The Gallager Converse

Abbas El Gamal
Director, Information Systems Laboratory
Department of Electrical Engineering
Stanford University

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

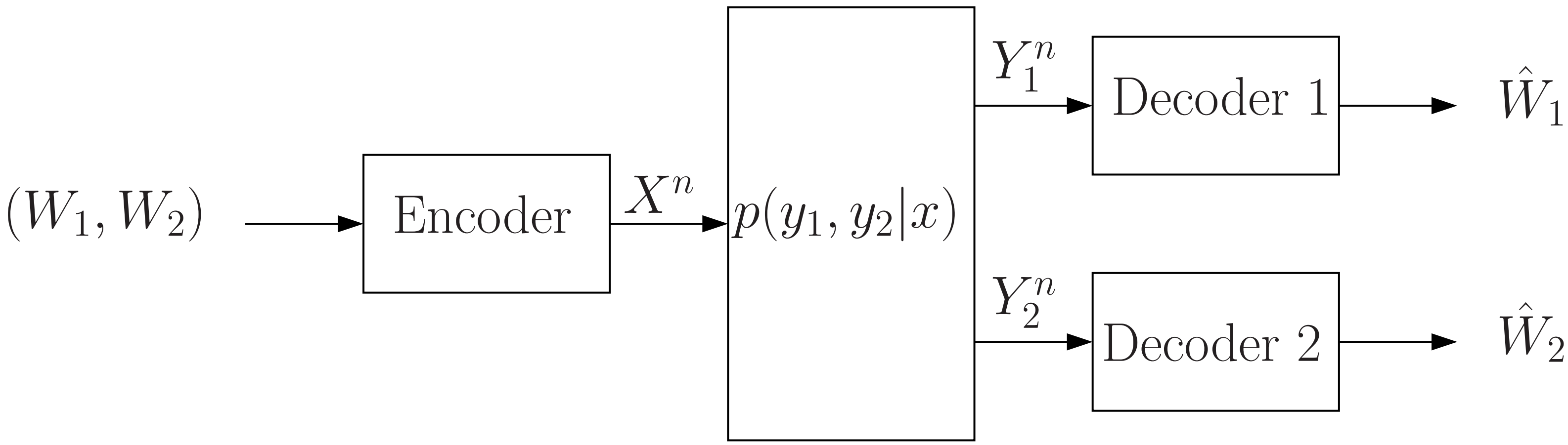
Information Theoretic Limits

- 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

T. M. Cover, "Broadcast Channels," *IEEE Trans. Info. Theory,* vol. IT-18, pp. 2-14, Jan. 1972

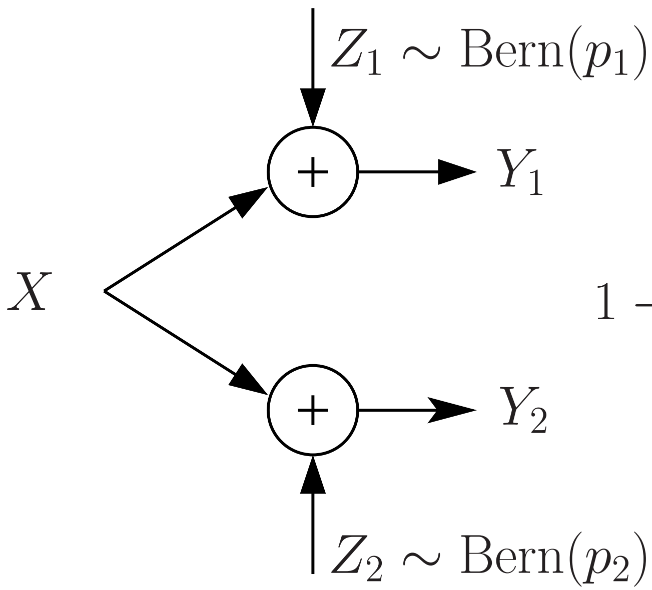
- Discrete memoryless (DM) *broadcast channel* $(\mathcal{X}, p(y_1, y_2|x), \mathcal{Y}_1, \mathcal{Y}_2)$:

