The Gallager Converse

Abbas El Gamal
Director, Information Systems Laboratory
Department of Electrical Engineering
Stanford University
Establishing information theoretic limits, e.g., channel capacity, requires:

- Finding a single-letter expression for the limit
- Proving achievability
- Proving a converse

While achievability tells us about how to improve system design, converse is necessary to prove optimality.

Proving a converse is typically harder and there are very few tools available, e.g., Fano’s inequality, data processing inequality, convexity.
Information Theoretic Limits

• Establishing information theoretic limits, e.g., channel capacity, requires:
  ◦ Finding a single-letter expression for the limit
  ◦ Proving achievability
  ◦ Proving a converse

• While achievability tells us about how to improve system design, converse is necessary to prove optimality

• Proving a converse is typically harder and there are very few tools available, e.g., Fano’s inequality, data processing inequality, convexity

• Gallager’s identification of the auxiliary random variable:
  ◦ Crucial step in proof of the degraded broadcast channel converse
  ◦ Has been used in the proof of almost all subsequent converses in multi-user information theory
  ◦ I used it many times in my papers
Broadcast Channel


- Discrete memoryless (DM) broadcast channel \((\mathcal{X}, p(y_1, y_2|x), \mathcal{Y}_1, \mathcal{Y}_2)\):

- Send independent message \(W_j \in [1, 2^{nR_j}]\) to receiver \(j = 1, 2\)
- Average probability of error: \(P_e(n) = P\{\hat{W}_1 \neq W_1 \text{ or } \hat{W}_2 \neq W_2\}\)
- \((R_1, R_2)\) achievable if there exists a sequence of codes with \(P_e(n) \to 0\)
- The capacity region is the closure of the set of achievable rates
Example 1: Binary Symmetric BC

- Assume that $p_1 \leq p_2 < 1/2$, $H(a)$, $a \in [0, 1]$ binary entropy function

\[ Z_1 \sim \text{Bern}(p_1) \quad Y_1 \]
\[ Z_2 \sim \text{Bern}(p_2) \quad Y_2 \]

$X$ \quad $R_1$

\[ 1 - H(p_2) \quad 1 - H(p_1) \]

\[ \text{time-sharing} \]
Example 1: Binary Symmetric BC

- Assume that $p_1 \leq p_2 < 1/2$

  \[
  \begin{align*}
  Z_1 &\sim \text{Bern}(p_1) \\
  + &\quad Y_1 \\
  X &\quad \rightarrow \\
  + &\quad Y_2 \\
  Z_2 &\sim \text{Bern}(p_2)
  \end{align*}
  \]

- **Superposition coding**: Use random coding. Let $0 \leq \beta \leq 1/2$; $U \sim \text{Bern}(1/2)$ and $X' \sim \text{Bern}(\beta)$ independent, $X = U + X'$
  
  - Generate $2^{nR_2}$ i.i.d. $U^n(w_2)$, $w_2 \in [1, 2^{nR_2}]$ (cloud centers)
  - Generate $2^{nR_1}$ i.i.d. $X'^n(w_1)$, $w_1 \in [1, 2^{nR_1}]$
  - Send $X^n(w_1, w_2) = U^n(w_2) + X'^n(w_1)$ (satellite codeword)
Example 1: Binary Symmetric BC

- Assume that \( p_1 \leq p_2 < 1/2 \)

\[
\begin{align*}
Z_1 &\sim \text{Bern}(p_1) \\
Z_2 &\sim \text{Bern}(p_2)
\end{align*}
\]

- Decoding:
  - Decoder 2 decodes \( U^n(W_2) \) \( \Rightarrow \) \( R_2 < 1 - H(\beta * p_2) \)
  - Decoder 1 first decodes \( U^n(W_2) \), subtracts it off, then decodes \( X^n(W_1) \) \( \Rightarrow \) \( R_1 < H(\beta * p_1) - H(p_1) \)
Example 2: AWGN BC

