Elements of Network Information Theory

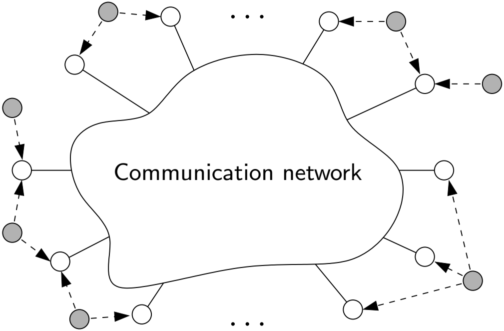
Abbas El Gamal and Young-Han Kim

Stanford University and UC San Diego

Tutorial, ISIT 2011

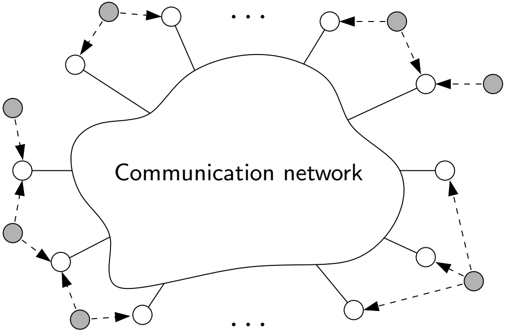
Slides available at http://isl.stanford.edu/~abbas

Networked Information Processing System



- System: Internet, peer-to-peer network, sensor network, . . .
- Sources: Data, speech, music, images, video, sensor data
- Nodes: Handsets, base stations, processors, servers, sensor nodes, . . .
- Network: Wired, wireless, or a hybrid of the two
- Task: Communicate the sources, or compute/make decision based on them

Network Information Flow Questions



- What is the limit on the amount of communication needed?
- What are the coding scheme/techniques that achieve this limit?
- Shannon (1948): Noisy point-to-point communication
- Ford–Fulkerson, Elias–Feinstein–Shannon (1956): Graphical unicast networks

Network Information Theory

- Simplistic model of network as graph with point-to-point links and forwarding nodes does not capture many important aspects of real-world networks:
    - Networked systems have multiple sources and destinations
    - The network task is often to compute a function or to make a decision
    - Many networks allow for feedback and interactive communication
    - The wireless medium is a shared broadcast medium
    - Network security is often a primary concern
    - Source–channel separation does not hold for networks
    - Data arrival and network topology evolve dynamically
- Network information theory aims to answer the information flow questions while capturing some of these aspects of real-world networks

Brief History

- First paper: Shannon (1961) “Two-way communication channels”
    - He didn’t find the optimal rates (capacity region)
    - The problem remains open
- Significant research activities in 70s and early 80s with many new results and techniques, but
    - Many basic problems remained open
    - Little interest from information and communication theorists
- Wireless communications and the Internet revived interest in mid 90s
    - Some progress on old open problems and many new models and problems
    - Coding techniques, such as successive cancellation, superposition, Slepian–Wolf, Wyner–Ziv, successive refinement, writing on dirty paper, and network coding, beginning to impact real-world networks

Network Information Theory Book

- The book provides a comprehensive coverage of key results, techniques, and open problems in network information theory
- The organization balances the introduction of new techniques and new models
- The focus is on discrete memoryless and Gaussian network models
- We discuss extensions (if any) to many users and large networks
- The proofs use elementary tools and techniques
- We use clean and unified notation and terminology

Book Organization

Part I. Preliminaries (Chapters 2,3): Review of basic information measures, typicality, Shannon's theorems. Introduction of key lemmas

Part II. Single-hop networks (Chapters 4 to 14): Networks with single-round, one-way communication

- Independent messages over noisy channels
- Correlated (uncompressed) sources over noiseless links
- Correlated sources over noisy channels

Part III. Multihop networks (Chapters 15 to 20): Networks with relaying and multiple communication rounds

- Independent messages over graphical networks
- Independent messages over general networks
- Correlated sources over graphical networks

Part IV. Extensions (Chapters 21 to 24): Extensions to distributed computing, secrecy, wireless fading channels, and information theory and networking

Tutorial Objectives

- Focus on elementary and unified approach to coding schemes
    - Typicality and simple “universal” lemmas for DM models

- Lossless source coding as a corollary of lossy source coding

- Extending achievability proofs from DM to Gaussian models

- Illustrate the approach through proofs of several classical coding theorems

Outline

1. Typical Sequences
2. Point-to-Point Communication
3. Multiple Access Channel
4. Broadcast Channel

◂ 10-minute break

5. Lossy Source Coding
6. Wyner–Ziv Coding
7. Gelfand–Pinsker Coding

◂ 10-minute break

8. Wiretap Channel
9. Relay Channel
10. Multicast Network

Typical Sequences

- Empirical pmf (or type) of $x^n \in \mathcal{X}^n$:

$$\pi(x|x^n) = \frac{\left|\{i\colon x_i = x\}\right|}{n} \quad \text{for } x \in \mathcal{X}$$

- Typical set (Orlitsky–Roche 2001): For $X \sim p(x)$ and $\epsilon > 0$,

$$\mathcal{T}_\epsilon^{(n)}(X) = \left\{x^n\colon \left|\pi(x|x^n) - p(x)\right| \le \epsilon \cdot p(x) \text{ for all } x \in \mathcal{X}\right\} = \mathcal{T}_\epsilon^{(n)}$$

Typical Average Lemma

Let $x^n \in \mathcal{T}_\epsilon^{(n)}(X)$ and $g(x) \ge 0$. Then

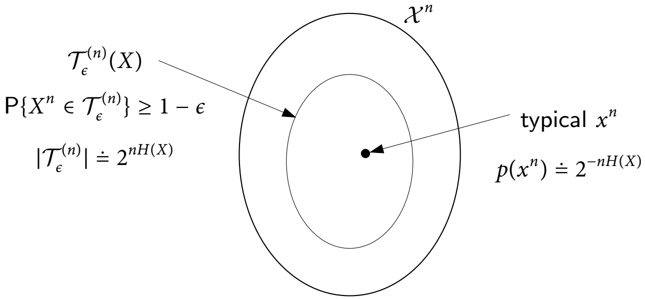
$$(1-\epsilon)\,\mathrm{E}(g(X)) \le \frac{1}{n}\sum_{i=1}^{n} g(x_i) \le (1+\epsilon)\,\mathrm{E}(g(X))$$

Properties of Typical Sequences

- Let $x^n \in \mathcal{T}_\epsilon^{(n)}(X)$ and $p(x^n) = \prod_{i=1}^n p_X(x_i)$. Then

  $$2^{-n(H(X)+\delta(\epsilon))} \le p(x^n) \le 2^{-n(H(X)-\delta(\epsilon))},$$

  where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$ (Notation: $p(x^n) \doteq 2^{-nH(X)}$)
- $|\mathcal{T}_\epsilon^{(n)}(X)| \doteq 2^{nH(X)}$ for n sufficiently large
- Let $X^n \sim \prod_{i=1}^n p_X(x_i)$. Then by the LLN, $\lim_{n\to\infty} \mathsf{P}\{X^n \in \mathcal{T}_\epsilon^{(n)}\} = 1$



$\mathcal{X}^n$

$\mathcal{T}_\epsilon^{(n)}(X)$

$\mathsf{P}\{X^n \in \mathcal{T}_\epsilon^{(n)}\} \ge 1 - \epsilon$

$|\mathcal{T}_\epsilon^{(n)}| \doteq 2^{nH(X)}$

typical x^n

$p(x^n) \doteq 2^{-nH(X)}$

Jointly Typical Sequences

- Joint type of $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$:

$$\pi(x, y | x^n, y^n) = \frac{\left|\{i \colon (x_i, y_i) = (x, y)\}\right|}{n} \text{ for } (x, y) \in \mathcal{X} \times \mathcal{Y}$$

