Outstanding Reading

- Read all chapters assigned from IT-I (EE514, Winter 2012).
- Read chapter 8 in the book.
- Read chapter 9 in the book.
- Read chapter 10 in the book (chapter on rate distortion theory).
- Read chapter 14 in the book (Kolmogorov complexity).
- Read chapter 13, section on Lempel Ziv compression, in the book.
- Read chapter 15 in C&T.
Please do use our discussion board (https://catalyst.uw.edu/gopost/board/bilmes/27386/) for all questions, comments, so that all will benefit from them being answered.
On Final Presentations

- Your task is to give a 15-20 minute presentation that summarizes 2-3 related and significant papers that come from IEEE Transactions on Information Theory (or a very related area).
On Final Presentations

- Your task is to give a 15-20 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.
On Final Presentations

- Your task is to give a 15-20 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).
On Final Presentations

- Your task is to give a 15-20 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 papers must have been published in the last 10 years (so no old or classic papers).
On Final Presentations

- Your task is to give a 15-20 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 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.
On Final Presentations

- Your task is to give a 15-20 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 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.
- 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.
On Final Presentations

- Your task is to give a 15-20 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 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.
- 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 dropbox
(https://catalyst.uw.edu/collectit/dropbox/bilmes/21171)

- Wed, May 2nd: Candidate proposed papers submitted. Include short at most 1-page write up: 1) why you chose these papers; 2) why they are important to pure IT; and 3) how they are fundamental and/or deep, and 4) how will you summarize them in a simple and precise way.
Final Presentation Milestones

All submissions done in PDF file format via our dropbox (https://catalyst.uw.edu/collectit/dropbox/bilmes/21171)

- Wed, May 2nd: Candidate proposed papers submitted. Include short at most 1-page write up: 1) why you chose these papers; 2) why they are important to pure IT; and 3) how they are fundamental and/or deep, and 4) how will you summarize them in a simple and precise way.

- Friday, May 11th: Updated list of proposed papers decided, based on feedback. Updated write up (noting progress)
Final Presentation Milestones

All submissions done in PDF file format via our dropbox (https://catalyst.uw.edu/collectit/dropbox/bilmes/21171)

- Wed, May 2nd: Candidate proposed papers submitted. Include short at most 1-page write up: 1) why you chose these papers; 2) why they are important to pure IT; and 3) how they are fundamental and/or deep, and 4) how will you summarize them in a simple and precise way.

- Friday, May 11th: Updated list of proposed papers decided, based on feedback. Updated write up (noting progress)

- Friday, May 18th: short write up on more details of how you will present the ideas in a simple fashion.
Final Presentation Milestones

All submissions done in PDF file format via our dropbox (https://catalyst.uw.edu/collectit/dropbox/bilmes/21171)

- Wed, May 2nd: Candidate proposed papers submitted. Include short at most 1-page write up: 1) why you chose these papers; 2) why they are important to pure IT; and 3) how they are fundamental and/or deep, and 4) how will you summarize them in a simple and precise way.
- Friday, May 11th: Updated list of proposed papers decided, based on feedback. Updated write up (noting progress)
- Friday, May 18th: short write up on more details of how you will present the ideas in a simple fashion.
- Friday, May 25th: updated short write up on more details of how you will present the ideas in a simple fashion.
Final Presentation Milestones

All submissions done in PDF file format via our dropbox (https://catalyst.uw.edu/collectit/dropbox/bilmes/21171)

- Wed, May 2nd: Candidate proposed papers submitted. Include short at most 1-page write up: 1) why you chose these papers; 2) why they are important to pure IT; and 3) how they are fundamental and/or deep, and 4) how will you summarize them in a simple and precise way.

- Friday, May 11th: Updated list of proposed papers decided, based on feedback. Updated write up (noting progress)

- Friday, May 18th: short write up on more details of how you will present the ideas in a simple fashion.

- Friday, May 25th: updated short write up on more details of how you will present the ideas in a simple fashion.

- Final presentations: Monday, June 4th in the afternoon late/evening (currently scheduled for 8:30am but that is too early). What to turn in: your slides and a short at most 4 page summary of the papers.
Problem sets 1 and 2

All problem sets must be turned in via PDF files via our dropbox (https://catalyst.uw.edu/collectit/dropbox/bilmes/21171)

- Problem set 1, due tonight at 11:00pm, see the problems listed on pdf page 161 of http://j.ee.washington.edu/~bilmes/classes/ee515a_spring_2012/lecture28_presented.pdf.

- Problem set 2, was due next Monday, May 21st, 11:45pm: Do book problems: 8.1, 8.8, 9.1, 9.2, 9.6, 10.5, 10.6, 13.5, 13.6, 14.3, 14.4, 14.5
General Network Information Theory

• Very important part of modern IT (still currently being actively researched).
Very important part of modern IT (still currently being actively researched).

The most general case (first). We have an arbitrary network:

\[ X_1, Y_1 \quad X_2, Y_2 \]
\[ X_m, Y_m \quad X_3, Y_3 \]
\[ \ldots \ldots \quad X_4, Y_4 \]
Very important part of modern IT (still currently being actively researched).

The most 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 $\{Y_i\}_i$)
Very important part of modern IT (still currently being actively researched).

The most 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 $\vec{C}(\Pr(x_1, x_2, \ldots, x_m))$. 
General Network Information Theory

- The goal is to compute the achievable region of capacities. I.e., a collective vector-valued function $\tilde{C}(Pr(x_1, x_2, \ldots, x_m))$.

- This is the capacity over which the sources can communicate without error (as $n \to \infty$).
General Network Information Theory

- 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 over which the sources can communicate without error (as $n \to \infty$).
- More generally, let $V = \{1, 2, \ldots, m\} = [m]$, and let $S \subseteq V$. 

General Network Information Theory

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

$$\sum_{s \in S} R_s \leq C(S) \quad \forall S \subseteq V \tag{1}$$
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 over 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 want a function $C : 2^V \to \mathbb{R}_+$ that gives constraints on the rate limits for communicating sources in $S$. I.e., constraints would be of the form:

$$\sum_{s \in S} R_S \leq C(S) \quad \forall S \subseteq V$$

General communication network is then:

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

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

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

$$-\frac{1}{n} \log \Pr(X_1^S) = -\frac{1}{n} \sum_{i=1}^{n} \log \Pr(X_i^S) \to H(X^S)$$  \hspace{1cm} (3)

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.)
WLLN and typicality

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

$$-\frac{1}{n} \log \Pr(X_{1:n}^S) = -\frac{1}{n} \sum_{i=1}^{n} \log \Pr(X_i^S) \to H(X^S)$$  \hspace{1cm} (3)

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: $\forall S \subseteq V$

$$A_\epsilon^{(n)}(S) = \left\{ (x_{1:n}^S) : \left| -\frac{1}{n} \log \Pr(x_{1:n}^{S'}) - H(X^{S'}) \right| < \epsilon, \ \forall S' \subseteq S \right\}$$  \hspace{1cm} (4)
WLLN and typicality

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

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

(3)

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: $\forall S \subseteq V$

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

(4)

- Note that this notion of typicality on $S$ requires typicality to hold for all subsets of $S$. 

Prof. Jeff Bilmes
the weak law of large numbers, again, says that $\forall S \subseteq V$:

$$-rac{1}{n} \log \Pr(X_{1:n}^S) = -\frac{1}{n} \sum_{i=1}^{n} \log \Pr(X_i^S) \to H(X^S)$$  \hspace{1cm} (3)

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: $\forall S \subseteq V$

$$A_\epsilon^{(n)}(S) = \left\{ (x_{1:n}^S) : \left| -\frac{1}{n} \log \Pr(x_{1:n}^{S'}) - H(X^{S'}) \right| < \epsilon, \ \forall S' \subseteq S \right\}$$  \hspace{1cm} (4)

Note that this notion of typicality on $S$ requires typicality to hold for all subsets of $S$.

Note, however, that $S = \emptyset$ or $S' = \emptyset$ is vacuous.
Typicality

- Notation: $a_n \doteq 2^n(b \pm \epsilon) \iff \left| \frac{1}{n} \log a_n - b \right| < \epsilon$. 
Typicality

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

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

Theorem 2.1 (Typicality)

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

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

**Theorem 2.1 (Typicality)**

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

1. $\Pr(A_\epsilon^n(S)) \geq 1 - \epsilon \text{ for all } S \subseteq V$
Typicality

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

Theorem 2.1 (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. If $x_{1:n}^S \in A_\epsilon^{(n)}(S)$ then $\Pr(x_{1:n}^S) \doteq 2^{-n(H(X^S) \pm \epsilon)}$
Typicality

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

Theorem 2.1 (Typicality)

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

1. \( \Pr(A^{(n)}_\epsilon(S)) \geq 1 - \epsilon \text{ for all } S \subseteq V \)
2. If \( x_{1:n}^S \in A^{(n)}_\epsilon(S) \) then \( \Pr(x_{1:n}^S) \doteq 2^{-n(H(X^S) \pm \epsilon)} \)
3. \( |A^{(n)}_\epsilon(S)| \doteq 2^n(H(X^S) \pm \epsilon) \)
Typicality

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

**Theorem 2.1 (Typicality)**

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

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

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

Theorem 2.1 (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. If \( x_{1:n}^S \in A_\epsilon^{(n)}(S) \text{ then } \Pr(x_{1:n}^S) = 2^{-n(H(X^S) \pm \epsilon)} \)
3. \( |A_\epsilon^{(n)}(S)| = 2^n(H(X^S) \pm \epsilon) \)
4. For \( S_1, S_2 \subseteq V \), if \( x_{1:n}^{S_1 \cup S_2} \in A_\epsilon^{(n)}(S_1 \cup S_2) \text{ then} \)
   \( \Pr(x_{1:n}^{S_1} | x_{1:n}^{S_2}) = 2^{-n(H(X^{S_1} | X^{S_2}) \pm 2\epsilon)} \).

Proof.