- Assume $N_1 \leq N_2$, average power constraint $P$ on $X$

  $Z_1 \sim \mathcal{N}(0, N_1)$
  $Z_2 \sim \mathcal{N}(0, N_2)$

  $Y_1 = X + Z_1$
  $Y_2 = X + Z_2$

- **Superposition coding**: Let $\alpha \in [0, 1]$, $U \sim \mathcal{N}(0, (1-\alpha)P)$, $X' \sim \mathcal{N}(0, \alpha P)$ independent, $X = U + X'$
  
  - Decoder 2 decodes cloud center $\Rightarrow R_2 \leq C \left((1-\alpha)P/(\alpha P + N_2)\right)$
  - Decoder 1 decodes clound center, subtracts it off, then decode the sattelite codeword $\Rightarrow R_1 \leq C \left(\alpha P/N_1\right)$
Degraded Broadcast Channel

• **Stochastically degraded** BC [Cover 1972]: There exists \( p(y_2|y_1) \) such that

\[
p(y_2|x) = \sum_{y_1} p(y_1|x)p(y_2|y_1)
\]

• Special case of channel inclusion in:

Degraded Broadcast Channel

- **Stochastically degraded** BC [Cover 1972]: There exists $p(y_2|y_1)$ such that

$$p(y_2|x) = \sum_{y_1} p(y_1|x)p(y_2|y_1)$$

Special case of channel inclusion in:


- **Physically degraded** version [Bergmans 73]:

- Since the capacity region of *any* BC depends only on marginals $p(y_1|x)$, $p(y_2|x)$ $\Rightarrow$ The capacity region of the degraded broadcast channel is the same as that of the physically degraded version
Degraded Broadcast Channel

• Cover conjectured that the capacity region is set of all \((R_1, R_2)\) such that

\[
R_2 \leq I(U; Y_2),
\]
\[
R_1 \leq I(X; Y_1|U),
\]

for some \(p(u, x)\)

• First time an auxiliary random variable is used in characterizing an information theoretic limit
**Achievability**


- Use superposition coding: Fix $p(u)p(x|u)$

- Decoder 2 decodes the cloud center $U^n \Rightarrow R_2 \leq I(U;Y_2)$
- Decoder 1 decodes first decodes the cloud center, then the satellite codeword $X^n \Rightarrow R_1 \leq I(X;Y_1|U)$

- Proved weak converse for binary symmetric broadcast channel (used Mrs. Gerber’s Lemma)
The Converse


- Proved weak converse for binary symmetric broadcast channel (used Mrs. Gerber’s Lemma)


- Proved the converse for the AWGN BC (used Entropy Power Inequality—very similar to Wyner’s proof for binary symmetric case)
The Converse


- Proved weak converse for binary symmetric broadcast channel (used Mrs. Gerber’s Lemma)


- Proved converse for AWGN BC (used Entropy Power Inequality)


- Proved the weak converse for the discrete-memoryless degraded BC—identification of auxiliary random variable
- Established bound on cardinality of the auxiliary random variable
Weak Converse for Shannon Channel Capacity

- Shannon capacity: \( C = \max_{p(x)} I(X; Y) \) (only uses “channel” variables)
- Weak converse: Show that for any sequence of \((n, R)\) codes with \( P_e^{(n)} \to 0, \ R \leq C \)
  - For each code form the empirical joint pmf:
    \((W, X^n, Y^n) \sim p(w)p(x^n|w)\prod_{i=1}^n p(y_i|x_i)\)
  - Fano’s inequality: \( H(W|Y^n) \leq nRP_e^{(n)} + H(P_e^{(n)}) = n\epsilon_n \)
  - Now consider
    \[
    nR = I(W; Y^n) + H(W|Y^n) \\
    \leq I(W; Y^n) + n\epsilon_n \\
    \leq I(X^n; Y^n) + n\epsilon_n \quad \text{data processing inequality} \\
    = H(Y^n) - H(Y^n|X^n) + n\epsilon_n \\
    \leq \sum_{i=1}^n I(X_i; Y_i) + n\epsilon_n \quad \text{convexity, memorylessness} \\
    \leq n(C + \epsilon_n)
    \]
Gallager Converse

- Show that given a sequence of \((n, R_1, R_2)\) codes with \(P_e^{(n)} \to 0\),
  \(R_1 \leq I(X; Y_1|U)\), \(R_2 \leq I(U; Y_2)\) for some \(p(u, x)\) such that
  \(U \to X \to (Y_1, Y_2)\)

