Seminar – 7/28/2011 – Univ. of Cambridge

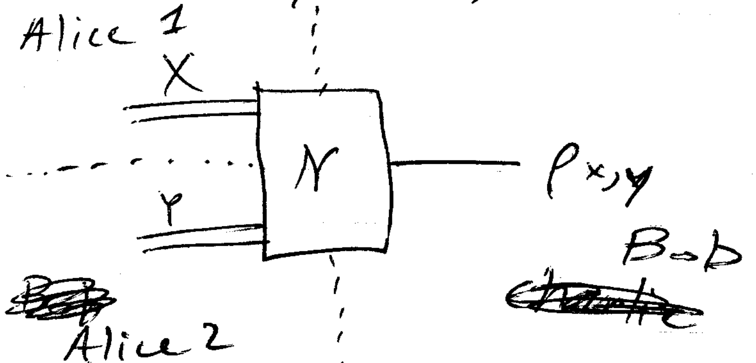
The Quest for a Quantum Simultaneous Decoder

joint with Fawzi, Hayden, Sen, Savov

Setting 1102.2624, 1102.2955

A classical-classical-quantum multiple access channel (use this setting to simplify things a bit)

$x, y \to \rho_{x,y}$



What are the maximal rates at which Alice + Bob can send classical information error free?

– Winter (9807019) proved that the classical capacity of this channel is given by the following rate region:

$R_1, R_2 \geq 0$ such that

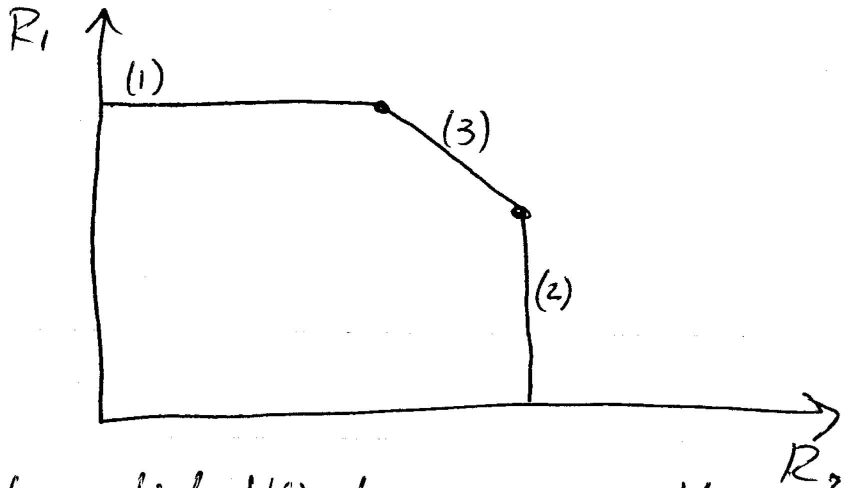
$$R_1 \leq I(X;B|Y)_\rho \quad (1)$$
$$R_2 \leq I(Y;B|X)_\rho \quad (2)$$
$$R_1 + R_2 \leq I(XY;B)_\rho \quad (3)$$ where

7/28/2011

+ $\rho^{XYB} \equiv \sum_{x,y} p(x)\, p(y)\ |x\rangle\langle x|^X \otimes |y\rangle\langle y|^Y \otimes \rho^B_{x,y}$

+ ~~take~~ union over all possible distributions $p(x)\,p(y)$

for a particular distribution $p(x)\,p(y)$ region is



How did Winter prove this result?
Two parts to any capacity theorem:
achievability + the converse part (optimality)

The converse part is similar to classical converse, so we won't review (see 1001.3404)

7/28/2011

Achievability follows from a strategy called "successive decoding." This means that the strategy is to employ "point-to-point" codes.

Review: HSW coding theorem for single-sender, single-receiver $x \rightarrow \rho_x$

1) Have the sender choose codewords independently + randomly accding to the product distributon $p_{X^n}(x^n) = \prod_{i=1}^{n} p_X(x_i)$

So, for first message, choose codeword $x^n(1)$ in this way. ~~Fors~~ For second message, choose $x^n(2)$ in this way. Repeat this prodedure unti'l reaching the last message M.

If sender transmits message m, codeword is $x^n(m)$ + receiver gets $\rho_{x^n(m)} = \rho_{x_1(m)} \otimes \cdots \otimes \rho_{x_n(m)}$

7/28/2011

2) Need to construct decoding POVM $\{\Lambda_m\}$ such that Λ_m detects $\rho_{x^n(m)}$ w/ high probability. Prob. of correct detection is

$$\mathrm{Tr}\{\Lambda_m \rho_{x^n(m)}\}$$

So prob. of incorrect detection is

$$\mathrm{Tr}\{(I - \Lambda_m)\rho_{x^n(m)}\}$$

Insteda analyze average error prob.

$$\frac{1}{M}\sum_m \mathrm{Tr}\{(I - \Lambda_m)\rho_{x^n(m)}\}$$

In fact, just analyze expectation of average error prob.

$$\mathbb{E}_{X^n}\left\{\frac{1}{M}\sum_m \mathrm{Tr}\{(I - \Lambda_m)\rho_{X^n(m)}\}\right\} \quad (*)$$

