delta$

eg. $\Omega = \{1, 2, 3, 4\}$, $p(a) = \frac{a}{10}$, $n = 20$.

$$C = 33344\ 21\ 322\ \ 24343\ 42443.$$

Then

$q(1) = 1/20$ $\quad$ $p(1) = 1/10$

$q(2) = 5/20$ $\quad$ $p(2) = 2/10$

$q(3) = 7/20$ $\quad$ $\vdots$

$q(4) = 7/20$

$$\|p - q\|_1 = \sum_x |p(x) - q(x)|$$

$$= \left|\frac{1}{20} - \frac{1}{10}\right| + \left|\frac{5}{20} - \frac{2}{10}\right| + \left|\frac{7}{20} - \frac{3}{10}\right| + \left|\frac{7}{20} - \frac{4}{10}\right|$$

$$= 0.2$$

So $C \in \$_{20, 0.2}$.

A2: Strongly typical $\Rightarrow$ typical

$$\not\Leftarrow$$

NB: For large n, strongly typical set is a high prob set.

pf idea: suffices to show $\Pr\left(\forall x\ |q(x) - p(x)| \leq \frac{\delta}{|\Omega|}\right)$ is large.

$$\text{prob the above fails} \underset{\substack{\uparrow \\ \text{union bdd}}}{\leq} |\Omega| \underbrace{\Pr\left(|p(x) - q(x)| > \frac{\varepsilon}{|\Omega|}\right)}_{\text{small by LLN.}}$$

Intuition for the direct coding theorem :

For large n, what does each γ_i look like?

① Each γ_i is form by n iid draws of $\{p(x), \rho_x\}$.

By Schumacher compression, all γ_i lives in a subspace of $\approx n\, S(\sum_x p(x)\, \rho_x)$ qubits.

② But for specific i, we know C_i, and γ_i occupies much less space.

Let $I_x = \{j : x_{ij} = x\}$.

eg $C_i = 33344\ 21322\ 24343\ 42443$, $n = 20$ as before

$I_1 = \{7\}$

$I_2 = \{6, 9, 10, 11, 17\}$

$I_3 = \{1, 2, 3, 8, 13, 15, 20\}$

$I_4 = \{4, 5, 12, 14, 16, 18, 19\}$.

We can compress $\rho_1^{\otimes 1}$ on systems $\in I_1$ to $\approx \underbrace{1 \cdot S(\rho_1)}_{n\, q(x)\, S(\rho_x) \text{ qubits}}$ qubits

$\rho_2^{\otimes 5}$ on systems $\in I_2$ -- $5 \cdot S(\rho_2)$ - - -

$\rho_3^{\otimes 7}$ - - - I_3 $7 \cdot S(\rho_3)$ - -

$\rho_4^{\otimes 7}$ - - I_4 - - $7 \cdot S(\rho_4)$ - - -

Effective # of qubits occupied by γ_i $\approx \sum_x n\, p(x) \cdot S(\rho_x)$

③ Output space of Bob $\approx n \cdot S\left(\sum_x p(x)\, \rho_x\right)$ (from ①).

④ If the γ_i's don't overlap too much, we can fit

$$2^{n S(\sum_x p(x) \rho_x)} \Big/ 2^{n \sum_x p(x) \rho_x} \approx 2^{n \chi(\{p_x, \rho_x\})}$$

$2^{n S(\sum_x p(x) \rho_x)}$: dim of output space

$2^{n \sum_x p(x) \rho_x}$: dim of γ_i

$2^{n \chi(\{p_x, \rho_x\})}$: messages in the output space.