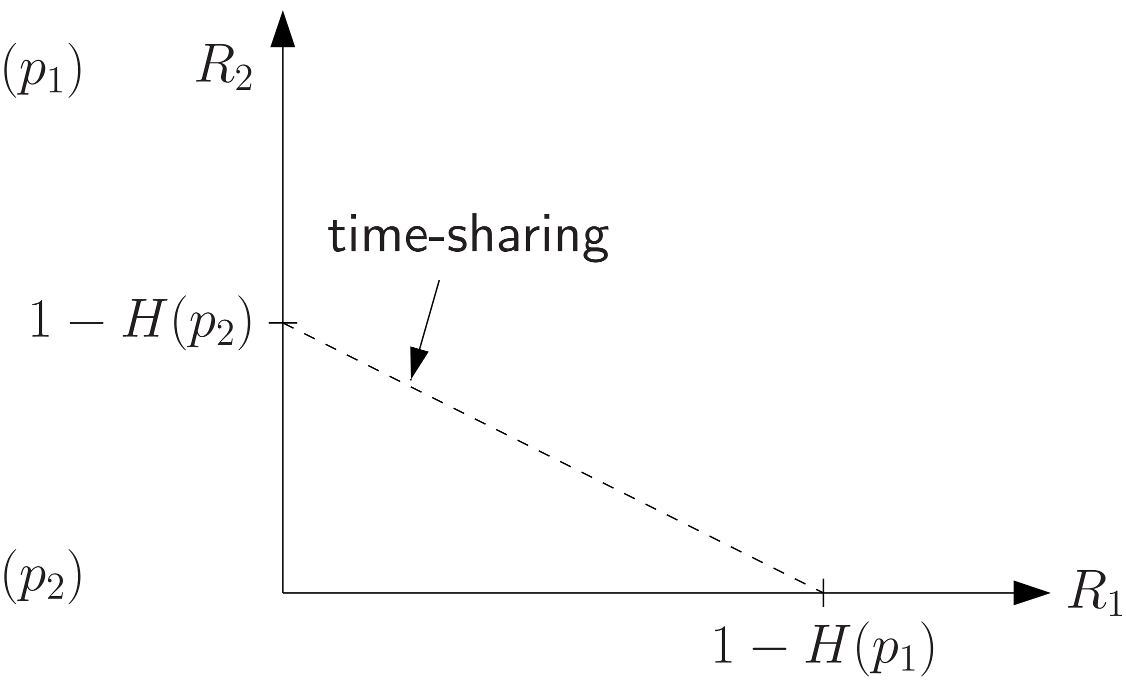


- Send independent message $W_j \in [1, 2^{nR_j}]$ to receiver $j = 1, 2$
- Average probability of error: $P_e^{(n)} = \mathrm{P}\{\hat{W}_1 \neq W_1$ or $\hat{W}_2 \neq W_2\}$
- (R_1, R_2) achievable if there exists a sequence of codes with $P_e^{(n)} \to 0$
- The *capacity region* is the closure of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

Example 1: Binary Symmetric BC

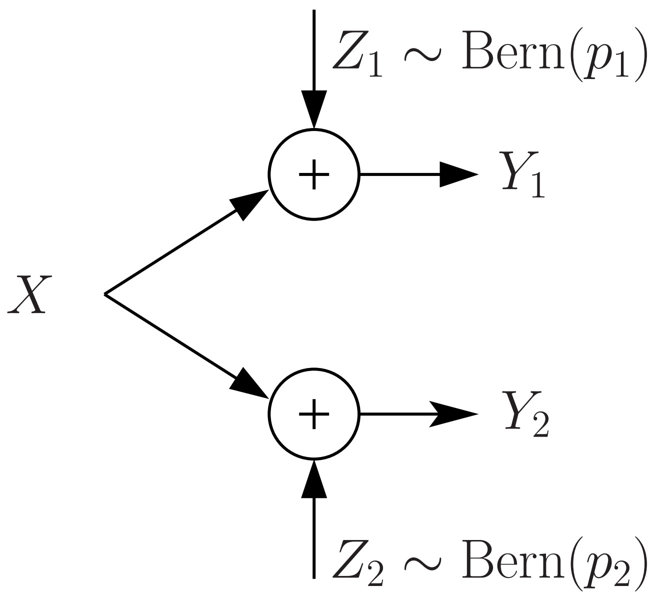
- Assume that $p_1 \leq p_2 < 1/2$, $H(a)$, $a \in [0, 1]$ binary entropy function

