Multi-User Information Theory December 6th, 2011

Lecture 6

Lecturer: Haim Permuter *Scribe: Pavel Shanin*

I. CUT - SET BOUND

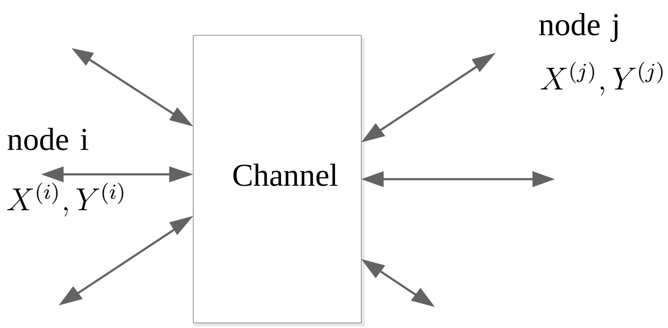

Fig. 1. General multiterminal channel with transmitters and receivers. Where each node has input transmitted variable $X^{(i)}$ and a received variable $Y^{(i)}$.

Consider a general multiterminal network with transmitters and receivers as depicted in Fig. 1. We derive upper bounds on the achievable rates for any multiterminal network.

Definition 1 A general multiterminal network with m nodes consists of:

- $X_k^{(i)}$ - denote an output variable of node (i) at time k, $Y_k^{(i)}$ - denote an input variable of node (i) at time k.
- Memoryless channel with transition function $P(y_k^{(1)}, y_k^{(2)}, ..., y_k^{(m)}|x_k^{(1)}, x_k^{(2)}, ..., x_k^{(m)})$
- $m_{ij} \in \{1, ..., 2^{nR_{ij}}\}$ is a message that goes from node i to node j at rate R_{ij}.
- all messages m_{ij} $i \in \{1, ..., m\}, j \in \{1, ..., m\}$, are uniformly distributed and independent of each other.

Definition 2 An encoding scheme of block length n for general multiterminal network consists of a set of encoding and decoding functions. For node $i \in \{1, ..., m\}$ we have:

- *Encoders:* $X_k^{(i)}(m_{i1},...,m_{im},Y_1^{(i)},Y_2^{(i)},...,Y_{k-1}^{(m)})$ - the encoder maps the messages and past received symbols.
- *Decoders:* The decoder j at node i maps the received symbols in each block and his own transmitted information to form estimates of the messages intended for him from node $j \in \{1,...,m\}$: $\hat{m}^{(ji)}(Y_1^{(i)},...,Y_n^{(i)},m_{i1},...,m_{im})$

- The probability of error for every pair of nodes are: $Pe^{(ij)} = \Pr(M_{ij} \neq \hat{M}_{ij}(Y^{(j),n},m_{j1},...,m_{jm}))$

Definition 3 Achievable region $\{R_{ij}\}_{i=1,j=1}^{m,m}$, if $\sum_{i=1,j=1}^{m,m} Pe^{(ij)} \to 0$ as $n \to \infty$

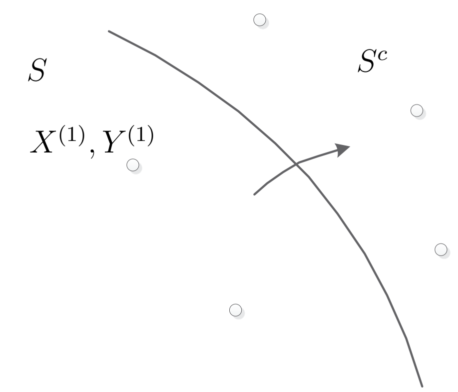

Fig. 2. General Multiterminal Network with a cut-set bound which divides the nodes into two sets S and S^c.

In the Fig. 2 a cut-set bound divides the nodes into two sets, S and S^c -the complement set.

Theorem 1 If $\{R_{ij}\}_{i=1,j=1}^{m,m}$ are achievable, then there exists some joint probability distribution $p(x^S,x^{S^c})$, where x^S - denote the input symbols in the set S, and x^{S^c} - denote the input symbols in the complement set S^c, such that

$$\sum_{i \in S, j \in S^c} R_{ij} \leq I(X^{(S)};Y^{(S^c)}|X^{(S^c)}) \tag{1}$$

for all possible S, where Y is given by channel.