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

**Theorem 2.2**

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

$$A^{(n)}(X_{1:n}^{S_1} | x_{1:n}^{S_2}) = \left\{ (x_{1:n}^{S_1} : x_{1:n}^{S_1 \cup S_2} \in A^{(n)}(S_1 \cup S_2) \right\}$$  \hspace{1cm} (5)

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

$$\left| A^{(n)}(X_{1:n}^{S_1} | x_{1:n}^{S_2}) \right| \leq 2^n (H(X^{S_1} | X^{S_2}) + 2 \epsilon)$$  \hspace{1cm} (6)
Typicality: we also have

**Theorem 2.2**

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

$$A_\epsilon^{(n)}(X_{1:n}^{S_1} | x_{1:n}^{S_2}) = \left\{ (x_{1:n}^{S_1} : x_{1:n}^{S_1 \cup S_2} \in A_\epsilon^{(n)}(S_1 \cup S_2) \right\}$$  \hspace{1cm} (5)

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

$$\left| A_\epsilon^{(n)}(X_{1:n}^{S_1} | x_{1:n}^{S_2}) \right| \leq 2^n(H(X^{S_1} | X^{S_2})+2\epsilon)$$  \hspace{1cm} (6)

And also,

$$(1 - \epsilon)2^n(H(X^{S_1} | X^{S_2})-2\epsilon) \leq \sum_{x_{1:n}^{S_2}} \text{Pr}(x_{1:n}^{S_2}) \left| A_\epsilon^{(n)}(X_{1:n}^{S_1} | x_{1:n}^{S_2}) \right|$$  \hspace{1cm} (7)
Typicality: we also have

**Theorem 2.2**

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

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

(5)

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

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

(6)

And also,

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

(7)

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 
\[
p((x, y) \in A_{\epsilon}^{(n)}) \approx 2^{-nI(X;Y)}
\]
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.
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 $p((x, y) \in A_{\epsilon}^{(n)}) \approx 2^{-n I(X;Y)}$).

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

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

(8)
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 $p((x, y) \in A^{(n)}_\epsilon) \approx 2^{-nI(X;Y)}$)

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

$$\Pr(x^{S_1 \cup S_2 \cup S_3}_{1:n}) = \prod_{i=1}^{n} p(x^{S_1 | x^{S_2}})p(x^{S_2 | x^{S_3}})p(x^{S_3}) \quad (8)$$

**Theorem 2.3**

$$\Pr(x^{S_1 \cup S_2 \cup S_3}_{1:n} \in A^{(n)}_\epsilon(S_1 \cup S_2 \cup S_3)) \approx 2^{-n(I(S_1;S_2 | S_3) \pm 6\epsilon)} \quad (9)$$
Multiple Access Channel

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

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

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

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

Clearly, $I(X_1, X_2; Y)$ is the rate of transmission but we can’t maximize over $p(x_1, x_2)$ since that would just be point-to-point and would require communication between $X_1$ and $X_2$. We want $X_1 \perp \perp X_2$. 
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 $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}$, 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 $X_1, X_2, Y$, 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 $p(w_1, w_2) = \frac{1}{|W_1||W_2|}$ uniform.
Multiple Access Channel

- We want to know the 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 \( g : Y^n \rightarrow W_1 \times W_2 \) (we decode both simultaneously).

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

- **Probability of error:**

\[
P_e^{(n)} = \frac{1}{2^n(R_1+R_2)} \sum_{w_1,w_2} \Pr(g(Y_{1:n}) \neq (w_1, w_2) | (w_1, w_2) \text{ sent}) \quad (10)
\]
Multiple Access Channel

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

- **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.
Multiple Access Channel

- **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 2.4**

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

\[
R_1 \leq I(X_1; Y | X_2) \quad (11)
\]

\[
R_2 \leq I(X_2; Y | X_1) \quad (12)
\]

\[
R_1 + R_2 \leq I(X_1, X_2; Y) \quad (13)
\]

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

We can view one instance of this as a polytope (or more simply a pentagon) in $\mathbb{R}^2$ since $\max \{I(X_1; Y|X_2), I(X_2; Y|X_1)\} \leq I(X_1X_2; Y) \leq I(X_1; Y|X_2) + I(X_2; Y|X_1)$.
Multiple Access Channel

- We can view one instance of this as a polytope (or more simply a pentagon) in $\mathbb{R}^2$ since $\max \{ I(X_1; Y | X_2), I(X_2; Y | X_1) \} \leq I(X_1X_2; Y) \leq I(X_1; Y | X_2) + I(X_2; Y | X_1)$
- For a particular $p_1(x_1), p_2(x_2)$ pair, we have:
Multiple Access Channel

- We can view one instance of this as a polytope (or more simply a pentagon) in $\mathbb{R}^2$ since $\max \{I(X_1; Y|X_2), I(X_2; Y|X_1)\} \leq I(X_1X_2; Y) \leq I(X_1; Y|X_2) + I(X_2; Y|X_1)$

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

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

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

\[ f(s) = I(S; Y | V S) \]

Since,

\[ I(A; B | C) = H(A, C) + H(B, C) - H(C) - H(A, B, C) \]

we have

\[ I(S; Y | V S) = H(V) - H(V S) - H(Y, V) + H(Y, V S) \]

so no immediately apparent nice structure here.

\[ L = \text{const} + H(Y, V S) - H(V S) \]

\[ \leq \text{const} + H(Y | V S) \]

\[ \geq \text{const} + \sum_{s \in V(S)} H(s) \]
Since, \( I(A; B|C) = H(A, C) + H(B, C) - H(C) - H(A, B, C) \) we have \( I(S; Y|V \setminus S) = H(V) - H(V \setminus S) - H(Y, V) + H(Y, V \setminus S) \) so no immediately apparent nice 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).
Multiple Access Channel

- Since, $I(A; B|C) = H(A, C) + H(B, C) - H(C) - H(A, B, C)$ we have $I(S; Y|V \setminus S) = H(V) - H(V \setminus S) - H(Y, V) + H(Y, V \setminus S)$ so no immediately apparent nice 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 fact, we want to find $p(x_1), p(x_2)$ to make the region as large as possible, so that we have capacity constraints $C_1, C_2, \text{ and } C_{12}$. 
Since, \( I(A; B|C) = H(A, C) + H(B, C) - H(C) - H(A, B, C) \) we have \( I(S; Y|V \setminus S) = H(V) - H(V \setminus S) - H(Y, V) + H(Y, V \setminus S) \), so no immediately apparent nice 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 fact, we want to find \( p(x_1), p(x_2) \) to make the region as large as possible, so that we have capacity constraints \( C_1, C_2, \) and \( C_{12} \).

But the achievable region in general
Some simple examples: Suppose we have two independent BSCs with no interference between channels.
Some simple examples: Suppose we have two independent BSCs with no interference between channels.

Each channel is a BSC and so has rate $R_i = 1 - H(p_i)$ for $i \in \{1, 2\}$. 
Some simple examples: Suppose we have two independent BSCs with no interference between channels.

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.

\[
\begin{align*}
I(X_2; Y | X_1) &= 1 - H(P_2) \\
&= C_2
\end{align*}
\]

\[
\begin{align*}
I(X_1; Y | X_2) &= 1 - H(P_1) = C_1
\end{align*}
\]

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

Suppose we have two binary multiplier channels, $Y = X_1X_2$. 

Max $H(Y)$ = 1 and gives limit on rate, so $R_1 + R_2 = 1$. Thus, we have triangle shaped polytope:
Multiple Access Channel

- Suppose we have two binary multiplier channels, $Y = X_1 X_2$.
- If $X_1 = 1$, then $X_2$ sent to $Y$ at rate $1 - H(0) = 1$. 

Max $H(Y)$ = 1 and gives limit on rate, so $R_1 + R_2 = 1$. Thus, we have triangle shaped polytope:
Multiple Access Channel

- Suppose we have two binary multiplier channels, $Y = X_1 X_2$.
- If $X_1 = 1$, then $X_2$ sent to $Y$ at rate $1 - H(0) = 1$.
- If $X_2 = 1$, then $X_1$ sent to $Y$ at a rate $1 - H(0) = 1$. 

Max $H(Y)$ = 1 and gives limit on rate, so $R_1 + R_2 = 1$.
Multiple Access Channel

- Suppose we have two binary multiplier channels, $Y = X_1 X_2$.
- If $X_1 = 1$, then $X_2$ sent to $Y$ at rate $1 - H(0) = 1$.
- If $X_2 = 1$, then $X_1$ sent to $Y$ at a rate $1 - H(0) = 1$.
- Max $H(Y) = 1$ and gives limit on rate, so $R_1 + R_2 = 1$. 
Multiple Access Channel

- Suppose we have two binary multiplier channels, $Y = X_1 X_2$.
- If $X_1 = 1$, then $X_2$ sent to $Y$ at rate $1 - H(0) = 1$.
- If $X_2 = 1$, then $X_1$ sent to $Y$ at a rate $1 - H(0) = 1$.
- Max $H(Y) = 1$ and gives limit on rate, so $R_1 + R_2 = 1$.
- Thus, we have triangle shaped polytope:
Binary Erasure Channel

$X \rightarrow 0$ with probability $1 - \alpha$

$X \rightarrow e$ with probability $\alpha$

$X \rightarrow 1$ with probability $1 - \alpha$

$Y \rightarrow 0$

$Y \rightarrow e$

$Y \rightarrow 1$

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

\[ X \xrightarrow{1 - \alpha} 0 \]
\[ X \xrightarrow{\alpha} e \]
\[ X \xrightarrow{1 - \alpha} 1 \]
\[ e \xrightarrow{1 - \alpha} 0 \]
\[ e \xrightarrow{\alpha} Y \]
\[ Y \xrightarrow{1 - \alpha} 1 \]
**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 $\alpha = 0$. 

$$X \xrightarrow{1-\alpha} 0 \xleftarrow{\alpha} e \xrightarrow{\alpha} 1 \xleftarrow{1-\alpha} Y$$
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 $\alpha = 0$.  

\[
\begin{array}{ccc}
X & \overset{1-\alpha}{\rightarrow} & 0 \\
\alpha & \rightarrow & e \\
\alpha & \rightarrow & Y \\
1 & \overset{1-\alpha}{\rightarrow} & 1
\end{array}
\]
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 $\alpha = 0$.

\[ C = \max_{p(x)} I(X; Y) \]

(15)
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 $\alpha = 0$.