- Jointly typical set: For $(X, Y) \sim p(x, y)$ and $\epsilon > 0$,

$$\begin{aligned}\mathcal{T}_\epsilon^{(n)}(X, Y) &= \{(x^n, y^n) \colon |\pi(x, y | x^n, y^n) - p(x, y)| \le \epsilon \cdot p(x, y) \text{ for all } (x, y)\} \\ &= \mathcal{T}_\epsilon^{(n)}((X, Y))\end{aligned}$$

- Let $(x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y)$ and $p(x^n, y^n) = \prod_{i=1}^n p_{X,Y}(x_i, y_i)$. Then
  - $x^n \in \mathcal{T}_\epsilon^{(n)}(X)$ and $y^n \in \mathcal{T}_\epsilon^{(n)}(Y)$
  - $p(x^n) \doteq 2^{-nH(X)}$, $p(y^n) \doteq 2^{-nH(Y)}$, and $p(x^n, y^n) \doteq 2^{-nH(X,Y)}$
  - $p(x^n | y^n) \doteq 2^{-nH(X|Y)}$ and $p(y^n | x^n) \doteq 2^{-nH(Y|X)}$

Conditionally Typical Sequences

- Conditionally typical set: For $x^n \in \mathcal{X}^n$,

$$\mathcal{T}_\epsilon^{(n)}(Y|x^n) = \{y^n : (x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}(X,Y)\}$$

- $|\mathcal{T}_\epsilon^{(n)}(Y|x^n)| \le 2^{n(H(Y|X)+\delta(\epsilon))}$

Conditional Typicality Lemma

Let $(X,Y) \sim p(x,y)$. If $x^n \in \mathcal{T}_{\epsilon'}^{(n)}(X)$ and $Y^n \sim \prod_{i=1}^n p_{Y|X}(y_i|x_i)$, then for $\epsilon > \epsilon'$,

$$\lim_{n\to\infty} \mathsf{P}\{(x^n, Y^n) \in \mathcal{T}_\epsilon^{(n)}(X,Y)\} = 1$$

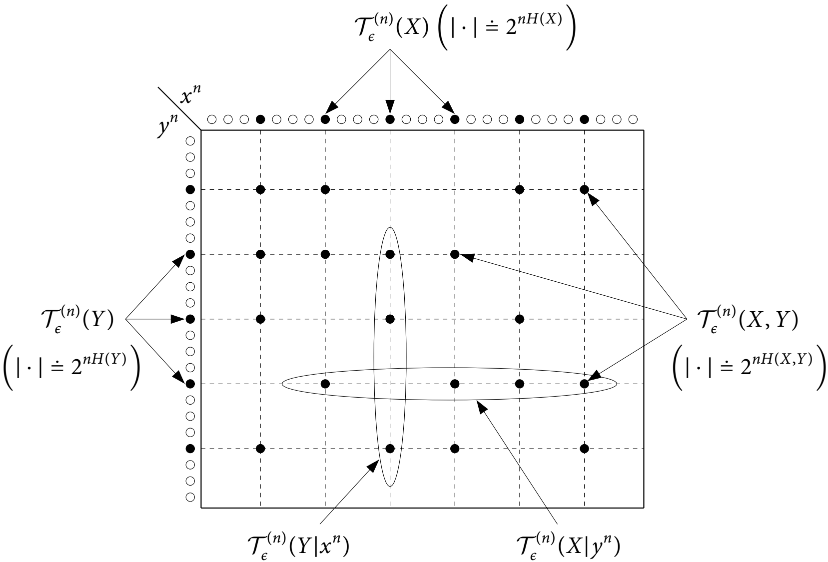
- If $x^n \in \mathcal{T}_{\epsilon'}^{(n)}(X)$ and $\epsilon > \epsilon'$, then for n sufficiently large,

$$|\mathcal{T}_\epsilon^{(n)}(Y|x^n)| \ge 2^{n(H(Y|X)-\delta(\epsilon))}$$

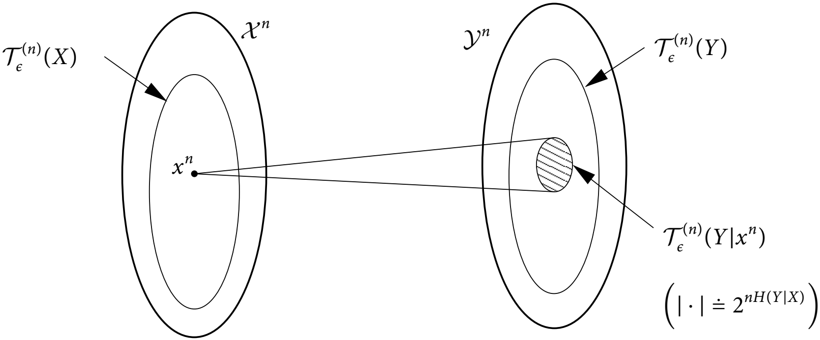
- Let $X \sim p(x)$, $Y = g(X)$, and $x^n \in \mathcal{T}_\epsilon^{(n)}(X)$. Then

$$y^n \in \mathcal{T}_\epsilon^{(n)}(Y|x^n) \quad \text{iff} \quad y_i = g(x_i),\ i \in [1:n]$$

Illustration of Joint Typicality



Another Illustration of Joint Typicality



Joint Typicality for Random Triples

- Let $(X, Y, Z) \sim p(x, y, z)$. The set of typical sequences is

$$\mathcal{T}_{\epsilon}^{(n)}(X, Y, Z) = \mathcal{T}_{\epsilon}^{(n)}((X, Y, Z))$$

Joint Typicality Lemma

Let $(X, Y, Z) \sim p(x, y, z)$ and $\epsilon' < \epsilon$. Then for some $\delta(\epsilon) \to 0$ as $\epsilon \to 0$:

- If $(\tilde{x}^n, \tilde{y}^n)$ is arbitrary and $\tilde{Z}^n \sim \prod_{i=1}^{n} p_{Z|X}(\tilde{z}_i|\tilde{x}_i)$, then

$$\mathsf{P}\{(\tilde{x}^n, \tilde{y}^n, \tilde{Z}^n) \in \mathcal{T}_{\epsilon}^{(n)}(X, Y, Z)\} \le 2^{-n(I(Y;Z|X)-\delta(\epsilon))}$$

- If $(x^n, y^n) \in \mathcal{T}_{\epsilon'}^{(n)}$ and $\tilde{Z}^n \sim \prod_{i=1}^{n} p_{Z|X}(\tilde{z}_i|x_i)$, then for n sufficiently large,

$$\mathsf{P}\{(x^n, y^n, \tilde{Z}^n) \in \mathcal{T}_{\epsilon}^{(n)}(X, Y, Z)\} \ge 2^{-n(I(Y;Z|X)+\delta(\epsilon))}$$

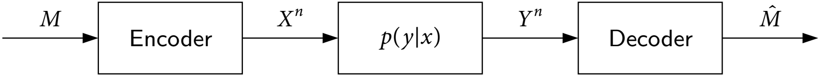
Summary

1. Typical Sequences
2. Point-to-Point Communication
3. Multiple Access Channel
4. Broadcast Channel
5. Lossy Source Coding
6. Wyner–Ziv Coding
7. Gelfand–Pinsker Coding
8. Wiretap Channel
9. Relay Channel
10. Multicast Network

- Typical average lemma
- Conditional typicality lemma
- Joint typicality lemma

Discrete Memoryless Channel (DMC)

- Point-to-point communication system

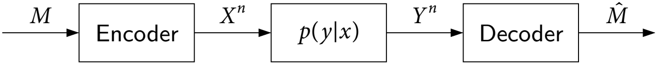


- Assume a discrete memoryless channel (DMC) model $(\mathcal{X}, p(y|x), \mathcal{Y})$
  - Discrete: Finite-alphabet
  - Memoryless: When used over n transmissions with message M and input X^n,

$$p(y_i|x^i, y^{i-1}, m) = p_{Y|X}(y_i|x_i)$$

  When used without feedback, $p(y^n|x^n, m) = \prod_{i=1}^{n} p_{Y|X}(y_i|x_i)$

- A $(2^{nR}, n)$ code for the DMC:
  - Message set $[1:2^{nR}] = \{1, 2, \ldots, 2^{\lceil nR \rceil}\}$
  - Encoder: a codeword $x^n(m)$ for each $m \in [1:2^{nR}]$
  - Decoder: an estimate $\hat{m}(y^n) \in [1:2^{nR}] \cup \{e\}$ for each y^n