Proof: We will define $\mathcal{T} = \{m_{lj}\}_{l\in S, j\in S^c}$ all the messages that are sent from S to S^c.

$$
\begin{aligned}
n\sum_{l\in S, j\in S^c} R_{lj} &\overset{(a)}{=} H(\mathcal{T}) \\
&\overset{(b)}{=} H(\mathcal{T}|\mathcal{T}^c) \\
&= H(\mathcal{T}|\mathcal{T}^c) - H(\mathcal{T}|\mathcal{T}^c, Y^{(S^c),n}) + H(\mathcal{T}|\mathcal{T}^c, Y^{(S^c),n}) \\
&\overset{(c)}{\leq} I(\mathcal{T}; Y^{(S^c),n}|\mathcal{T}^c) + n\varepsilon_n \\
&\overset{(d)}{=} \sum_{i=1}^{n} I(\mathcal{T}; Y_i^{(S^c)}|\mathcal{T}^c, Y^{(S^c),i-1}) + n\varepsilon_n \\
&\overset{(e)}{\leq} \sum_{i=1}^{n} I(\mathcal{T}, Y^{(S^c),i-1}; Y_i^{(S^c)}|\mathcal{T}^c, Y^{(S^c),i-1}) + n\varepsilon_n \\
&\overset{(f)}{=} \sum_{i=1}^{n} I(\mathcal{T}, Y^{(S^c),i-1}, X_i^{(s)}; Y_i^{(S^c)}|\mathcal{T}^c, Y^{(S^c),i-1}, X_i^{(S^c)}) + n\varepsilon_n \\
&= \sum_{i=1}^{n} H(Y_i^{(S^c)}|\mathcal{T}^c, Y^{(S^c),i-1}, X_i^{(S^c)}) - H(Y_i^{(S^c)}|\mathcal{T}^c, Y^{(S^c),i-1}, X_i^{(S^c)}, \mathcal{T}, Y^{(S^c),i-1}, X_i^{(s)}) \\
&+ n\varepsilon_n \\
&\overset{(g)}{\leq} \sum_{i=1}^{n} H(Y_i^{(S^c)}|X_i^{(S^c)}) - H(Y_i^{(S^c)}|X_i^{(S^c)}, X_i^{(S)}) + n\varepsilon_n
\end{aligned}
$$

Where:

(a) - follows from the fact that the messages m_{ij} are uniformly distributed over $\{1, 2, ..., 2^{nR_{ij}}\}$.

(b) - from the independence of the messages.

(c) - Fano's inequality $H(P_e) + P_e \log_2(|(\mathcal{T})| - 1) \geq H(\mathcal{T}|\hat{\mathcal{T}})$

$$
\begin{aligned}
H(\mathcal{T}|\mathcal{T}^c, Y^{(S^c),n}) &\leq H(\mathcal{T}|Y^{(S^c),n}) \\
&= H(\mathcal{T}|Y^{(S^c),n}, \hat{\mathcal{T}} = g(Y^{(S^c),n})) \leq H(\mathcal{T}|\hat{\mathcal{T}}) \leq n\varepsilon_n(P_e)
\end{aligned}
$$

Where $\varepsilon_n(P_e) \to 0$ as $n \to \infty$

(d) - chain rule.

(e) - follows from the properties of mutual information.