(a) Gentle measurement lemma

(Winter IEEE TIT 45 (7) p 2481-2485 1999)

Let $\rho \geq 0$, $tr(\rho) \leq 1$, $0 \leq E \leq I$

If $tr(\rho E) \geq 1-\eta$

then $\| E^{\frac{1}{2}} \rho E^{\frac{1}{2}} - \rho \|_1 \leq \sqrt{3\eta}$

Pf: $\underbrace{\text{see original paper}}_{\text{3 pages}}$ or $\underbrace{\text{F2012 lec 9.}}_{\text{4 pages}}$

(b) Pretty good measurement (Belavkin 75)

Let $\sigma_1, \ldots, \sigma_k \in \mathrm{Pos}(\mathbb{C}^d)$

The PGM for $\{\sigma_i\}_{i=1}^k$ has POVM elements:

- $M_i = \Gamma^{-\frac{1}{2}} \sigma_i \Gamma^{-\frac{1}{2}}$ for $i = 1, \ldots, k$
- $M_{k+1} = I - \sum_{i=1}^{k} M_i$

where $\Gamma = \sum_{i=1}^{k} \sigma_i$ which is positive semi def, so with spectral decomp

$$= \sum_j \lambda_j |\phi_j\rangle\langle\phi_j| \qquad (\lambda_j > 0,\ |\phi_j\rangle \text{ orthonormal})$$

and $\Gamma^{-\frac{1}{2}} = \sum_j \lambda_j^{-\frac{1}{2}} |\phi_j\rangle\langle\phi_j|$

ie we only invert $\Gamma^{\frac{1}{2}}$ on its support.

- Checking the above defines a POVM:

Let $\Pi = \sum_j |\phi_j\rangle\langle\phi_j| = $ proj onto $\mathrm{supp}(\Gamma)$.

Then $\sum_{i=1}^{k} M_i = \sum_{i=1}^{k} \Gamma^{-\frac{1}{2}} \sigma_i \Gamma^{-\frac{1}{2}} = \Pi$.

$\therefore\ M_{k+1} = I - \Pi \geq 0$.

Also $M_i \geq 0 \quad \forall\, i = 1, 2, \ldots, k$.

Def: Let $S = \{\rho_1, \rho_2, \dots, \rho_K\}$

The "distinguishability error of S" is defined as — min average error

$$d_e(S) := \min_{\text{meas}} \mathbb{E}_i \Pr(\text{outcome} \neq i \mid \text{state} = \rho_i)$$

NB: Each specific measurement gives an upper bd for $d_e(S)$.

© The Packing Lemma

Let $\{r(x)\}$ be a prob distribution, ξ_x be states, $\xi = \sum_{x=1}^{\ell} r(x)\,\xi_x$

If there exist projectors Π, Π_x s.t. $\forall x$:

① $\mathrm{Tr}(\xi_x \Pi) \geq 1-\eta_1$

② $\mathrm{Tr}(\xi_x \Pi_x) \geq 1-\eta_2$

③ $\mathrm{Tr}(\Pi_x) \leq d_1$

④ $\Pi \xi \Pi \leq \frac{I}{d_0}$

Then $S = \{\gamma_1, \gamma_2, \dots, \gamma_K\}$ formed by drawing each γ_i iid from $\{r(x), \xi_x\}$ with $K = \frac{d_0}{d_1} f$ has

$$\mathbb{E}_S\, d_e(S) \leq 2(\eta_2 + \sqrt{3\eta_1}) + 4f$$

NB ① Conditions only need to hold for most x's.

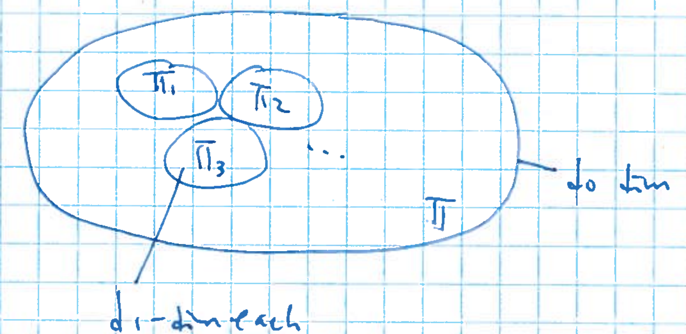
② it's common to have multiple codes for the direct coding theorem. Any proof that's airtight will do.