- Assume $M \sim \mathrm{Unif}[1:2^{nR}]$
- Average probability of error: $P_e^{(n)} = \mathsf{P}\{\hat{M} \neq M\}$
- Assume cost $b(x) \geq 0$ with $b(x_0) = 0$
- Average cost constraint:

$$\sum_{i=1}^{n} b(x_i(m)) \leq nB \quad \text{for every } m \in [1:2^{nR}]$$

- R achievable if $\exists$ $(2^{nR}, n)$ codes that satisfy the cost constraint with $\lim_{n\to\infty} P_e^{(n)} = 0$
- Capacity–cost function $C(B)$ of the DMC $p(y|x)$ with average cost constraint B on X is the supremum of all achievable rates

Channel Coding Theorem (Shannon 1948)

$$C(B) = \max_{p(x):\mathsf{E}(b(X))\leq B} I(X;Y)$$

Proof of Achievability

- We use random coding and joint typicality decoding

- Codebook generation:
  - Fix $p(x)$ that attains $C(B/(1+\epsilon))$
  - Randomly and independently generate 2^{nR} sequences $x^n(m) \sim \prod_{i=1}^n p_X(x_i)$, $m \in [1:2^{nR}]$

- Encoding:
  - To send message m, the encoder transmits $x^n(m)$ if $x^n(m) \in \mathcal{T}_\epsilon^{(n)}$ (by the typical average lemma, $\sum_{i=1}^n b(x_i(m)) \le nB$)
  - Otherwise it transmits $(x_0, \dots, x_0)$

- Decoding:
  - Decoder declares that $\hat{m}$ is sent if it is unique message such that $(x^n(\hat{m}), y^n) \in \mathcal{T}_\epsilon^{(n)}$
  - Otherwise it declares an error

Analysis of the Probability of Error

- Consider the probability of error $\mathsf{P}(\mathcal{E})$ averaged over M and codebooks
- Assume $M = 1$ (symmetry of codebook generation)
- The decoder makes an error iff one or both of the following events occur:

$$\begin{aligned} \mathcal{E}_1 &= \{(X^n(1), Y^n) \notin \mathcal{T}_\epsilon^{(n)}\} \\ \mathcal{E}_2 &= \{(X^n(m), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m \neq 1\} \end{aligned}$$

Thus, by the union of events bound

$$\begin{aligned} \mathsf{P}(\mathcal{E}) &= \mathsf{P}(\mathcal{E}|M = 1) \\ &= \mathsf{P}(\mathcal{E}_1 \cup \mathcal{E}_2) \\ &\leq \mathsf{P}(\mathcal{E}_1) + \mathsf{P}(\mathcal{E}_2) \end{aligned}$$

Analysis of the Probability of Error

- Consider the first term

$$
\begin{aligned}
\mathsf{P}(\mathcal{E}_1) &= \mathsf{P}\{(X^n(1), Y^n) \notin \mathcal{T}_\epsilon^{(n)}\} \\
&= \mathsf{P}\{X^n(1) \in \mathcal{T}_\epsilon^{(n)}, (X^n(1), Y^n) \notin \mathcal{T}_\epsilon^{(n)}\} + \mathsf{P}\{X^n(1) \notin \mathcal{T}_\epsilon^{(n)}, (X^n(1), Y^n) \notin \mathcal{T}_\epsilon^{(n)}\} \\
&\le \sum_{x^n \in \mathcal{T}_\epsilon^{(n)}} \prod_{i=1}^{n} p_X(x_i) \sum_{y^n \notin \mathcal{T}_\epsilon^{(n)}(Y|x^n)} \prod_{i=1}^{n} p_{Y|X}(y_i|x_i) + \mathsf{P}\{X^n(1) \notin \mathcal{T}_\epsilon^{(n)}\} \\
&\le \sum_{(x^n, y^n) \notin \mathcal{T}_\epsilon^{(n)}} \prod_{i=1}^{n} p_X(x_i) p_{Y|X}(y_i|x_i) + \mathsf{P}\{X^n(1) \notin \mathcal{T}_\epsilon^{(n)}\}
\end{aligned}
$$

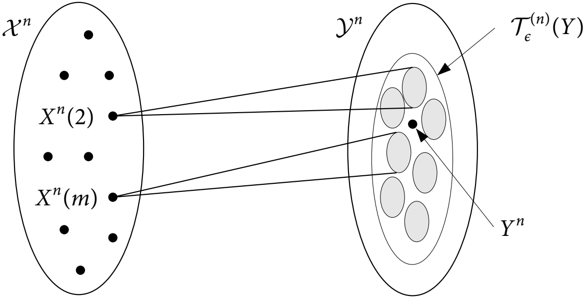
By the LLN, each term $\to 0$ as $n \to \infty$

Analysis of the Probability of Error

- Consider the second term

$$\mathsf{P}(\mathcal{E}_2) = \mathsf{P}\{(X^n(m), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m \neq 1\}$$

For $m \neq 1$, $X^n(m) \sim \prod_{i=1}^n p_X(x_i)$, independent of $Y^n \sim \prod_{i=1}^n p_Y(y_i)$



- To bound $\mathsf{P}(\mathcal{E}_2)$, we use the packing lemma

Packing Lemma

- Let $(U, X, Y) \sim p(u, x, y)$
- Let $(\tilde{U}^n, \tilde{Y}^n) \sim p(\tilde{u}^n, \tilde{y}^n)$ be arbitrarily distributed
- Let $X^n(m) \sim \prod_{i=1}^n p_{X|U}(x_i|\tilde{u}_i)$, $m \in \mathcal{A}$, where $|\mathcal{A}| \leq 2^{nR}$, be pairwise conditionally independent of $\tilde{Y}^n$ given $\tilde{U}^n$

Packing Lemma

There exists $\delta(\epsilon) \to 0$ as $\epsilon \to 0$ such that

$$\lim_{n\to\infty} \mathrm{P}\{(\tilde{U}^n, X^n(m), \tilde{Y}^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m \in \mathcal{A}\} = 0,$$

if $R < I(X; Y|U) - \delta(\epsilon)$

Analysis of the Probability of Error

- Consider the second term

  $$\mathsf{P}(\mathcal{E}_2) = \mathsf{P}\{(X^n(m), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m \neq 1\}$$

  For $m \neq 1$, $X^n(m) \sim \prod_{i=1}^n p_X(x_i)$, independent of $Y^n \sim \prod_{i=1}^n p_Y(y_i)$

- Hence, by the packing lemma with $\mathcal{A} = [2:2^{nR}]$ and $U = \emptyset$, $\mathsf{P}(\mathcal{E}_2) \to 0$ if

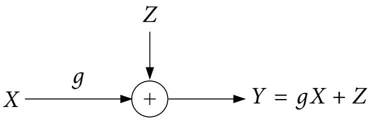
  $$R < I(X;Y) - \delta(\epsilon) = C(B/(1+\epsilon)) - \delta(\epsilon)$$

- Since $\mathsf{P}(\mathcal{E}) \to 0$ as $n \to \infty$, there must exist a sequence of $(2^{nR}, n)$ codes with $\lim_{n\to\infty} P_e^{(n)} = 0$ if $R < C(B/(1+\epsilon)) - \delta(\epsilon)$

- By the continuity of $C(B)$ in B, $C(B/(1+\epsilon)) \to C(B)$ as $\epsilon \to 0$, which implies the achievability of every rate $R < C(B)$

Gaussian Channel

- Discrete-time additive white Gaussian noise channel



  - g: channel gain (path loss)
  - $\{Z_i\}$: WGN($N_0/2$) process, independent of M