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

(14)

(15)
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 $\alpha = 0$.

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

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

(14)
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 $\alpha = 0$.

\[
C = \max_{p(x)} I(X; Y) = \max_{p(x)} \left( H(Y) - H(Y|X) \right) \tag{14}
\]

\[
= \max_{p(x)} H(Y) - H(\alpha) \tag{15}
\]

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

\[ C = \max_{p(x)} H(Y) - H(\alpha) \]  
\[ = \max_{\pi} \left( (1 - \alpha)H(\pi) + H(\alpha) \right) - H(\alpha) \]  
\[ = \max_{\pi} (1 - \alpha)H(\pi) = 1 - \alpha \]
Then we get

\[ C = \max_{p(x)} H(Y) - H(\alpha) \] (16)

\[ = \max_{\pi} \left( (1 - \alpha)H(\pi) + H(\alpha) \right) - H(\alpha) \] (17)

\[ = \max_{\pi} (1 - \alpha)H(\pi) = 1 - \alpha \] (18)

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

\[ C = \max_p \left( H(Y) - H(\alpha) \right) \]

\[ = \max_\pi \left( (1 - \alpha)H(\pi) + H(\alpha) \right) - H(\alpha) \]  \hfill (17)

\[ = \max_\pi (1 - \alpha)H(\pi) = 1 - \alpha \]  \hfill (18)

- Best capacity when \( \pi = 1/2 = \Pr(X = 1) = \Pr(X = 0) \).
- This makes sense, loose \( \alpha\% \) of the bits of original capacity.
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.
Binary erasure MAC

- Channel Description: \( Y = X_1 + X_2, \ |X_1| = |X_2| = 1 \) 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: $Y = X_1 + X_2$, $|\mathcal{X}_1| = |\mathcal{X}_2| = 1$ 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.
Binary erasure MAC

- Channel Description: $Y = X_1 + X_2$, $|\mathcal{X}_1| = |\mathcal{X}_2| = 1$ 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 $X_1 \rightarrow Y$ may have $R_1 = 1$
Binary erasure MAC

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

- If \( X_2 \equiv 0 \) then \( X_1 \rightarrow Y \) 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 \( X_2 \rightarrow Y \) may have \( R_1 = 1 \), etc.
- Thus, we may achieve the two on-axis extreme points \((0, 1)\) and \((1, 0)\) in the following:
Binary erasure MAC

- Lets assume $R_1 = 1$ so that $X_1 \sim \text{Bernoulli}(1/2)$.
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 $Y$. 

\[ R_2 \quad R_1 \quad C_2 = 1 \quad C_1 = 1 \quad 1 \quad 0 \quad 2 \quad 1 \quad 2 \]

We can "cheat" with TDMA to get any of the other points (but clever & more computationally demanding coding can also do this).
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 $Y$.

In fact, this turns $X_2 \rightarrow Y$’s channel into a binary erasure channel with $\alpha = 1/2$ and which (thus) has capacity $C_2 = 1 - 1/2 = 1/2$. 
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:
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

- 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

- **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.
Multiple Access Channel

- **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 3.4**

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

\[
R_1 \leq I(X_1; Y|X_2) \tag{11}
\]

\[
R_2 \leq I(X_2; Y|X_1) \tag{12}
\]

\[
R_1 + R_2 \leq I(X_1, X_2; Y) \tag{13}
\]

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 $p(x_1, x_2) = p(x_1)p(x_2)$.
Theorem: Achievability

Theorem 3.1

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. \(P_e^{(n)} \to 0 \text{ for } n \to \infty\).

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: Achievability

Theorem 3.1

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: Achievability

Theorem 3.1

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 \(x_{1:n}^k(i) \sim \prod_{i=1}^{n} p_k(x_j^i)\) 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: Achievability

**Theorem 3.1**

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. $P_e(n) \rightarrow 0$ for $n \rightarrow \infty$.

**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 $x_{1:n}^k(i) \sim \prod_{i=1}^{n} p_k(x^j_i)$ 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.
Theorem: Achievability

Theorem 3.1

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 \((x_{1:n}^1, x_{1:n}^2, y_{1:n}) \in A_{\epsilon}^{(n)}\) if it exists, and otherwise error will occur.
- Note: no TDMA required.
Theorem: Achievability

proof of Theorem 3.1 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).
Theorem: Achievability

proof of Theorem 3.1 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).
Theorem: Achievability

proof of Theorem 3.1 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: \(E_{ij} = \left\{ (x_{1:n}^1(i), x_{1:n}^2(j), y_{1:n}) \in A_{\epsilon(n)} \right\} \).
Theorem: Achievability

proof of Theorem 3.1 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:

\[
P_e^{(n)} = \Pr\left( E_{11}^c \cup \bigcup_{(i,j) \neq (1,1)} E_{ij} \right)
\leq \Pr(E_{11}^c) + \sum_{j=1, i \neq 1} \Pr(E_{i1}) + \sum_{i=1, j \neq 1} \Pr(E_{1j}) + \sum_{i \neq 1, j \neq 1} \Pr(E_{ij})
\]

where inequality is by the union bound.

...
proof of Theorem 3.1 continued.

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

proof of Theorem 3.1 continued.

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

\[
\Pr(E_{i1}) = \Pr((x_{1:n}^1(i), x_{1:n}^2(1), y_{1:n}) \in A^{(n)}_\epsilon) \tag{20}
\]
\[
= \Pr(\text{indep. events } x_{1:n}^1(i) \text{ and } x_{1:n}^2(1), y_{1:n} \text{ jointly typical}) \tag{21}
\]
\[
= \sum \Pr(x_{1:n}^1(i)) \Pr(x_{1:n}^2(1), y_{1:n}) \tag{22}
\]
\[
\leq |A^{(n)}_\epsilon| 2^{-n(H(X)_-\epsilon)} 2^{-n(H(X_2,Y)_-\epsilon)} \tag{23}
\]
\[
\leq 2^{-n(-H(X_1,X_2,Y)+H(X_1)+H(X_2,Y)-3\epsilon)} \tag{24}
\]
\[
= 2^{-n(I(X_1;X_2,Y)-3\epsilon)} \tag{25}
\]
\[
= 2^{-n(I(X_1;Y|X_2)-3\epsilon)} \quad \text{since } X_1 \perp X_2 \tag{26}
\]
Theorem: Achievability

proof of Theorem 3.1 continued.

- Also,

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

and

\[ \Pr(E_{i, j}) \leq 2^{-n(I(X_1, X_2; Y) - 4\epsilon)} \]
proof of Theorem 3.1 continued.

- Also,

\[
\Pr(E_{1j}) \leq 2^{-n(I(X_2;Y|X_1) - 3\epsilon)} \tag{28}
\]

and

\[
\Pr(E_{ij}) \leq 2^{-n(I(X_1,X_2;Y) - 4\epsilon)} \tag{29}
\]

- thus, we have

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

\[
+ 2^{n(R_1+R_2)}2^{-n(I(X_1,X_2;Y) - 4\epsilon)} \tag{30}
\]

\[\rightarrow 0 \text{ as } n \rightarrow \infty \tag{31}\]

for the given constraints on $R_1, R_2$. 

...
Achievability, discussion

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

Decoder first declares $w_2$ sent if $(x_2(w_2), y) \in A(n)$ (error if not) and then, for that $w_2$, declares $w_1$ sent if $(x_1(w_1), x_2(w_2), y) \in A(n)$ – this achieves one extreme point (other is symmetric case).
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 $I(X_1; Y|X_2)$.

- Decoder first declares $w_2$ sent if $(x_2(w_2), y) \in A_\epsilon(n)$ (error if not) and then, for that $w_2$, declares $w_1$ sent if $(x_1(w_1), x_2(w_2), y) \in A_\epsilon(n)$ – this achieves one extreme point (other is symmetric case).
Theorem: Converse

- The converse of Theorem 3.1 states that any given sequence of 
  \((2^nR_1, 2^nR_2, n)\) codes for a MAC with \(\lim_{n \to \infty} P_e^{(n)} = 0\) is such 
  that we must have \((R_1, R_2) \in \mathcal{C}\) where \(\mathcal{C}\) is the convex hull of those 
  \((R_1, R_2)\) that lie within polytopes for various different \(p_1(x_1)p_2(x_2)\).
Theorem: Converse

The converse of Theorem 3.1 states that any given sequence of 
\((2^nR_1, 2^nR_2), n\) codes for a MAC with \(\lim_{n \to \infty} P_e^n = 0\) is such that we must have \((R_1, R_2) \in \mathcal{C}\) where \(\mathcal{C}\) is the convex hull of those \((R_1, R_2)\) that lie within polytopes for various different \(p_1(x_1)p_2(x_2)\).

Each code induces a joint pmf as follows, w. message r.v.s \((W_1, W_2)\).

\[ (W_1, W_2, X_{1:n}, X_{1:n}, Y_{1:n}) \]
\[ \sim 2^{-n(R_1+R_2)} p(x_{1:n}^1|w_1)p(x_{1:n}^2|w_2) \prod_{i=1}^{n} p(y_i|x_i^1, x_i^2) \]  \( (32) \)
Theorem: Converse

- The converse of Theorem 3.1 states that any given sequence of $((2^{nR_1}, 2^{nR_2}), n)$ codes for a MAC with $\lim_{n \to \infty} P_e^{(n)} = 0$ is such that we must have $(R_1, R_2) \in \mathcal{C}$ where $\mathcal{C}$ is the convex hull of those $(R_1, R_2)$ that lie within polytopes for various different $p_1(x_1)p_2(x_2)$.