③ value of ℓ does not appear explicitly in the proof

$$\underbrace{\{\xi_1, \xi_2, \dots, \xi_\ell\}}_{\text{signal states}} \longrightarrow \underbrace{\{\gamma_1, \gamma_2, \dots, \gamma_K\}}_{\text{codewords}} = S$$

Interpretation of the packing lemma:

- Conditions ②, ③ ensure each ρ_x lives in some d_1-dim space, up to approximations.
- Condition ① says all ρ_x lives in some common space



- Condition ④ controls how distinguishable the ρ_x's are.

Since $\mathrm{tr}(\rho_x \Pi) \geq 1-\eta$

by GML, $\Pi \rho_x \Pi \approx \rho_x$ and $\Pi \rho \Pi \approx \rho \overset{④}{\leq} \frac{\Pi}{d_0}$

Condition ④ says $\underbrace{\text{max eig value of } \rho}_{\lambda_{max}(\rho)} \leq \frac{1}{d_0}$ $\left[\begin{array}{l} \Rightarrow \mathrm{rank}(\rho) \geq d_0 \\ \Pi : \text{at least } d_0\text{-dim} \end{array}\right]$

$\because \rho = \sum_x p(x) \rho_x$

more distinguishable ρ_x $\Rightarrow$ smaller $\lambda_{max}(\rho)$ $\Rightarrow$ d_0 larger.

- The packing lemma says, to communicate with a Q box emitting $\{\rho_x\}$ satisfying conditions ①-④, a random code S with $K = \frac{d_0}{d_1} f$ messages can be decoded whp over the choice of S.

f: a fraction of being fully packed

$X_1, X_2, \ldots, X_K$ iid $\sim p(x)$

Pf: Let $\gamma_i = \rho_{x_i}$, $\Gamma_i = \Pi\, \Pi_{x_i}\, \Pi$

Intuition:
Π_{x_i}: identifies ρ_{x_i}
Add Π to keep everything within the right space but otherwise still identifies γ_i

Define a "PGM" based on $\{\Gamma_i\}$

Handles the overlap of Γ_i's

Let $\Gamma = \sum_{i=1}^{K} \Gamma_i$

$$M_i = \Gamma^{-\frac{1}{2}}\, \Gamma_i\, \Gamma^{-\frac{1}{2}} \quad \text{for } i = 1, 2, \ldots K$$

$$M_{K+1} = I - \sum_{i=1}^{K} M_i .$$

(*) Then $d_e(S) \leq \frac{1}{K} \sum_i \operatorname{Tr} \gamma_i (I - M_i)$

RHS: upp bound of error of PGM.

Trick: use op ineq $\underbrace{I - (X+Y)^{-\frac{1}{2}}\, X\, (X+Y)^{-\frac{1}{2}}}_{M_i} \leq 2(I-X) + 4Y$, with $Y = \sum_{j\neq i} \Gamma_j$, $X = \Gamma_i$, $X+Y = \Gamma$

Pf: Ex

$$\therefore\ I - M_i = I - \Gamma^{-\frac{1}{2}}\, \Gamma_i\, \Gamma^{-\frac{1}{2}}$$
$$= I - \Big(\Gamma_i + \sum_{j\neq i} \Gamma_j\Big)^{-\frac{1}{2}} \Gamma_i \Big(\Gamma_i + \sum_{j\neq i} \Gamma_j\Big)^{-\frac{1}{2}}$$
$$\leq 2\,(I - \Gamma_i) + 4 \sum_{j\neq i} \Gamma_j$$

Purpose: "unwind" $I - M_i$ & remove $\Gamma^{-\frac{1}{2}}$.
Also, we break $I - M_i$ into $I - \Gamma_i$ & $\sum_{j\neq i} \Gamma_j$.
} need linearity