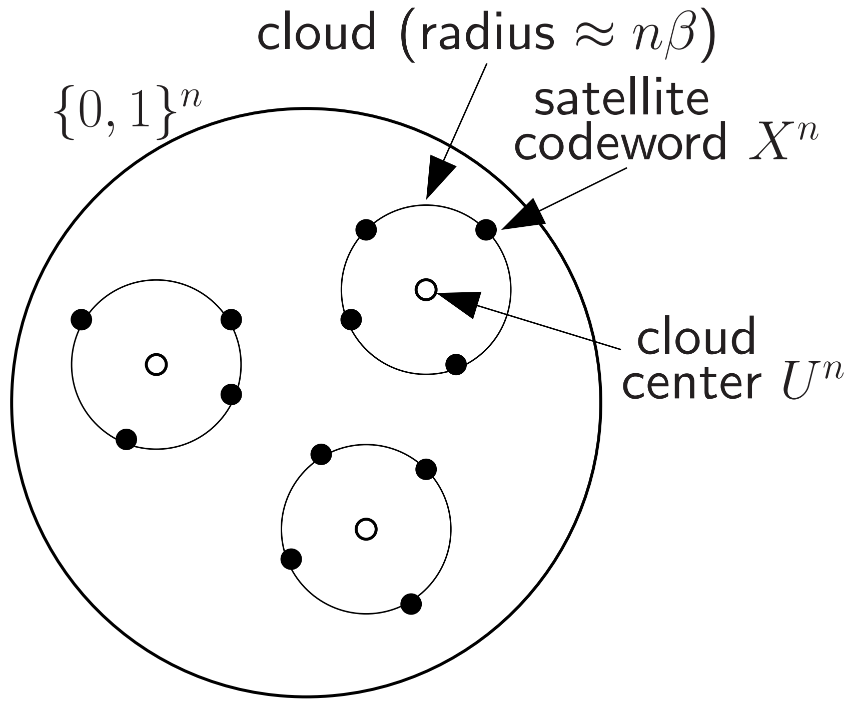




Example 1: Binary Symmetric BC

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

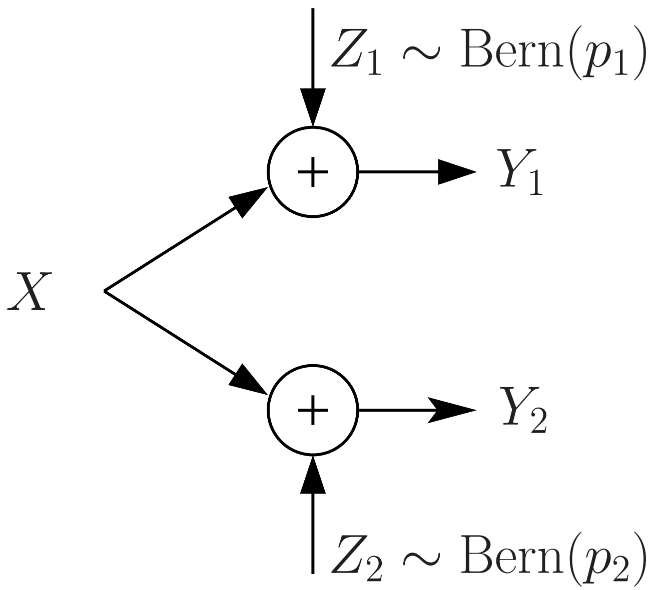


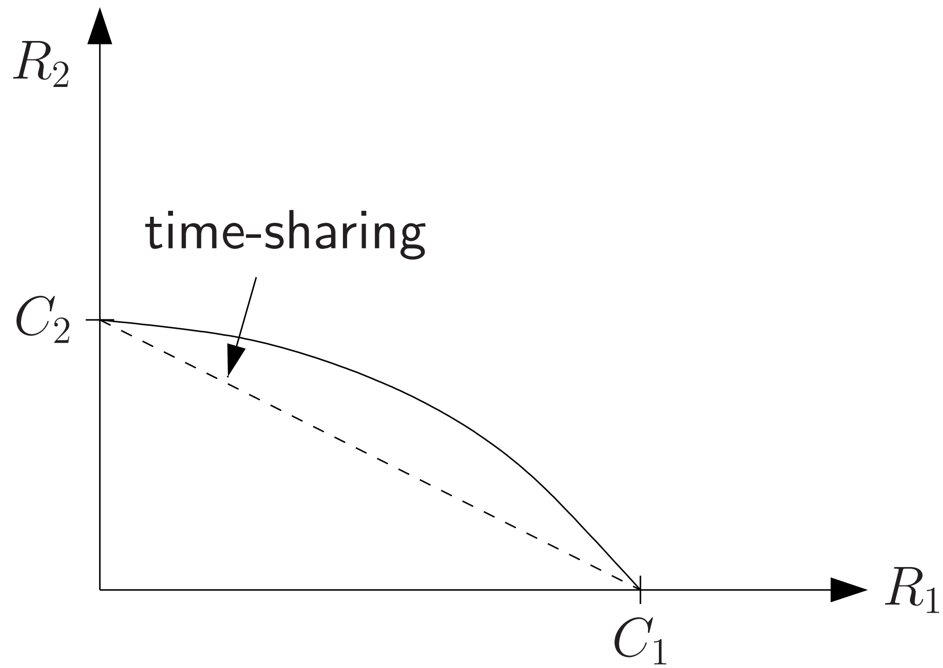


- **Superposition coding**: Use random coding. Let $0 \leq \beta \leq 1/2$; $U \sim \mathrm{Bern}(1/2)$ and $X' \sim \mathrm{Bern}(\beta)$ independent, $X = U + X'$
  - Generate 2^{nR_2} i.i.d. $U^n(w_2)$, $w_2 \in [1, 2^{nR_2}]$ (cloud centers)
  - Generate 2^{nR_1} i.i.d. $X'^n(w_1)$, $w_1 \in [1,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

Example 1: Binary Symmetric BC

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

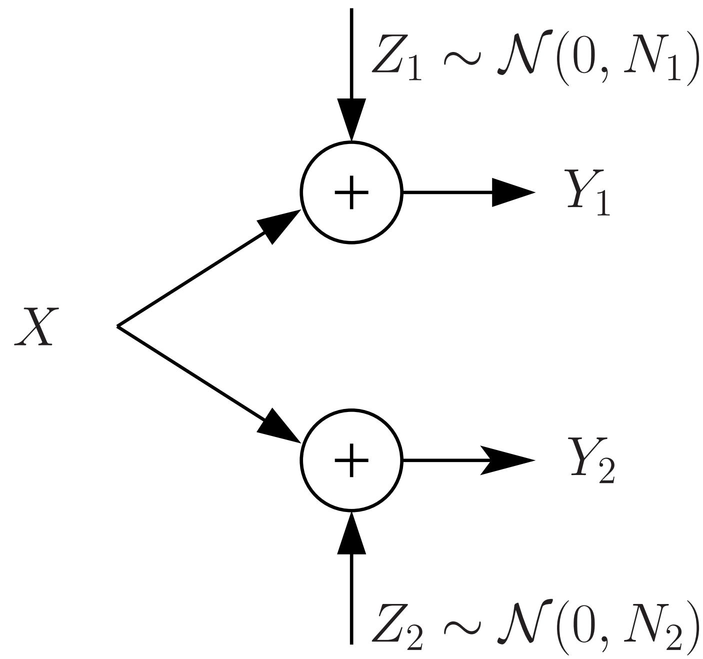




- **Decoding**:
  - Decoder 2 decodes $U^n(W_2) \Rightarrow R_2 < 1 - H(\beta * p_2)$