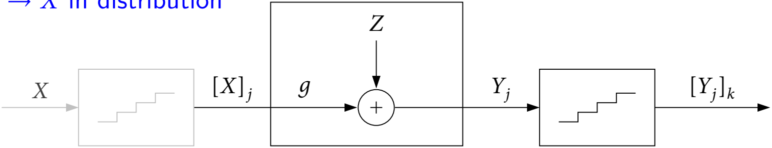
- Average power constraint: $\sum_{i=1}^{n} x_i^2(m) \le nP$ for every m
  - Assume $N_0/2 = 1$ and label received power g^2P as S (SNR)

Theorem (Shannon 1948)

$$C = \max_{F(x):\mathrm{E}(X^2)\le P} I(X;Y) = \frac{1}{2}\log(1+S)$$

Proof of Achievability

- We extend the proof for DMC using a discretization procedure (McEliece 1977)
- First note that the capacity is attained by $X \sim \mathrm{N}(0, P)$, i.e., $I(X; Y) = C$
- Let $[X]_j$ be a finite quantization of X such that $\mathsf{E}([X]_j^2) \le \mathsf{E}(X^2) = P$ and $[X]_j \to X$ in distribution



- Let $Y_j = g[X]_j + Z$ and $[Y_j]_k$ be a finite quantization of Y_j
- By the achievability proof for the DMC, $I([X]_j; [Y_j]_k)$ is achievable for every j, k
- By the data processing inequality and the maximum differential entropy lemma,

$$I([X]_j; [Y_j]_k) \le I([X]_j; Y_j) = h(Y_j) - h(Z) \le h(Y) - h(Z) = I(X; Y)$$

- By the weak convergence and the dominated convergence theorem,

$$\liminf_{j\to\infty} \lim_{k\to\infty} I([X]_j; [Y_j]_k) = \liminf_{j\to\infty} I([X]_j; Y_j) \ge I(X; Y)$$

- Combining the two bounds $I([X]_j; [Y_j]_k) \to I(X; Y)$ as $j, k \to \infty$

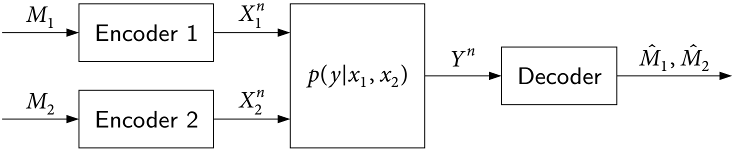
Summary

1. Typical Sequences
2. Point-to-Point Communication
3. Multiple Access Channel
4. Broadcast Channel
5. Lossy Source Coding
6. Wyner–Ziv Coding
7. Gelfand–Pinsker Coding
8. Wiretap Channel
9. Relay Channel
10. Multicast Network

- Random coding
- Joint typicality decoding
- Packing lemma
- Discretization procedure for Gaussian

DM Multiple Access Channel (MAC)

- Multiple access communication system (uplink)



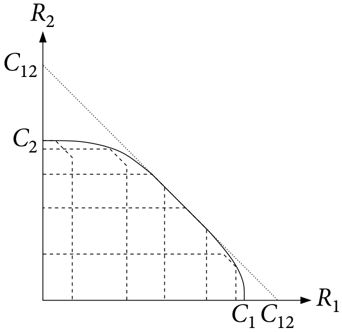
- Assume a 2-sender DM-MAC model $(\mathcal{X}_1 \times \mathcal{X}_2, p(y|x_1, x_2), \mathcal{Y})$
- A $(2^{nR_1}, 2^{nR_2}, n)$ code for the DM-MAC:
  - Message sets: $[1:2^{nR_1}]$ and $[1:2^{nR_2}]$
  - Encoder $j = 1, 2$: $x_j^n(m_j)$
  - Decoder: $(\hat{m}_1(y^n), \hat{m}_2(y^n))$
- Assume $(M_1, M_2) \sim \text{Unif}([1:2^{nR_1}] \times [1:2^{nR_2}])$: $x_1^n(M_1)$ and $x_2^n(M_2)$ independent
- Average probability of error: $P_e^{(n)} = \mathsf{P}\{(\hat{M}_1, \hat{M}_2) \neq (M_1, M_2)\}$
- (R_1, R_2) achievable: if $\exists$ $(2^{nR_1}, 2^{nR_2}, n)$ codes with $\lim_{n\to\infty} P_e^{(n)} = 0$
- Capacity region: closure of the set of achievable (R_1, R_2)

Theorem (Ahlswede 1971, Liao 1972, Slepian–Wolf 1973b)

Capacity region of DM-MAC $p(y|x_1, x_2)$ is the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\le I(X_1; Y | X_2, Q), \\ R_2 &\le I(X_2; Y | X_1, Q), \\ R_1 + R_2 &\le I(X_1, X_2; Y | Q) \end{aligned}$$

for some pmf $p(q)p(x_1|q)p(x_2|q)$, where Q is an auxiliary (time-sharing) r.v.



- Individual capacities:
  $C_1 = \max_{p(x_1),\, x_2} I(X_1; Y | X_2 = x_2)$
  $C_2 = \max_{p(x_2),\, x_1} I(X_2; Y | X_1 = x_1)$
- Sum-capacity:
  $C_{12} = \max_{p(x_1)p(x_2)} I(X_1, X_2; Y)$

Proof of Achievability (Han–Kobayashi 1981)

- We use simultaneous decoding and coded time sharing
- Codebook generation:
  - Fix $p(q)p(x_1|q)p(x_2|q)$
  - Randomly generate a time-sharing sequence $q^n \sim \prod_{i=1}^n p_Q(q_i)$
  - Randomly and conditionally independently generate 2^{nR_1} sequences $x_1^n(m_1) \sim \prod_{i=1}^n p_{X_1|Q}(x_{1i}|q_i)$, $m_1 \in [1:2^{nR_1}]$
  - Similarly generate 2^{nR_2} sequences $x_2^n(m_2) \sim \prod_{i=1}^n p_{X_2|Q}(x_{2i}|q_i)$, $m_2 \in [1:2^{nR_2}]$
- Encoding:
  - To send (m_1, m_2), transmit $x_1^n(m_1)$ and $x_2^n(m_2)$
- Decoding:
  - Find the unique message pair $(\hat{m}_1, \hat{m}_2)$ such that $(q^n, x_1^n(\hat{m}_1), x_2^n(\hat{m}_2), y^n) \in \mathcal{T}_\epsilon^{(n)}$

Analysis of the Probability of Error

- Assume $(M_1, M_2) = (1, 1)$
- Joint pmfs induced by different (m_1, m_2)

m_1	m_2	Joint pmf
1	1	$p(q^n)p(x_1^n \vert q^n)p(x_2^n \vert q^n)p(y^n \vert x_1^n, x_2^n, q^n)$
$*$	1	$p(q^n)p(x_1^n \vert q^n)p(x_2^n \vert q^n)p(y^n \vert x_2^n, q^n)$
1	$*$	$p(q^n)p(x_1^n \vert q^n)p(x_2^n \vert q^n)p(y^n \vert x_1^n, q^n)$
$*$	$*$	$p(q^n)p(x_1^n \vert q^n)p(x_2^n \vert q^n)p(y^n \vert q^n)$

- We divide the error events into the following 4 events:

$$\mathcal{E}_1 = \{(Q^n, X_1^n(1), X_2^n(1), Y^n) \notin \mathcal{T}_\epsilon^{(n)}\}$$

$$\mathcal{E}_2 = \{(Q^n, X_1^n(m_1), X_2^n(1), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_1 \neq 1\}$$

$$\mathcal{E}_3 = \{(Q^n, X_1^n(1), X_2^n(m_2), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_2 \neq 1\}$$

$$\mathcal{E}_4 = \{(Q^n, X_1^n(m_1), X_2^n(m_2), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_1 \neq 1, m_2 \neq 1\}$$

