Class Road Map - IT-I

- L19 (1/6): Overview, Communications, Gaussian Channel
- L20 (1/8): Gaussian Channel, band limitation, parallel channels, optimization and duality
- L21 (1/13): parallel channels, colored noise, feedback, matrix inequalities
- L22 (1/15): matrix inequalities, rate distortion.
- – (1/20): Monday holiday
- L23 (1/22): rate distortion for Bernoulli, Gaussian, and Multiple Gaussians with unequal noise
- L24 (1/27): main rate distortion theorem, geometry
- L25 (1/29): computing $R(D)$
- L26 (2/3): computing $R(D)$, alternating minimization
- L27 (2/5): Kolmogorov complexity
- L28 (2/10): algorithmic randomness, universal prob.,
- L29 (2/12): universal compression, LZ compression
- – (2/17): Monday, Holiday
- L30 (2/19): LZ compression,
- L31 (2/24): Info measures
- L32 (2/26): Info measures, Info inequalities
- L33 (3/3): Info inequalities, start NIT
- L34 (3/5): NIT, MAC
- L35 (3/10):
- L36 (3/12):

Cumulative Outstanding Reading

- Read Ch. 15 in our book (Cover & Thomas, “Information Theory”).
- Read Ch. 13 in our book (Cover & Thomas, “Information Theory”).
- Read Ch. 14 in our book (Cover & Thomas, “Information Theory”).
- Read Ch. 10 in our book (Cover & Thomas, “Information Theory”).
- Read Ch. 17 in our book (Cover & Thomas, “Information Theory”) on matrix inequalities.
- Read Ch. 9 in our book (Cover & Thomas, “Information Theory”).
- Read Ch. 5 in Boyd and Vandenberghe’s Convex Optimization book.
- Read all readings assigned in EE514a, Fall 2013. (see later lectures on our previous web page (http://j.ee.washington.edu/~bilmes/classes/ee514a_fall_2013/)).
Announcements

- Office hours on Mondays, 3:30-4:30.
- As always, email me if you want to skype/google hangout rather than come to office hours, also at different times.
On Final Presentations

- Your task is to give a 10-15 minute presentation that summarizes 2-3 related and significant papers that come from IEEE Transactions on Information Theory (or a very related area).
- The papers must not be ones that we covered in class, although they can be related.
- You need to do the research to find the papers yourself (i.e., that is part of the assignment).
- The majority of the papers must have been published in the last 10 years (so no old or classic papers).
- Your grade will be based on how clear, understandable, and accurate your presentation is (and also milestones).
- This is a real challenge and will require significant work! Many of the papers are complex. To get a good grade, you will need to work very hard to present very complex ideas in an extremely simple yet still precise way.
- Again, don’t expect this to be easy, you might need to try a few topics until you find one that is suitable.
Final Presentation Milestones

All submissions done in PDF file format via our assignment dropbox (https://canvas.uw.edu/courses/880971/assignments)

- Monday, March 10th, 11:45pm: updated short (≤ 1 page) writeup on more details of how you will present the ideas in a simple fashion.
- Final presentations: Monday, March 17, 2014, 2:30–4:20pm, LOW 102. What to turn in: your slides and a short at most 4 page summary of the papers.
Homework

- Next slide summarizes remaining assignments.
Upcoming assignments summary

- Three one-page summary of assigned papers to read, due Friday at 5pm,
Upcoming assignments summary

- Three one-page summary of assigned papers to read, due Friday at 5pm,
- Final project status update 3, due Mar 10 at 11:45pm,
Upcoming assignments summary

- Three one-page summary of assigned papers to read, due **Friday at 5pm**, 
- Final project status update 3, due **Mar 10 at 11:45pm**, 
- Graded 1 page writeups, due **Mar 11 at 11:45pm**
Upcoming assignments summary

- Three one-page summary of assigned papers to read, due Friday at 5pm,
- Final project status update 3, due Mar 10 at 11:45pm,
- Graded 1 page writeups, due Mar 11 at 11:45pm
- Three one-page summaries of assigned papers (second round), due Mar 14 at 5pm,
Upcomming assignments summary

- Three one-page summary of assigned papers to read, due Friday at 5pm,
- Final project status update 3, due Mar 10 at 11:45pm,
- Graded 1 page writeups, due Mar 11 at 11:45pm
- Three one-page summaries of assigned papers (second round), due Mar 14 at 5pm,
- Final slides and 4 page paper, no lates accepted! due Mar 17 at 1pm,
Upcoming assignments summary

- Three one-page summary of assigned papers to read, due Friday at 5pm,
- Final project status update 3, due Mar 10 at 11:45pm,
- Graded 1 page writeups, due Mar 11 at 11:45pm
- Three one-page summaries of assigned papers (second round), due Mar 14 at 5pm,
- Final slides and 4 page paper, no lates accepted! due Mar 17 at 1pm,
- Final presentations, March 17th, 2:30pm in this room.
Upcomming assignments summary

- Three one-page summary of assigned papers to read, due **Friday at 5pm**, 
- Final project status update 3, due **Mar 10 at 11:45pm**, 
- Graded 1 page writeups, due **Mar 11 at 11:45pm** 
- Three one-page summaries of assigned papers (second round), due **Mar 14 at 5pm**, 
- Final slides and 4 page paper, no lates accepted! due **Mar 17 at 1pm**, 
- Final presentations, **March 17th, 2:30pm** in this room. 
- Graded 1 page writeups (round 2), due **Mar 18 at 11:45pm**
Information Inequalities, and Shannon Type

- Most generally, an “information inequality” is given by defining a set of $k$ sets $\mathcal{C} = \{C_1, \ldots, C_k\}$ and a set of values $\{\alpha_1, \ldots, \alpha_k\}$ with $\alpha_i \in \mathbb{R}$. An information inequality takes the form:

$$\sum_i \alpha_i H(X_{C_i}) \geq 0 \quad (34.1)$$

- E.g., non-negativity of mutual information $I(X_{C_1}; X_{C_2})$ takes the form $H(X_{C_1}) + H(X_{C_2}) - H(X_{C_1 \cup C_2}) \geq 0$ when $C_1 \cap C_2 = \emptyset$.

- As we’ve seen, simple inequalities of the form $H(X) \geq 0$, $H(X|Y) \geq 0$, $I(X; Y) \geq 0$ are all derivable by $I(A; B|C) \geq 0$.