- **Key is identifying** \(U\)
  - The converses for binary symmetric and AWGN BCs are direct and do not explicitly identify \(U\)

- As before, by Fano’s inequality
  \[
  H(W_1|Y_1^n) \leq nR_1P_e^{(n)} + 1 = n\epsilon_{1n}
  \]
  \[
  H(W_2|Y_2^n) \leq nR_2P_e^{(n)} + 1 = n\epsilon_{2n}
  \]

  So, we have
  \[
  nR_1 \leq I(W_1; Y_1^n) + n\epsilon_{1n},
  \]
  \[
  nR_2 \leq I(W_2; Y_2^n) + n\epsilon_{2n}
  \]
• **A First attempt:** \( U = W_2 \) (satisfies \( U \rightarrow X_i \rightarrow (Y_{i1}, Y_{2i}) \))

\[
I(W_1; Y_1^n) \leq I(W_1; Y_1^n|W_2) = I(W_1; Y_1^n|U)
\]

\[
= \sum_{i=1}^{n} I(W_1; Y_{1i}|U, Y_1^{i-1}) \quad \text{chain rule}
\]

\[
\leq \sum_{i=1}^{n} \left( H(Y_{1i}|U) - H(Y_{1i}|U, Y_1^{i-1}, W_1) \right)
\]

\[
\leq \sum_{i=1}^{n} \left( H(Y_{1i}|U) - H(Y_{1i}|U, Y_1^{i-1}, W_1, X_i) \right)
\]

\[
= \sum_{i=1}^{n} I(X_i; Y_{1i}|U)
\]

So far so good.
• **A First attempt:** $U = W_2$ (satisfies $U \rightarrow X_i \rightarrow (Y_{i1}, Y_{2i})$)

\[
I(W_1; Y_1^n) \leq I(W_1; Y_1^n|W_2) = I(W_1; Y_1^n|U) = \sum_{i=1}^{n} I(W_1; Y_{1i}|U, Y_{1i}^{i-1}) \leq \sum_{i=1}^{n} (H(Y_{1i}|U) - H(Y_{1i}|U, Y_{1i}^{i-1}, W_1)) \leq \sum_{i=1}^{n} (H(Y_{1i}|U) - H(Y_{1i}|U, Y_{1i}^{i-1}, W_1, X_i)) = \sum_{i=1}^{n} I(X_i; Y_{1i}|U)
\]

So far so good. Now let’s try the second inequality

\[
I(W_2; Y_2^n) = \sum_{i=1}^{n} I(W_2; Y_{2i}|Y_{2i}^{i-1}) = \sum_{i=1}^{n} I(U; Y_{2i}|Y_{2i}^{i-1})
\]

But $I(U; Y_{2i}|Y_{2i}^{i-1})$ is not necessarily $\leq I(U; Y_{2i})$, so $U = W_2$ does not work.
Gallager: Ok, let’s try $U_i = W_2, Y_1^{i-1}$ ($U_i \rightarrow X_i \rightarrow (Y_{1i}, Y_{2i})$), so

$$I(W_1; Y_1^n | W_2) \leq \sum_{i=1}^{n} I(X_i; Y_{1i} | U_i)$$

Now, let’s consider the other term

$$I(W_2; Y_2^n) \leq \sum_{i=1}^{n} I(W_2, Y_2^{i-1}; Y_{2i})$$

$$\leq \sum_{i=1}^{n} I(W_2, Y_2^{i-1}, Y_1^{i-1}; Y_{2i})$$

$$= \sum_{i=1}^{n} \left( H(Y_{2i}) - H(Y_{2i} | W_2, Y_2^{i-1}, Y_1^{i-1}) \right)$$