- Each code induces a joint pmf as follows, w. message r.v.s $(W_1, W_2)$.

$$ (W_1, W_2, X_1^{1:n}, X_2^{1:n}, Y_1^{1:n}) $$

$$ \sim 2^{-n(R_1+R_2)} p(x_1^{1:n} | w_1) p(x_2^{1:n} | w_2) \prod_{i=1}^{n} p(y_i | x_1^i, x_2^i) \quad (32) $$

- Fano’s inequality, in this context, states that:

$$ H(W_1, W_2 | Y_1^{1:n}) \leq n(R_1 + R_2) P_e^{(n)} + 1 = n((R_1 + R_2) P_e^{(n)} + \frac{1}{n}) = n \epsilon_n $$

(33)

where $\epsilon \to 0$ as $n \to \infty$. 
Theorem: Proof of Converse

- We can bound the sum of the rates as follows (which mimics converse of standard channel coding theorem)

\begin{align*}
\sum_{i=1}^{n} (R_1 + R_2) &= H(W_1, W_2) = I(W_1, W_2; Y_{1:n}) + H(W_1, W_2 | Y_{1:n}) \\
&\leq I(W_1, W_2; Y_{1:n}) + n\epsilon/n \\
&= H(Y_{1:n}) - H(Y_{1:n} | X_{1:n}(W_1), X_{1:n}(W_2)) + n\epsilon/n \\
&\leq \sum_{i=1}^{n} I(X_{1:i}(W_1), X_{1:i}(W_2); Y_i) + n\epsilon/n
\end{align*}
Theorem: Proof of Converse

- We can bound the sum of the rates as follows (which mimics converse of standard channel coding theorem)

\[(R_1 + R_2)\]
Theorem: Proof of Converse

- We can bound the sum of the rates as follows (which mimics converse of standard channel coding theorem)

\[ n(R_1 + R_2) = H(W_1, W_2) \]
**Theorem: Proof of Converse**

- We can bound the sum of the rates as follows (which mimics converse of standard channel coding theorem)

\[
\begin{align*}
\log(R_1 + R_2) &= H(W_1, W_2) = I(W_1, W_2; Y_{1:n}) + H(W_1, W_2|Y_{1:n}) \\
&\leq I(W_1, W_2; Y_{1:n}) + n\epsilon
\end{align*}
\]  

(34)
**Theorem: Proof of Converse**

- We can bound the sum of the rates as follows (which mimics converse of standard channel coding theorem)

\[ n(R_1 + R_2) = H(W_1, W_2) = I(W_1, W_2; Y_{1:n}) + H(W_1, W_2|Y_{1:n}) \]  
\[ \leq I(W_1, W_2; Y_{1:n}) + n\epsilon_n \]  
\[ \uparrow \]  
\[ \text{Pan o} \]
Theorem: Proof of Converse

We can bound the sum of the rates as follows (which mimics converse of standard channel coding theorem)

\[ n(R_1 + R_2) = H(W_1, W_2) = I(W_1, W_2; Y_1:n) + H(W_1, W_2 | Y_1:n) \] (34)

\[ \leq I(W_1, W_2; Y_1:n) + n\epsilon_n \leq I(X_1^{1:n}(W_1), X_2^{1:n}(W_2); Y_1:n) + n\epsilon_n \] (38)
Theorem: Proof of Converse

- We can bound the sum of the rates as follows (which mimics converse of standard channel coding theorem)

\[ n(R_1 + R_2) = H(W_1, W_2) = I(W_1, W_2; Y_{1:n}) + H(W_1, W_2|Y_{1:n}) \]  \hspace{1cm} (34)

\[ \leq I(W_1, W_2; Y_{1:n}) + n\epsilon_n \leq I(X_{1:n}^1(W_1), X_{1:n}^2(W_2); Y_{1:n}) + n\epsilon_n \]

\[ = H(Y_{1:n}) - H(Y_{1:n}|X_{1:n}^1(W_1), X_{1:n}^2(W_2)) + n\epsilon_n \]  \hspace{1cm} (35)

\[ \leq n\sum_{i=1}^{\infty} H(Y_i) - n\sum_{i=1}^{\infty} H(Y_i|X_{1:n}^1(W_1), X_{1:n}^2(W_2)) + n\epsilon_n \]  \hspace{1cm} (36)

\[ \leq n\sum_{i=1}^{\infty} I(X_{1:n}^1(W_1), X_{1:n}^2(W_2); Y_i) + n\epsilon_n \]  \hspace{1cm} (38)
Theorem: Proof of Converse

- We can bound the sum of the rates as follows (which mimics converse of standard channel coding theorem)

\[ n(R_1 + R_2) = H(W_1, W_2) = I(W_1, W_2; Y_{1:n}) + H(W_1, W_2 | Y_{1:n}) \]

\[ \leq I(W_1, W_2; Y_{1:n}) + n\epsilon_n \leq I(X^1_{1:n}(W_1), X^2_{1:n}(W_2); Y_{1:n}) + n\epsilon_n \]

\[ = H(Y_{1:n}) - H(Y_{1:n} | X^1_{1:n}(W_1), X^2_{1:n}(W_2)) + n\epsilon_n \]

\[ = H(Y_{1:n}) - \sum_{i=1}^{n} H(Y_i | Y_{1:i-1}, X^1_{1:n}(W_1), X^2_{1:n}(W_2)) + n\epsilon_n \]

(34) (35) (36) (37) (38)
Theorem: Proof of Converse

We can bound the sum of the rates as follows (which mimics converse of standard channel coding theorem)

\[ n(R_1 + R_2) = H(W_1, W_2) = I(W_1, W_2; Y_{1:n}) + H(W_1, W_2|Y_{1:n}) \]  \hspace{1cm} (34)

\[ \leq I(W_1, W_2; Y_{1:n}) + n\epsilon_n \leq I(X^1_{1:n}(W_1), X^2_{1:n}(W_2); Y_{1:n}) + n\epsilon_n \]

\[ = H(Y_{1:n}) - H(Y_{1:n}|X^1_{1:n}(W_1), X^2_{1:n}(W_2)) + n\epsilon_n \]  \hspace{1cm} (35)

\[ = H(Y_{1:n}) - \sum_{i=1}^{n} H(Y_i|Y_{1:i-1}, X^1_{1:n}(W_1), X^2_{1:n}(W_2)) + n\epsilon_n \]

\[ = H(Y_{1:n}) - \sum_{i=1}^{n} H(Y_i|X_i^1, X_i^2) + n\epsilon_n \]  \hspace{1cm} (36)

\[ \text{proof}

\[ \text{channel} \]
Theorem: Proof of Converse

- We can bound the sum of the rates as follows (which mimics converse of standard channel coding theorem)

\[ n(R_1 + R_2) = H(W_1, W_2) = I(W_1, W_2; Y_{1:n}) + H(W_1, W_2|Y_{1:n}) \]  \( (34) \)

\[ \leq I(W_1, W_2; Y_{1:n}) + n\epsilon_n \leq I(X_{1:n}^1(W_1), X_{1:n}^2(W_2); Y_{1:n}) + n\epsilon_n \]

\[ = H(Y_{1:n}) - H(Y_{1:n}|X_{1:n}^1(W_1), X_{1:n}^2(W_2)) + n\epsilon_n \]  \( (35) \)

\[ = H(Y_{1:n}) - \sum_{i=1}^{n} H(Y_i|Y_{1:i-1}, X_{1:n}^1(W_1), X_{1:n}^2(W_2)) + n\epsilon_n \]

\[ = H(Y_{1:n}) - \sum_{i=1}^{n} H(Y_i|X_i^1, X_i^2) + n\epsilon_n \]  \( (36) \)

\[ \leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i^1, X_i^2) + n\epsilon_n \]

\[ \leq \sum_{i=1}^{n} I(X_i^1, X_i^2) + n\epsilon_n \]  \( (38) \)
Theorem: Proof of Converse

We can bound the sum of the rates as follows (which mimics converse of standard channel coding theorem)

\[ n(R_1 + R_2) = H(W_1, W_2) = I(W_1, W_2; Y_{1:n}) + H(W_1, W_2|Y_{1:n}) \]

\[ \leq I(W_1, W_2; Y_{1:n}) + n\epsilon_n \leq I(X^1_{1:n}(W_1), X^2_{1:n}(W_2); Y_{1:n}) + n\epsilon_n \]

\[ = H(Y_{1:n}) - H(Y_{1:n}|X^1_{1:n}(W_1), X^2_{1:n}(W_2)) + n\epsilon_n \]

\[ = H(Y_{1:n}) - \sum_{i=1}^{n} H(Y_i|Y_{1:i-1}, X^1_{1:n}(W_1), X^2_{1:n}(W_2)) + n\epsilon_n \]

\[ = H(Y_{1:n}) - \sum_{i=1}^{n} H(Y_i|X^1_i, X^2_i) + n\epsilon_n \]

\[ \leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X^1_i, X^2_i) + n\epsilon_n \]

\[ = \sum_{i=1}^{n} I(X^1_i, X^2_i; Y_i) + n\epsilon_n \]
Theorem: Proof of Converse

- And hence, we have

\[ R_1 + R_2 \leq \frac{1}{n} \sum_{i=1}^{n} I(X_1^i, X_2^i; Y_i) + \epsilon_n \]  

(39)
Theorem: Proof of Converse

- And hence, we have

\[ R_1 + R_2 \leq \frac{1}{n} \sum_{i=1}^{n} I(X^1_i, X^2_i; Y_i) + \epsilon_n \]  

(39)

- Next, we need to bound the individual rates \( R_1 \) and \( R_2 \).
Theorem: Proof of Converse

And hence, we have

\[ R_1 + R_2 \leq \frac{1}{n} \sum_{i=1}^{n} I(X_1^i, X_2^i; Y_i) + \epsilon_n \]  (39)

Next, we need to bound the individual rates $R_1$ and $R_2$.