HSW showed that as long as

$$\text{rate } R = \frac{\log(M)}{n} = \frac{\#\text{of bits}}{\text{channel use}} \approx I(X;B)$$

where

$$\sum_x p(x)|x\rangle\langle x| \otimes \rho_x^B$$

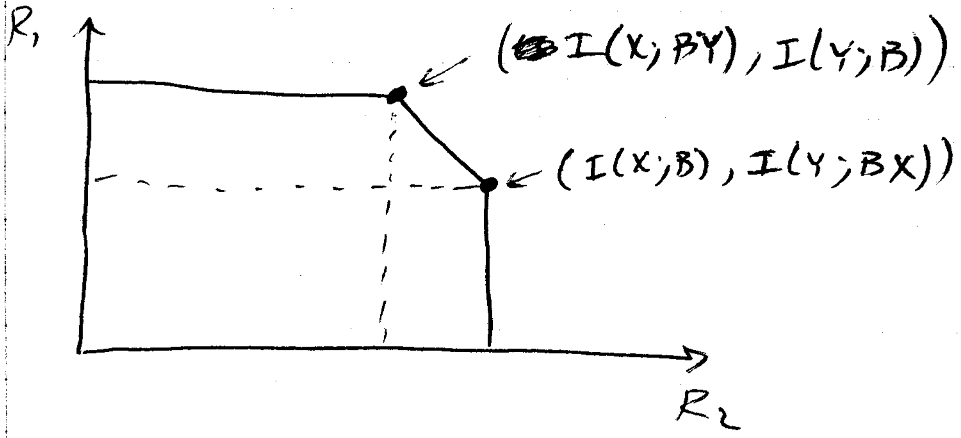
7/28/2011

then $(*) \leq \epsilon$

$\forall \epsilon > 0$ + sufficiently large n.

(Also, measurement does not disturb the state by much)

Back to MAC,



Winter showed achievability of corner points by using point-to-point codes. How does this work?

If Alice 1 + Bob pretend that Alice 2's input is noise, then they can achieve $R_1 \approx I(X;B)_\rho$

where $\sum_x p(x)|x\rangle\langle x| \otimes \underbrace{\sum_y p(y)\,\rho^B_{x,y}}_{\text{averaging over represents extra mixing.}}$

7/28/2011

After Bob decodes X, state doesn't change by very much, & X is then available as "side information" which is helpful for decoding Y. Alice 2 & Bob employ HSW code of rate $R_2 \approx I(Y;BX)$ for decoding Y.

This gets one corner point. Can get the other corner point w/ a symmetric strategy. Get all points on the sum rate bound by time-sharing and all others by resource wasting.

This strategy is called "successive decoding". We are not interested in it for the purposes of this work, but instead interested in a simultaneous decoder because such a decoder should have many applications in quantum Shannon theory.

7/28/2011

Simultaneous decoding

Would like to prove there is a POVM $\{\Lambda_{l,m}\}$ + codebooks $\{x^n(l)\}_{l\in[L]}$ $\{y^n(m)\}_{m\in[M]}$ such that

$$\mathbb{E}_{X^nY^n}\left\{\frac{1}{LM}\sum_{l,m}\text{Tr}\{(I-\Lambda_{l,m})\rho_{x^n(l),y^n(m)}\}\right\}$$
$$\le \epsilon$$

$\forall \epsilon>0$ + sufficiently large n as long as

$$R_1 \le I(X;B|Y)$$
$$R_2 \le I(Y;B|X)$$
$$R_1+R_2 \le I(XY;B)$$

(We demand that this should hold for any choice of R_1 + R_2 in the above region.)

8

~~Quantum effects might play some unexpected role for the quantum interference channel and allow us to achieve a rate region that is superior to the well-known Han-Kobayashi rate region.~~

~~Finally, it could be that quantum simultaneous decoding is not necessary in order to achieve the Han-Kobayashi region. In fact, our first attempt at the proof of Theorem 12 was to quantize the successive decoding method from Ref. [59], by exploiting the coding techniques from Refs. [70, 17] tailored for classical communication. But we found an issue with the technique in Ref. [59] even for the classical interference channel because rate-splitting at the convenience of one receiver affects the other receiver's decoding abilities. Thus, it remains open to determine if a successive decoding strategy can achieve the Han-Kobayashi rate region.~~

~~We acknowledge discussions with Charlie Bennett, Kamil Brádler, Nilanjana Datta, Frédéric Dupuis, Saikat Guha, Aram Harrow, Min-Hsiu Hsieh, Debbie Leung, Will Matthews, Marco Piani, Eren Şaşoğlu, Graeme Smith, John Smolin, Jon Tyson, Mai Vu, Andreas Winter, and Shen Chen Xu. P. Hayden acknowledges support from the Canada Research Chairs program, the Perimeter Institute, CIFAR, FQRNT's INTRIQ, MITACS, NSERC, ONR through grant N000140811249, and QuantumWorks. M. M. Wilde acknowledges support from the MDEIE (Québec) PSR-SIIRI international collaboration grant. I. Savov acknowledges support from FQRNT and NSERC.~~

A Typical Sequences and Typical Subspaces

Consider a density operator ρ with the following spectral decomposition:

$$\rho = \sum_x p_X(x) \left|x\right\rangle \left\langle x\right| .$$

The weakly typical subspace is defined as the span of all vectors such that the sample entropy $\overline{H}(x^n)$ of their classical label is close to the true entropy $H(X)$ of the distribution $p_X(x)$ [47, 68]:

$$T_\delta^{X^n} \equiv \text{span}\left\{\left|x^n\right\rangle : \left|\overline{H}(x^n) - H(X)\right| \leq \delta\right\},$$

where

$$\overline{H}(x^n) \equiv -\frac{1}{n}\log\left(p_{X^n}(x^n)\right),$$
$$H(X) \equiv -\sum_x p_X(x)\log p_X(x).$$