(f) - $X_i^{(S^c)}(\mathcal{T}^c, Y^{(S^c),i-1})$, $X_i^{(S)}(\mathcal{T}, Y^{(S),i-1})$

(g) - follows from Markov chain $(Y^{(m),i-1}, X^{(m),i-1}, \mathcal{T}^c, \mathcal{T}) - X_i^{(m)} - Y_i^{(m)}$ and properties of entropy.

$$\begin{aligned}
\sum_{l\in S, j\in S^c} R_{lj} &\overset{(a)}{\leq} \frac{1}{n}\sum_{i=1}^{n} I(X_i^S; Y_i^{S^c}|X_i^{S^c}) \\
&\overset{(b)}{=} \sum_{i=1}^{n} P(Q=1) I(X_Q^S; Y_Q^{S^c}|X_Q^{S^c}, Q=i) \\
&\overset{(c)}{=} I(X_Q^S; Y_Q^{S^c}|X_Q^{S^c}) \\
&\leq H(Y_Q^{S^c}|X_Q^{S^c}, Q) - H(Y_Q^{S^c}|X_Q^S, X_Q^{S^c}, Q) \\
&\overset{(d)}{\leq} H(Y_Q^{S^c}|X_Q^{S^c}) - H(Y_Q^{S^c}|X_Q^S, X_Q^{S^c}) \\
&= I(X_Q^S; Y_Q^{S^c}|X_Q^{S^c})
\end{aligned}$$

Where:

(a) - follows from the Fano's inequality

(b) - We introduce a new random variable Q, distributed $Q \sim Unif(1,...,n), P(Q) = \frac{1}{n}$, and is independent of X^S, X^{S^c} and Y^{S^c}.

(c) - expectation

(d) - properties of entropy and the fact that the channel is $P(y_Q^{S^c}|x_Q^S, x_Q^{S^c}) = P(y_i^{S^c}|x_i^S, x_i^{S^c}, Q=i)$ and therefore $Y_Q^{S^c}$ depends only on the inputs X_Q^S and $X_Q^{S^c}$ and is conditionally independent of Q. ■

Remark 1 The cut set doesn't provide information about joint probability distribution $P(x^S, x^{S^c})$.

Example 1 Relay Channel

In Lecture 4 we saw that the upper bound for this channel is:

$$C \leq \max_{p(x,x_1)} \min\left(I(X;Y,Y_1|X_1), I(X,X_1;Y)\right) \tag{2}$$

Using cut-set bound 1 we get:

For the following cutting lines in Fig. 3 we have:

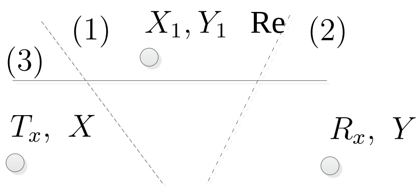

Fig. 3. Cut set for Relay Channel

1) For the cut set (1) we have: $R \leq I(X; Y, Y_1|X_1)$
2) For the cut set (2) the rate is: $R \leq I(X, X_1; Y)$
3) For the cut set (3): $R = 0$ From the definition of rate $R_{ij} = \frac{\log m_{ij}}{n} = 0$ no messages are sent from relay.

For some distribution $p(x, x_1)p(y|x, x_1)$.

Example 2 MAC - Multiple Access Channel

In this example two senders are send to base station. For the following cutting lines in Fig. 5 we have:

1) $R_{13} \leq I(X_1; Y|X_2)$
2) $R_{23} \leq I(X_2; Y|X_1)$
3) $R_{13} + R_{23} \leq I(X_1, X_2; Y)$

Note: It is worth noting that cut-set bound are not always achievable. The cut-set

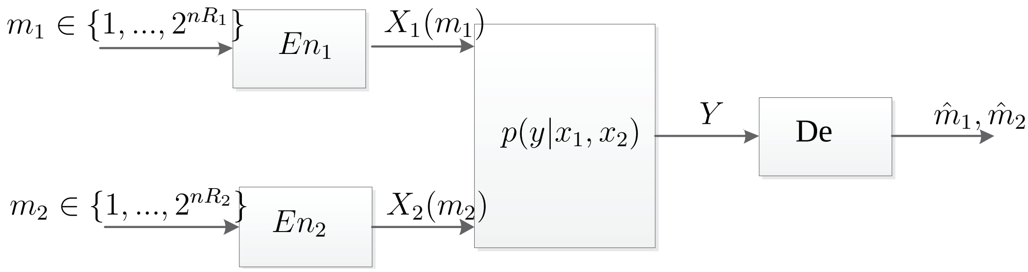

Fig. 4. MAC Channel

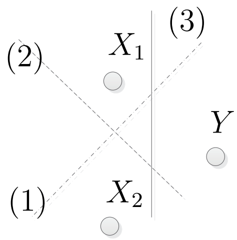

Fig. 5. Cut set for MAC Channel

bound for multiple access channel takes the same form as the capacity, but fails to provide a restriction on the independence of the input distribution $p(x_1, x_2)p(y|x_1, x_2)$. The multiple access channel capacity region is defined with the input distribution of the form $p(x_1)p(x_2)p(y|x_1, x_2)$.