$\because \gamma_i \geq 0$, by (*) & op ineq:

$$d_e(S) \leq \frac{1}{K} \sum_i \operatorname{Tr} \gamma_i \Big[2(I - \Gamma_i) + 4 \sum_{j\neq i} \Gamma_j \Big]$$

$$= \frac{2}{K} \sum_i \underbrace{\operatorname{Tr} \gamma_i (I - \Gamma_i)}_{(a)} + \frac{4}{K} \sum_i \underbrace{\sum_{j\neq i} \operatorname{Tr} \gamma_i \Gamma_j}_{(b)}$$

(a) ← see next page → (b)

$$\mathbb{E}_S\, d_e(S) \leq 2\,(\eta_2 + 2\sqrt{3\eta_1}) + 4f$$

If a fraction $1-g$ of x's satisfies ①, we have instead:

γ_i state, $I - \Gamma_i \leq I$

$$\mathbb{E}_S\, d_e(S) \leq 2(1-g)(\eta_2 + 2\sqrt{3\eta_1}) + 2g + 4f.$$

(a) $= 1 - \operatorname{Tr} \gamma_i \Gamma_i$

$= 1 - \operatorname{Tr}(\rho_{x_i} \cdot \Pi \, \Pi_{x_i} \Pi)$

$= 1 - \operatorname{Tr}(\Pi \rho_{x_i} \Pi \cdot \Pi_{x_i})$

$\leq 1 - \operatorname{Tr}(\rho_{x_i} \Pi_{x_i}) + 2\sqrt{3\eta_1} \underset{(2)}{\leq} \eta_2 + 2\sqrt{3\eta_1}$

If $\|A - B\|_1 \leq \delta$

then $\forall \; 0 \leq P \leq I$

$|\operatorname{tr} P(A-B)| \leq 2\delta$

Here $A = \Pi \rho_{x_i} \Pi$, $B = \rho_{x_i}$, $P = \Pi_{x_i}$

$\delta = \sqrt{3\eta_1}$ by GML & (1)

Given $\operatorname{Tr}(\rho_x \Pi) \geq 1 - \eta_1$

Then $\| \Pi^{\frac{1}{2}} \rho_x \Pi^{\frac{1}{2}} - \rho_x \|_1 \leq \sqrt{3\eta_1}$

Very mechanical way to flesh out our intuitions...

(b) take $\underset{S}{\mathbb{E}} \sum_{j \neq i} \operatorname{Tr} \gamma_i \Gamma_j$

$= \sum_{j \neq i} \operatorname{Tr} \underbrace{\left(\underset{S}{\mathbb{E}} \rho_{x_i} \right)}_{\rho} \cdot \Pi \left(\underset{S}{\mathbb{E}} \Pi_{x_j} \right) \Pi$ (use the fact γ_i, Γ_j indep) (only place $i \neq j$ used)

$\leq \sum_j \operatorname{Tr}(\Pi \rho \Pi) \left(\underset{S}{\mathbb{E}} \Pi_{x_j} \right)$ (ith term added back)

$\underset{S}{\mathbb{E}} \Pi_{x_i} \geq 0$, (4) $\leq \sum_j \operatorname{Tr}\left(\frac{\Pi}{d_0} \right) \left(\underset{S}{\mathbb{E}} \Pi_{x_j} \right)$ (distinguishability of ρ_{x_i}'s)

$\underset{S}{\mathbb{E}} \Pi_{x_i} \geq 0$, $\Pi \leq I$ $\leq \sum_j \operatorname{Tr}\left(\frac{I}{d_0} \right) \left(\underset{S}{\mathbb{E}} \Pi_{x_j} \right) \leq \frac{1}{d_0} \sum_j d_1 = K \frac{d_1}{d_0} = f$

(3) $\operatorname{tr} \Pi_{x_j} \leq d_1$

Now apply the packing lemma to the random code.