The projector $\Pi^n_{\rho,\delta}$ onto the typical subspace of ρ is defined as

$$\Pi^n_{\rho,\delta} \equiv \sum_{x^n \in T_\delta^{X^n}} \left|x^n\right\rangle\left\langle x^n\right|,$$

where we have "overloaded" the symbol $T_\delta^{X^n}$ to refer also to the set of δ-typical sequences:

$$T_\delta^{X^n} \equiv \left\{x^n : \left|\overline{H}(x^n) - H(X)\right| \leq \delta\right\}.$$

The three important properties of the typical projector are as follows:

$$\text{Tr}\left\{\Pi^n_{\rho,\delta}\rho^{\otimes n}\right\} \geq 1-\epsilon,$$
$$\text{Tr}\left\{\Pi^n_{\rho,\delta}\right\} \leq 2^{n[H(X)+\delta]},$$
$$2^{-n[H(X)+\delta]}\Pi^n_{\rho,\delta} \leq \Pi^n_{\rho,\delta}\rho^{\otimes n}\Pi^n_{\rho,\delta} \leq 2^{-n[H(X)-\delta]}\Pi^n_{\rho,\delta},$$

where the first property holds for arbitrary $\epsilon, \delta > 0$ and sufficiently large n.

Consider an ensemble $\{p_X(x), \rho_x\}_{x \in \mathcal{X}}$ of states. Suppose that each state ρ_x has the following spectral decomposition:

$$\rho_x = \sum_y p_{Y|X}(y|x) \, |y_x\rangle \langle y_x| \, .$$

Consider a density operator ρ_{x^n} which is conditional on a classical sequence $x^n \equiv x_1 \cdots x_n$:

$$\rho_{x^n} \equiv \rho_{x_1} \otimes \cdots \otimes \rho_{x_n} .$$

We define the weak conditionally typical subspace as the span of vectors (conditional on the sequence x^n) such that the sample conditional entropy $\overline{H}(y^n|x^n)$ of their classical labels is close to the true conditional entropy $H(Y|X)$ of the distribution $p_{Y|X}(y|x) p_X(x)$ [47, 68]:

$$T_\delta^{Y^n|x^n} \equiv \text{span}\left\{ |y_{x^n}^n\rangle : \left| \overline{H}(y^n|x^n) - H(Y|X) \right| \leq \delta \right\},$$

where

$$\overline{H}(y^n|x^n) \equiv -\frac{1}{n} \log \left(p_{Y^n|X^n}(y^n|x^n) \right),$$
$$H(Y|X) \equiv -\sum_x p_X(x) \sum_y p_{Y|X}(y|x) \log p_{Y|X}(y|x) .$$

The projector $\Pi_{\rho_{x^n},\delta}$ onto the weak conditionally typical subspace of ρ_{x^n} is as follows:

$$\Pi_{\rho_{x^n},\delta} \equiv \sum_{y^n \in T_\delta^{Y^n|x^n}} |y_{x^n}^n\rangle \langle y_{x^n}^n| ,$$

where we have again overloaded the symbol $T_\delta^{Y^n|x^n}$ to refer to the set of weak conditionally typical sequences:

$$T_\delta^{Y^n|x^n} \equiv \left\{ y^n : \left| \overline{H}(y^n|x^n) - H(Y|X) \right| \leq \delta \right\} .$$

The three important properties of the weak conditionally typical projector are as follows:

$$\mathbb{E}_{X^n}\left\{ \text{Tr}\left\{ \Pi_{\rho_{X^n},\delta} \rho_{X^n} \right\} \right\} \geq 1 - \epsilon,$$
$$\text{Tr}\left\{ \Pi_{\rho_{x^n},\delta} \right\} \leq 2^{n[H(Y|X)+\delta]},$$
$$2^{-n[H(Y|X)+\delta]} \, \Pi_{\rho_{x^n},\delta} \leq \Pi_{\rho_{x^n},\delta} \, \rho_{x^n} \, \Pi_{\rho_{x^n},\delta} \leq 2^{-n[H(Y|X)-\delta]} \, \Pi_{\rho_{x^n},\delta},$$

where the first property holds for arbitrary $\epsilon, \delta > 0$ and sufficiently large n, and the expectation is with respect to the distribution $p_{X^n}(x^n)$.

B Gentle Operator Lemma

Lemma 15 (Gentle Operator Lemma for Ensembles [69, 48, 68]). *Given an ensemble $\{p_X(x), \rho_x\}$ with expected density operator $\rho \equiv \sum_x p_X(x) \rho_x$, suppose that the operator Λ such that $I \geq \Lambda \geq 0$ succeeds with high probability on the state ρ:*

$$Tr\{\Lambda \rho\} \geq 1 - \epsilon.$$

Then the subnormalized state $\sqrt{\Lambda} \rho_x \sqrt{\Lambda}$ is close in expected trace distance to the original state ρ_x:

$$\mathbb{E}_X \left\{ \left\| \sqrt{\Lambda} \rho_X \sqrt{\Lambda} - \rho_X \right\|_1 \right\} \leq 2\sqrt{\epsilon}.$$

~~*and* $R_2 = \frac{1}{n}\log_2(M) + \delta$ *(where* $\delta > 0$*) satisfy*~~

$$\cancel{\begin{aligned} R_1 &\leq I(X;B)_\rho, \\ R_2 &\leq I(Y;B|X)_\rho, \end{aligned}}$$

~~*where the Holevo information quantities are with respect to a classical-quantum state of the form*~~

$$\cancel{\rho^{XYB} \equiv \sum_{x,y} p_X(x) p_Y(y) |x\rangle\langle x|^X \otimes |y\rangle\langle y|^Y \otimes \rho^B_{x,y}.} \tag{2}$$