- Then $\mathrm{P}(\mathcal{E}) \leq \mathrm{P}(\mathcal{E}_1) + \mathrm{P}(\mathcal{E}_2) + \mathrm{P}(\mathcal{E}_3) + \mathrm{P}(\mathcal{E}_4)$

m_1	m_2	Joint pmf
1	1	$p(q^n)p(x_1^n\|q^n)p(x_2^n\|q^n)p(y^n\|x_1^n,x_2^n,q^n)$
$*$	1	$p(q^n)p(x_1^n\|q^n)p(x_2^n\|q^n)p(y^n\|x_2^n,q^n)$
1	$*$	$p(q^n)p(x_1^n\|q^n)p(x_2^n\|q^n)p(y^n\|x_1^n,q^n)$
$*$	$*$	$p(q^n)p(x_1^n\|q^n)p(x_2^n\|q^n)p(y^n\|q^n)$

$$\mathcal{E}_1 = \{(Q^n, X_1^n(1), X_2^n(1), Y^n) \notin \mathcal{T}_\epsilon^{(n)}\}$$

$$\mathcal{E}_2 = \{(Q^n, X_1^n(m_1), X_2^n(1), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_1 \neq 1\}$$

$$\mathcal{E}_3 = \{(Q^n, X_1^n(1), X_2^n(m_2), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_2 \neq 1\}$$

$$\mathcal{E}_4 = \{(Q^n, X_1^n(m_1), X_2^n(m_2), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_1 \neq 1, m_2 \neq 1\}$$

- By the LLN, $\mathrm{P}(\mathcal{E}_1) \to 0$ as $n \to \infty$
- By the packing lemma ($\mathcal{A} = [2:2^{nR_1}]$, $U \leftarrow Q$, $X \leftarrow X_1$, $Y \leftarrow (X_2, Y)$), $\mathrm{P}(\mathcal{E}_2) \to 0$ as $n \to \infty$ if $R_1 < I(X_1; X_2, Y|Q) - \delta(\epsilon) = I(X_1; Y|X_2, Q) - \delta(\epsilon)$
- Similarly, $\mathrm{P}(\mathcal{E}_3) \to 0$ as $n \to \infty$ if $R_2 < I(X_2; Y|X_1, Q) - \delta(\epsilon)$

Packing Lemma

- Let $(U, X, Y) \sim p(u, x, y)$
- Let $(\tilde{U}^n, \tilde{Y}^n) \sim p(\tilde{u}^n, \tilde{y}^n)$ be arbitrarily distributed
- Let $X^n(m) \sim \prod_{i=1}^n p_{X|U}(x_i|\tilde{u}_i)$, $m \in \mathcal{A}$, where $|\mathcal{A}| \le 2^{nR}$, be pairwise conditionally independent of $\tilde{Y}^n$ given $\tilde{U}^n$

Packing Lemma

There exists $\delta(\epsilon) \to 0$ as $\epsilon \to 0$ such that

$$\lim_{n\to\infty} \mathsf{P}\{(\tilde{U}^n, X^n(m), \tilde{Y}^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m \in \mathcal{A}\} = 0,$$

if $R < I(X; Y|U) - \delta(\epsilon)$

m_1	m_2	Joint pmf
1	1	$p(q^n)p(x_1^n \vert q^n)p(x_2^n \vert q^n)p(y^n \vert x_1^n, x_2^n, q^n)$
$*$	1	$p(q^n)p(x_1^n \vert q^n)p(x_2^n \vert q^n)p(y^n \vert x_2^n, q^n)$
1	$*$	$p(q^n)p(x_1^n \vert q^n)p(x_2^n \vert q^n)p(y^n \vert x_1^n, q^n)$
$*$	$*$	$p(q^n)p(x_1^n \vert q^n)p(x_2^n \vert q^n)p(y^n \vert q^n)$

$$\mathcal{E}_1 = \{(Q^n, X_1^n(1), X_2^n(1), Y^n) \notin \mathcal{T}_\epsilon^{(n)}\}$$
$$\mathcal{E}_2 = \{(Q^n, X_1^n(m_1), X_2^n(1), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_1 \neq 1\}$$
$$\mathcal{E}_3 = \{(Q^n, X_1^n(1), X_2^n(m_2), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_2 \neq 1\}$$
$$\mathcal{E}_4 = \{(Q^n, X_1^n(m_1), X_2^n(m_2), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_1 \neq 1, m_2 \neq 1\}$$

- By the LLN, $\mathrm{P}(\mathcal{E}_1) \to 0$ as $n \to \infty$
- By the packing lemma ($\mathcal{A} = [2:2^{nR_1}]$, $U \leftarrow Q$, $X \leftarrow X_1$, $Y \leftarrow (X_2, Y)$), $\mathrm{P}(\mathcal{E}_2) \to 0$ as $n \to \infty$ if $R_1 < I(X_1; X_2, Y|Q) - \delta(\epsilon) = I(X_1; Y|X_2, Q) - \delta(\epsilon)$
- Similarly, $\mathrm{P}(\mathcal{E}_3) \to 0$ as $n \to \infty$ if $R_2 < I(X_2; Y|X_1, Q) - \delta(\epsilon)$

m_1	m_2	Joint pmf
1	1	$p(q^n)p(x_1^n\|q^n)p(x_2^n\|q^n)p(y^n\|x_1^n,x_2^n,q^n)$
$*$	1	$p(q^n)p(x_1^n\|q^n)p(x_2^n\|q^n)p(y^n\|x_2^n,q^n)$
1	$*$	$p(q^n)p(x_1^n\|q^n)p(x_2^n\|q^n)p(y^n\|x_1^n,q^n)$
$*$	$*$	$p(q^n)p(x_1^n\|q^n)p(x_2^n\|q^n)p(y^n\|q^n)$

$$\mathcal{E}_1 = \{(Q^n, X_1^n(1), X_2^n(1), Y^n) \notin \mathcal{T}_\epsilon^{(n)}\}$$

$$\mathcal{E}_2 = \{(Q^n, X_1^n(m_1), X_2^n(1), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_1 \neq 1\}$$

$$\mathcal{E}_3 = \{(Q^n, X_1^n(1), X_2^n(m_2), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_2 \neq 1\}$$

$$\mathcal{E}_4 = \{(Q^n, X_1^n(m_1), X_2^n(m_2), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_1 \neq 1, m_2 \neq 1\}$$

- By the packing lemma ($\mathcal{A} = [2:2^{nR_1}] \times [2:2^{nR_2}]$, $U \leftarrow Q$, $X \leftarrow (X_1, X_2)$), $\mathrm{P}(\mathcal{E}_4) \to 0$ as $n \to \infty$ if $R_1 + R_2 < I(X_1, X_2; Y|Q) - \delta(\epsilon)$
- Remark: $(X_1^n(m_1), X_2^n(m_2))$, $m_1 \neq 1$, $m_2 \neq 1$, are not mutually independent but each of them is pairwise independent of Y^n (given Q^n)

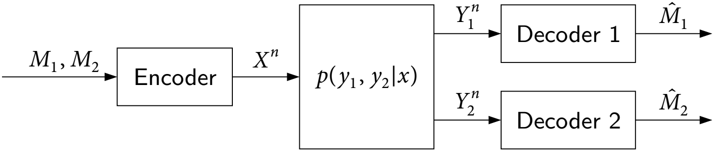
Summary

1. Typical Sequences
2. Point-to-Point Communication
3. Multiple Access Channel
4. Broadcast Channel
5. Lossy Source Coding
6. Wyner–Ziv Coding
7. Gelfand–Pinsker Coding
8. Wiretap Channel
9. Relay Channel
10. Multicast Network

- Coded time sharing
- Simultaneous decoding
- Systematic procedure for decomposing error event

DM Broadcast Channel (BC)

- Broadcast communication system (downlink)



- Assume a 2-receiver DM-BC model $(\mathcal{X}, p(y_1, y_2|x), \mathcal{Y}_1 \times \mathcal{Y}_2)$
- A $(2^{nR_1}, 2^{nR_2}, n)$ code for the DM-BC:
  - Message sets: $[1:2^{nR_1}]$ and $[1:2^{nR_2}]$
  - Encoder: $x^n(m_1, m_2)$
  - Decoder $j = 1, 2$: $\hat{m}_j(y_j^n)$
- Assume $(M_1, M_2) \sim \text{Unif}([1:2^{nR_1}] \times [1:2^{nR_2}])$
- Average probability of error: $P_e^{(n)} = \mathsf{P}\{(\hat{M}_1, \hat{M}_2) \neq (M_1, M_2)\}$
- (R_1, R_2) achievable: if $\exists$ $(2^{nR_1}, 2^{nR_2}, n)$ codes with $\lim_{n\to\infty} P_e^{(n)} = 0$
- Capacity region: closure of the set of achievable (R_1, R_2)