- General “shannon type” inequality takes the form:

$$\sum_i \alpha_i I(A_i; B_i|C_i) \geq 0 \quad (34.2)$$

where $\alpha_i \geq 0$ for all $i$. 
There are others that do not have this form. We in fact saw one already (the one not necessarily satisfied by all polymatroid functions), for four variables $A, B, C, D$ i.e.:

$$2I(C; D) \leq I(A; B) + I(A; C, D) + 3I(C; D|A) + I(C; D|B) \quad (34.1)$$

There are others as well (Dougherty et al., 2006), i.e.,:

$$2I(A; B) \leq 3I(A; B|C) + 3I(A; c|B) + 3I(B; C|A) + 2I(A; D) + 2I(B; C|D) \quad (34.2)$$
$$2I(A; B) \leq 4I(A; B|C) + I(A; C|B) + 2I(B; C|A) + 3I(A; B|D) + I(B; D|A) + 2I(C; D) \quad (34.3)$$
$$2I(A; B) \leq 3I(A; B|C) + 2I(A; C|B) + 4I(B; C|A) + 2I(A; C|D) + I(A; D|C) + 2I(B; D) + I(C; D|A) \quad (34.4)$$
$$2I(A; B) \leq 5I(A; B|C) + 3I(A; C|B) + I(B; C|A) + 2I(A; D) + 2I(B; C|D) \quad (34.5)$$
$$2I(A; B) \leq 4I(A; B|C) + 4I(A; C|B) + I(B; C|A) + 2I(A; D) + 3I(B; C|D) + I(C; D|B) \quad (34.6)$$
$$2I(A; B) \leq 3I(A; B|C) + 2I(A; C|B) + 2I(B; C|A) + 2I(A; B|D) + I(A; D|B) + I(B; D|A) + 2I(C; D) \quad (34.7)$$

These were apparently discovered using computer search.
Degree based inequalities

**Theorem 34.2.1 (Madiman & Tetali 2010)**

Let $X_1, X_2, \ldots, X_n$ be any arbitrary set of discrete random variables jointly distributed according to some distribution. Let $C$ be any set of subsets of indices of the random variables, and assume that for all $v \in [n]$, $r(v) > 0$ (non-zero degree), then we have:

$$\sum_{C \in C} \frac{H(X_C | X_{C^c})}{r_+(C)} \leq H(X_V) \leq \sum_{C \in C} \frac{H(X_C)}{r_-(C)} \quad (34.2)$$

- This is a very general and potentially very powerful set of inequalities, and generalizes a number of results, some of which we already know.
- These are still Shannon! (hold for polymatroid functions).
- This is generalizable to the notion of fractional covers and fractional packings (and continuous entropy for fractional partitions).
Other aggregation functions

- Recall from lecture 2, events $E_k$ each occur with probability $p_k$, $I_k = I(E_k) = -\log p(E_k)$ is the self information of the event, and $H = \sum_k p_k I(E_k)$ is the weighted mean of the self information.

- Consider an “aggregation function” $\varphi$, then we have $\varphi^{-1}(\sum_k p_k \varphi(I(E_k)))$ is a generalized mean, and we get entropy $H$ if $\varphi(x) = x$ is the identity function.

- For any $\varphi$, we still want additivity of independent events, i.e., want:

$$
\varphi^{-1} \left[ \sum_h \sum_k p_h q_k \varphi(I_h + J_k) \right] = \varphi^{-1} \left[ \sum_h p_h \varphi(I_h) \right] + \varphi^{-1} \left[ \sum_k q_k \varphi(J_k) \right]
$$

(34.10)

where $I_k$ and $J_k$ are the information, respectively, of $p_k$ and $q_k$.

- The identity $\varphi(x) = x$ has this property (as entropy is additive for independent events and hence random variables $H(X; Y) = H(X) + H(Y)$ when $X \perp \perp Y$.)
Rényi Entropy

- If instead we use $\varphi(x) = 2^{(1-\alpha)x}$ for $\alpha \neq 1$ we still have activity.
- Doing so, gets us the Rényi Entropy, i.e.,

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \sum_{i=1}^{n} p_i^\alpha$$  \hspace{1cm} (34.10)

for $\alpha > 0$, $\alpha \neq 1$.
- Note $H_\alpha(X)$ is monotonic decreasing with increasing $\alpha$, and

$$\log n \geq H_\alpha(X) \geq -\log p_{\text{max}}$$  \hspace{1cm} (34.11)

where the bounds are achieved at $\alpha = 0$ and $\alpha \to \infty$.
- Also, $\lim_{\alpha \to 1} H_\alpha(X) = H(X)$, so this generalizes Shannon’s entropy.
- There are other entropies as well (Daroczy’s, Quadratic, R-norm, Havrda-Charvát, etc.), all of which are studied both for their mathematical as well as their practical properties.
Rates and Capacities

Definition 34.2.3 (information flow)

The rate of information flow through a channel is given by $I(X; Y)$, the mutual information between $X$ and $Y$, in units of bits per channel use.

Definition 34.2.4 (capacity)

The information capacity of a channel is the maximum information flow.

$$C \triangleq \max_{p(x) \in \Delta} I(X; Y)$$ (34.3)

where $\Delta$ is the set of all possible probability distributions over source alphabet $\mathcal{X}$. Thus, $C$ is the maximum number of bits sent over the channel per channel use.

Definition 34.2.5 (rate)

The rate $R$ of a code is measured in the number of bits per channel use.
For communication, lower bound on probability of error becomes bounded away from 0 as the rate of the code $R$ goes above a fundamental quantity $C$. Note, $P_e \propto e^{-nE(R)}$.

That is, we have a “dual” situation to entropy compression, i.e.,

We will show: only way to get low error is $R < C$. Something funny happens at the point $C$, the channel capacity.

Note that $C$ is not 0, so can still achieve “perfect” communication over a noisy channel as long as $R < C$. 
General Network Information Theory

- Very important part of modern IT (still currently being actively researched).
- A very general case (first). We have an arbitrary network:

Each sender $X_i$ is trying to communicate simultaneously with each receiver $Y_i$ (i.e., for all $i$, $X_i$ is sending to $\{Y_i\}_i$).
- The $X_i$ are not necessarily independent.
The goal is to compute the achievable region of capacities. I.e., a collective vector-valued function \( \vec{C}(\Pr(x_1, x_2, \ldots, x_m)) \).

This is the capacity rates below which the sources can communicate without error (as \( n \to \infty \)).

More generally, let \( V = \{1, 2, \ldots, m\} = [m] \), and let \( S \subseteq V \).

We might want a function \( C : 2^V \to \mathbb{R}_+ \) that gives constraints on the rate limits for communicating sources in \( S \). I.e., constraints might be of the form:

\[
\sum_{s \in S} R_s \leq C(S) \quad \forall S \subseteq V
\]  

(34.10)

Is this polyhedral form most general? No, as we shall see.

General communication network, given by conditional distribution:

\[
\Pr(y^1, y^2, \ldots, y^m | x^1, x^2, \ldots, x^m)
\]  

(34.11)

so a single overall rate is not specific enough.
WLLN and typicality

- The weak law of large numbers, again, says that $\forall S \subseteq V$:

$$-\frac{1}{n} \log \Pr(X_{1:n}^S) = -\frac{1}{n} \sum_{i=1}^{n} \log \Pr(X_i^S) \rightarrow H(X^S) \quad (34.12)$$

when $x_i^S \sim \Pr(x^S)$, and this is true for all $S \subseteq V$ (note again, there are $2^{|V|}$ such subsets here.)