~~*Then there exist two POVMs* $\{\Lambda_l\}$ *and* $\left\{\Gamma^{(l)}_m\right\}$ *acting in successive order such that the expectation of the average probability of correct detection is arbitrarily close to one:*~~

$$\cancel{\mathbb{E}_{X^n,Y^n}\left\{\frac{1}{LM}\sum_{l,m} Tr\left\{\sqrt{\Gamma^{(l)}_m}\sqrt{\Lambda_l}\rho_{X^n(l),Y^n(m)}\sqrt{\Lambda_l}\sqrt{\Gamma^{(l)}_m}\right\}\right\} \geq 1-\epsilon.}$$

5.2 Quantum Simultaneous Decoding

Another approach to achieve the capacity of the multiple access channel is for the receiver to use a simultaneous decoder (sometimes referred to as a jointly typical decoder in the IID setting), which decodes the messages of all senders at the same time rather than in succession [13, 20]. On the one hand, simultaneous decoding is more complex than successive decoding because it considers all tuples of messages, but on the other hand, it is more powerful than a successive decoding strategy because it can decode at any rates provided that the rates are in the capacity region.

Note that the receiver can exploit a simultaneous decoder to achieve any point in the capacity region of a multiple access channel without invoking time-sharing. With such a strategy and for two senders, there are four different types of errors that can occur—one of these we can bound with a standard typicality argument and the other three correspond to the bounds on the capacity region of the channel. This strategy is our approach below, and we can prove that a quantum simultaneous decoder exists for multiple access channels with two classical inputs and one quantum output. Though, for a three-sender quantum multiple access channel, we are only able to prove that a quantum simultaneous decoder exists in the special case where the averaged output states commute. Thus, we leave the general case stated as a conjecture.

5.2.1 Two-Sender Quantum Simultaneous Decoding

This section contains the proof of the two-sender quantum simultaneous decoder. We should mention that Sen independently arrived at this result by a different technique [58].

Theorem 2 (Two-Sender Quantum Simultaneous Decoding). *Let* $x, y \to \rho_{x,y}$ *be a ccq channel from two senders to a single receiver. Let* $p_X(x)$ *and* $p_Y(y)$ *be respective input distributions that each sender uses to create random codebooks of the form* $\{X^n(l)\}_{l\in[1,\ldots,L]}$ *and* $\{Y^n(m)\}_{m\in[1,\ldots,M]}$. *Suppose that the rates* $R_1 = \frac{1}{n}\log_2(L) + \delta$ *and* $R_2 = \frac{1}{n}\log_2(M) + \delta$ *(where* $\delta > 0$*) satisfy the following inequalities:*

$$R_1 \leq I(X;B|Y), \tag{3}$$

$$R_2 \leq I(Y;B|X)_\rho, \tag{4}$$

$$R_1 + R_2 \leq I(XY;B)_\rho, \tag{5}$$

where the entropies are with respect to a state of the form in (2). Then there exists a simultaneous decoding POVM $\{\Lambda_{l,m}\}$ *such that the expectation of the average probability of error is bounded above by* ϵ *for all* $\epsilon > 0$ *and sufficiently large* n.

Proof. Suppose that the channel is a ccq channel of the form $x, y \to \rho_{x,y}$ and that the two senders have independent distributions $p_X(x)$ and $p_Y(y)$. These distributions induce the following averaged output states:

$$\rho_x \equiv \sum_y p_Y(y)\, \rho_{x,y}, \tag{6}$$

$$\rho_y \equiv \sum_x p_X(x)\, \rho_{x,y}, \tag{7}$$

$$\rho \equiv \sum_{x,y} p_X(x)\, p_Y(y)\, \rho_{x,y}. \tag{8}$$

Codeword Selection. Senders 1 and 2 choose codewords $\{X^n(l)\}_{l\in\{1,\ldots,L\}}$ and $\{Y^n(m)\}_{m\in\{1,\ldots,M\}}$ independently and randomly according to the distributions $p_{X^n}(x^n)$ and $p_{Y^n}(y^n)$.

POVM Construction. Let $\Pi^n_{\rho,\delta}$ be the typical projector for the tensor power state $\rho^{\otimes n}$ defined by (8). Let $\Pi^n_{\rho_{y^n},\delta}$ be the conditionally typical projector for the tensor product state ρ_{y^n} defined by (7) for n uses of the channel. Let $\Pi^n_{\rho_{x^n},\delta}$ be the conditionally typical projector for the tensor product state ρ_{x^n} defined by (6) for n uses of the channel. Let $\Pi^n_{\rho_{x^n,y^n},\delta}$ be the conditionally typical projector for the tensor product state ρ_{x^n,y^n} defined as the output of the n channels when codewords x^n and y^n are input. (We are using the "weak" definitions of these projectors as defined in the appendix.) In what follows, we make the following abbreviations:

$$\begin{aligned}
\Pi &\equiv \Pi^n_{\rho,\delta}, \\
\Pi_{y^n} &\equiv \Pi^n_{\rho_{y^n},\delta}, \\
\Pi_{x^n} &\equiv \Pi^n_{\rho_{x^n},\delta}, \\
\Pi_{x^n,y^n} &\equiv \Pi^n_{\rho_{x^n,y^n},\delta}.
\end{aligned}$$

The detection POVM $\{\Lambda_{l,m}\}$ has the following form:

$$\Lambda_{l,m} \equiv \left(\sum_{l',m'} \Pi'_{l',m'}\right)^{-\frac{1}{2}} \Pi'_{l,m} \left(\sum_{l',m'} \Pi'_{l',m'}\right)^{-\frac{1}{2}}, \tag{9}$$

$$\Pi'_{l,m} \equiv \Pi\, \Pi_{X^n(l)}\, \Pi_{X^n(l),Y^n(m)}\, \Pi_{X^n(l)}\, \Pi.$$

(Observe that the operator $\Pi'_{l,m}$ is a positive operator and thus $\{\Lambda_{l,m}\}$ is a valid POVM.)