  - Decoder 1 first decodes $U^n(W_2)$, subtracts it off, then decodes $X'^n(W_1) \Rightarrow R_1 < H(\beta * p_1) - H(p_1)$

Example 2: AWGN BC

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



- **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 $\Rightarrow$ $R_2 \le \mathcal{C}\left((1-\alpha)P/(\alpha P + N_2)\right)$
  - Decoder 1 decodes clound center, subtracts it off, then decode the sattelite codeword $\Rightarrow$ $R_1 \le \mathcal{C}\left(\alpha P/N_1\right)$

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:

  C. E. Shannon, "A note on a partial ordering for communication channels," *Information and Control*, vol. 1, pp. 390-397, 1958

Degraded Broadcast Channel

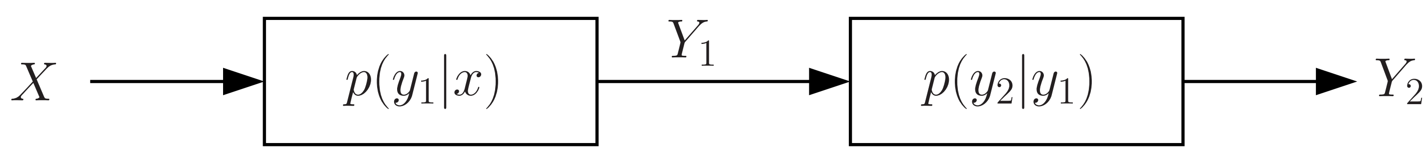
- **Stochastically degraded** BC [Cover 1972]: There exists $p(y_2|y_1)$ such that