- Define: $\forall S \subseteq V$

$$A_{\epsilon}^{(n)}(S) = \left\{ (x_{1:n}^S) : \left|-\frac{1}{n} \log \Pr(x_{1:n}^{S'}) - H(X^{S'}) \right| < \epsilon, \ \forall S' \subseteq S \right\} \quad (34.13)$$

- Note that this notion of typicality on $S$ requires typicality to hold for all subsets $S'$ of $S$.
- Note, however, that $S = \emptyset$ or $S' = \emptyset$ is vacuous.
Typicality

- Notation: \( a_n \doteq 2^{n(b \pm \epsilon)} \Leftrightarrow \left| \frac{1}{n} \log a_n - b \right| < \epsilon \). Stated another way, \( a_n = \text{poly}(n)2^{n(b \pm \epsilon)} \)

**Theorem 34.2.2 (Typicality)**

\( \forall \epsilon > 0, \exists n_0 \text{ s.t. for } n > n_0, \text{ we have:} \)

1. \( \Pr(A_{\epsilon}^{(n)}(S)) \geq 1 - \epsilon \text{ for all } S \subseteq V \)
2. \( \text{If } x_{1:n}^S \in A_{\epsilon}^{(n)}(S) \text{ then } \Pr(x_{1:n}^S) \doteq 2^{-n(H(X^S) \pm \epsilon)} \)
3. \( |A_{\epsilon}^{(n)}(S)| \doteq 2^{n(H(X^S) \pm \epsilon)} \)
4. \( \text{For } S_1, S_2 \subseteq V, \text{ if } x_{1:n}^{S_1 \cup S_2} \in A_{\epsilon}^{(n)}(S_1 \cup S_2) \text{ then } \Pr(x_{1:n}^{S_1} \mid x_{1:n}^{S_2}) \doteq 2^{-n(H(X^{S_1} \mid X^{S_2}) \pm 2\epsilon)} \).

**Proof.**

Obvious from previous proofs of typicality.
Typicality: we also have

Theorem 34.2.2

For all $S_1, S_2 \subseteq V$ and for all $\epsilon > 0$, we have

$$A^{(n)}_\epsilon(X_{1:n}^{S_1}|x_{1:n}^{S_2}) = \left\{ (x_{1:n}^{S_1} : x_{1:n}^{S_1\cup S_2} \in A^{(n)}_\epsilon(S_1 \cup S_2) \right\} \quad (34.13)$$

(i.e., the set of $S_1$ sequences jointly-typical with a given $S_2$ sequence $x_{1:2}^{S_2}$). Then, if $x_{1:n}^{S_2} \in A^{(n)}_\epsilon(S_2)$, then for large enough $n$, we have:

$$\left| A^{(n)}_\epsilon(X_{1:n}^{S_1}|x_{1:n}^{S_2}) \right| \leq 2^n(H(X^{S_1}|X^{S_2})+2\epsilon) \quad (34.14)$$

And also,

$$(1 - \epsilon)2^n(H(X^{S_1}|X^{S_2})-2\epsilon) \leq \sum_{x_{1:n}^{S_2}} \Pr(x_{1:n}^{S_2}) \left| A^{(n)}_\epsilon(X_{1:n}^{S_1}|x_{1:n}^{S_2}) \right| \quad (34.15)$$

Proof is again obvious given what we’ve done previously.
Conditional Independence and Typicality

- Before we wanted the probability that independent $X, Y$ were jointly typical (i.e., if $(X, Y) \sim p(x)p(y)$ generated from marginals $p(x)p(y)$ of $p(x, y)$, we found that $p((x, y) \in A_{\epsilon}^{(n)}) \approx 2^{-nI(X;Y)}$
- Here, we do a similar thing but use conditional independence.
- I.e., we have $S_1, S_2, S_3 \subseteq V$. If $X^{S_1} \perp \perp X^{S_2} | X^{S_3}$, then $X^{S_1} \rightarrow X^{S_3} \rightarrow X^{S_2}$ forms a Markov chain, and

$$\Pr(x_{1:n}^{S_1 \cup S_2 \cup S_3}) = \prod_{i=1}^{n} p(x^{S_1} | x^{S_2}) p(x^{S_2} | x^{S_3}) p(x^{S_3}) \quad (34.13)$$

**Theorem 34.2.2**

The probability that such conditionally independent variables are typical is:

$$\Pr(x_{1:n}^{S_1 \cup S_2 \cup S_3} \in A_{\epsilon}^{(n)}(S_1 \cup S_2 \cup S_3)) \triangleright 2^{-n(I(S_1;S_2|S_3) \pm 6\epsilon)} \quad (34.14)$$
Multiple Access Channel

- Multiple senders to one receiver, goal is to have the rate of information between the multiple senders and single receiver be as large as possible.

- Senders can’t cooperate/communicate!

- More importantly, goal is to understand the achievable region: what set of rate vectors is achievable (such that as block length gets large, error probability goes to zero).

- Visualized:

\[
p(y \mid x_1, x_2) \quad W_1 \rightarrow X_1 \rightarrow p(y \mid x_1, x_2) \rightarrow Y \rightarrow (\hat{W}_1, \hat{W}_2) \\
W_2 \rightarrow X_2
\]
Multiple Access Channel

- Clearly, $I(X_1, X_2; Y)$ is the rate of transmission but we can’t maximize over $p(x_1, x_2)$, as in $\max_{p(x_1, x_2)} I(X_1, X_2; Y)$, since that would just be point-to-point and would require communication between $X_1$ and $X_2$. We want $X_1 \perp \perp X_2$.

- Senders must deal with noise between each sender and the receiver, but the senders are like noise to each other and must therefore communicate in the presence of this (additional) noise.

- In general, senders can’t communicate (so no chance for cooperate between senders, but we’ll visit this again in TDMA case later).
Multiple Access Channel

- We want to know relationship between $I(X_1, X_2; Y)$, and $R_1, R_2$, and also a coding/decoding algorithm, so that the two senders need not communicate with each other while sending in a way that we can still achieve capacity.
Multiple Access Channel

- We want to know relationship between $I(X_1, X_2; Y)$, and $R_1, R_2$, and also a coding/decoding algorithm, so that the two senders need not communicate with each other while sending in a way that we can still achieve capacity.