Error Analysis. The average error probability of the code has the following form:

$$\overline{p}_e \equiv \frac{1}{LM} \sum_{l,m} \text{Tr}\left\{(I - \Lambda_{l,m})\, \rho_{X^n(l),Y^n(m)}\right\}. \tag{10}$$

We instead analyze the expectation of the average error probability, where the expectation is with respect to the random choice of code:

$$\begin{aligned}
\mathbb{E}_{X^n,Y^n}\left\{\overline{p}_e\right\} &\equiv \mathbb{E}_{X^n,Y^n}\left\{\frac{1}{LM} \sum_{l,m} \text{Tr}\left\{(I - \Lambda_{l,m})\, \rho_{X^n(l),Y^n(m)}\right\}\right\} \\
&= \frac{1}{LM} \sum_{l,m} \mathbb{E}_{X^n,Y^n}\left\{\text{Tr}\left\{(I - \Lambda_{l,m})\, \rho_{X^n(l),Y^n(m)}\right\}\right\}.
\end{aligned}$$

Due to the symmetry of the code construction (the fact that the expectation $\mathbb{E}_{X^n,Y^n}\left\{\text{Tr}\left\{(I - \Lambda_{l,m})\, \rho_{X^n(l),Y^n(m)}\right\}\right\}$ is independent of the particular message pair (l, m)), it suffices to analyze the expectation of the average error probability for the first message pair $(1, 1)$:

$$\mathbb{E}_{X^n,Y^n}\left\{\overline{p}_e\right\} = \mathbb{E}_{X^n,Y^n}\left\{\text{Tr}\left\{(I - \Lambda_{1,1})\, \rho_{X^n(1),Y^n(1)}\right\}\right\}.$$

12

We now begin our error analysis. We first bound this error probability from above as

$$\begin{aligned}\mathbb{E}_{X^n,Y^n}\left\{\overline{p}_e\right\} &\leq \mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\left(I-\Lambda_{1,1}\right)\ \Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{Y^n(1)}\right\}\right\} \\ &\quad + \mathbb{E}_{X^n,Y^n}\left\{\left\|\Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{Y^n(1)} - \rho_{X^n(1),Y^n(1)}\right\|_1\right\} \qquad (11) \\ &\leq \mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\left(I-\Lambda_{1,1}\right)\ \Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{Y^n(1)}\right\}\right\} + 2\sqrt{\epsilon}, \qquad (12)\end{aligned}$$

where the first inequality follows from the inequality

$$\mathrm{Tr}\left\{\Lambda\rho\right\} \leq \mathrm{Tr}\left\{\Lambda\sigma\right\} + \left\|\rho-\sigma\right\|_1, \qquad (13)$$

which holds for all ρ, σ, and Λ such that $0 \leq \rho, \sigma, \Lambda \leq I$. The second inequality follows from the properties of weak conditionally typical subspaces and the Gentle Operator Lemma for ensembles, by taking n to be sufficiently large (a discussion of these properties is in the appendix).

The Hayashi-Nagaoka operator inequality applies to a positive operator T and an operator S where $0 \leq S \leq I$ [32, 31]:

$$I - (S+T)^{-\frac{1}{2}}\, S\, (S+T)^{-\frac{1}{2}} \leq 2\,(I-S) + 4T.$$

Choosing

$$\begin{aligned} S &= \Pi'_{1,1}, \\ T &= \sum_{(l,m)\neq(1,1)} \Pi'_{l,m}, \end{aligned}$$

we can apply the above operator inequality to bound the first term on the RHS of (12) as

$$\begin{aligned}&\mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\left(I-\Lambda_{1,1}\right)\ \Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{Y^n(1)}\right\}\right\} \\ &\qquad \leq 2\,\mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\left(I-\Pi'_{1,1}\right)\ \Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{Y^n(1)}\right\}\right\} \\ &\qquad\qquad + 4\sum_{(l,m)\neq(1,1)} \mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\Pi'_{l,m}\ \Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{Y^n(1)}\right\}\right\}. \qquad (14)\end{aligned}$$

We first consider bounding the first term on the RHS above. Consider that

$$\begin{aligned}&\mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\Pi'_{1,1}\ \Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{Y^n(1)}\right\}\right\} \\ &= \mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\Pi\ \Pi_{X^n(1)}\ \Pi_{X^n(1),Y^n(1)}\ \Pi_{X^n(1)}\ \Pi\ \Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{Y^n(1)}\right\}\right\} \\ &\geq \mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\Pi_{X^n(1),Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\right\}\right\} \\ &\quad - \mathbb{E}_{X^n,Y^n}\left\{\left\|\Pi\ \rho_{X^n(1),Y^n(1)}\ \Pi - \rho_{X^n(1),Y^n(1)}\right\|_1\right\} \\ &\quad - \mathbb{E}_{X^n,Y^n}\left\{\left\|\Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{Y^n(1)} - \rho_{X^n(1),Y^n(1)}\right\|_1\right\} \\ &\quad - \mathbb{E}_{X^n,Y^n}\left\{\left\|\Pi_{X^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{X^n(1)} - \rho_{X^n(1),Y^n(1)}\right\|_1\right\} \\ &\geq 1 - \epsilon - 6\sqrt{\epsilon}. \qquad (15)\end{aligned}$$