Superposition Coding Inner Bound

- Capacity region of the DM-BC is not known in general
- There are several inner and outer bounds tight in some cases

Superposition Coding Inner Bound (Cover 1972, Bergmans 1973)

A rate pair (R_1, R_2) is achievable for the DM-BC $p(y_1, y_2|x)$ if

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

for some pmf $p(u, x)$, where U is an auxiliary random variable

- This bound is tight for several special cases, including
  - Degraded: $X \to Y_1 \to Y_2$ physically or stochastically
  - Less noisy: $I(U; Y_1) \geq I(U; Y_2)$ for all $p(u, x)$
  - More capable: $I(X; Y_1) \geq I(X; Y_2)$ for all $p(x)$
  - Degraded $\Rightarrow$ Less noisy $\Rightarrow$ More capable

Proof of Achievability

- We use superposition coding and simultaneous nonunique decoding

- Codebook generation:
  - Fix $p(u)p(x|u)$
  - Randomly and independently generate 2^{nR_2} sequences (cloud centers) $u^n(m_2) \sim \prod_{i=1}^n p_U(u_i)$, $m_2 \in [1:2^{nR_2}]$
  - For each $m_2 \in [1:2^{nR_2}]$, randomly and conditionally independently generate 2^{nR_1} sequences (satellite codewords) $x^n(m_1, m_2) \sim \prod_{i=1}^n p_{X|U}(x_i|u_i(m_2))$, $m_1 \in [1:2^{nR_1}]$

- Encoding:
  - To send (m_1, m_2), transmit $x^n(m_1, m_2)$

- Decoding:
  - Decoder 2 finds the unique message $\hat{m}_2$ such that $(u^n(\hat{m}_2), y_2^n) \in \mathcal{T}_\epsilon^{(n)}$ (by the packing lemma, $\mathrm{P}(\mathcal{E}_2) \to 0$ as $n \to \infty$ if $R_2 < I(U;Y_2) - \delta(\epsilon)$)
  - Decoder 1 finds the unique message $\hat{m}_1$ such that

$$(u^n(m_2), x^n(\hat{m}_1, m_2), y_1^n) \in \mathcal{T}_\epsilon^{(n)} \quad \text{for some } m_2$$

Analysis of the Probability of Error for Decoder 1

- Assume $(M_1, M_2) = (1, 1)$
- Joint pmfs induced by different (m_1, m_2)

m_1	m_2	Joint pmf
1	1	$p(u^n, x^n)p(y_1^n \| x^n)$
$*$	1	$p(u^n, x^n)p(y_1^n \| u^n)$
$*$	$*$	$p(u^n, x^n)p(y_1^n)$
1	$*$	$p(u^n, x^n)p(y_1^n)$

- The last case does not result in an error

  So we divide the error event into the following 3 events:

$$\mathcal{E}_{11} = \{(U^n(1), X^n(1,1), Y_1^n) \notin \mathcal{T}_\epsilon^{(n)}\}$$

$$\mathcal{E}_{12} = \{(U^n(1), X^n(m_1,1), Y_1^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_1 \neq 1\}$$

$$\mathcal{E}_{13} = \{(U^n(m_2), X^n(m_1,m_2), Y_1^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_1 \neq 1,\ m_2 \neq 1\}$$

- Then $\mathrm{P}(\mathcal{E}_1) \leq \mathrm{P}(\mathcal{E}_{11}) + \mathrm{P}(\mathcal{E}_{12}) + \mathrm{P}(\mathcal{E}_{13})$

m_1	m_2	Joint pmf
1	1	$p(u^n, x^n)p(y_1^n \| x^n)$
$*$	1	$p(u^n, x^n)p(y_1^n \| u^n)$
$*$	$*$	$p(u^n, x^n)p(y_1^n)$
1	$*$	$p(u^n, x^n)p(y_1^n)$

$$\mathcal{E}_{11} = \{(U^n(1), X^n(1,1), Y_1^n) \notin \mathcal{T}_\epsilon^{(n)}\}$$
$$\mathcal{E}_{12} = \{(U^n(1), X^n(m_1,1), Y_1^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_1 \neq 1\}$$
$$\mathcal{E}_{13} = \{(U^n(m_2), X^n(m_1,m_2), Y_1^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_1 \neq 1,\ m_2 \neq 1\}$$

- By the packing lemma ($\mathcal{A} = [2:2^{nR_1}]$), $\mathrm{P}(\mathcal{E}_{12}) \to 0$ as $n \to \infty$ if $R_1 < I(X;Y_1|U) - \delta(\epsilon)$
- By the packing lemma ($\mathcal{A} = [2:2^{nR_1}] \times [2:2^{nR_2}],\ U \leftarrow \emptyset,\ X \leftarrow (U,X)$), $\mathrm{P}(\mathcal{E}_{13}) \to 0$ as $n \to \infty$ if $R_1 + R_2 < I(U,X;Y_1) - \delta(\epsilon) = I(X;Y_1) - \delta(\epsilon)$
- Remark: $\mathrm{P}(\mathcal{E}_{14}) = \mathrm{P}\{(U^n(m_2), X^n(1,m_2), Y_1^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_2 \neq 1\} \to 0$ as $n \to \infty$ if $R_2 < I(U,X;Y_1) - \delta(\epsilon) = I(X;Y_1) - \delta(\epsilon)$

  Hence, the inner bound continues to hold when decoder 1 is also to recover M_2

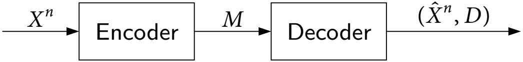
Summary

1. Typical Sequences
2. Point-to-Point Communication
3. Multiple Access Channel
4. Broadcast Channel
5. Lossy Source Coding
6. Wyner–Ziv Coding
7. Gelfand–Pinsker Coding
8. Wiretap Channel
9. Relay Channel
10. Multicast Network

- Superposition coding
- Simultaneous nonunique decoding

Lossy Source Coding

- Point-to-point compression system



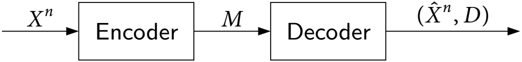
- Assume a discrete memoryless source (DMS) $(\mathcal{X}, p(x))$

  a distortion measure $d(x, \hat{x})$, $(x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}$

- Average per-letter distortion between x^n and $\hat{x}^n$:

$$d(x^n, \hat{x}^n) = \frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i)$$

- A $(2^{nR}, n)$ lossy source code:
  - Encoder: an index $m(x^n) \in [1:2^{nR}) := \{1, 2, \ldots, 2^{\lfloor nR \rfloor}\}$
  - Decoder: an estimate (reconstruction sequence) $\hat{x}^n(m) \in \hat{\mathcal{X}}^n$



- Expected distortion associated with the $(2^{nR}, n)$ code:

$$D = \mathsf{E}\big(d(X^n, \hat{X}^n)\big) = \sum_{x^n} p(x^n) d(x^n, \hat{x}^n(m(x^n)))$$

- (R, D) achievable if $\exists$ $(2^{nR}, n)$ codes with $\limsup_{n\to\infty} \mathsf{E}(d(X^n, \hat{X}^n)) \le D$