We have $H(W_1|Y_1:n, W_2) \leq H(W_1, W_2|Y_1:n) \leq n \epsilon_n$ from Fano. Hence,
Theorem: Proof of Converse

\[ nR_1 \]

\[ n \sum_{i=1}^{\infty} I(W_1; Y_i | Y_1:i-1, W_2) + n \epsilon \]

(46)
Theorem: Proof of Converse

\[ nR_1 = H(W_1) \]
Theorem: Proof of Converse

\[ nR_1 = H(W_1) = H(W_1|W_2) \]
Theorem: Proof of Converse

\[ nR_1 = H(W_1) = H(W_1|W_2) = I(W_1; Y_{1:n}|W_2) + H(W_1|Y_{1:n}, W_2) \] (40)
Theorem: Proof of Converse

\[ nR_1 = H(W_1) = H(W_1|W_2) = I(W_1; Y_1:n|W_2) + H(W_1|Y_1:n, W_2) \]  \hspace{1cm} (40)

\[ \leq I(W_1; Y_1:n|W_2) + n\epsilon_n \]
Theorem: Proof of Converse

\[ nR_1 = H(W_1) = H(W_1|W_2) = I(W_1; Y_{1:n}|W_2) + H(W_1|Y_{1:n}, W_2) \quad (40) \]

\[ \leq I(W_1; Y_{1:n}|W_2) + n\epsilon_n = \sum_{i=1}^{n} I(W_1; Y_i|Y_{1:i-1}, W_2) + n\epsilon_n \quad (41) \]
Theorem: Proof of Converse

\[ nR_1 = H(W_1) = H(W_1|W_2) = I(W_1; Y_{1:n}|W_2) + H(W_1|Y_{1:n}, W_2) \]  
(40)

\[ \leq I(W_1; Y_{1:n}|W_2) + n\epsilon_n a = \sum_{i=1}^{n} I(W_1; Y_i|Y_{1:i-1}, W_2) + n\epsilon_n \]  
(41)

\[ = \sum_{i=1}^{n} I(W_1; Y_i|Y_{1:i-1}, W_2, X_i^2) + n\epsilon_n \]  
(42)

\[ W_2 \rightarrow X_i^2 \quad X_i^2 \text{ is func. at } W_2 \]  
(46)
Theorem: Proof of Converse

\[ nR_1 = H(W_1) = H(W_1|W_2) = I(W_1; Y_1:n|W_2) + H(W_1|Y_1:n, W_2) \]  

(40)

\[ \leq I(W_1; Y_1:n|W_2) + n\epsilon_n a = \sum_{i=1}^{n} I(W_1; Y_i|Y_1:i-1, W_2) + n\epsilon_n \]  

(41)

\[ = \sum_{i=1}^{n} I(W_1; Y_i|Y_1:i-1, W_2, X_i^2) + n\epsilon_n \]  

(42)

\[ \leq \sum_{i=1}^{n} I(W_1, W_2, Y_1:i-1; Y_i|X_i^2) + n\epsilon_n \]  

(43)
Theorem: Proof of Converse

\[ nR_1 = H(W_1) = H(W_1|W_2) = I(W_1; Y_1:n|W_2) + H(W_1|Y_1:n, W_2) \] (40)

\[ \leq I(W_1; Y_1:n|W_2) + n\epsilon_n a = \sum_{i=1}^{n} I(W_1; Y_i|Y_{1:i-1}, W_2) + n\epsilon_n \] (41)

\[ = \sum_{i=1}^{n} I(W_1; Y_i|Y_{1:i-1}, W_2, X_i^2) + n\epsilon_n \] (42)

\[ \leq \sum_{i=1}^{n} I(W_1, W_2, Y_{1:i-1}; Y_i|X_i^2) + n\epsilon_n \] (43)

\[ = \sum_{i=1}^{n} I(X_i^1, W_1, W_2, Y_{1:i-1}; Y_i|X_i^2) + n\epsilon_n \] (44)

(46)
Theorem: Proof of Converse

\[ nR_1 = H(W_1) = H(W_1|W_2) = I(W_1; Y_{1:n}|W_2) + H(W_1|Y_{1:n}, W_2) \]  \hspace{1cm} (40)

\[ \leq I(W_1; Y_{1:n}|W_2) + n\epsilon_n a = \sum_{i=1}^{n} I(W_1; Y_i|Y_{1:i-1}, W_2) + n\epsilon_n \]  \hspace{1cm} (41)

\[ = \sum_{i=1}^{n} I(W_1; Y_i|Y_{1:i-1}, W_2, X_i^2) + n\epsilon_n \]  \hspace{1cm} (42)

\[ \leq \sum_{i=1}^{n} I(W_1, W_2, Y_{1:i-1}; Y_i|X_i^2) + n\epsilon_n \]  \hspace{1cm} (43)

\[ = \sum_{i=1}^{n} I(X_i^1, W_1, W_2, Y_{1:i-1}; Y_i|X_i^2) + n\epsilon_n \]  \hspace{1cm} (44)

\[ = \sum_{i=1}^{n} I(X_i^1; Y_i|X_i^2) + \sum_{i=1}^{n} I(W_1, W_2, Y_{1:i-1}; Y_i|X_i^1, X_i^2) + n\epsilon_n \]  \hspace{1cm} (45)

\[ \hspace{1cm} (46) \]
Theorem: Proof of Converse

\[ nR_1 = H(W_1) = H(W_1|W_2) = I(W_1; Y_{1:n}|W_2) + H(W_1|Y_{1:n}, W_2) \]  \hspace{1cm} (40)

\[ \leq I(W_1; Y_{1:n}|W_2) + n\epsilon_n a = \sum_{i=1}^{n} I(W_1; Y_i|Y_{1:i-1}, W_2) + n\epsilon_n \]  \hspace{1cm} (41)

\[ = \sum_{i=1}^{n} I(W_1; Y_{i}|Y_{1:i-1}, W_2, X_{i}^2) + n\epsilon_n \]  \hspace{1cm} (42)

\[ \leq \sum_{i=1}^{n} I(W_1, W_2, Y_{1:i-1}; Y_i|X_{i}^2) + n\epsilon_n \]  \hspace{1cm} (43)

\[ = \sum_{i=1}^{n} I(X_{i}^1, W_1, W_2, Y_{1:i-1}; Y_i|X_{i}^2) + n\epsilon_n \]  \hspace{1cm} (44)

\[ = \sum_{i=1}^{n} I(X_{i}^1; Y_i|X_{i}^2) + \sum_{i=1}^{n} I(W_1, W_2, Y_{1:i-1}; Y_i|X_{i}^1, X_{i}^2) + n\epsilon_n \]  \hspace{1cm} (45)

\[ = \sum_{i=1}^{n} I(X_{i}^1; Y_i|X_{i}^2) + n\epsilon_n \]  \hspace{1cm} (46)
Theorem: Proof of Converse

So, to summarize what we have so far:

\[ R_1 \leq \frac{1}{n} \sum_{i=1}^{n} I(X_i^1; Y_i | X_i^2) + \epsilon_n \] (47)

\[ R_2 \leq \frac{1}{n} \sum_{i=1}^{n} I(X_i^2; Y_i | X_i^1) + \epsilon_n \] (48)

\[ R_1 + R_2 \leq \frac{1}{n} \sum_{i=1}^{n} I(X_i^1, X_i^2; Y_i) + \epsilon_n \] (49)
Theorem: Proof of Converse

- So, to summarize what we have so far:

\[
R_1 \leq \frac{1}{n} \sum_{i=1}^{n} I(X_i^1; Y_i | X_i^2) + \epsilon_n \quad (47)
\]

\[
R_2 \leq \frac{1}{n} \sum_{i=1}^{n} I(X_i^2; Y_i | X_i^1) + \epsilon_n \quad (48)
\]

\[
R_1 + R_2 \leq \frac{1}{n} \sum_{i=1}^{n} I(X_i^1, X_i^2; Y_i) + \epsilon_n \quad (49)
\]

- Note that \(X_i^1(W_1) \perp \perp X_i^2(W_2)\) since \(W_1 \perp \perp W_2\).
Theorem: Proof of Converse

So, to summarize what we have so far:

\[ R_1 \leq \frac{1}{n} \sum_{i=1}^{n} I(X^1_i; Y_i | X^2_i) + \epsilon_n \] (47)

\[ R_2 \leq \frac{1}{n} \sum_{i=1}^{n} I(X^2_i; Y_i | X^1_i) + \epsilon_n \] (48)

\[ R_1 + R_2 \leq \frac{1}{n} \sum_{i=1}^{n} I(X^1_i, X^2_i; Y_i) + \epsilon_n \] (49)

Note that \( X^1_i(W_1) \perp \perp X^2_i(W_2) \) since \( W_1 \perp \perp W_2 \).

Note that each of the three above equations gives a capacity if we maximize, respectively, over \( p(x_1) \) (for fixed \( x_2 \)), \( p(x_2) \) (for fixed \( x_1 \)), and jointly over \( p(x_1)p(x_2) \) giving capacities \( C_1 \), \( C_2 \), and \( C_{12} \).
Theorem: Proof of Converse

So, to summarize what we have so far:

\[ R_1 \leq \frac{1}{n} \sum_{i=1}^{n} I(X_1^i; Y_i | X_2^i) + \epsilon_n \]  \hspace{1cm} (47)

\[ R_2 \leq \frac{1}{n} \sum_{i=1}^{n} I(X_2^i; Y_i | X_1^i) + \epsilon_n \]  \hspace{1cm} (48)

\[ R_1 + R_2 \leq \frac{1}{n} \sum_{i=1}^{n} I(X_1^i, X_2^i; Y_i) + \epsilon_n \]  \hspace{1cm} (49)

Note that \( X_1^i(W_1) \perp \perp X_2^i(W_2) \) since \( W_1 \perp \perp W_2 \).

Note that each of the three above equations gives a capacity if we maximize, respectively, over \( p(x_1) \) (for fixed \( x_2 \)), \( p(x_2) \) (for fixed \( x_1 \)), and jointly over \( p(x_1) p(x_2) \) giving capacities \( C_1 \), \( C_2 \), and \( C_{12} \).