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

  Special case of channel inclusion in:

  C. E. Shannon, "A Note on a Partial Ordering for Communication Channels," *Information and Control*, vol. 1, pp. 390-397, 1958

- **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

$$\begin{aligned} R_2 &\le I(U; Y_2), \\ R_1 &\le I(X; Y_1|U), \end{aligned}$$

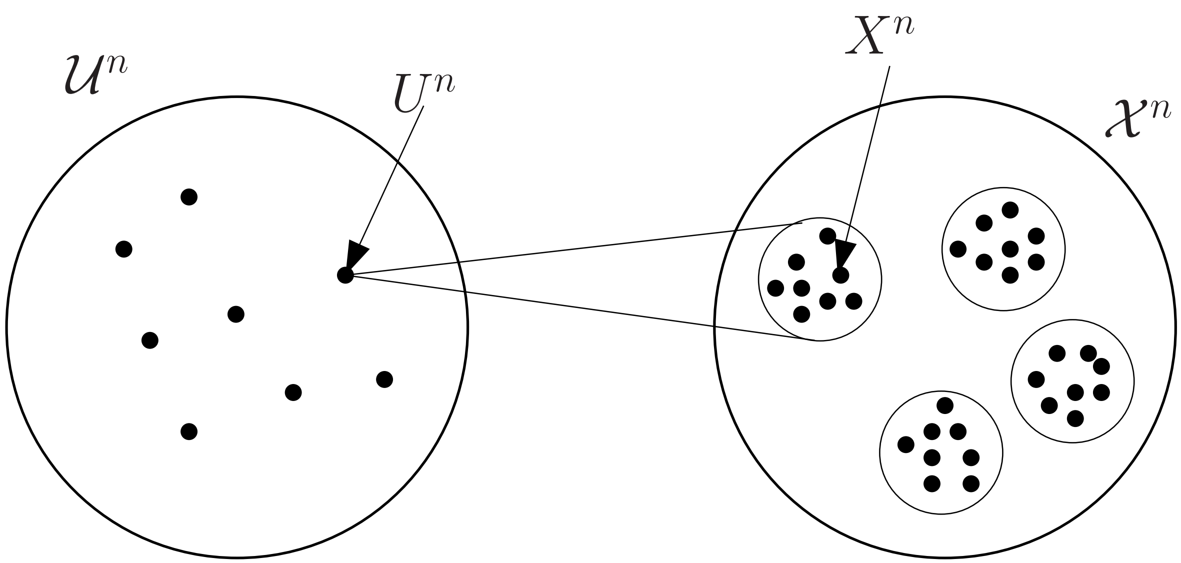
  for some $p(u, x)$

- First time an auxiliary random variable is used in characterizing an information theoretic limit

Achievability

P. P. Bergmans, "Random Coding Theorem for Broadcast Channels with Degraded Components," *IEEE Trans. Info. Theory*, vol. IT-19, pp. 197-207, Mar. 1973

- 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 sattelite codeword X^n $\Rightarrow R_1 \leq I(X;Y_1|U)$

The Converse

A. Wyner, "A Theorem on the Entropy of Certain Binary Sequences and Applications: Part II" *IEEE Trans. Info. Theory*, vol. IT-10, pp. 772-777, Nov. 1973

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

The Converse

A. Wyner, "A Theorem on the Entropy of Certain Binary Sequences and Applications: Part II" *IEEE Trans. Info. Theory*, vol. IT-10, pp. 772-777, Nov. 1973

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

P.P. Bergmans, "A Simple Converse for Broadcast Channels with Additive White Gaussian Noise (Corresp.)," *IEEE Trans. Info. Theory*, vol. IT-20, pp. 279-280, Mar. 1974

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

The Converse

A. Wyner, "A Theorem on the Entropy of Certain Binary Sequences and Applications: Part II" *IEEE Trans. Info. Theory*, vol. IT-10, pp. 772-777, Nov. 1973

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

P.P. Bergmans, "A Simple Converse for Broadcast Channels with Additive White Gaussian Noise (Corresp.)," *IEEE Trans. Info. Theory*, vol. IT-20, pp. 279-280, Mar. 1974

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

R. Gallager, "Capacity and Coding for Degraded Broadcast Channels," *Problemy Peredaci Informaccii*, vol. 10, no. 3, pp. 3-14, July-Sept 1974