- Rate–distortion function $R(D)$: infimum of R such that (R, D) is achievable

Lossy Source Coding Theorem (Shannon 1959)

$$R(D) = \min_{p(\hat{x}|x):\mathsf{E}(d(x,\hat{x}))\le D} I(X; \hat{X})$$

for $D \ge D_{\min} = \mathsf{E}[\min_{\hat{x}(x)} d(X, \hat{x}(X))]$

Proof of Achievability

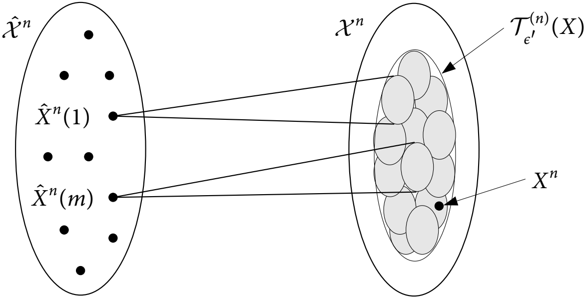
- We use random coding and joint typicality encoding
- Codebook generation:
  - Fix $p(\hat{x}|x)$ that attains $R(D/(1+\epsilon))$ and compute $p(\hat{x}) = \sum_x p(x)p(\hat{x}|x)$
  - Randomly and independently generate sequences $\hat{x}^n(m) \sim \prod_{i=1}^n p_{\hat{X}}(\hat{x}_i)$, $m \in [1:2^{nR}]$
- Encoding:
  - Find an index m such that $(x^n, \hat{x}^n(m)) \in \mathcal{T}_\epsilon^{(n)}$
  - If more than one, choose the smallest index among them
  - If none, choose $m = 1$
- Decoding:
  - Upon receiving m, set the reconstruction sequence $\hat{x}^n = \hat{x}^n(m)$

Analysis of Expected Distortion

- We bound the expected distortion averaged over codebooks
- Define the "encoding error" event

$$\mathcal{E} = \{(X^n, \hat{X}^n(M)) \notin \mathcal{T}_\epsilon^{(n)}\} = \{(X^n, \hat{X}^n(m)) \notin \mathcal{T}_\epsilon^{(n)} \text{ for all } m \in [1:2^{nR}]\}$$

$\hat{X}^n(m) \sim \prod_{i=1}^n p_{\hat{X}}(\hat{x}_i)$, independent of each other and of $X^n \sim \prod_{i=1}^n p_X(x_i)$



- To bound $\mathrm{P}(\mathcal{E})$, we use the covering lemma

Covering Lemma

- Let $(U, X, \hat{X}) \sim p(u, x, \hat{x})$ and $\epsilon' < \epsilon$
- Let $(U^n, X^n) \sim p(u^n, x^n)$ be arbitrarily distributed such that

$$\lim_{n\to\infty} \mathrm{P}\{(U^n, X^n) \in \mathcal{T}_{\epsilon'}^{(n)}(U, X)\} = 1$$

- Let $\hat{X}^n(m) \sim \prod_{i=1}^n p_{\hat{X}|U}(\hat{x}_i|u_i)$, $m \in \mathcal{A}$, where $|\mathcal{A}| \ge 2^{nR}$, be conditionally independent of each other and of X^n given U^n

Covering Lemma

There exists $\delta(\epsilon) \to 0$ as $\epsilon \to 0$ such that

$$\lim_{n\to\infty} \mathrm{P}\{(U^n, X^n, \hat{X}^n(m)) \notin \mathcal{T}_{\epsilon}^{(n)} \text{ for all } m \in \mathcal{A}\} = 0,$$

if $R > I(X; \hat{X}|U) + \delta(\epsilon)$

Analysis of Expected Distortion

- We bound the expected distortion averaged over codebooks
- Define the "encoding error" event

$$\mathcal{E} = \{(X^n, \hat{X}^n(M)) \notin \mathcal{T}_\epsilon^{(n)}\} = \{(X^n, \hat{X}^n(m)) \notin \mathcal{T}_\epsilon^{(n)} \text{ for all } m \in [1:2^{nR}]\}$$

$\hat{X}^n(m) \sim \prod_{i=1}^n p_{\hat{X}}(\hat{x}_i)$, independent of each other and of $X^n \sim \prod_{i=1}^n p_X(x_i)$

- By the covering lemma ($U = \emptyset$), $\mathsf{P}(\mathcal{E}) \to 0$ as $n \to \infty$ if

$$R > I(X;\hat{X}) + \delta(\epsilon) = R(D/(1+\epsilon)) + \delta(\epsilon)$$

- Now, by the law of total expectation and the typical average lemma,

$$\begin{aligned} \mathsf{E}[d(X^n, \hat{X}^n)] &= \mathsf{P}(\mathcal{E})\,\mathsf{E}[d(X^n, \hat{X}^n)|\mathcal{E}] + \mathsf{P}(\mathcal{E}^c)\,\mathsf{E}[d(X^n, \hat{X}^n)|\mathcal{E}^c] \\ &\le \mathsf{P}(\mathcal{E})\ d_{\max} + \mathsf{P}(\mathcal{E}^c)(1+\epsilon)\,\mathsf{E}(d(X,\hat{X})) \end{aligned}$$

- Hence, $\limsup_{n\to\infty} \mathsf{E}[d(X^n, \hat{X}^n)] \le D$ and there must exist a sequence of $(2^{nR}, n)$ codes that satisfies the asymptotic distortion constraint
- By the continuity of $R(D)$ in D, $R(D/(1+\epsilon)) + \delta(\epsilon) \to R(D)$ as $\epsilon \to 0$

Lossless Source Coding

- Suppose we wish to reconstruct X^n losslessly, i.e., $\hat{X}^n = X^n$
- R achievable if $\exists$ $(2^{nR}, n)$ codes with $\lim_{n\to\infty} \mathsf{P}\{\hat{X}^n \neq X^n\} = 0$
- Optimal rate R^*: infimum of achievable R

Lossless Source Coding Theorem (Shannon 1948)

$$R^* = H(X)$$

- We prove this theorem as a corollary of the lossy source coding theorem
- Consider the lossy source coding problem for a DMS X, $\hat{\mathcal{X}} = \mathcal{X}$, and Hamming distortion measure ($d(x,\hat{x}) = 0$ if $x = \hat{x}$, and $d(x,\hat{x}) = 1$ otherwise)
- At $D = 0$, the rate–distortion function is $R(0) = H(X)$
- We now show operationally $R^* = R(0)$ without using the fact that $R^* = H(X)$

Proof of the Lossless Source Coding Theorem

- Proof of $R^* \geq R(0)$:
  - First note that

$$\lim_{n\to\infty} \mathsf{E}(d(X^n, \hat{X}^n)) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathsf{P}\{\hat{X}_i \neq X_i\} \leq \lim_{n\to\infty} \mathsf{P}\{\hat{X}^n \neq X^n\}$$

  - Hence, any sequence of $(2^{nR}, n)$ codes with $\lim_{n\to\infty} \mathsf{P}\{\hat{X}^n \neq X^n\} = 0$ achieves $D = 0$
- Proof of $R^* \leq R(0)$:
  - We can still use random coding and joint typicality encoding!
  - Fix $p(\hat{x}|x) = 1$ if $x = \hat{x}$ and 0 otherwise $\big(p(\hat{x}) = p_X(\hat{x})\big)$
  - As before, generate a random code $\hat{x}^n(m)$, $m \in [1:2^{nR}]$
  - Then $\mathsf{P}(\mathcal{E}) = \mathsf{P}\{(X^n, \hat{X}^n) \notin \mathcal{T}_\epsilon^{(n)}\} \to 0$ as $n \to \infty$ if $R > I(X;\hat{X}) + \delta(\epsilon) = R(0) + \delta(\epsilon)$
  - Now recall that if $(x^n, \hat{x}^n) \in \mathcal{T}_\epsilon^{(n)}$, then $\hat{x}^n = x^n$ (or if $\hat{x}^n \neq x^n$, then $(x^n, \hat{x}^n) \notin \mathcal{T}_\epsilon^{(n)}$)
  - Hence, $\mathsf{P}\{\hat{X}^n \neq X^n\} \to 0$ as $n \to \infty$ if $R > R(0) + \delta(\epsilon)$