II. Relay Channel - Decode and Forward via Binning

We saw in previous lectures and in Example 1 that for relay channel we have:
Upper bound:

$$C \leq \max_{p(x,x_1)} \min \left(I(X; Y, Y_1|X_1), I(X, X_1; Y) \right) \tag{3}$$

We will prove the lower bound with decode and forward via binning:

Theorem 2 [Decode and Forward via binning]

$$R \leq \max_{p(x,x_1)} \min \left(I(X; Y_1|X_1), I(X, X_1; Y) \right) \tag{4}$$

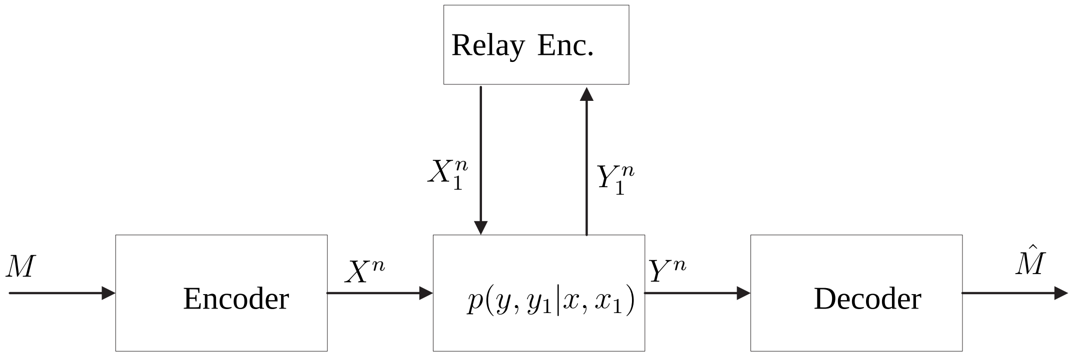

Fig. 6. Relay Channel

Proof: We will use Block Markov coding. Encoding is performed in B blocks each of length n and $b \in \{1, ..., B\}$ is a block index. We generate a separate code book for each block. The idea is that in each block b we decode message m_b and it's bin number l_b and forwarding the bin number to the next block $b+1$. (In block $b+1$, we use l_b as a super bin number in order to decode m_{b+1}.)
l_b - is the function of m_b, that is sent cooperatively by both senders in block $b+1$ to help the receiver decode the message m_{b+1}.

Code design(for block b): Fix $p(x, x_1)$ that achieves the lower bound. Randomly and independently generate $2^{nR'}$ sequences $x_1^n(l_{b-1}) \sim p(x_1^n(l_{b-1})) = \prod_{i=1}^n P_{X_1}(x_{1i})$. In block $b = 1$, $l_{b-1} = l_0 = 1$. $l_{b-1} \in \{1, ..., 2^{nR'}\}$ - is the index of superbin.
For each $x_1^n(l_{b-1})$ superbin generate 2^{nR} sequences $x^n(m_b|l_{b-1})$, $m_b \in \{1, ..., 2^{nR}\}$ according to $P(x^n) = \prod_{i=1}^n P_{X|X_1}(x_i|x_{1i})$. And they are further uniformly divided into $2^{nR'}$ equal size bins. In each bin there are $2^{n(R-R')}$ codewords.

Remark 2 In each supebin we have the same division of messages into bins, that means that if we know $m_b \in \{1, ..., 2^{nR}\}$ we also know $l \in \{1, ..., 2^{nR'}\}$ - the index of bin. In other words we can find l from $m_b(l) \in B(l) = [(l-1)2^{n(R-R')} + 1 : l2^{n(R-R')}]$.