- Discrete Memoryless Multi-Access Channel (MAC), is $X_1, X_2, Y$, and $p(y|x_1, x_2)$. 
Multiple Access Channel

- We want to know relationship between $I(X_1, X_2; Y)$, and $R_1, R_2$, and also a coding/decoding algorithm, so that the two senders need not communicate with each other while sending in a way that we can still achieve capacity.
- Discrete Memoryless Multi-Access Channel (MAC), is $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}$, and $p(y|x_1, x_2)$.
- **Definition:** A $((2^{nR_1}, 2^{nR_2}), n)$ code for a MAC is the pair of message indices $W_1 = \{1, \ldots, 2^{nR_1}\}$, $W_2 = \{1, \ldots, 2^{nR_2}\}$;
Multiple Access Channel

- We want to know relationship between \( I(X_1, X_2; Y) \), and \( R_1, R_2 \), and also a coding/decoding algorithm, so that the two senders need not communicate with each other while sending in a way that we can still achieve capacity.

- Discrete Memoryless Multi-Access Channel (MAC), is \( \mathcal{X}_1, \mathcal{X}_2, \mathcal{Y} \), and \( p(y|x_1, x_2) \).

- **Definition:** A \(((2^{nR_1}, 2^{nR_2}), n)\) code for a MAC is the pair of message indices \( W_1 = \{1, \ldots, 2^{nR_1}\}, W_2 = \{1, \ldots, 2^{nR_2}\} \); encoders \( X_1 : W_1 \rightarrow \mathcal{X}_1^n \) and \( X_2 : W_2 \rightarrow \mathcal{X}_2^n \) (so to length-\( n \) strings);
Multiple Access Channel

- We want to know relationship between $I(X_1, X_2; Y)$, and $R_1, R_2$, and also a coding/decoding algorithm, so that the two senders need not communicate with each other while sending in a way that we can still achieve capacity.
- Discrete Memoryless Multi-Access Channel (MAC), is $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}$, and $p(y|x_1, x_2)$.
- **Definition:** A $((2^{nR_1}, 2^{nR_2}), n)$ code for a MAC is the pair of message indices $W_1 = \{1, \ldots, 2^{nR_1}\}, W_2 = \{1, \ldots, 2^{nR_2}\}$; encoders $X_1 : W_1 \rightarrow \mathcal{X}_1^n$ and $X_2 : W_2 \rightarrow \mathcal{X}_2^n$ (so to length-$n$ strings); a decoding function $g : \mathcal{Y}^n \rightarrow W_1 \times W_2$ (we decode both simultaneously).
Multiple Access Channel

- We want to know relationship between $I(X_1, X_2; Y)$, and $R_1, R_2$, and also a coding/decoding algorithm, so that the two senders need not communicate with each other while sending in a way that we can still achieve capacity.

- Discrete Memoryless Multi-Access Channel (MAC), is $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}$, and $p(y|x_1, x_2)$.

- **Definition:** A $((2^{nR_1}, 2^{nR_2}), n)$ code for a MAC is the pair of message indices $W_1 = \{1, \ldots, 2^{nR_1}\}$, $W_2 = \{1, \ldots, 2^{nR_2}\}$; encoders $X_1 : W_1 \rightarrow \mathcal{X}_1^n$ and $X_2 : W_2 \rightarrow \mathcal{X}_2^n$ (so to length-$n$ strings); a decoding function $g : \mathcal{Y}^n \rightarrow W_1 \times W_2$ (we decode both simultaneously).

- Assume $p(w_1, w_2) = \frac{1}{|W_1||W_2|}$ independent and uniform.
Multiple Access Channel

- We want to know relationship between \( I(X_1, X_2; Y) \), and \( R_1, R_2 \), and also a coding/decoding algorithm, so that the two senders need not communicate with each other while sending in a way that we can still achieve capacity.

- Discrete Memoryless Multi-Access Channel (MAC), is \( \mathcal{X}_1, \mathcal{X}_2, \mathcal{Y} \), and \( p(y|x_1, x_2) \).

- **Definition:** A \( (2^{nR_1}, 2^{nR_2}), n \) code for a MAC is the pair of message indices \( W_1 = \{ 1, \ldots, 2^{nR_1} \} \), \( W_2 = \{ 1, \ldots, 2^{nR_2} \} \); encoders \( X_1 : W_1 \rightarrow \mathcal{X}_1^n \) and \( X_2 : W_2 \rightarrow \mathcal{X}_2^n \) (so to length-\( n \) strings); a decoding function \( g : \mathcal{Y}^n \rightarrow W_1 \times W_2 \) (we decode both simultaneously).

- Assume \( p(w_1, w_2) = \frac{1}{|W_1||W_2|} \) independent and uniform.

- **Probability of error:**

\[
P_e^{(n)} = \frac{1}{2^{n(R_1+R_2)}} \sum_{w_1,w_2} \Pr(g(Y_{1:n}) \neq (w_1, w_2) | (w_1, w_2) \text{ sent})
\]  

\text{(34.1)}
Multiple Access Channel: Main Theorem

- **Definition:** a pair \((R_1, R_2)\) is achievable for a MAC if there exists a sequence \(((2^nR_1, 2^nR_2), n)\) of codes with \(P_e^{(n)} \to 0\) as \(n \to \infty\).
Multiple Access Channel: Main Theorem

- **Definition:** a pair \((R_1, R_2)\) is achievable for a MAC if there exists a sequence \(((2^nR_1, 2^nR_2), n)\) of codes with \(P_e(n) \rightarrow 0\) as \(n \rightarrow \infty\).

- **Definition:** The Capacity region is the set of achievable \((R_1, R_2)\) pairs.
Multiple Access Channel: Main Theorem

- **Definition:** a pair $(R_1, R_2)$ is achievable for a MAC if there exists a sequence $((2^nR_1, 2^nR_2), n)$ of codes with $P_e(n) \to 0$ as $n \to \infty$.
- **Definition:** The Capacity region is the set of achievable $(R_1, R_2)$ pairs.

**Theorem 34.3.1**

The MAC capacity of a channel is the closure of the convex hull of all $(R_1, R_2)$ satisfying:

\[
R_1 \leq I(X_1; Y|X_2) \tag{34.2}
\]
\[
R_2 \leq I(X_2; Y|X_1) \tag{34.3}
\]
\[
R_1 + R_2 \leq I(X_1, X_2; Y) \tag{34.4}
\]

under a given product distribution $p(x_1)p(x_2)$. 
Multiple Access Channel

For a particular pair \( p_1(x_1)p_2(x_2) \), this defines a polytope (or more simply a pentagon) in \( \mathbb{R}^2 \) since

\[
\max \{ I(X_1; Y|X_2), I(X_2; Y|X_1) \} \leq I(X_1, X_2; Y) \leq I(X_1; Y|X_2) + I(X_2; Y|X_1),
\]

achievable region
Multiple Access Channel

For a particular pair \( p_1(x_1)p_2(x_2) \), this defines a polytope (or more simply a pentagon) in \( \mathbb{R}^2 \) since \( \max \{ I(X_1; Y|X_2), I(X_2; Y|X_1) \} \leq I(X_1, X_2; Y) \leq I(X_1; Y|X_2) + I(X_2; Y|X_1) \), r.h.s. since \( X_1 \perp \perp X_2 \).
Multiple Access Channel

For a particular pair $p_1(x_1)p_2(x_2)$, this defines a polytope (or more simply a pentagon) in $\mathbb{R}^2$ since $\max \{I(X_1; Y|X_2), I(X_2; Y|X_1)\} \leq I(X_1, X_2; Y) \leq I(X_1; Y|X_2) + I(X_2; Y|X_1)$, r.h.s. since $X_1 \perp \perp X_2$.