- 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 $P_e^{(n)} \to 0$, $R \le C$
  - 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: $H(W|Y^n) \le nRP_e^{(n)} + H(P_e^{(n)}) = n\epsilon_n$
  - Now consider

$$\begin{aligned}
nR &= I(W;Y^n) + H(W|Y^n) \\
&\le I(W;Y^n) + n\epsilon_n \\
&\le I(X^n;Y^n) + n\epsilon_n && \text{data processing inequality} \\
&= H(Y^n) - H(Y^n|X^n) + n\epsilon_n \\
&\le \sum_{i=1}^n I(X_i;Y_i) + n\epsilon_n && \text{convexity, memorylessness} \\
&\le n(C + \epsilon_n)
\end{aligned}$$

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

$$H(W_1|Y_1^n) \leq nR_1 P_e^{(n)} + 1 = n\epsilon_{1n}$$
$$H(W_2|Y_2^n) \leq nR_2 P_e^{(n)} + 1 = n\epsilon_{2n}$$

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 $U \to X_i \to (Y_{i1}, Y_{2i})$)

$$\begin{aligned} I(W_1; Y_1^n) &\le I(W_1; Y_1^n | W_2) \\ &= I(W_1; Y_1^n | U) \\ &= \sum_{i=1}^{n} I(W_1; Y_{1i} | U, Y_1^{i-1}) \qquad \text{chain rule} \\ &\le \sum_{i=1}^{n} \big(H(Y_{1i}|U) - H(Y_{1i}|U, Y_1^{i-1}, W_1) \big) \\ &\le \sum_{i=1}^{n} \big(H(Y_{1i}|U) - H(Y_{1i}|U, Y_1^{i-1}, W_1, X_i) \big) \\ &= \sum_{i=1}^{n} I(X_i; Y_{1i}|U) \end{aligned}$$

So far so good.

- **A First attempt**: $U = W_2$ (satisfies $U \to X_i \to (Y_{i1}, Y_{2i})$)

$$\begin{aligned} I(W_1; Y_1^n) &\le I(W_1; Y_1^n | W_2) \\ &= I(W_1; Y_1^n | U) \\ &= \sum_{i=1}^{n} I(W_1; Y_{1i} | U, Y_1^{i-1}) \\ &\le \sum_{i=1}^{n} \big(H(Y_{1i}|U) - H(Y_{1i}|U, Y_1^{i-1}, W_1) \big) \\ &\le \sum_{i=1}^{n} \big(H(Y_{1i}|U) - H(Y_{1i}|U, Y_1^{i-1}, W_1, X_i) \big) \\ &= \sum_{i=1}^{n} I(X_i; Y_{1i}|U) \end{aligned}$$

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

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

But $I(U; Y_{2i}|Y_2^{i-1})$ is not necessarily $\le I(U; Y_{2i})$, so $U = W_2$ does not work

- **Gallager**: Ok, let's try $U_i = W_2, Y_1^{i-1}$ ($U_i \to X_i \to (Y_{1i}, Y_{2i})$), 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

$$\begin{aligned} I(W_2; Y_2^n) &\leq \sum_{i=1}^{n} I(W_2, Y_2^{i-1}; Y_{2i}) \\ &\leq \sum_{i=1}^{n} I(W_2, Y_2^{i-1}, Y_1^{i-1}; Y_{2i}) \\ &= \sum_{i=1}^{n} \left(H(Y_{2i}) - H(Y_{2i} | W_2, Y_2^{i-1}, Y_1^{i-1}) \right) \end{aligned}$$

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 $U_i = W_2, Y_1^{i-1}$ $(U_i \to X_i \to (Y_{1i}, Y_{2i}))$, so

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

Now, let's consider the other term

$$\begin{aligned} I(W_2; Y_2^n) &\le \sum_{i=1}^{n} I(W_2, Y_2^{i-1}; Y_{2i}) \\ &\le \sum_{i=1}^{n} I(W_2, Y_2^{i-1}, Y_1^{i-1}; Y_{2i}) \\ &= \sum_{i=1}^{n} \left(H(Y_{2i}) - H(Y_{2i} | W_2, Y_2^{i-1}, Y_1^{i-1}) \right) \end{aligned}$$

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 $X \to Y_1 \to Y_2$, thus $Y_2^{i-1} \to (W_2, Y_1^{i-1}) \to Y_{2i}$, and

$$I(W_2; Y_2^n) \le \sum_{i=1}^{n} I(U_i; Y_{2i}) \qquad \text{Eureka !}$$

- 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 $U_i \to X_i \to (Y_{1i}, Y_{2i})$). Let's try $U_i = W_2, Y_2^{i-1}$