But $H(Y_{2i} | W_2, Y_2^{i-1}, Y_1^{i-1})$ is not in general equal to $H(Y_{2i} | W_2, Y_1^{i-1})$
• **Gallager**: Ok, let’s try $U_i = W_2, Y_1^{i-1}$ ($U_i \rightarrow X_i \rightarrow (Y_{1i}, Y_{2i})$), so

$$I(W_1; Y_1^n|W_2) \leq \sum_{i=1}^{n} I(X_i; Y_{1i}|U_i)$$

Now, let’s consider the other term

$$I(W_2; Y_2^n) \leq \sum_{i=1}^{n} I(W_2, Y_2^{i-1}; Y_{2i})$$

$$\leq \sum_{i=1}^{n} I(W_2, Y_2^{i-1}, Y_1^{i-1}; Y_{2i})$$

$$= \sum_{i=1}^{n} \left( H(Y_{2i}) - H(Y_{2i}|W_2, Y_2^{i-1}, Y_1^{i-1}) \right)$$

But $H(Y_{2i}|W_2, Y_2^{i-1}, Y_1^{i-1})$ is not in general equal to $H(Y_{2i}|W_2, Y_1^{i-1})$

• **Key insight**: Since capacity is the same as physically degraded, can assume $X \rightarrow Y_1 \rightarrow Y_2$, thus $Y_2^{i-1} \rightarrow (W_2, Y_1^{i-1}) \rightarrow Y_{2i}$, and

$$I(W_2; Y_2^n) \leq \sum_{i=1}^{n} I(U_i; Y_{2i})$$

**Eureka!**
• Note: Proof also works with $U_i = W_2, Y_2^{i-1}$ or $U_i = W_2, Y_1^{i-1}, Y_2^{i-1}$ (both satisfy $U_i \rightarrow X_i \rightarrow (Y_{1i}, Y_{2i})$). Let’s try $U_i = W_2, Y_2^{i-1}$

First consider

$$I(W_2; Y_2^n) \leq \sum_{i=1}^{n} I(W_2, Y_2^{i-1}; Y_{2i}) = \sum_{i=1}^{n} I(U_i; Y_{2i})$$

Now consider

$$I(W_1; Y_1^n|W_2) = \sum_{i=1}^{n} I(W_1; Y_{1i}|W_2, Y_1^{i-1})$$

$$\leq \sum_{i=1}^{n} I(X_i; Y_{1i}|W_2, Y_1^{i-1})$$

$$= \sum_{i=1}^{n} \left( H(Y_{1i}|W_2, Y_1^{i-1}) - H(Y_{1i}|X_i) \right)$$

$$= \sum_{i=1}^{n} \left( H(Y_{1i}|W_2, Y_1^{i-1}, Y_2^{i-1}) - H(Y_{1i}|X_i, U_i) \right)$$

$$\leq \sum_{i=1}^{n} \left( H(Y_{1i}|W_2, Y_2^{i-1}) - H(Y_{1i}|X_i, U_i) \right) = \sum_{i=1}^{n} I(X_i; Y_{1i}|U_i)$$
Thus we finally have

\[ nR_1 \leq \sum_{i=1}^{n} I(X_i; Y_{1i}|U_i) + n\epsilon_{1n} \]

\[ nR_2 \leq \sum_{i=1}^{n} I(U_i; Y_{2i}) + n\epsilon_{2n} \]

By convexity and letting \( n \to \infty \), it follows that there exists \( U \to X \to Y_1 \to Y_2 \) such that

\[ R_1 \leq I(X; Y_1|U), \]

\[ R_2 \leq I(U; Y_2) \]
Thus we finally have

\[ nR_1 \leq \sum_{i=1}^{n} I(X_i; Y_1 | U_i) + n\epsilon_{1n} \]

\[ nR_2 \leq \sum_{i=1}^{n} I(U_i; Y_2) + n\epsilon_{1n} \]

By convexity and letting \( n \to \infty \), it follows that there exists \( U \to X \to Y_1 \to Y_2 \) such that

\[ R_1 \leq I(X; Y_1 | U) + \epsilon_{1n}, \]

\[ R_2 \leq I(U; Y_2) + \epsilon_{2n} \]

- But there is still a problem here; \( U \) can have unbounded cardinality, so this is \textit{not} a computable expression

- Gallager then argues that \( |U| \leq \min\{|\mathcal{X}|, |\mathcal{Y}_1|, |\mathcal{Y}_2|\} \) suffices