The above inequalities follow by employing the Gentle Operator Lemma for ensembles, (13), and the below inequalities that follow from the discussion in the appendix:

$$\mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\{\Pi_{X^n(1)}\ \rho_{X^n(1),Y^n(1)}\}\right\} \geq 1-\epsilon, \qquad (16)$$

$$\mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\{\Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\}\right\} \geq 1-\epsilon, \qquad (17)$$

$$\mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\{\Pi\ \rho_{X^n(1),Y^n(1)}\}\right\} \geq 1-\epsilon. \qquad (18)$$

$$\mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\{\Pi_{X^n(1),Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\}\right\} \geq 1-\epsilon. \qquad (19)$$

This bound then implies that

$$\mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\left(I-\Pi'_{1,1}\right)\ \Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{Y^n(1)}\right\}\right\} \leq \epsilon + 6\sqrt{\epsilon}. \qquad (20)$$

Gallager 1974

Capacity of Degraded Broadcast Channel

Probability Paradaci Russian

The bound in (14) reduces to the following one after applying (20):

$$\overline{p}_e \leq 2\left(\epsilon + 6\sqrt{\epsilon}\right) + 4 \sum_{(l,m)\neq(1,1)} \mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\Pi'_{l,m}\ \Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{Y^n(1)}\right\}\right\}.$$

We can expand the doubly-indexed sum on the RHS above:

$$\begin{aligned}\sum_{(l,m)\neq(1,1)} \mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\Pi'_{l,m}\ \Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{Y^n(1)}\right\}\right\} = &\\ \sum_{l\neq 1}\mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\Pi'_{l,1}\ \Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{Y^n(1)}\right\}\right\}&\\ +\sum_{m\neq 1}\mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\Pi'_{1,m}\ \Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{Y^n(1)}\right\}\right\}&\\ +\sum_{l\neq 1,\ m\neq 1}\mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\Pi'_{l,m}\ \Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{Y^n(1)}\right\}\right\}.&\quad(21)\end{aligned}$$

We begin by bounding the first term on the RHS above. Consider the following chain of inequalities:

$$\begin{aligned}&\sum_{l\neq 1}\mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\Pi'_{l,1}\Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{Y^n(1)}\right\}\right\}\\ &=\sum_{l\neq 1}\mathbb{E}_{Y^n}\left\{\mathrm{Tr}\left\{\Pi\ \mathbb{E}_{X^n}\left\{\Pi_{X^n(l)}\ \Pi_{X^n(l),Y^n(1)}\ \Pi_{X^n(l)}\right\}\ \Pi\ \Pi_{Y^n(1)}\mathbb{E}_{X^n}\left\{\rho_{X^n(1),Y^n(1)}\right\}\Pi_{Y^n(1)}\right\}\right\}\\ &=\sum_{l\neq 1}\mathbb{E}_{Y^n}\left\{\mathrm{Tr}\left\{\Pi\ \mathbb{E}_{X^n}\left\{\Pi_{X^n(l)}\ \Pi_{X^n(l),Y^n(1)}\ \Pi_{X^n(l)}\right\}\ \Pi\ \Pi_{Y^n(1)}\ \rho_{Y^n(1)}\ \Pi_{Y^n(1)}\right\}\right\}\\ &\leq 2^{-n[H(B|Y)-\delta]}\sum_{l\neq 1}\mathbb{E}_{Y^n}\left\{\mathrm{Tr}\left\{\Pi\ \mathbb{E}_{X^n}\left\{\Pi_{X^n(l)}\ \Pi_{X^n(l),Y^n(1)}\ \Pi_{X^n(l)}\right\}\ \Pi\ \Pi_{Y^n(1)}\right\}\right\}\\ &=2^{-n[H(B|Y)-\delta]}\sum_{l\neq 1}\mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\Pi_{X^n(l),Y^n(1)}\ \Pi_{X^n(l)}\ \Pi\ \Pi_{Y^n(1)}\ \Pi\ \Pi_{X^n(l)}\right\}\right\}\\ &\leq 2^{-n[H(B|Y)-\delta]}\sum_{l\neq 1}\mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\Pi_{X^n(l),Y^n(1)}\right\}\right\}\\ &\leq 2^{-n[H(B|Y)-\delta]}\ 2^{n[H(B|XY)+\delta]}\ L\\ &=2^{-n[I(X;B|Y)-2\delta]}\ L\end{aligned}\quad(22)$$

The first equality follows by substitution and because $X^n\left(l\right)$ and $X^n\left(1\right)$ are independent—the senders choose the code randomly in such a way that this is true. The second equality follows because $\mathbb{E}_{X^n}\left\{\rho_{X^n(1),Y^n(1)}\right\} = \rho_{Y^n(1)}$. The first inequality follows by applying the following operator inequality for weak conditionally typical subspaces:

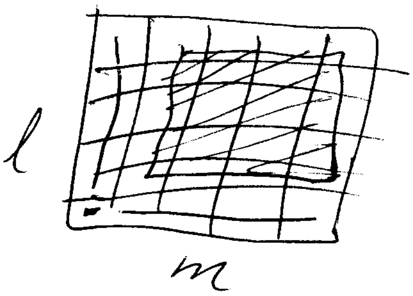
$$\Pi_{y^n}\ \rho_{y^n}\ \Pi_{y^n} \leq 2^{-n[H(B|Y)-\delta]}\ \Pi_{y^n}.$$

The third equality is from cyclicity of trace. The second inequality is from

$$\Pi_{x^n}\ \Pi\ \Pi_{y^n}\ \Pi\ \Pi_{x^n} \leq \Pi_{x^n}\ \Pi\ \Pi_{x^n} \leq \Pi_{x^n} \leq I.$$

The final inequality follows from the bound on the rank of the conditionally typical projector.