For a particular $p_1(x_1), p_2(x_2)$ pair, we have:

\[ R_1 + R_2 = I(X_1, X_2; Y) \]

achievable region
For a particular pair $p_1(x_1)p_2(x_2)$, this defines a polytope (or more simply a pentagon) in $\mathbb{R}^2$ since $\max \{I(X_1; Y | X_2), I(X_2; Y | X_1)\} \leq I(X_1, X_2; Y) \leq I(X_1; Y | X_2) + I(X_2; Y | X_1)$, r.h.s. since $X_1 \perp \perp X_2$.

For a particular $p_1(x_1), p_2(x_2)$ pair, we have:

- Any pair of rates $(R_1, R_2)$ within polytope is achievable.
Multiple Access Channel

- We have $I(X_1; Y|X_2)$ and $I(X_2; Y|X_1)$, $X_1 = \{X_1, X_2\} \setminus X_2$. 
Multiple Access Channel

- We have $I(X_1; Y|X_2)$ and $I(X_2; Y|X_1)$, $X_1 = \{X_1, X_2\} \setminus X_2$.
- Since, $I(A; B|C) = H(A, C) + H(B, C) - H(C) - H(A, B, C)$ we have that $I(S; Y|S^c) =$
  
  $$I(S; Y|V \setminus S) = H(V) - H(V \setminus S) - H(Y, V) + H(Y, V \setminus S)$$

so no immediately apparent nice polyhedral structure here.
Multiple Access Channel

- We have $I(X_1; Y|X_2)$ and $I(X_2; Y|X_1)$, $X_1 = \{X_1, X_2\} \setminus X_2$.
- Since, $I(A; B|C) = H(A, C) + H(B, C) - H(C) - H(A, B, C)$ we have that $I(S; Y|S^c) =\]

$$I(S; Y|V \setminus S) = H(V) - H(V \setminus S) - H(Y, V) + H(Y, V \setminus S)$$

so no immediately apparent nice polyhedral structure here.
- However, the function $f(S) = I(S; Y|V \setminus S) = \text{const.} + H(Y|V \setminus S)$ in fact is polymatroidal (non-negative, monotone non-decreasing, submodular) under the MAC model ($X_i$’s are independent). In other words, the function:

$$f(A) = I(X_A; Y|X_V\setminus A) \quad (34.5)$$

is polymatroidal under the Mac assumption (Han 1979).
We have $I(X_1; Y|X_2)$ and $I(X_2; Y|X_1)$, $X_1 = \{X_1, X_2\} \setminus X_2$.

Since, $I(A; B|C) = H(A, C) + H(B, C) - H(C) - H(A, B, C)$ we have that $I(S; Y|S^c) =$

$$I(S; Y|V \setminus S) = H(V) - H(V \setminus S) - H(Y, V) + H(Y, V \setminus S)$$

so no immediately apparent nice polyhedral structure here.

However, the function $f(S) = I(S; Y|V \setminus S) = \text{const.} + H(Y|V \setminus S)$ in fact is polymatroidal (non-negative, monotone non-decreasing, submodular) under the MAC model ($X_i$'s are independent). In other words, the function:

$$f(A) = I(X_A; Y|X_{V\setminus A}) \quad (34.5)$$

is polymatroidal under the Mac assumption (Han 1979).

In fact, we want to find $p(x_1), p(x_2)$ to make the region as large as possible, so that we might capacity constraints $C_1, C_2, \text{ and } C_{12}$ . . .
Multiple Access Channel

- We have \( I(X_1; Y|X_2) \) and \( I(X_2; Y|X_1), X_1 = \{X_1, X_2\} \setminus X_2. \)
- Since, \( I(A; B|C) = H(A, C) + H(B, C) - H(C) - H(A, B, C) \) we have that \( I(S; Y|S^c) = \)

\[
I(S; Y|V \setminus S) = H(V) - H(V \setminus S) - H(Y, V) + H(Y, V \setminus S)
\]
so no immediately apparent nice polyhedral structure here.
- However, the function \( f(S) = I(S; Y|V \setminus S) = \text{const.} + H(Y|V \setminus S) \)
in fact is polymatroidal (non-negative, monotone non-decreasing, submodular) under the MAC model (\( X_i \)'s are independent). In other words, the function:

\[
f(A) = I(X_A; Y|X_V \setminus A)
\]  \hspace{1cm} (34.5)

is polymatroidal under the Mac assumption (Han 1979).
- In fact, we want to find \( p(x_1), p(x_2) \) to make the region as large as possible, so that we might capacity constraints \( C_1, C_2, \) and \( C_{12} \ldots \)
- Or at least some characterization of the achievable region in general.
Multiple Access Channel

- Simple ex: Suppose we have two independent BSCs, $X_1 \rightarrow Y_1$, $X_2 \rightarrow Y_2$, no interference between channels, $Y = (Y_1, Y_2)$.
Multiple Access Channel

- Simple ex: Suppose we have two independent BSCs, $X_1 \rightarrow Y_1$, $X_2 \rightarrow Y_2$, no interference between channels, $Y = (Y_1, Y_2)$.
- Each channel is a BSC and so has rate $R_i = 1 - H(p_i)$ for $i \in \{1, 2\}$. 
**Multiple Access Channel**

- Simple ex: Suppose we have two independent BSCs, \( X_1 \rightarrow Y_1, \) \( X_2 \rightarrow Y_2, \) no interference between channels, \( Y = (Y_1, Y_2). \)
- Each channel is a BSC and so has rate \( R_i = 1 - H(p_i) \) for \( i \in \{1, 2\}. \)
- The achievable region is a square.

\[
I(X_1; Y | X_2) = 1 - H(P_1) = C_1
\]

\[
I(X_2; Y | X_1) = 1 - H(P_2) = C_2
\]

\[
R_1 + R_2 \leq C_1 + C_2
\]
Multiple Access Channel

Suppose we have a binary multiplier channel, i.e., $Y = X_1 X_2$. 
Multiple Access Channel

- Suppose we have a binary multiplier channel, i.e., $Y = X_1X_2$.
- If $X_1 = 1$, then $X_2$ sends to $Y$ at rate $R_2 = 1 - H(0) = 1$ while $R_1 = 0$. 
Multiple Access Channel