- Remark: Proof extends to more than 2 receivers
Subsequent Broadcast Channel Results

- Direct application of Gallager converse:
  - Physically degraded BC with feedback [El Gamal 1978]
  - Product and sum of degraded and reversely degraded BC [Poltyrev 1977, El Gamal 1980]
Subsequent Broadcast Channel Results

- Direct application of Gallager converse:
  - Physically degraded BC with feedback [ElGamal 1978]
  - Product and sum of degraded and reversely degraded BC [Poltyrev 1977, ElGamal 1980]
- Gallager converse + Csizsar Lemma:
  - Less noisy [Korner, Marton 1975]: \( I(U; Y_1) \leq I(U; Y_2) \) for all \( U \rightarrow X \rightarrow (Y, Z) \)
  - BC with degraded message sets [Korner, Marton 1977]
  - More capable [ElGamal 1979] \( I(X; Y_1) \geq I(X; Y_2) \) for all \( p(x) \)
  - Semi-deterministic BC [Gelfand-Pinsker 1980]
  - Nair-El Gamal outer bound for BC [2006]
- All BC channel converses have used Gallager converse
Degraded Broadcast Channel With Feedback

• Assume at time $i$: $X_i = f_i(W_1, W_2, Y_1^{i-1}, Y_2^{i-1})$

• Capacity region can now depend on joint pmf $p(y_1, y_2|x) \Rightarrow$ capacity of stochastically degraded BC with feedback not necessarily equal to that of physically degraded counterpart

• My first result was showing that feedback doesn’t increase capacity of physically degraded BC [El Gamal 78]

• Converse:

\[
    nR_2 \leq I(W_2; Y_2^n) + n\epsilon_n \\
    \leq \sum_{i=1}^{n} I(W_2, Y_2^{i-1}; Y_2^i) + n\epsilon_n \\
    \leq \sum_{i=1}^{n} I(W_2, Y_1^{i-1}, Y_2^{i-1}; Y_2^i) + n\epsilon_n
\]

But because of feedback, $Y_2^{i-1}, (W_2, Y_1^{i-1}), Y_2^i$ do not necessarily form a Markov chain, thus cannot in general get rid of $Y_2^{i-1}$
Let $U_i = W, Y_1^{i-1}, Y_2^{i-1}$ ($U_i \rightarrow X_i \rightarrow Y_{1i} \rightarrow Y_{2i}$)

Now consider

$$nR_1 \leq I(W_1; Y_1^n | W_2) + n\epsilon_n$$

$$\leq I(W_1; Y_1^n, Y_2^n | W_2) + n\epsilon_n$$

$$= \sum_{i=1}^{n} I(W_1; Y_{1i}, Y_{2i} | W_2, Y_1^{i-1}, Y_2^{i-1}) + n\epsilon_n$$

$$= \sum_{i=1}^{n} I(W_1; Y_{1i}, Y_{2i} | U_i) + n\epsilon_n$$

$$\leq \sum_{i=1}^{n} I(X_i; Y_{1i}, Y_{2i} | U_i) + n\epsilon_n$$

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

$$= \sum_{i=1}^{n} I(X_i; Y_{1i} | U_i) + n\epsilon_n, \ (U_i \rightarrow X_i \rightarrow Y_{1i} \rightarrow Y_{2i})$$

- In general, feedback can enlarge the capacity region of degraded BC
- Capacity region of degraded BC with feedback is not known
Broadcast Channel with Degraded Message Sets

- Consider a general DM broadcast channel \((\mathcal{X}, p(y_1, y_2|x), \mathcal{Y}_1, \mathcal{Y}_2)\)
- Common message \(W_0\) is to be sent to both \(Y_1\) and \(Y_2\), while message \(W_1\) is to be sent only to \(Y_1\)
- Korner-Marton 77: The capacity region is given by the set \(C_1\) of all \((R_0, R_1)\) such that:

\[
\begin{align*}
R_0 &\leq I(U; Y_2) \\
R_1 &\leq I(X; Y_1|U) \\
R_0 + R_1 &\leq I(X; Y_1) \text{, for some } p(u, x)
\end{align*}
\]