Simple TDMA can achieve any point on the line between \((0, C_1)\) and \((C_1, 0)\) so this constitutes an inner bound of \( \mathcal{C} \).
Theorem: Proof of Converse

On the other hand, $R_1 < C_1$, $R_2 < C_2$ and $R_1 + R_2 < C_{12}$ constitute an outer bound.
Theorem: Proof of Converse

- On the other hand, $R_1 < C_1$, $R_2 < C_2$ and $R_1 + R_2 < C_{12}$ constitute an outer bound.
- This is shown in the following:
To continue the proof, and show convex hull, we will use an auxiliary integer-valued random variable $Q \sim \text{Uniform}[1 : n]$ independent of $(X_1^{1:n}, X_2^{1:n}, Y_1^{1:n})$. 

Theorem: Proof of Converse
Theorem: Proof of Converse

To continue the proof, and show convex hull, we will use an auxiliary integer-valued random variable $Q \sim \text{Uniform}[1 : n]$ independent of $(X_1^{1:n}, X_1^{2:n}, Y_1^{n})$.

We can thus write

(53)
Theorem: Proof of Converse

To continue the proof, and show convex hull, we will use an auxiliary integer-valued random variable $Q \sim \text{Uniform}[1 : n]$ independent of $(X_1^{1:n}, X_2^{1:n}, Y_1^{1:n})$.

We can thus write

$$R_1$$  

(53)
Theorem: Proof of Converse

- To continue the proof, and show convex hull, we will use an auxiliary integer-valued random variable $Q \sim \text{Uniform}[1:n]$ independent of $(X_1^1, X_2^2, Y_1^2)$.
- We can thus write

$$R_1 \leq \frac{1}{n} \sum_{i=1}^{n} I(X_i^1; Y_i | X_i^2) + \epsilon_n$$  \hspace{1cm} (50)
Theorem: Proof of Converse

- To continue the proof, and show convex hull, we will use an auxiliary integer-valued random variable $Q \sim \text{Uniform}[1 : n]$ independent of $(X^1_{1:n}, X^2_{1:n}, Y_{1:n})$.
- We can thus write

$$R_1 \leq \frac{1}{n} \sum_{i=1}^{n} I(X^1_i; Y_i|X^2_i) + \epsilon_n$$  \hfill (50)

$$= \frac{1}{n} \sum_{i=1}^{n} I(X^1_i; Y_i|X^2_i, Q = i) + \epsilon_n$$  \hfill (51)
Theorem: Proof of Converse

- To continue the proof, and show convex hull, we will use an auxiliary integer-valued random variable $Q \sim \text{Uniform}[1 : n]$ independent of $(X_{1:n}^1, X_{1:n}^2, Y_{1:n})$.
- We can thus write

$$R_1 \leq \frac{1}{n} \sum_{i=1}^{n} I(X_i^1; Y_i | X_i^2) + \epsilon_n$$

(50)

$$= \frac{1}{n} \sum_{i=1}^{n} I(X_i^1; Y_i | X_i^2, Q = i) + \epsilon_n$$

(51)

$$= \sum_{i=1}^{n} I(X_i^1; Y_i | X_i^2, Q = i) \Pr(Q = i) + \epsilon_n$$

(52)

$$= \sum_{i=1}^{n} I(X_i^1; Y_i | X_i^2, Q = i) \Pr(Q = i) + \epsilon_n$$

(53)
Theorem: Proof of Converse

- To continue the proof, and show convex hull, we will use an auxiliary integer-valued random variable $Q \sim \text{Uniform}[1:n]$ independent of $(X^{1}_{1:n}, X^{2}_{1:n}, Y_{1:n})$.

- We can thus write

$$R_1 \leq \frac{1}{n} \sum_{i=1}^{n} I(X^1_i; Y_i | X^2_i) + \epsilon_n$$  \hspace{1cm} (50)

$$= \frac{1}{n} \sum_{i=1}^{n} I(X^1_i; Y_i | X^2_i, Q = i) + \epsilon_n$$  \hspace{1cm} (51)

$$= \sum_{i=1}^{n} I(X^1_i; Y_i | X^2_i, Q = i) \Pr(Q = i) + \epsilon_n$$  \hspace{1cm} (52)

$$= I(X^1_Q; Y_Q | X^2_Q, Q) + \epsilon_n$$  \hspace{1cm} (53)
Theorem: Proof of Converse

Notice that

\[
\Pr(Y_Q = y | X_Q^1 = x^1, X_Q^2 = x^2) = \sum_{i=1}^{n} \Pr(Y_{Q=i} = y | X_{Q=i}^1 = x^1, X_{Q=i}^2 = x^2) \Pr(Q = i)
\]

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

(54)

(55)

(56)
Theorem: Proof of Converse

Notice that

\[
\Pr(Y_Q = y | X_Q^1 = x^1, X_Q^2 = x^2) = \sum_{i=1}^{n} \Pr(\text{Q}_i = y | \text{Q}_1 = x^1, \text{Q}_2 = x^2) \Pr(\text{Q} = i)
\]

\[
= \Pr(y | x^1, x^2) \quad \text{(even when Q i) not written.}
\]

therefore, we can use \( X^1 \) for \( X_Q^1 \) and \( X^2 \) for \( X_Q^2 \) (and in fact lets just use subscripts and hope there is no confusion) and we get

\[
R_1 \leq I(X_1; Y | X_2; Q) + \epsilon_n
\]

and similarly bound for \( R_2 \) and \( R_1 + R_2 \).
Theorem: Proof of Converse

That is, we get that the rate pair \((R_1, R_2)\) must lie within the region

\[
R_1 \leq I(X_1; Y|X_2, Q) + \epsilon_n \tag{58}
\]

\[
R_2 \leq I(X_2; Y|X_1, Q) + \epsilon_n \tag{59}
\]

\[
R_1 + R_2 \leq I(X_1, X_2; Y|Q) + \epsilon_n \tag{60}
\]
Theorem: Proof of Converse

- That is, we get that the rate pair \((R_1, R_2)\) must lie within the region

\[
R_1 \leq I(X_1; Y|X_2, Q) + \epsilon_n \tag{58}
\]
\[
R_2 \leq I(X_2; Y|X_1, Q) + \epsilon_n \tag{59}
\]
\[
R_1 + R_2 \leq I(X_1, X_2; Y|Q) + \epsilon_n \tag{60}
\]

- When \(n \to \infty\), since we assumed \(P_e^{(n)} \to 0\), this becomes

\[
R_1 \leq I(X_1; Y|X_2, Q) \tag{61}
\]
\[
R_2 \leq I(X_2; Y|X_1, Q) \tag{62}
\]
\[
R_1 + R_2 \leq I(X_1, X_2; Y|Q) \tag{63}
\]

which we’ll call \(\mathcal{C}'\), with \(Q\) uniform on \([n]\) as mentioned,
Theorem: Proof of Converse

- That is, we get that the rate pair \((R_1, R_2)\) must lie within the region

\[
R_1 \leq I(X_1; Y|X_2, Q) + \epsilon_n \tag{58}
\]
\[
R_2 \leq I(X_2; Y|X_1, Q) + \epsilon_n \tag{59}
\]
\[
R_1 + R_2 \leq I(X_1, X_2; Y|Q) + \epsilon_n \tag{60}
\]

- When \(n \to \infty\), since we assumed \(P_e^{(n)} \to 0\), this becomes

\[
R_1 \leq I(X_1; Y|X_2, Q) \tag{61}
\]
\[
R_2 \leq I(X_2; Y|X_1, Q) \tag{62}
\]
\[
R_1 + R_2 \leq I(X_1, X_2; Y|Q) \tag{63}
\]

which we’ll call \(C'\), with \(Q\) uniform on \([n]\) as mentioned, and this must hold for joint distribution \(p(q)p(x_1|q)p(x_2|q)\) where again \(p(q)\) is uniform.
Theorem: Proof of Converse

- That is, we get that the rate pair \((R_1, R_2)\) must lie within the region

\[
R_1 \leq I(X_1; Y|X_2, Q) + \epsilon_n \\
R_2 \leq I(X_2; Y|X_1, Q) + \epsilon_n \\
R_1 + R_2 \leq I(X_1, X_2; Y|Q) + \epsilon_n
\] (58-60)

- When \(n \to \infty\), since we assumed \(P_e^{(n)} \to 0\), this becomes

\[
R_1 \leq I(X_1; Y|X_2, Q) \\
R_2 \leq I(X_2; Y|X_1, Q) \\
R_1 + R_2 \leq I(X_1, X_2; Y|Q)
\] (61-63)

which we’ll call \(C'\), with \(Q\) uniform on \([n]\) as mentioned, and this must hold for joint distribution \(p(q)p(x_1|q)p(x_2|q)\) where again \(p(q)\) is uniform.

- Note that \(Q \to (X_1, X_2) \to Y\) forms a Markov chain.
Theorem: Proof of Converse

- We are done if we can show that $C' = C$. 

---

Prof. Jeff Bilmes  
Theorem: Proof of Converse

- We are done if we can show that $\mathcal{C}' = \mathcal{C}$.

- First note, $\mathcal{C} \subseteq \mathcal{C}'$ since any polytope region for $R_1, R_2$ for a given $p(x_1)p(x_2)$ can be achieved using the right $p(q)p(x_1|q)p(x_2|q)$. 
Theorem: Proof of Converse

- We are done if we can show that $\mathcal{C}' = \mathcal{C}$.
- First note, $\mathcal{C} \subseteq \mathcal{C}'$ since any polytope region for $R_1, R_2$ for a given $p(x_1)p(x_2)$ can be achieved using the right $p(q)p(x_1|q)p(x_2|q)$.
- Moreover, we can achieve the convex hull of such regions since $Q$ acts as a “mixing” variable.
Theorem: Proof of Converse

- We are done if we can show that $\mathcal{C}' = \mathcal{C}$.
- First note, $\mathcal{C} \subseteq \mathcal{C}'$ since any polytope region for $R_1, R_2$ for a given $p(x_1)p(x_2)$ can be achieved using the right $p(q)p(x_1|q)p(x_2|q)$.
- Moreover, we can achieve the convex hull of such regions since $Q$ acts as a “mixing” variable.
- That is, let $\mathcal{R}(p(x_1)p(x_2))$ be one such polytope. Then if $(R_1^a, R_2^a) \in \mathcal{R}(p^a(x_1)p^a(x_2))$ and $(R_1^b, R_2^b) \in \mathcal{R}(p^b(x_1)p^b(x_2))$ then for any $\lambda \in [0, 1]$, 
  
  $$(\lambda R_1^a + (1 - \lambda) R_1^b, \lambda R_2^a + (1 - \lambda) R_2^b) \in \mathcal{R}(p^b(x_1|q)p^b(x_2|q)p(q))$$
  
  for the right $p(q)$ by the definition of the quantities $I(X_1; Y|X_2, Q)$, $I(X_2; Y|X_1, Q)$, and $I(X_1, X_2; Y|Q)$. 