  First consider

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

  Now consider

$$\begin{aligned}
I(W_1; Y_1^n | W_2) &= \sum_{i=1}^{n} I(W_1; Y_{1i} | W_2, Y_1^{i-1}) \\
&\le \sum_{i=1}^{n} I(X_i; Y_{1i} | W_2, Y_1^{i-1}) \\
&= \sum_{i=1}^{n} \big(H(Y_{1i} | W_2, Y_1^{i-1}) - H(Y_{1i} | X_i) \big) \\
&= \sum_{i=1}^{n} \big(H(Y_{1i} | W_2, Y_1^{i-1}, Y_2^{i-1}) - H(Y_{1i} | X_i, U_i) \big) \\
&\le \sum_{i=1}^{n} \big(H(Y_{1i} | W_2, Y_2^{i-1}) - H(Y_{1i} | X_i, U_i) \big) = \sum_{i=1}^{n} I(X_i; Y_{1i} | U_i)
\end{aligned}$$

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

$$nR_1 \leq \sum_{i=1}^{n} I(X_i; Y_{1i}|U_i) + n\epsilon_{1n}$$
$$nR_2 \leq \sum_{i=1}^{n} I(U_i; Y_{2i}) + n\epsilon_{1n}$$

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 *not* a computable expression
- Gallager then argues that $|\mathcal{U}| \leq \min\{|\mathcal{X}|, |\mathcal{Y}_1|, |\mathcal{Y}_2|\}$ 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 [El Gamal 1978]
  - Product and sum of degraded and reversely degraded BC [Poltyrev 1977, El Gamal 1980]
- Gallager converse+Csizsar Lemma:
  - Less noisy [Korner, Marton 1975]: $I(U;Y_1) \leq I(U;Y_2)$ for all $U \to X \to (Y,Z)$
  - BC with degraded message sets [Korner, Marton 1977]
  - More capable [El Gamal 1979] $\big(I(X;Y_1) \geq I(X;Y_2)$ for all $p(x)\big)$
  - 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:

$$\begin{aligned} nR_2 &\le I(W_2; Y_2^n) + n\epsilon_n \\ &\le \sum_{i=1}^n I(W_2, Y_2^{i-1}; Y_{2i}) + n\epsilon_n \\ &\le \sum_{i=1}^n I(W_2, Y_1^{i-1}, Y_2^{i-1}; Y_{2i}) + n\epsilon_n \end{aligned}$$

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

Let $U_i = W_2, Y_1^{i-1}, Y_2^{i-1}$ $(U_i \to X_i \to Y_{1i} \to Y_{2i})$

Now conisder

$$
\begin{aligned}
nR_1 &\le I(W_1; Y_1^n | W_2) + n\epsilon_n \\
&\le I(W_1; Y_1^n, Y_2^n | W_2) + n\epsilon_n \\
&= \sum_{i=1}^{n} I(W_1; Y_{1i}, Y_{2i} | W_2, Y_1^{i-1}, Y_2^{i-1}) + n\epsilon_n \\
&= \sum_{i=1}^{n} I(W_1; Y_{1i}, Y_{2i} | U_i) + n\epsilon_n \\
&\le \sum_{i=1}^{n} I(X_i; Y_{1i}, Y_{2i} | U_i) + n\epsilon_n \\
&= \sum_{i=1}^{n} I(X_i; Y_{1i} | U_i) + \sum_{i=1}^{n} I(X_i; Y_{2i} | Y_{1i}, U_i) + n\epsilon_n \\
&= \sum_{i=1}^{n} I(X_i; Y_{1i} | U_i) + n\epsilon_n, \;\; (U_i \to X_i \to Y_{1i} \to Y_{2i})
\end{aligned}
$$

- 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 $\mathcal{C}_1$ of all (R_0, R_1) such that:

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

- 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 $\mathcal{C}_2$ 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 $\mathcal{C}_2 \subseteq \mathcal{C}_1$
- Weak converse for $\mathcal{C}_2$: Define $U_i = W_0, Y_1^{i-1}, Y_{2(i+1)}^n$!!
- It is not difficult to show that

$$R_0 \leq \frac{1}{n}\min\left\{\sum_{i=1}^n I(U_i;Y_{1i}), \sum_{i=1}^n I(U_i;Y_{2i})\right\}+\epsilon_n, \; R_0+R_1 \leq \frac{1}{n}\sum_{i=1}^n I(X_i;Y_{1i})+\epsilon_n$$