- Achievability: Superposition coding
- Converse: Gallager-type converse doesn’t seem to work:
  - The first inequality holds with either \(U_i = W_0, Y_2^{i-1}\) or \(U_i = W_0, Y_1^{i-1}, Y_2^{i-1}\), but neither works for the second inequality
  - The second inequality works with \(U_i = W_0, Y_1^{i-1}\), but the first doesn’t
BC with Degraded Message Sets

- A Gallager type converse works, but needs a new ingredient [El Gamal]
- Define the set $C_2$ of all $(R_0, R_1)$ such that:
  
  $$R_0 \leq \min \{I(U; Y_1), I(U; Y_2)\}$$
  $$R_0 + R_1 \leq \min \{I(X; Y_1), I(X; Y_1|U) + I(U; Y_2)\}, \text{ for some } p(u, x)$$

- It is easy to show that $C_2 \subseteq C_1$
- Weak converse for $C_2$: Define $U_i = W_0, Y_1^{i-1}, Y_2^{n(2i+1)}$
- It is not difficult to show that
  
  $$R_0 \leq \frac{1}{n} \min \left\{ \sum_{i=1}^{n} I(U_i; Y_{1i}), \sum_{i=1}^{n} I(U_i; Y_{2i}) \right\} + \epsilon_n, \quad R_0 + R_1 \leq \frac{1}{n} \sum_{i=1}^{n} I(X_i; Y_{1i}) + \epsilon_n$$

- The difficult step is to show that
  
  $$R_0 + R_1 \leq \frac{1}{n} \sum_{i=1}^{n} (I(X_i; Y_{1i}|U_i) + I(U_i; Y_{2i})) + \epsilon_n$$
• To show this, consider

\[ n(R_0 + R_1) \leq I(W_1; Y_1^n|W_0) + I(W_0; Y_2^n) + n\epsilon_n \]

\[ = \sum_{i=1}^{n} \left( I(W_1; Y_1 i|W_0, Y_1^i-1) + I(W_0; Y_2 i|Y_2^n_{(i+1)}) \right) + n\epsilon_n \]

\[ \leq \sum_{i=1}^{n} \left( I(W_1, Y_2^n_{(i+1)}; Y_1 i|W_0, Y_1^i-1) + I(W_0, Y_2^n_{(i+1)}; Y_2 i) \right) + n\epsilon_n \]

\[ \leq \sum_{i=1}^{n} \left( I(X; Y_1 i|U_i) + I(Y_2^n_{(i+1)}; Y_1 i|W_0, Y_1^i-1) \right) \]

\[ + \sum_{i=1}^{n} \left( I(U; Y_2 i) - I(Y_1^i-1; Y_2 i|W_0, Y_2^n_{(i+1)}) \right) + n\epsilon_n. \]
To show this, consider

\[ n(R_0 + R_1) \leq I(W_1; Y_1^n | W_0) + I(W_0; Y_2^n) + n\epsilon_n \]

\[ = \sum_{i=1}^{n} \left( I(W_1; Y_{1i} | W_0, Y_1^{i-1}) + I(W_0; Y_{2i} | Y_2^n_{2(i+1)}) \right) + n\epsilon_n \]

\[ \leq \sum_{i=1}^{n} \left( I(W_1, Y_{2(i+1)}^n; Y_{1i} | W_0, Y_1^{i-1}) + I(W_0, Y_{2(i+1)}^n; Y_{2i}) \right) + n\epsilon_n \]

\[ \leq \sum_{i=1}^{n} \left( I(X_i; Y_{1i} | U_i) + I(Y_{2(i+1)}^n; Y_{1i} | W_0, Y_1^{i-1}) \right) \]

\[ + \sum_{i=1}^{n} \left( I(U_i; Y_{2i}) - I(Y_1^{i-1}; Y_{2i} | W_0, Y_2^n_{2(i+1)}) \right) + n\epsilon_n. \]

Now using the Csiszar lemma [Csiszar-Korner 1978]

\[ \sum_{i=1}^{n} I(Y_1^{i-1}; Y_{2i} | W_0, Y_{2(i+1)}^n) = \sum_{i=1}^{n} I(Y_{2(i+1)}^n; Y_{1i} | W_0, Y_1^{i-1}), \]

the second and fourth terms cancel and the converse is proved.