We employ a different argument to bound the second term in (21). Consider the following chain of



inequalities:

$$
\begin{aligned}
&\sum_{m\neq 1} \mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\Pi'_{1,m}\ \Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{Y^n(1)}\right\}\right\} \\
&= \sum_{m\neq 1} \mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\Pi\ \Pi_{X^n(1)}\ \Pi_{X^n(1),Y^n(m)}\ \Pi_{X^n(1)}\ \Pi\ \Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{Y^n(1)}\right\}\right\} \\
&\leq 2^{n[H(B|XY)+\delta]} \sum_{m\neq 1} \mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\Pi\ \Pi_{X^n(1)}\ \rho_{X^n(1),Y^n(m)}\ \Pi_{X^n(1)}\ \Pi\ \Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{Y^n(1)}\right\}\right\} \\
&= 2^{n[H(B|XY)+\delta]} \sum_{m\neq 1} \mathbb{E}_{X^n}\left\{\mathrm{Tr}\left\{\Pi\ \Pi_{X^n(1)}\ \mathbb{E}_{Y^n}\left\{\rho_{X^n(1),Y^n(m)}\right\}\ \Pi_{X^n(1)}\ \Pi\ \mathbb{E}_{Y^n}\left\{\Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{Y^n(1)}\right\}\right\}\right\} \\
&= 2^{n[H(B|XY)+\delta]} \sum_{m\neq 1} \mathbb{E}_{X^n}\left\{\mathrm{Tr}\left\{\Pi\ \Pi_{X^n(1)}\ \rho_{X^n(1)}\ \Pi_{X^n(1)}\ \Pi\ \mathbb{E}_{Y^n}\left\{\Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{Y^n(1)}\right\}\right\}\right\} \qquad (23)
\end{aligned}
$$

The first equality follows by substitution. The first inequality follows from the following operator inequality:

$$\Pi_{x^n,y^n} \leq 2^{n[H(B|XY)+\delta]}\ \Pi_{x^n,y^n}\ \rho_{x^n,y^n}\ \Pi_{x^n,y^n} \leq 2^{n[H(B|XY)+\delta]}\ \rho_{x^n,y^n}.$$

The second equality follows from the fact that $Y^n(m)$ and $Y^n(1)$ are independent, and the third equality follows because $\mathbb{E}_{Y^n}\left\{\rho_{X^n(1),Y^n(m)}\right\} = \rho_{X^n(1)}$. Continuing, we have

$$
\begin{aligned}
&\leq 2^{n[H(B|XY)+\delta]}\ 2^{-n[H(B|X)-\delta]} \sum_{m\neq 1} \mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\Pi\ \Pi_{X^n(1)}\ \Pi\ \Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{Y^n(1)}\right\}\right\} \\
&= 2^{n[H(B|XY)+\delta]}\ 2^{-n[H(B|X)-\delta]} \sum_{m\neq 1} \mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\Pi_{Y^n(1)}\ \Pi\ \Pi_{X^n(1)}\ \Pi\ \Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\right\}\right\} \\
&\leq 2^{n[H(B|XY)+\delta]}\ 2^{-n[H(B|X)-\delta]} \sum_{m\neq 1} \mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\rho_{X^n(1),Y^n(1)}\right\}\right\} \\
&\leq 2^{n[H(B|XY)+\delta]}\ 2^{-n[H(B|X)-\delta]}\ M \\
&= 2^{-n[I(Y;B|X)-2\delta]}\ M \qquad (24)
\end{aligned}
$$

The first inequality follows from the operator inequality

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

The first equality is cyclicity of trace, and the second inequality follows because

$$\Pi_{y^n}\ \Pi\ \Pi_{x^n}\ \Pi\ \Pi_{y^n} \leq \Pi_{y^n}\ \Pi\ \Pi_{y^n} \leq \Pi_{y^n} \leq I.$$

Finally, we obtain a bound on the last term in (21) with a slightly different argument:

$$
\begin{aligned}
&\sum_{l\neq 1,\ m\neq 1} \mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\Pi'_{l,m}\ \Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{Y^n(1)}\right\}\right\} \\
&= \sum_{l\neq 1,\ m\neq 1} \mathbb{E}_{X^n,Y^n}\left\{\mathrm{Tr}\left\{\Pi\ \Pi_{X^n(l)}\ \Pi_{X^n(l),Y^n(m)}\ \Pi_{X^n(l)}\ \Pi\ \Pi_{Y^n(1)}\ \rho_{X^n(1),Y^n(1)}\ \Pi_{Y^n(1)}\right\}\right\} \\
&= \sum_{l\neq 1,\ m\neq 1} \mathbb{E}_{Y^n}\left\{\mathrm{Tr}\left\{\Pi\ \mathbb{E}_{X^n}\left\{\Pi_{X^n(l)}\ \Pi_{X^n(l),Y^n(m)}\ \Pi_{X^n(l)}\right\}\ \Pi\ \Pi_{Y^n(1)}\ \mathbb{E}_{X^n}\left\{\rho_{X^n(1),Y^n(1)}\right\}\ \Pi_{Y^n(1)}\right\}\right\} \\
&= \sum_{l\neq 1,\ m\neq 1} \mathbb{E}_{Y^n}\left\{\mathrm{Tr}\left\{\Pi\ \mathbb{E}_{X^n}\left\{\Pi_{X^n(l)}\ \Pi_{X^n(l),Y^n(m)}\ \Pi_{X^n(l)}\right\}\ \Pi\ \Pi_{Y^n(1)}\ \rho_{Y^n(1)}\ \Pi_{Y^n(1)}\right\}\right\} \\
&\leq \sum_{l\neq 1,\ m\neq 1} \mathbb{E}_{Y^n}\left\{\mathrm{Tr}\left\{\Pi\ \mathbb{E}_{X^n}\left\{\Pi_{X^n(l)}\ \Pi_{X^n(l),Y^n(m)}\ \Pi_{X^n(l)}\right\}\ \Pi\ \rho_{Y^n(1)}\right\}\right\} \\
&= \sum_{l\neq 1,\ m\neq 1} \mathrm{Tr}\left\{\Pi\ \mathbb{E}_{X^n}\left\{\Pi_{X^n(l)}\ \mathbb{E}_{Y^n}\left\{\Pi_{X^n(l),Y^n(m)}\right\}\ \Pi_{X^n(l)}\right\}\ \Pi\ \mathbb{E}_{Y^n}\left\{\rho_{Y^n(1)}\right\}\right\} \qquad (25)
\end{aligned}
$$