Encoder: In block b sends $x^n(m_b|l_{b-1})$, that means m_b is sent from superbin l_{b-1}.

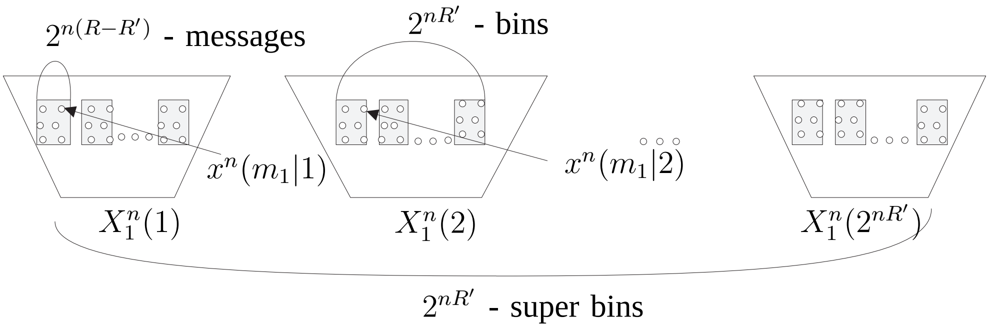

Fig. 7. Distribution of Codewords in Super Bins and Bins in Relay Channel

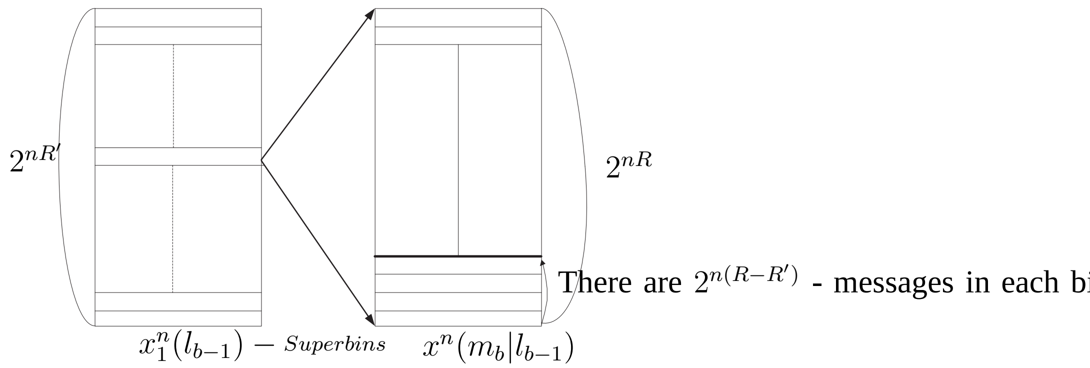

Fig. 8. Coding scheme of Block Markov Coding ,Decode and Forward via Binning. For each codeword x_1^n generate 2^{nR} sequences $x^n(m_b|l_{b-1})$ and divide them into bins

Relay Encoder: In block b: Assume $\tilde{m}_{b-1}$ - estimation of m_{b-1} in relay is known, therefore we know $l_{b-1}(\tilde{m}_{b-1}) = \tilde{l}(m_{b-1})$. (another notation $\tilde{m}_{b-1} \in B(\tilde{l}_{b-1})$). The relay transmits $x_1^n(\tilde{l}_{b-1})$.

Relay Decoder: Upon receiving y_{1b}^n declares at the end of block b that $\tilde{m}_b$ is sent if it is the unique message such that:

$$\left(X^n(\tilde{m}_b|\tilde{l}_{b-1}), X_1^n(\tilde{l}_{b-1}), Y_{1b}^n\right) \in \mathcal{T}_\epsilon^{(n)}(X, X_1, Y) \tag{5}$$

Decoder: First assume that in block $b-1$ we know $\hat{l}_{b-2}$ - that was estimated in the block $b-2$. At the end of block $b-1$ it looks for $\hat{m}_{b-1}$ such that:

$$\left(X^n(\hat{m}_{b-1}|\hat{l}_{b-2}), X_1^n(\hat{l}_{b-2}), Y_{b-1}^n\right) \in \mathcal{T}_\epsilon^{(n)}(X, X_1, Y) \tag{6}$$

and $\hat{m}_{b-1} \in B(\hat{l}_{b-1})$.
Upon receiving y_b^n the receiver declares that $\hat{l}_{b-1}$ is sent if:

$$\left(X_1^n(\hat{l}_{b-1}), Y_b^n\right) \in \mathcal{T}_\epsilon^{(n)}(X_1, Y) \tag{7}$$

It uses $\hat{l}_{b-1}$ in block b for decoding $\hat{m}_b$.