Prof. Jeff Bilmes
Theorem: Proof of Converse

Next, we must show $\mathcal{C}' \subseteq \mathcal{C}$. 

Note, any $p(X_1|Q)p(X_2|Q)$ gives a polytope (pentagon), so we need only check that the extreme points of the polytope are in $\mathcal{C}$. Consider the point: $(I(X_1;Y|Q), I(X_2;Y|X_1, Q))$. This point is a (finite) convex combination of points of the form $(I(X_1;Y|Q), I(X_2;Y|X_1, Q))$ which come from distribution $p(X_1|Q)p(X_2|Q)$, and since $\mathcal{C}$ is the convex hull of such points, clearly $(I(X_1;Y|Q), I(X_2;Y|X_1, Q)) \in \mathcal{C}$. The same can be done for the other points. Thus, $\mathcal{C}' = \mathcal{C}$. In fact, $Q$ need only have $|Q| = 3$ since by Carathéodory's theorem, any point in convex closure of a compact set $A$ in $d$-dimensional Euclidean space can be represented as a convex combination of $d + 1$ or fewer points from $A$. In fact, it can be shown that for this particular shape, it is sufficient to have $|Q| \leq 2$. 

Prof. Jeff Bilmes
Theorem: Proof of Converse

- Next, we must show \( \mathcal{C}' \subseteq \mathcal{C} \).
- Note, any \( p(q)p(x_1|q)p(x_2|q) \) gives a polytope (pentagon), so we need only check that the extreme points of the polytope are in \( \mathcal{C} \).
Theorem: Proof of Converse

Next, we must show $\mathcal{C}' \subseteq \mathcal{C}$.

Note, any $p(q)p(x_1|q)p(x_2|q)$ gives a polytope (pentagon), so we need only check that the extreme points of the polytope are in $\mathcal{C}$.

Consider the point: $(I(X_1; Y|Q), I(X_2; Y|X_1, Q))$. 

In fact, $Q$ need only have $|Q| = 3$ since by Carathéodory's theorem, any point in convex closure of a compact set $\mathcal{A}$ in $d$-dimensional Euclidean space can be represented as a convex combination of $d+1$ or fewer points from $\mathcal{A}$.
Theorem: Proof of Converse

- Next, we must show $\mathcal{C}' \subseteq \mathcal{C}$.
- Note, any $p(q)p(x_1|q)p(x_2|q)$ gives a polytope (pentagon), so we need only check that the extreme points of the polytope are in $\mathcal{C}$.
- Consider the point: $(I(X_1; Y|Q), I(X_2; Y|X_1, Q))$.
- This point is a (finite) convex combination of points of the form $(I(X_1; Y|Q = q), I(X_2; Y|X_1, Q = q))$ which come from distribution $p(x_1|q)p(x_2|q)$, and since $\mathcal{C}$ is the convex hull of such points, clearly $(I(X_1; Y|Q), I(X_2; Y|X_1, Q)) \in \mathcal{C}$. 

The same can be done for the other points. Thus, $\mathcal{C}' = \mathcal{C}$. In fact, $Q$ need only have $|Q| = 3$ since by Carathéodory's theorem, any point in convex closure of a compact set $A$ in $d$-dimensional Euclidean space can be represented as a convex combination of $d + 1$ or fewer points from $A$. In fact, it can be shown that for this particular shape, it is sufficient to have $|Q| \leq 2$. 

Prof. Jeff Bilmes
Theorem: Proof of Converse

Next, we must show \( C' \subseteq C \).

Note, any \( p(q)p(x_1|q)p(x_2|q) \) gives a polytope (pentagon), so we need only check that the extreme points of the polytope are in \( C \).

Consider the point: \( (I(X_1; Y|Q), I(X_2; Y|X_1, Q)) \).

This point is a (finite) convex combination of points of the form \( (I(X_1; Y|Q = q), I(X_2; Y|X_1, Q = q)) \) which come from distribution \( p(x_1|q)p(x_2|q) \), and since \( C \) is the convex hull of such points, clearly \( (I(X_1; Y|Q), I(X_2; Y|X_1, Q)) \in C \).

The same can be done for the other points.
Theorem: Proof of Converse

- Next, we must show $C' \subseteq C$.
- Note, any $p(q)p(x_1|q)p(x_2|q)$ gives a polytope (pentagon), so we need only check that the extreme points of the polytope are in $C$.
- Consider the point: $(I(X_1;Y|Q), I(X_2;Y|X_1,Q))$.
- This point is a (finite) convex combination of points of the form $(I(X_1;Y|Q=q), I(X_2;Y|X_1,Q=q))$ which come from distribution $p(x_1|q)p(x_2|q)$, and since $C$ is the convex hull of such points, clearly $(I(X_1;Y|Q), I(X_2;Y|X_1,Q)) \in C$.
- The same can be done for the other points.
- Thus, $C' = C$. 
Theorem: Proof of Converse

- Next, we must show $C' \subseteq C$.
- Note, any $p(q)p(x_1|q)p(x_2|q)$ gives a polytope (pentagon), so we need only check that the extreme points of the polytope are in $C$.
- Consider the point: $(I(X_1; Y|Q), I(X_2; Y|X_1, Q))$.
- This point is a (finite) convex combination of points of the form $(I(X_1; Y|Q = q), I(X_2; Y|X_1, Q = q))$ which come from distribution $p(x_1|q)p(x_2|q)$, and since $C$ is the convex hull of such points, clearly $(I(X_1; Y|Q), I(X_2; Y|X_1, Q)) \in C$.
- The same can be done for the other points.
- Thus, $C' = C$.
- In fact, $Q$ need only have $|Q| = 3$ since by Carathéodory’s theorem, any point in convex closure of a compact set $A$ in $d$-dimensional Euclidean space can be represented as a convex combination of $d + 1$ or fewer points from $A$.
Theorem: Proof of Converse

Next, we must show \( C' \subseteq C \).

Note, any \( p(q)p(x_1|q)p(x_2|q) \) gives a polytope (pentagon), so we need only check that the extreme points of the polytope are in \( C \).

Consider the point: \( (I(X_1; Y|Q), I(X_2; Y|X_1, Q)) \).

This point is a (finite) convex combination of points of the form \( (I(X_1; Y|Q = q), I(X_2; Y|X_1, Q = q)) \) which come from distribution \( p(x_1|q)p(x_2|q) \), and since \( C \) is the convex hull of such points, clearly \( (I(X_1; Y|Q), I(X_2; Y|X_1, Q)) \in C \).

The same can be done for the other points.

Thus, \( C' = C \).

In fact, \( Q \) need only have \( |Q| = 3 \) since by Carathéodory’s theorem, any point in convex closure of a compact set \( A \) in \( d \)-dimensional Euclidean space can be represented as a convex combination of \( d + 1 \) or fewer points from \( A \).

In fact, it can be shown that for this particular shape, it is sufficient to have \( |Q| \leq 2 \).
Theorem: Proof of Converse

- $Q$ can be seen as a “time-sharing” variable, where $Q$ gives the proportion of time to allocate each to $X_1$ and $X_2$ to do TDMA.
Theorem: Proof of Converse

• $Q$ can be seen as a “time-sharing” variable, where $Q$ gives the proportion of time to allocate each to $X_1$ and $X_2$ to do TDMA.

• On the other hand, we saw that by doing joint typical decoding, we do not need to do a time sharing approach (although this again only an AEP style proof, and practical coding considerations might be such that time sharing is more efficient in a variety of ways).
The capacity region of the $m$-user multiple access channel is the closure of the convex hull of the rate vectors satisfying:

$$R(S) \leq I(X^S; Y|X^{\bar{S}}) = f(S)$$

(64)

for some product distribution $\prod_i p_i(x_i)$ where $R(S) = \sum_{s \in S} R_s$ and $X^S = (X^s : s \in S)$.

- We state this as a theorem, the proof is quite similar to the $m = 2$ case.
The capacity region of the $m$-user multiple access channel is the closure of the convex hull of the rate vectors satisfying:

$$R(S) \leq I(X^S; Y | X^\bar{S})$$

for some product distribution $\prod_i p_i(x_i)$ where $R(S) = \sum_{s \in S} R_s$ and $X^S = (X^s : s \in S)$.

- We state this as a theorem, the proof is quite similar to the $m = 2$ case.
- Note again, for a particular $\prod_i p_i(x_i)$, Equation (64) defines an $m$-dimensional polymatroid (thanks to independence of the $X_i$’s, but once we take closure of convex hull, it no longer is).
Gaussian MAC

- Here $Y = g_1 X_1 + g_2 X_2 + Z$ where $g_i$ are channel gains and $Z \sim \mathcal{N}(0, N/2)$ is Gaussian noise.
Gaussian MAC

- Here $Y = g_1X_1 + g_2X_2 + Z$ where $g_i$ are channel gains and $Z \sim \mathcal{N}(0, N/2)$ is Gaussian noise.
- Transmission time $i$, we thus have

$$Y_i = g_1X_i^1 + g_2X_i^2 + Z_i \tag{65}$$

where $Z_i$ are independent of channel inputs.
Here \( Y = g_1 X_1 + g_2 X_2 + Z \) where \( g_i \) are channel gains and \( Z \sim \mathcal{N}(0, N/2) \) is Gaussian noise.