- Suppose we have a binary multiplier channel, i.e., $Y = X_1X_2$.
- If $X_1 = 1$, then $X_2$ sends to $Y$ at rate $R_2 = 1 - H(0) = 1$ while $R_1 = 0$.
- If $X_2 = 1$, then $X_1$ sends to $Y$ at a rate $R_1 = 1 - H(0) = 1$ while $R_2 = 0$. 
Multiple Access Channel

- Suppose we have a binary multiplier channel, i.e., $Y = X_1 X_2$.
- If $X_1 = 1$, then $X_2$ sends to $Y$ at rate $R_2 = 1 - H(0) = 1$ while $R_1 = 0$.
- If $X_2 = 1$, then $X_1$ sends to $Y$ at a rate $R_1 = 1 - H(0) = 1$ while $R_2 = 0$.
- $\text{Max } H(Y) = 1$ and gives limit on rate, so $R_1 + R_2 = 1$. 
Multiple Access Channel

- Suppose we have a binary multiplier channel, i.e., \( Y = X_1 X_2 \).
- If \( X_1 = 1 \), then \( X_2 \) sends to \( Y \) at rate \( R_2 = 1 - H(0) = 1 \) while \( R_1 = 0 \).
- If \( X_2 = 1 \), then \( X_1 \) sends to \( Y \) at a rate \( R_1 = 1 - H(0) = 1 \) while \( R_2 = 0 \).
- \( \text{Max } H(Y) = 1 \) and gives limit on rate, so \( R_1 + R_2 = 1 \).
- Thus, we have triangle shaped polytope:
The next three slides are from Lecture 13.
Binary Erasure Channel

$e$ is an erasure symbol, if that happens we don’t have access to the transmitted bit.

The probability of dropping a bit is then $\alpha$.

We want to compute capacity. Obviously, $C = 1$ if $\alpha = 0$.

\[
C = \max_{p(x)} I(X; Y) = \max_{p(x)} (H(Y) - H(Y|X)) \tag{34.7}
\]

\[
= \max_{p(x)} H(Y) - H(\alpha) \tag{34.8}
\]

So while $H(Y) \leq \log 3$, we want actual value of the capacity.
Binary Erasure Channel

Let $E = \{Y = e\}$. Then

$$H(Y) = H(Y, E) = H(E) + H(Y | E)$$

Let $\pi = \Pr(X = 1)$. Then

$$H(Y) = H((1 - \pi)(1 - \alpha), \alpha, \pi(1 - \alpha))$$

$$= H(\alpha) + (1 - \alpha)H(\pi) \quad (34.7)$$

This last equality follows since $H(E) = H(\alpha)$, and

$$H(Y | E) = \alpha H(Y | Y = e) + (1 - \alpha)H(Y | Y \neq e) = \alpha \cdot 0 + (1 - \alpha)H(\pi)$$
Then we get

\[ C = \max_{p(x)} \left( H(Y) - H(\alpha) \right) \tag{34.7} \]

\[ = \max_{\pi} \left( (1 - \alpha)H(\pi) + H(\alpha) \right) - H(\alpha) \tag{34.8} \]

\[ = \max_{\pi} (1 - \alpha)H(\pi) = 1 - \alpha \tag{34.9} \]

Best capacity when \( \pi = 1/2 = \Pr(X = 1) = \Pr(X = 0) \).

This makes sense, loose \( \alpha \% \) of the bits of original capacity.
Channel Description: $Y = X_1 + X_2$, $|\mathcal{X}_1| = |\mathcal{X}_2| = 2$ while $|\mathcal{Y}| = 3$, so a ternary output alphabet and two binary input alphabets.
Binary erasure MAC

- Channel Description: $Y = X_1 + X_2$, $|X_1| = |X_2| = 2$ while $|Y| = 3$, so a ternary output alphabet and two binary input alphabets.
- If $Y = 0$ then $X_1 = X_2 = 0$ and inputs are unambiguously decodable.
Binary erasure MAC

- Channel Description: \( Y = X_1 + X_2, |X_1| = |X_2| = 2 \) while \( |Y| = 3 \), so a ternary output alphabet and two binary input alphabets.
- If \( Y = 0 \) then \( X_1 = X_2 = 0 \) and inputs are unambiguously decodable.
- Also, if \( Y = 2 \) then \( X_1 = X_2 = 1 \), again inputs are unambiguous.
Binary erasure MAC

- **Channel Description**: $Y = X_1 + X_2$, $|\mathcal{X}_1| = |\mathcal{X}_2| = 2$ while $|\mathcal{Y}| = 3$, so a ternary output alphabet and two binary input alphabets.
- If $Y = 0$ then $X_1 = X_2 = 0$ and inputs are unambiguously decodable.
- Also, if $Y = 2$ then $X_1 = X_2 = 1$, again inputs are unambiguous.
- If $Y = 1$ then two possible values for senders, either $(X_1, X_2) = (0, 1)$ or $(1, 0)$.
Binary erasure MAC

- If $X_2 \equiv 0$ then $R_2 = 0$, $X_1 \rightarrow Y$ and may have $R_1 = 1$
Binary erasure MAC

- If $X_2 \equiv 0$ then $R_2 = 0$, $X_1 \rightarrow Y$ and may have $R_1 = 1$
- To get $R_1 = 1$ need $X_1 \sim \text{Bernoulli}(1/2)$.
Binary erasure MAC

- If $X_2 \equiv 0$ then $R_2 = 0$, $X_1 \rightarrow Y$ and may have $R_1 = 1$.
- To get $R_1 = 1$ need $X_1 \sim \text{Bernoulli}(1/2)$.
- Similarly, if $X_1 \equiv 0$ then $R_1 = 0$, $X_2 \rightarrow Y$ and may have $R_2 = 1$. 
Binary erasure MAC

- If $X_2 \equiv 0$ then $R_2 = 0$, $X_1 \to Y$ and may have $R_1 = 1$.
- To get $R_1 = 1$ need $X_1 \sim \text{Bernoulli}(1/2)$.
- Similarly, if $X_1 \equiv 0$ then $R_1 = 0$, $X_2 \to Y$ and may have $R_2 = 1$.
- Thus, we may achieve the two on-axis extreme points $(0, 1)$ and $(1, 0)$ in the following:
Binary erasure MAC

- Let's assume $R_1 = 1$ so that $X_1 \sim \text{Bernoulli}(1/2)$.
Binary erasure MAC

- Let's assume $R_1 = 1$ so that $X_1 \sim \text{Bernoulli}(1/2)$.
- Thus, $X_1$ looks like noise for $X_2$'s transmission to $Y$. 
Binary erasure MAC

- Lets assume $R_1 = 1$ so that $X_1 \sim \text{Bernoulli}(1/2)$.
- Thus, $X_1$ looks like noise for $X_2$'s transmission to $Y$.
- In fact, this turns $X_2 \rightarrow Y$'s channel into a binary erasure channel with $\alpha = 1/2$ and which (thus) has capacity $C_2 = 1 - 1/2 = 1/2$. 