For each $c = x_1 x_2 \dots x_n \in \mathcal{T}_{n,\delta}$ (let $\mathrm{prob}(c \in \mathcal{T}_{n,\delta}) = 1 - \varepsilon_3$)

Let $p(c) = \frac{1}{1-\varepsilon_3} p(x_1) \cdots p(x_n)$

$$\rho_c = \rho_{x_1} \otimes \rho_{x_2} \cdots \otimes \rho_{x_n}$$

some density matrix

Then $\rho = \frac{\sum_{c \in \mathcal{T}_{n,\delta}} p(c) \rho_c}{1-\varepsilon_3} = \frac{\rho^{\otimes n} - \varepsilon_3 \rho'}{1-\varepsilon_3}$ $\quad (\rho = \sum_x p(x) \rho_x)$

Let Π = δ_2-typical space of $\rho^{\otimes n}$, $\quad \mathrm{tr}(\Pi \rho^{\otimes n}) = 1 - \varepsilon_2$

$$\Pi_c = \bigotimes_{x \in \Omega} \underbrace{\delta_1\text{-typical space of } \rho_x^{\otimes n g(x)}}$$

Empirical dist^n of x in c

on systems in I_x

take n large enough so each typical space has prob $\geq 1-\varepsilon_1$

Then

② $\mathrm{Tr}(\rho_c \Pi_c) \geq (1-\varepsilon_1)^{|\Omega|}$ $\quad$ $\therefore \eta_2 = 1 - (1-\varepsilon_1)^{|\Omega|}$

need $|\Omega|$ const here.

③ $\mathrm{Tr}(\Pi_c) \leq \prod_{x \in \Omega} 2^{n g(x)(S(\rho_x) + \delta_1)}$

$$\leq 2^{n \sum_{x \in \Omega} g(x)(S(\rho_x) + \delta_1)}$$

$c \in \mathcal{T}_{n,\delta}$

$$\leq 2^{n \sum_{x \in \Omega} \left[p(x) S(\rho_x) + |g(x) - p(x)| S(\rho_x) + g(x) \delta_1 \right]}$$

$$\leq 2^{n \left[\sum_{x \in \Omega} [p(x) S(\rho_x)] + \delta \log d + \delta_1 \right]} =: d_1$$

For (4):

By QAEP: $\Pi \rho^{\otimes n} \Pi \leq 2^{-n[S(\rho)-\delta_2]} \Pi$

To obtain $\Pi \xi \Pi \leq \frac{\Pi}{d_0}$,

note $\Pi \xi \Pi = \frac{\Pi(\rho^{\otimes n} - \varepsilon_3 \xi')\Pi}{1-\varepsilon_3}$

$$\leq \frac{2^{-n(S(\rho)-\delta_2)}}{1-\varepsilon_3} \Pi = \frac{\Pi}{(1-\varepsilon_3)\, 2^{n(S(\rho)-\delta_2)}}$$

$\therefore d_0 \geq (1-\varepsilon_3) \cdot 2^{n(S(\rho)-\delta_2)}$

For (1):

$$\operatorname{tr}(\rho^{\otimes n} \Pi) = \sum_{c \in T_{n,\delta}} p(c) \operatorname{tr}(\xi_c \Pi) + \varepsilon_3 \operatorname{tr}(\xi' \Pi) \geq 1-\varepsilon_2$$

need each $\operatorname{tr} \xi_c \Pi$ to be high. Fix:

$$\frac{\sum_{c \in T_{n,\delta}} p(c) \operatorname{tr}(\xi_c \Pi)}{1-\varepsilon_3} \geq \frac{1-\varepsilon_2-\varepsilon_3}{1-\varepsilon_3} = 1-d_4$$

Let a fraction g of $c \in T_{n,\delta}$ has $\operatorname{tr}(\xi_c \Pi) \leq 1-\sqrt{d_4}$

Then $(1-g) \cdot 1 + g(1-\sqrt{d_4}) \geq \sum_{c \in T_{n,\delta}} \frac{p(c)}{1-\varepsilon_3} \operatorname{tr}(\xi_c \Pi) \geq 1-d_4$

$\therefore \quad g \leq \sqrt{d_4}$

* General technique to handle a bad fraction of outcomes.