$$2^{-nH(B|XY)}\ \Pi_{xy} \leq \Pi_{xy}\ \rho_{xy}\ \Pi_{xy} \leq 2^{-nH(B|XY)}\ \Pi_{xy}$$

$$\leq \rho_{xy}$$

15

The first equality follows by substitution. The second equality follows from the independence of $X^n(l)$ and $X^n(1)$. The third equality follows because $\mathbb{E}_{X^n}\left\{\rho_{X^n(1),Y^n(1)}\right\} = \rho_{Y^n(1)}$. The first inequality follows from the fact that ρ_{y^n} and Π_{y^n} commute and thus $\Pi_{y^n}\,\rho_{y^n}\,\Pi_{y^n} = \sqrt{\rho_{y^n}}\,\Pi_{y^n}\,\sqrt{\rho_{y^n}} \leq \rho_{y^n}$. The fourth equality follows from the independence of $Y^n(m)$ and $Y^n(1)$. Continuing, we have

$$\begin{aligned}
&= \sum_{l\neq 1,\ m\neq 1} \operatorname{Tr}\left\{\mathbb{E}_{X^n}\left\{\Pi_{X^n(l)}\,\mathbb{E}_{Y^n}\left\{\Pi_{X^n(l),Y^n(m)}\right\}\,\Pi_{X^n(l)}\right\}\,\Pi\,\rho^{\otimes n}\,\Pi\right\} \\
&\leq 2^{-n[H(B)-\delta]} \sum_{l\neq 1,\ m\neq 1} \operatorname{Tr}\left\{\mathbb{E}_{X^n}\left\{\Pi_{X^n(l)}\,\mathbb{E}_{Y^n}\left\{\Pi_{X^n(l),Y^n(m)}\right\}\,\Pi_{X^n(l)}\right\}\,\Pi\right\} \\
&= 2^{-n[H(B)-\delta]} \sum_{l\neq 1,\ m\neq 1} \mathbb{E}_{X^n,Y^n}\left\{\operatorname{Tr}\left\{\Pi_{X^n(l),Y^n(m)}\,\Pi_{X^n(l)}\,\Pi\,\Pi_{X^n(l)}\right\}\right\} \\
&\leq 2^{-n[H(B)-\delta]} \sum_{l\neq 1,\ m\neq 1} \mathbb{E}_{X^n,Y^n}\left\{\operatorname{Tr}\left\{\Pi_{X^n(l),Y^n(m)}\right\}\right\} \\
&\leq 2^{-n[H(B)-\delta]}\,2^{n[H(B|XY)+\delta]}\,LM \\
&= 2^{-n[I(XY;B)-2\delta]}\,LM.
\end{aligned} \tag{26}$$

The first equality follows because $\mathbb{E}_{Y^n}\left\{\rho_{Y^n(1)}\right\} = \rho^{\otimes n}$ and from cyclicity of trace. The first inequality is from the following operator inequality:

$$\Pi\,\rho^{\otimes n}\,\Pi \leq 2^{-n[H(B)-\delta]}\Pi.$$

The second equality is from cyclicity of trace and factoring out the expectations. The second inequality is from the operator inequality

$$\Pi_{x^n}\,\Pi\,\Pi_{x^n} \leq \Pi_{x^n} \leq I.$$

The final inequality is from the bound on the rank of the weak conditionally typical projector.

Combining everything together, we get the following bound on the expectation of the average error probability:

$$\mathbb{E}_{X^n,Y^n}\left\{\overline{p}_e\right\} \leq 2\left(\epsilon + 7\sqrt{\epsilon}\right) + 4\,\left(L\,2^{-n[I(X;B|Y)-2\delta]} + M\,2^{-n[I(Y;B|X)-2\delta]} + LM\,2^{-n[I(XY;B)-2\delta]}\right).$$

Thus, we can choose the message sizes to be as follows:

$$\begin{aligned}
L &= 2^{n[R_1-3\delta]}, \\
M &= 2^{n[R_2-3\delta]},
\end{aligned}$$

so that the expectation of the average error probability vanishes in the asymptotic limit whenever the rates R_1 and R_2 obey the following inequalities:

$$\begin{aligned}
R_1 - \delta &< I(X;B|Y), \\
R_2 - \delta &< I(Y;B|X), \\
R_1 + R_2 - 4\delta &< I(XY;B).
\end{aligned}$$

□

A casual glance at the above proof might lead one to believe it is just a straightforward extension of the "usual" proofs of the HSW theorem [35, 57, 15, 38, 68], but it differs from these and extends them non trivially in several regards. First, we choose the square-root POVM in (9) in a particular way—specifically, the layering of projectors is such that the projector of size $\approx 2^{nH(B|XY)}$ is surrounded by the projector of size $\approx 2^{H(B|X)}$, which itself is surrounded by the projector of size $\approx 2^{nH(B)}$. If one were to place the projector of size $\approx 2^{nH(B|Y)}$ somewhere in the square-root POVM, this leads to difficulties with non-commutative projectors (discussed in earlier versions of this paper on the arXiv). So, our second observation

Mark Wilde