Lecture 34 - Mar 5th, 2014

Binary erasure MAC

- Lets assume $R_1 = 1$ so that $X_1 \sim \text{Bernoulli}(1/2)$.
- Thus, $X_1$ looks like noise for $X_2$’s transmission to $Y$.
- In fact, this turns $X_2 \to Y$’s channel into a binary erasure channel with $\alpha = 1/2$ and which (thus) has capacity $C_2 = 1 - 1/2 = 1/2$.

Thus, we may achieve the additional extra points $(1, 1/2)$ and $(1/2, 1)$ in the following:
**Binary erasure MAC**

- Lets assume $R_1 = 1$ so that $X_1 \sim \text{Bernoulli}(1/2)$.
- Thus, $X_1$ looks like noise for $X_2$‘s transmission to $Y$.
- In fact, this turns $X_2 \rightarrow Y$‘s channel into a binary erasure channel with $\alpha = 1/2$ and which (thus) has capacity $C_2 = 1 - 1/2 = 1/2$.

Thus, we may achieve the additional extra points $(1, 1/2)$ and $(1/2, 1)$ in the following:

- We can “cheat” with TDMA to get any of the other points (but clever & more computationally demanding coding can also do this).
**Multiple Access Channel: Main Theorem**

- **Definition:** A pair \((R_1, R_2)\) is achievable for a MAC if there exists a sequence \(((2^nR_1, 2^nR_2), n)\) of codes with \(P_e(n) \to 0\) as \(n \to \infty\).
- **Definition:** The Capacity region is the set of achievable \((R_1, R_2)\) pairs.

**Theorem 34.4.1**

The MAC capacity of a channel is the closure of the convex hull of all \((R_1, R_2)\) satisfying:

\[
R_1 \leq I(X_1; Y | X_2) \tag{34.2}
\]

\[
R_2 \leq I(X_2; Y | X_1) \tag{34.3}
\]

\[
R_1 + R_2 \leq I(X_1, X_2; Y) \tag{34.4}
\]

under a given product distribution \(p(x_1)p(x_2)\).
The MAC capacity of a channel is the closure of the convex hull (let's call it $C$) of all $(R_1, R_2)$ satisfying:

\[ R_1 \leq I(X_1; Y | X_2) \]
\[ R_2 \leq I(X_2; Y | X_1) \]
\[ R_1 + R_2 \leq I(X_1, X_2; Y) \]

for $p(x_1, x_2) = p(x_1)p(x_2)$. 

The general shape of the achievable region is illustrated in the diagram.
Theorem: MAC Achievability

**Theorem 34.4.1**

For all rate pairs \((R_1, R_2)\) satisfying for some \(p(x_1, x_2) = p(x_1)p(x_2)\), \(R_1 < I(X_1; Y|X_2)\), \(R_2 < I(X_2; Y|X_1)\), and \(R_1 + R_2 < I(X_1, X_2; Y)\), then there exists a code s.t. \(P_e^{(n)} \to 0\) for \(n \to \infty\).

**Proof.**

- Randomly generate \(2^{nR_k}\) independent codewords \(x_{1:n}^k(i)\) for \(i = 1, \ldots, 2^{nR_k}\) of length \(n\) so that \(x_{1:n}^k(i) \sim \prod_{i=1}^{n} p_k(x_{i}^j)\) for \(k = 1, 2\).
Theorem: MAC Achievability

**Theorem 34.4.1**

For all rate pairs \((R_1, R_2)\) satisfying for some \(p(x_1, x_2) = p(x_1)p(x_2)\), \(R_1 < I(X_1; Y|X_2)\), \(R_2 < I(X_2; Y|X_1)\), and \(R_1 + R_2 < I(X_1, X_2; Y)\), then there exists a code s.t. \(P_e^{(n)} \to 0\) for \(n \to \infty\).

**Proof.**

- Randomly generate \(2^{nR_k}\) independent codewords \(x_{1:n}^k(i)\) for \(i = 1, \ldots, 2^{nR_k}\) of length \(n\) so that \(x_{1:n}^k(i) \sim \prod_{i=1}^n p_k(x_i^j)\) for \(k = 1, 2\).
- Codebooks known to both senders and the receiver.
Theorem: MAC Achievability

**Theorem 34.4.1**

For all rate pairs \((R_1, R_2)\) satisfying for some \(p(x_1, x_2) = p(x_1)p(x_2)\), \(R_1 < I(X_1; Y|X_2)\), \(R_2 < I(X_2; Y|X_1)\), and \(R_1 + R_2 < I(X_1, X_2; Y)\), then there exists a code s.t. \(P_e^{(n)} \to 0\) for \(n \to \infty\).

**Proof.**

- Randomly generate \(2^{nR_k}\) independent codewords \(x_{1:n}^k(i)\) for \(i = 1, \ldots, 2^{nR_k}\) of length \(n\) so that \(x_{1:n}^k(i) \sim \prod_{i=1}^n p_k(x_i^j)\) for \(k = 1, 2\).
- Codebooks known to both senders and the receiver.
- Encoding: Sender \(k\) sending message \(i\) sends \(x_{1:n}^k(i)\) over channel.
Theorem: MAC Achievability

**Theorem 34.4.1**

For all rate pairs \((R_1, R_2)\) satisfying for some \(p(x_1, x_2) = p(x_1)p(x_2)\), \(R_1 < I(X_1; Y|X_2)\), \(R_2 < I(X_2; Y|X_1)\), and \(R_1 + R_2 < I(X_1, X_2; Y)\), then there exists a code s.t. \(P_e^{(n)} \to 0\) for \(n \to \infty\).

**Proof.**

- Randomly generate \(2^{nR_k}\) independent codewords \(x_{1:n}^k(i)\) for \(i = 1, \ldots, 2^{nR_k}\) of length \(n\) so that \(x_{1:n}^k(i) \sim \prod_{i=1}^{n} p_k(x_{i}^j)\) for \(k = 1, 2\).
- Codebooks known to both senders and the receiver.
- Encoding: Sender \(k\) sending message \(i\) sends \(x_{1:n}^k(i)\) over channel.
- Decoding: \(A^{(n)}_\epsilon\) is the set of typical \((x_{1:n}^1, x_{1:n}^2, y_{1:n})\) sequences. Choose \((i, j)\) such that \((x_{1:n}^1(i), x_{1:n}^2(j), y_{1:n}) \in A^{(n)}_\epsilon\) if it exists, and otherwise error will occur.

Note: no TDMA required.
Theorem: MAC Achievability

Theorem 34.4.1

For all rate pairs \((R_1, R_2)\) satisfying for some \(p(x_1, x_2) = p(x_1)p(x_2)\), \(R_1 < I(X_1; Y|X_2)\), \(R_2 < I(X_2; Y|X_1)\), and \(R_1 + R_2 < I(X_1, X_2; Y)\), then there exists a code s.t. \(P_e^{(n)} \to 0\) for \(n \to \infty\).