	1	b
Relay En.	$x_1^n(1)$	$x_1^n(\tilde{l}_{b-1})$
En.	$x^n(m_1\|1)$	$x^n(m_b\|l_{b-1})$
Reley Dec.	–	finds $\tilde{m}_b$, $\tilde{l}_b$
Dec.	$\hat{l}_1 = 1$	estimates $\hat{l}_{b-1}$ and $\hat{m}_{b-1}$

Fig. 9. Encoding and decoding in Relay Channel for blocks 1 and b are explained with the help of the table.

Analysis of probability of error: Assume without loss of generality that $(L_{b-2}, L_{b-1}, M_b, \tilde{L}_{b-1}) = (1,1,1,1)$, where $\tilde{l}_{b-1}$ is the relay estimate of l_{b-1}. We can assume see Lecture 5, that there are no errors in the blocks that were decoded previously to block b.
We define the following events:

$$\begin{aligned}
E_1 &= \left\{(X^n(1|1), X_1^n(1), Y_{1b}^n) \notin \mathcal{T}_\epsilon^{(n)}\right\}, && (8)\\
E_{2,j} &= \left\{\exists \tilde{m}_b = j, j \neq 1 : (X^n(j|1), X_1^n(1), Y_{1b}^n) \in \mathcal{T}_\epsilon^{(n)}\right\}, && (9)\\
E_3 &= \left\{(X_1^n(1), Y_b^n) \notin \mathcal{T}_\epsilon^{(n)}\right\}, && (10)\\
E_{4,j} &= \left\{\exists \tilde{l}_{b-1} = j, j \neq 1 : (X_1^n(j), Y_b^n) \in \mathcal{T}_\epsilon^{(n)}\right\}, && (11)\\
E_5 &= \left\{(X^n(1|1), X_1^n(1), Y_{b-1}^n) \notin \mathcal{T}_\epsilon^{(n)}\right\}, && (12)\\
E_{6,j} &= \left\{\exists \hat{m}_{b-1} = j, j \neq 1 : (X^n(j|1), X_1^n(1), Y_{b-1}^n) \in \mathcal{T}_\epsilon^{(n)}, \hat{m}_{b-1} \in B(\hat{L}_{b-1})\right\}, && (13)\\
E_{7,j} &= \left\{\exists \hat{l}_{b-1} = j, j \neq 1 : (X_1^n(j), Y_b^n) \in \mathcal{T}_\epsilon^{(n)}\right\}.
\end{aligned}$$

(14)

By the union bound of events:

$$\begin{aligned} P_e^{(n)} &= Pr(E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5 \cup E_6 \cup E_7) \\ &\leq \sum_{i=1}^{7} P(E_i). \end{aligned}$$

We find the probability of error of each event:

- For $P(E_1) \to 0$, $P(E_3) \to 0$ and $P(E_5) \to 0$ as $n \to \infty$ by L.L.N.

- In order to bound the second term in the last equation, we will use properties of joint typical sequences. When we know that Y_1^n is generated according to $\sim p(y_1|x_1)$ and x^n is generated according to $\sim p(x|x_1)$. We will look at the probability that Y_1^n is jointly typical with x^n. The probability of this event is bounded by:

$$\begin{aligned} \Pr(\cup_j E_{2,j}) &\leq \sum_{j=2}^{2^{nR}} P(E_{2,j}) \\ &\leq \sum_{j=2}^{2^{nR}} \sum_{(x,x_1,y)\in\mathcal{T}_\epsilon^{(n)}} p(x,x_1,y) \\ &\leq \sum_{j=2}^{2^{nR}} \underbrace{2^{n(H(X,X_1,Y)+\epsilon)}}_{Number\ of\ elements\ in\ \mathcal{T}_\epsilon^{(n)}} p(x,x_1)p(y_1|x_1) \\ &\leq \sum_{j=2}^{2^{nR}} 2^{n(H(X,X_1,Y)+\epsilon)} 2^{-n(H(X,X_1)-\epsilon)} 2^{-n(H(Y_1|X_1)-\epsilon)} \\ &\leq \sum_{j=2}^{2^{nR}} 2^{-n(I(X;Y_1|X_1)-3\epsilon)} \\ &= 2^{nR} 2^{-n(I(X;Y_1|X_1)-\delta\epsilon)}. \end{aligned}$$

  $R < I(X;Y_1|X_1) - \delta(\epsilon)$ and $\delta(\epsilon) \to 0$ as $n \to \infty$.

- In order to bound the forth term in the last equation, we will use properties of joint typical sequences. When we look at the probability that y^n, which is generated according to $\sim p(y)$, is jointly typical with x_1^n, which is generated according to $\sim p(x_1)$. The probability of these event is bounded by:

$$
\begin{aligned}
\Pr(\cup_j E_{4,j}) &\leq \sum_{j=2}^{2^{nR'}} P(E_{4,j}) \\
&\leq \sum_{j=2}^{2^{nR'}} 2^{-n(I(X_1;Y)-\delta(\epsilon))} \\
&= 2^{nR'} 2^{-n(I(X_1;Y)-\delta(\epsilon))}.
\end{aligned}
$$

$R' < I(X_1;Y) - \delta(\epsilon)$ and $\delta(\epsilon) \to 0$ as $n \to \infty$.

- In order to bound the six term in the last equation, we look at the probability that y^n, which is generated according to $\sim p(y|x_1)$, is jointly typical with x^n which is generated according to $\sim p(x|x_1)$ where $x_1^n \in \mathcal{T}_\epsilon^{(n)}$. Because in block b-1 decoder know $\hat{l}_{b-1}$ from (7), the bin are known and in each bin there are $2^{n(R-R')}$ messages. The probability of this event is bounded by:

$$
\begin{aligned}
\Pr(\cup_j E_{6,j}) &\leq \sum_{j=2}^{2^{n(R-R')}} P(E_{6,j}) \\
&\leq \sum_{j=2}^{2^{n(R-R')}} 2^{-n(I(X;Y|X_1)-\delta(\epsilon))} \\
&= 2^{n(R-R')} 2^{-n(I(X;Y|X_1)-\delta(\epsilon))}.
\end{aligned}
$$

$R - R' < I(X;Y|X_1) - \delta(\epsilon)$ and $\delta(\epsilon) \to 0$ as $n \to \infty$.

- For the seventh term, similar to the forth term we become: $R' < I(X_1;Y) - \delta(\epsilon)$ and $\delta(\epsilon) \to 0$ as $n \to \infty$.

Therefore we have the following bounds:

$$\begin{cases} R' \leq I(X_1;Y) \\ R \leq R' - I(X;Y|X_1) \\ R \leq I(X;Y|X_1) \end{cases} \tag{15}$$

By using Fourier Motskin elimination (see Lecture 4) from first two equations of (15) we become:

$$I(X_1;Y) \geq R - I(X;Y|X_1) \tag{16}$$

$$R \leq I(X_1;Y) + I(X;Y|X_1) = I(X,X_1;Y) \tag{17}$$

From last equation of (15) and (17):

$$R \leq \max_{p(x,x_1)} \min\left(I(X;Y_1|X_1), I(X,X_1;Y)\right) \tag{18}$$

This completes the proof of achievability. ■

REFERENCES

[1] T. M. Cover and J. A. Thomas, ‘*Elements of Information Theory*‘. Wiley, New York, 2nd edition 2006

[2] T. M. Cover and A. El Gamal, ‘*Capacity theorems for the relay channel*‘. IEEE Trans. Inf. Theory, vol. 25, no. 5, pp. 572-584, Sept. 1979.