- The difficult step is to show that

$$R_0 + R_1 \leq \frac{1}{n}\sum_{i=1}^n \left(I(X_i;Y_{1i}|U_i) + I(U_i;Y_{2i})\right) + \epsilon_n$$

- To show this, consider

$$
\begin{aligned}
n(R_0 + R_1) &\le I(W_1; Y_1^n | W_0) + I(W_0; Y_2^n) + n\epsilon_n \\
&= \sum_{i=1}^{n} \Big(I(W_1; Y_{1i} | W_0, Y_1^{i-1}) + I(W_0; Y_{2i} | Y_{2(i+1)}^n) \Big) + n\epsilon_n \\
&\le \sum_{i=1}^{n} \Big(I(W_1, Y_{2(i+1)}^n; Y_{1i} | W_0, Y_1^{i-1}) + I(W_0, Y_{2(i+1)}^n; Y_{2i}) \Big) + n\epsilon_n \\
&\le \sum_{i=1}^{n} \Big(I(X_i; Y_{1i} | U_i) + I(Y_{2(i+1)}^n; Y_{1i} | W_0, Y_1^{i-1}) \Big) \\
&\quad + \sum_{i=1}^{n} \Big(I(U_i; Y_{2i}) - I(Y_1^{i-1}; Y_{2i} | W_0, Y_{2(i+1)}^n) \Big) + n\epsilon_n .
\end{aligned}
$$

- To show this, consider

$$
\begin{aligned}
n(R_0+R_1) &\le I(W_1;Y_1^n|W_0) + I(W_0;Y_2^n) + n\epsilon_n \\
&= \sum_{i=1}^n \Big(I(W_1;Y_{1i}|W_0,Y_1^{i-1}) + I(W_0;Y_{2i}|Y_{2(i+1)}^n) \Big) + n\epsilon_n \\
&\le \sum_{i=1}^n \Big(I(W_1,Y_{2(i+1)}^n;Y_{1i}|W_0,Y_1^{i-1}) + I(W_0,Y_{2(i+1)}^n;Y_{2i}) \Big) + n\epsilon_n \\
&\le \sum_{i=1}^n \Big(I(X_i;Y_{1i}|U_i) + I(Y_{2(i+1)}^n;Y_{1i}|W_0,Y_1^{i-1}) \Big) \\
&\qquad + \sum_{i=1}^n \Big(I(U_i;Y_{2i}) - I(Y_1^{i-1};Y_{2i}|W_0,Y_{2(i+1)}^n) \Big) + n\epsilon_n.
\end{aligned}
$$

- 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

$$\begin{aligned}
R_0 &\le \min\left\{I(W;Y_1), I(W;Y_2)\right\},\\
R_0 + R_1 &\le I(W,U;Y_1),\\
R_0 + R_2 &\le I(W,V;Y_2),\\
R_0 + R_1 + R_2 &\le \min\{I(W,U;Y_1) + I(V;Y_2|W), I(U;Y_1|W) + I(W,V;Y_2)\}\\
&\quad - I(U;V|W),
\end{aligned}$$

  for some $p(w,u,v,x)$

- Outer bound [Nair-El Gamal 2006]: Set of rate triples (R_1, R_2) such that

$$\begin{aligned}
R_0 &\le \min\left\{I(W;Y_1), I(W;Y_2)\right\}\\
R_0 + R_1 &\le I(W,U;Y_1)\\
R_0 + R_2 &\le I(W,V;Y_2)\\
R_0 + R_1 + R_2 &\le \min\{I(W,U;Y_1) + I(V;Y_2|W,U), I(W,V;Y_2)\\
&\quad + I(U;Y_1|W,V)\},
\end{aligned}$$

  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

$$\begin{aligned} R_1 &\le I(W,U;Y_1) \\ R_2 &\le I(W,V;Y_2) \\ R_1 + R_2 &\le \min\{I(W,U;Y_1) + I(V;Y_2|W), I(W,V;Y_2) + I(U;Y_1|V)\} \\ &\quad - I(U;V|W), \end{aligned}$$

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

$$\begin{aligned} R_1 &\le I(W,U;Y_1) \\ R_2 &\le I(W,V;Y_2) \\ R_1 + R_2 &\le \min\{I(W,U;Y_1) + I(V;Y_2|W,U), I(W,V;Y_2) + I(U;Y_1|W,V)\}, \end{aligned}$$

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 ?