So $\eta_1 = \sqrt{d_4}$. But contribution of (6) towards $\mathop{\mathbb{E}}_{c} \operatorname{tr}(S)$

is now $2(1-g)(\eta_2 + 2\sqrt{3\eta_1}) + 2g$

So, $\operatorname{pr} M = \frac{d_0}{d_1} f$

$$\geq (1-\varepsilon_3) \cdot \frac{2^{n(S(\rho)-\delta_2)}}{2^{n\left(\left(\sum_x p(x) S(\rho_x)\right) + \delta \log d + \delta_1\right)}} \cdot f$$

$$= (1-\varepsilon_3)\, 2^{n\left[\chi(\{p(x), \rho_x\}) - \delta_2 - \delta \log d - \delta_1\right]} \cdot f$$

$$\mathbb{E}_S \, de(S) \leq 2(1+q)\left(\eta_2 + 2\sqrt{3\eta_1}\right) + 2q + 4f$$

$$\leq 2\left(1+\sqrt{\delta_4}\right)\left(\left(1-(1-\varepsilon_1)^{1/2}\right) + 2\sqrt{3}\sqrt{\delta_4}\right) + 2\sqrt{\delta_4} + 4f$$

As $n \to \infty$, $\delta, \varepsilon, f \to 0$, $\exists \tilde{S}_n$ with $de(\tilde{S}_n) \to 0$.

Keep the better half of the code words in $\tilde{S}_n$ to bound the worst case error.

$\therefore$ Rate $\chi(\{p(x), \rho_x\})$ is achievable.

Converse:

Consider the state $\sum_{x_1 x_2 \dots x_n} p(x_1 x_2 \dots x_n) \underbrace{|x_1 \dots x_n\rangle\langle x_1 \dots x_n|}_{\text{on sys } X_1 \dots X_n} \otimes \underbrace{\rho_{x_1} \otimes \dots \otimes \rho_{x_n}}_{B_1 B_2 \dots B_n}$

arbitrary

eg $p(x_1 \dots x_n) = \begin{cases} \frac{1}{M} & \text{if } x_1 \dots x_n \text{ is a codeword} \\ 0 & \text{otherwise} \end{cases}$

$$\text{Then } nR \le I_{acc}(\{p(x_1 x_2 \dots x_n), \rho_{x_1} \otimes \dots \otimes \rho_{x_n}\})$$

$$\le \chi(\{p(x_1 x_2 \dots x_n), \rho_{x_1} \otimes \dots \otimes \rho_{x_n}\})$$

$$= S(X_1 X_2 \dots X_n : B_1 \dots B_n)$$

$$= S(B_1 B_2 \dots B_n) - S(B_1 \dots B_n | X_1 \dots X_n)$$

$$\overset{\text{SA}}{\le} \sum_{i=1}^{n} S(B_i) - \underbrace{\sum_{x_1 \dots x_n} p(x_1 \dots x_n) S(\rho_{x_1} \otimes \dots \otimes \rho_{x_n})}_{S(B_1 \dots B_n | X_1 \dots X_n)}$$

$\rho_{x_1} \otimes \dots \otimes \rho_{x_n}$ product state

$$= \sum_{i=1}^{n} S(B_i) - \sum_{x_1 \dots x_n} p(x_1 \dots x_n) \sum_{i=1}^{n} S(\rho_{x_i})$$

$$= \sum_{i=1}^{n} S(B_i) - \sum_{i=1}^{n} p(x_i) S(\rho_{x_i})$$

$\sum_{i=1}^{n} \underbrace{\sum_{x_1 \dots x_n} p(x_1 \dots x_n)}_{\sum_{x_j \ne i} \text{ gives } p(x_i)} \cdot S(\rho_{x_i})$

$$\le \sum_{i=1}^{n} \chi(\{p(x_i), \rho_{x_i}\}) \le n \max_{p(x)} \chi(\{p(x), \rho_x\}).$$

$S\left(\sum_{i=1}^{n} p(x_i) \rho_{x_i}\right)$