Note: This problem is open for more than two users.
Tightest Bounds on Capacity of Broadcast Channel

- **Inner bound [Marton 79]:** Set of rate triples \((R_0, R_1, R_2)\) such that
  \[
  R_0 \leq \min \{I(W; Y_1), I(W; Y_2)\},
  \]
  \[
  R_0 + R_1 \leq I(W, U; Y_1),
  \]
  \[
  R_0 + R_2 \leq I(W, V; Y_2),
  \]
  \[
  R_0 + R_1 + R_2 \leq \min \{I(W, U; Y_1) + I(V; Y_2|W), I(U; Y_1|W) + I(W, V; Y_2)\} - I(U; V|W),
  \]
  for some \(p(w, u, v, x)\)

- **Outer bound [Nair-El Gamal 2006]:** Set of rate triples \((R_1, R_2)\) such that
  \[
  R_0 \leq \min \{I(W; Y_1), I(W; Y_2)\}
  \]
  \[
  R_0 + R_1 \leq I(W, U; Y_1)
  \]
  \[
  R_0 + R_2 \leq I(W, V; Y_2)
  \]
  \[
  R_0 + R_1 + R_2 \leq \min \{I(W, U; Y_1) + I(V; Y_2|W, U), I(W, V; Y_2) + I(U; Y_1|W, V)\},
  \]
  for some \(p(u, v, w, x) = p(u)p(v)p(w|u, v)p(x|u, v, w)\)
Tightest Bounds for Independent Messages

- Inner bound: Set of all rate pairs \((R_1, R_2)\) such that

\[
\begin{align*}
R_1 &\leq I(W; U; Y_1) \\
R_2 &\leq I(W; V; Y_2) \\
R_1 + R_2 &\leq \min \{ I(W; U; Y_1) + I(V; Y_2|W), I(W; V; Y_2) + I(U; Y_1|V) \} \\
&\quad - I(U; V|W),
\end{align*}
\]

for some \(p(u, v, w, x) = p(w)p(u, v|w)p(x|u, v, w)\)

- Outer bound: Set of all rate pairs \((R_1, R_2)\) such that

\[
\begin{align*}
R_1 &\leq I(W; U; Y_1) \\
R_2 &\leq I(W; V; Y_2) \\
R_1 + R_2 &\leq \min \{ I(W; U; Y_1) + I(V; Y_2|W, U), I(W; V; Y_2) + I(U; Y_1|W, V) \},
\end{align*}
\]

for some joint distribution \(p(u, v, w, x) = p(u)p(v)p(w|u, v)p(x|u, v, w)\)
Broadcast Channel: Summary

• Marton region is tight for all classes of BC channels where capacity is known and contains all other achievable rate regions

• Nair-El Gamal region is tight for all classes of BC channels where capacity is known
  ○ Uses Gallager converse + Csizsar Lemma

• Capacity of BC is still not known in general
  ○ Is there a “single-letter” characterization?
  ○ Does it use auxiliary random variables?
  ○ If so, then the Gallager converse is again likely to play a key role
Gallager Converse: Other Multi-User Channels/Sources

- Lossless source coding with side information [Ahlswede-Korner 1975]
- Lossy source coding with side information [Wyner-Ziv 1976]
- Channel with state known at encoder
  [Gelfand-Pinsker 1980] — uses Gallager converse + Csiszar Lemma
- MAC with partially cooperating encoders [Willems 1983]
- MAC with cribbing encoders [Willems-van der Meulen, 1985]
- Rate distortion when side information may be absent
  [Heegard-Berger 1985]
- Multi-terminal source coding with one distortion criterion
  [Berger-Yeung 1989]
- Relay without delay [El Gamal-Hassanpour 2005]
- and, many others...
Conclusion

- Gallager converse has had a huge impact on the development of multi-user information theory
- Gallager converse has had a huge impact on my work
Conclusion

- Gallager converse has had a huge impact on the development of multi-user information theory
- Gallager converse has had a huge impact on my work
- But, Bob, what does the identification $U_i = W_2, Y_1^{i-1}$ really mean?