Proof.

- Randomly generate \(2^{nR_k}\) independent codewords \(x_{1:n}^k(i)\) for \(i = 1, \ldots, 2^{nR_k}\) of length \(n\) so that \(x_{1:n}^k(i) \sim \prod_{i=1}^{n} p_k(x_i^j)\) for \(k = 1, 2\).
- Codebooks known to both senders and the receiver.
- Encoding: Sender \(k\) sending message \(i\) sends \(x_{1:n}^k(i)\) over channel.
- Decoding: \(A_\epsilon^{(n)}\) is the set of typical \((x_{1:n}^1, x_{1:n}^2, y_{1:n})\) sequences.
  Choose \((i, j)\) such that \((x_{1:n}^1(i), x_{1:n}^2(j), y_{1:n}) \in A_\epsilon^{(n)}\) if it exists, and otherwise error will occur.
- Note: no TDMA required.
proof of Theorem 34.4.1 continued.

- **Symmetry:** Random code construction, so error does not depend on which index pair was sent (when sending an index pair, all possible codebooks are possible with non-zero probability, and we average them all out).
**Symmetry:** Random code construction, so error does not depend on which index pair was sent (when sending an index pair, all possible codebooks are possible with non-zero probability, and we average them all out).

**Therefore, assume** $(i, j) = (1, 1)$ (generalizing point-to-point case).
proof of Theorem 34.4.1 continued.

- Symmetry: Random code construction, so error does not depend on which index pair was sent (when sending an index pair, all possible codebooks are possible with non-zero probability, and we average them all out).
- Therefore, assume \((i, j) = (1, 1)\) (generalizing point-to-point case).
- Events, joint typicality: \(E_{ij} = \left\{ (x_{1:n}^1(i), x_{1:n}^2(j), y_{1:n}) \in A_{\epsilon}^{(n)} \right\} \).
proof of Theorem 34.4.1 continued.

- Symmetry: Random code construction, so error does not depend on which index pair was sent (when sending an index pair, all possible codebooks are possible with non-zero probability, and we average them all out).
- Therefore, assume \((i, j) = (1, 1)\) (generalizing point-to-point case).
- Events, joint typicality: 
  \[ E_{ij} = \left\{ (x_{1:n}^1(i), x_{1:n}^2(j), y_{1:n}) \in A_{\epsilon}^{(n)} \right\}. \]
- We can write and bound the probability of error:

\[
P_e^{(n)} = \Pr(E_{11}^c \cup \bigcup_{(i,j)\neq(1,1)} E_{ij}) \]
\[
\leq \Pr(E_{11}^c) + \sum_{j=1,i\neq1} \Pr(E_i) + \sum_{i=1,j\neq1} \Pr(E_{1j}) + \sum_{i\neq1,j\neq1} \Pr(E_{ij})
\]

where inequality is by the union bound.
proof of Theorem 34.4.1 continued.

- Clearly, $\Pr(E_{11}^c) \to 0$ by joint typicality.
Theorem: MAC Achievability

Proof of Theorem 34.4.1 continued.

- Clearly, \( \Pr(E_{11}^c) \to 0 \) by joint typicality.
- We next bound \( \Pr(E_{i1}) \).

\[
\Pr(E_{i1}) = \Pr((x_1^{1:n}(i), x_2^{1:n}(1), y_1:n) \in A_{\epsilon}^{(n)})
\]
\[
= \Pr(\text{indep. events } x_1^{1:n}(i) \text{ and } x_2^{1:n}(1), y_1:n \text{ jointly typical})
\]
\[
= \sum_{x_1^{1:n}(i), x_2^{1:n}(1) \in A_{\epsilon}^{(n)}} \Pr(x_1^{1:n}(i)) \Pr(x_2^{1:n}(1), y_1:n)
\]
\[
\leq |A_{\epsilon}^{(n)}| 2^{-n(H(X) - \epsilon)} 2^{-n(H(X_2,Y) - \epsilon)}
\]
\[
\leq 2^{-n(-H(X_1,X_2,Y) + H(X_1) + H(X_2,Y) - 3\epsilon)}
\]
\[
= 2^{-n(I(X_1;X_2,Y) - 3\epsilon)}
\]
\[
= 2^{-n(I(X_1;Y|X_2) - 3\epsilon)} \quad \text{since } X_1 \perp \!\!\!\!\!\! X_2
\]
proof of Theorem 34.4.1 continued.

Also,

\[ \Pr(E_{1j}) \leq 2^{-n(I(X_2;Y|X_1)-3\epsilon)} \]  \hspace{1cm} (34.15)

and

\[ \Pr(E_{ij}) \leq 2^{-n(I(X_1,X_2;Y)-4\epsilon)} \]  \hspace{1cm} (34.16)
Also,

\[ \Pr(E_{1j}) \leq 2^{-n(I(X_2;Y|X_1)-3\epsilon)} \]  

(34.15)

and

\[ \Pr(E_{ij}) \leq 2^{-n(I(X_1,X_2;Y)-4\epsilon)} \]  

(34.16)

thus, we have

\[ P_e^{(n)} \leq \Pr(E_{11}^c) + 2^nR_1 2^{-n(I(X_1;Y|X_2)-3\epsilon)} + 2^nR_2 2^{-n(I(X_2;Y|X_1)-3\epsilon)} \]

\[ + 2^n(R_1+R_2) 2^{-n(I(X_1,X_2;Y)-4\epsilon)} \]  

(34.17)

\[ \rightarrow 0 \text{ as } n \rightarrow \infty \]  

(34.18)

for the given constraints on \( R_1, R_2 \).
MAC Achievability, discussion

- First, recall that \( I(X_1; Y | X_2) + I(X_2; Y) = I(X_1, X_2; Y) \).
First, recall that $I(X_1; Y | X_2) + I(X_2; Y) = I(X_1, X_2; Y)$.

Why these extreme points of the polytope?
MAC Achievability, discussion

- First, recall that $I(X_1; Y|X_2) + I(X_2; Y) = I(X_1, X_2; Y)$.

Why these extreme points of the polytope?

- One way to do this is, say, have $X_2$ communicate at rate $I(X_2; Y)$ and $X_1$ communicate at rate $I(X_1; Y|X_2)$. 
First, recall that \( I(X_1; Y|X_2) + I(X_2; Y) = I(X_1, X_2; Y) \).

Why these extreme points of the polytope?

One way to do this is, say, have \( X_2 \) communicate at rate \( I(X_2; Y) \) and \( X_1 \) communicate at rate \( I(X_1; Y|X_2) \).

Decoder first declares \( w_2 \) sent if \( (x_2(w_2), y) \in A^{(n)}_\epsilon \) (error if not) and then, for that \( w_2 \), declares \( w_1 \) sent if \( (x_1(w_1), x_2(w_2), y) \in A^{(n)}_\epsilon \) – this achieves one extreme point (other is symmetric case).