Transmission time \( i \), we thus have

\[
Y_i = g_1 X_1^i + g_2 X_2^i + Z_i
\]  \hspace{1cm} (65)

where \( Z_i \) are independent of channel inputs.

- Power constraints: \( \sum_{i=1}^{n} (x_i^k)^2(w_k) \leq nP \) for all \( w_k \in [1 : 2^{nR_k}] \) for \( k \in \{1, 2\} \).
Gaussian MAC

- Here $Y = g_1X_1 + g_2X_2 + Z$ where $g_i$ are channel gains and $Z \sim \mathcal{N}(0, N/2)$ is Gaussian noise.

- Transmission time $i$, we thus have

$$Y_i = g_1X_i^1 + g_2X_i^2 + Z_i$$

(65)

where $Z_i$ are independent of channel inputs.

- Power constraints: $\sum_{i=1}^{n} (x_i^k)^2(w_k) \leq nP$ for all $w_k \in [1 : 2^{nR_k}]$ for $k \in \{1, 2\}$.

- We can extend Theorem 3.1 to the this Gaussian MAC channel as well with the same convex hull of polytopes, where here

$$I(X_1; Y|X_2) = \frac{1}{2} \log(1 + \frac{P_1}{N})$$

(66)

and similarly for $I(X_1; Y|X_2)$ and $I(X_1, X_2; Y)$.
Gaussian MAC

- Define $C(x) \triangleq \frac{1}{2} \log(1 + x)$,
  then we get rate bounds:

  \[
  R_1 \leq C \left( \frac{P_1}{N} \right) = C(S_1)
  \]

  \[
  R_2 \leq C \left( \frac{P_2}{N} \right) = C(S_2)
  \]

  \[
  R_1 + R_2 \leq C \left( \frac{P_1 + P_2}{N} \right)
  = C(S_1 + S_2)
  \]
Define $C(x) \triangleq \frac{1}{n} \log(1 + x)$, then we get rate bounds:

\[
R_1 \leq C \left( \frac{P_1}{N} \right) = C(S_1)
\]
\[
R_2 \leq C \left( \frac{P_2}{N} \right) = C(S_2)
\]
\[
R_1 + R_2 \leq C \left( \frac{P_1 + P_2}{N} \right) = C(S_1 + S_2)
\]

This gives the following:

\[
\begin{align*}
C(S_1) &\leq R_1 \\
C(S_2) &\leq R_2 \\
C(S_1/(1 + S_2)) &\leq C(S_1)
\end{align*}
\]
Gaussian MAC: Decoding Schemes (w.l.o.g. \( N = 1 \))

A: We could treat the other codeword as noise, and get:

\[
R_1 < C\left(\frac{S_1}{1 + S_2}\right) \quad \text{and} \quad R_2 < C\left(\frac{S_2}{1 + S_1}\right)
\]  

(67)
Gaussian MAC: Decoding Schemes (w.l.o.g. \( N = 1 \))

A: We could treat the other codeword as noise, and get:

\[
R_1 < C\left(\frac{S_1}{1 + S_2}\right) \quad \text{and} \quad R_2 < C\left(\frac{S_2}{1 + S_1}\right)
\]  

(67)

B: time-division multiple access (TDMA). Could also define \( \alpha \in [0, 1] \) to give

\[
R_1 < \alpha C(S_1) \quad \text{and} \quad R_2 < (1 - \alpha) C(S_2)
\]

(68)
A: We could treat the other codeword as noise, and get:

\[ R_1 < C\left(\frac{S_1}{1 + S_2}\right) \] and \[ R_2 < C\left(\frac{S_2}{1 + S_1}\right) \] (67)

B: time-division multiple access (TDMA). Could also define \( \alpha \in [0, 1] \) to give

\[ R_1 < \alpha C(S_1) \] and \[ R_2 < (1 - \alpha) C(S_2) \] (68)

C: time-division multiple access (TDMA) with power control. Bi-partition block of length \( n \) into block of length \( \alpha n \) (for sender 1, who sends at power \( P/\alpha \) while sender 2 waits) and \( (1 - \alpha)n \) (for sender 2 to sends with power \( P/(1 - \alpha) \) while sender 1 waits). This gets rates:

\[ R_1 < \alpha C(S_1/\alpha) \] and \[ R_2 < (1 - \alpha) C(S_2/(1 - \alpha)) \] (69)
D: Frequency division multiple access (FDMA), where

\[ R_1 < W_1 C \left( \frac{P_1}{NW_1} \right) \quad \text{and} \quad R_2 < W_2 C \left( \frac{P_2}{NW_2} \right) \]  

(70)

where \( W_i \) are the corresponding bandwidths of non-overlapping frequency bands.
Gaussian MAC: Decoding Schemes (w.l.o.g. $N = 1$)

**D:** Frequency division multiple access (FDMA), where

$$R_1 < W_1 C(P_1/NW_1) \text{ and } R_2 < W_2 C(P_2/NW_2)$$  \hspace{1cm} (70)

where $W_i$ are the corresponding bandwidths of non-overlapping frequency bands.

**E:** Successive cancellation (subtraction) decoding.
Gaussian MAC: Decoding Schemes (w.l.o.g. $N = 1$)

**D:** Frequency division multiple access (FDMA), where

$$R_1 < W_1 C(P_1/NW_1) \text{ and } R_2 < W_2 C(P_2/NW_2)$$

(70)

where $W_i$ are the corresponding bandwidths of non-overlapping frequency bands.

**E:** Successive cancellation (subtraction) decoding. 1) upon receiving

$$y_{1:n} = g_1 x_{1:n}^1(w_1) + g_2 x_{1:n}^2(w_2) + z_{1:n},$$

we decode $w_2$ treating $g_1 x_{1:n}^1(w_1)$ as noise, possible if $R_2 < C(S_2/(S_1 + 1))$. 
Gaussian MAC: Decoding Schemes (w.l.o.g. $N = 1$)

D: Frequency division multiple access (FDMA), where

$$R_1 < W_1 C(P_1/NW_1) \quad \text{and} \quad R_2 < W_2 C(P_2/NW_2) \quad (70)$$

where $W_i$ are the corresponding bandwidths of non-overlapping frequency bands.

E: Successive cancellation (subtraction) decoding. 1) upon receiving $y_{1:n} = g_1 x_{1:n}^1(w_1) + g_2 x_{1:n}^2(w_2) + z_{1:n}$, we decode $w_2$ treating $g_1 x_{1:n}^1(w_1)$ as noise, possible if $R_2 < C(S_2/(S_1 + 1))$. Then, we form $y_{1:n} - g_2 x_{1:n}^2(w_2) = g_1 x_{1:n}^1(w_1) + z_{1:n}$ and decode message $w_1$, possible if $R_1 < C(S_1)$. 
Gaussian MAC: Decoding Schemes (w.l.o.g. \( N = 1 \))

D: Frequency division multiple access (FDMA), where

\[
R_1 < W_1 C(P_1/NW_1) \quad \text{and} \quad R_2 < W_2 C(P_2/NW_2) \quad (70)
\]

where \( W_i \) are the corresponding bandwidths of non-overlapping frequency bands.

E: Successive cancellation (subtraction) decoding. 1) upon receiving \( y_{1:n} = g_1 x_{1:n}^1(w_1) + g_2 x_{1:n}^2(w_2) + z_{1:n} \), we decode \( w_2 \) treating \( g_1 x_{1:n}^1(w_1) \) as noise, possible if \( R_2 < C(S_2/(S_1 + 1)) \). Then, we form \( y_{1:n} - g_2 x_{1:n}^2(w_2) = g_1 x_{1:n}^1(w_1) + z_{1:n} \) and decode message \( w_1 \), possible if \( R_1 < C(S_1) \).
D: Frequency division multiple access (FDMA), where

\[ R_1 < W_1 C(P_1/NW_1) \]  \quad \text{and}  \quad \[ R_2 < W_2 C(P_2/NW_2) \]  \quad (70)

where \( W_i \) are the corresponding bandwidths of non-overlapping frequency bands.

E: Successive cancellation (subtraction) decoding. 1) upon receiving
\[ y_{1:n} = g_1 x_{1:n}^1(w_1) + g_2 x_{1:n}^2(w_2) + z_{1:n}, \]  we decode \( w_2 \) treating
\[ g_1 x_{1:n}^1(w_1) \] as noise, possible if \( R_2 < C(S_2/(S_1 + 1)) \). Then, we
form  \[ y_{1:n} - g_2 x_{1:n}^2(w_2) = g_1 x_{1:n}^1(w_1) + z_{1:n} \] and decode message \( w_1 \),
possible if \( R_1 < C(S_1) \).
Gaussian MAC: Decoding Schemes (w.l.o.g. $N = 1$)

**D:** Frequency division multiple access (FDMA), where

$$R_1 < W_1 C(P_1/NW_1) \text{ and } R_2 < W_2 C(P_2/NW_2) \quad (70)$$

where $W_i$ are the corresponding bandwidths of non-overlapping frequency bands.

**E:** Successive cancellation (subtraction) decoding. 1) upon receiving $y_{1:n} = g_1 x_{1:n}^1(w_1) + g_2 x_{1:n}^2(w_2) + z_{1:n}$, we decode $w_2$ treating $g_1 x_{1:n}^1(w_1)$ as noise, possible if $R_2 < C(S_2/(S_1 + 1))$. Then, we form $y_{1:n} - g_2 x_{1:n}^2(w_2) = g_1 x_{1:n}^1(w_1) + z_{1:n}$ and decode message $w_1$, possible if $R_1 < C(S_1)$.

Thus, either extreme point can be achieved, and time-sharing can get points in-between.
Gaussian MAC: rate regions with coding schemes

Left (a), high SNR. Right (b), low SNR.
Gaussian MAC: rate regions with coding schemes

Both TDMA with power control (TDP) and FDMA can get this. And successive cancellation/subtraction can achieve the entire region.