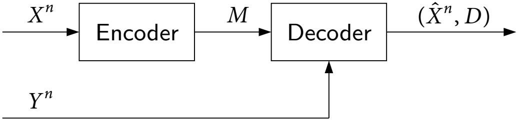
Summary

1. Typical Sequences
2. Point-to-Point Communication
3. Multiple Access Channel
4. Broadcast Channel
5. Lossy Source Coding
6. Wyner–Ziv Coding
7. Gelfand–Pinsker Coding
8. Wiretap Channel
9. Relay Channel
10. Multicast Network

- Joint typicality encoding
- Covering lemma
- Lossless as a corollary of lossy

Lossy Source Coding with Side Information at the Decoder

- Lossy compression system with side information



- Assume a 2-DMS $(\mathcal{X} \times \mathcal{Y}, p(x, y))$ and a distortion measure $d(x, \hat{x})$
- A $(2^{nR}, n)$ lossy source code with side information available at the decoder:
  - Encoder: $m(x^n)$
  - Decoder: $\hat{x}^n(m, y^n)$
- Expected distortion, achievability, rate–distortion function: defined as before

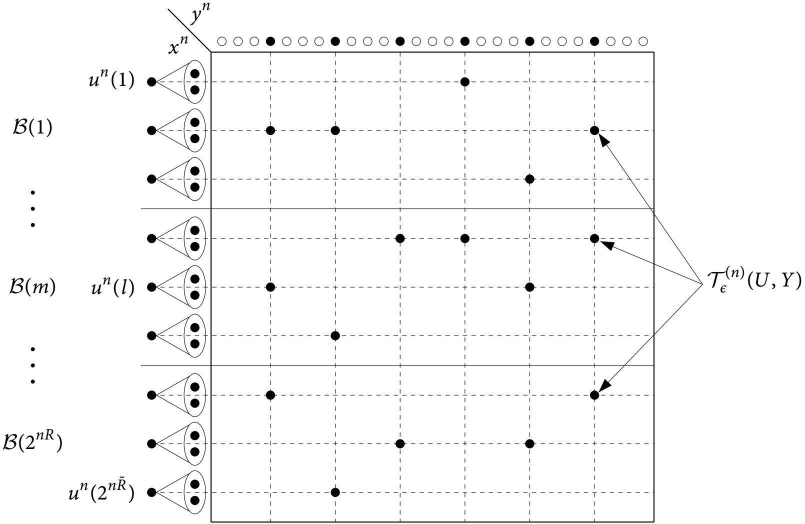
Theorem (Wyner–Ziv 1976)

$$R_{\text{SI-D}}(D) = \min\big(I(X;U) - I(Y;U)\big) = \min I(X;U|Y),$$

where the minimum is over all $p(u|x)$ and $\hat{x}(u, y)$ such that $\mathrm{E}(d(X, \hat{X})) \le D$

Proof of Achievability

- We use binning in addition to joint typicality encoding and decoding



Analysis of Expected Distortion

- We bound the distortion averaged over the random codebook and encoding
- Let (L, M) denote chosen indices and $\hat{L}$ be the index estimate at the decoder
- Define the "error" event

$$\mathcal{E} = \{(U^n(\hat{L}), X^n, Y^n) \notin \mathcal{T}_\epsilon^{(n)}\}$$

and consider

$$\begin{aligned}
\mathcal{E}_1 &= \{(U^n(l), X^n) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } l \in [1:2^{n\tilde{R}}]\} \\
\mathcal{E}_2 &= \{(U^n(L), X^n, Y^n) \notin \mathcal{T}_\epsilon^{(n)}\} \\
\mathcal{E}_3 &= \{(U^n(\tilde{l}), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } \tilde{l} \in \mathcal{B}(M),\ \tilde{l} \neq L\}
\end{aligned}$$

- The probability of "error" is bounded as

$$\mathsf{P}(\mathcal{E}) \le \mathsf{P}(\mathcal{E}_1) + \mathsf{P}(\mathcal{E}_1^c \cap \mathcal{E}_2) + \mathsf{P}(\mathcal{E}_3)$$

$$\mathcal{E}_1 = \{(U^n(l), X^n) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } l \in [1:2^{n\tilde{R}}]\}$$

$$\mathcal{E}_2 = \{(U^n(L), X^n, Y^n) \notin \mathcal{T}_{\epsilon}^{(n)}\}$$

$$\mathcal{E}_3 = \{(U^n(\tilde{l}), Y^n) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } \tilde{l} \in \mathcal{B}(M),\ \tilde{l} \neq L\}$$

$$\mathrm{P}(\mathcal{E}) \leq \mathrm{P}(\mathcal{E}_1) + \mathrm{P}(\mathcal{E}_1^c \cap \mathcal{E}_2) + \mathrm{P}(\mathcal{E}_3)$$

- By the covering lemma, $\mathrm{P}(\mathcal{E}_1) \to 0$ as $n \to \infty$ if $\tilde{R} > I(X;U) + \delta(\epsilon')$
- Since $\mathcal{E}_1^c = \{(U^n(L), X^n) \in \mathcal{T}_{\epsilon'}^{(n)}\}$, $\epsilon > \epsilon'$, and

  $$Y^n \,|\, \{U^n(L) = u^n, X^n = x^n\} \sim \prod_{i=1}^{n} p_{Y|U,X}(y_i|u_i, x_i) = \prod_{i=1}^{n} p_{Y|X}(y_i|x_i),$$

  by the conditional typicality lemma, $\mathrm{P}(\mathcal{E}_1^c \cap \mathcal{E}_2) \to 0$ as $n \to \infty$
- To bound $\mathrm{P}(\mathcal{E}_3)$, it can be shown that

  $$\mathrm{P}(\mathcal{E}_3) \leq \mathrm{P}\{(U^n(\tilde{l}), Y^n) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } \tilde{l} \in \mathcal{B}(1)\}$$

  Since each $U^n(\tilde{l}) \sim \prod_{i=1}^{n} p_U(u_i)$, independent of Y^n,

  by the packing lemma, $\mathrm{P}(\mathcal{E}_3) \to 0$ as $n \to \infty$ if $\tilde{R} - R < I(Y;U) - \delta(\epsilon)$
- Combining the bounds, we have shown that $\mathrm{P}(\mathcal{E}) \to 0$ as $n \to \infty$ if

  $R > I(X;U) - I(Y;U) + \delta(\epsilon) + \delta(\epsilon') = R_{\text{SI-D}}(D/(1+\epsilon)) + \delta(\epsilon) + \delta(\epsilon')$

Lossless Source Coding with Side Information

- What is the minimum rate $R^*_{\text{SI-D}}$ needed to recover X losslessly?

Theorem (Slepian–Wolf 1973a)

$$R^*_{\text{SI-D}} = H(X|Y)$$

- We prove the Slepian–Wolf theorem as a corollary of the Wyner–Ziv theorem
- Let d be the Hamming distortion measure and consider the case $D = 0$
- Then $R_{\text{SI-D}}(0) = H(X|Y)$
- As before, we can show operationally $R^*_{\text{SI-D}} = R_{\text{SI-D}}(0)$
  - $R^*_{\text{SI-D}} \geq R_{\text{SI-D}}(0)$ since $(1/n)\sum_{i=1}^n \mathsf{P}\{\hat{X}_i \neq X_i\} \leq \mathsf{P}\{\hat{X}^n \neq X^n\}$
  - $R^*_{\text{SI-D}} \leq R_{\text{SI-D}}(0)$ by Wyner–Ziv coding with $\hat{X} = U = X$

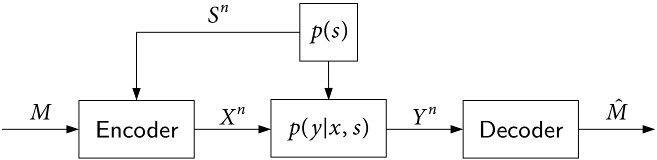
Summary

1. Typical Sequences
2. Point-to-Point Communication
3. Multiple Access Channel
4. Broadcast Channel
5. Lossy Source Coding
6. Wyner–Ziv Coding
7. Gelfand–Pinsker Coding
8. Wiretap Channel
9. Relay Channel
10. Multicast Network

- Binning
- Application of conditional typicality lemma
- Channel coding techniques in source coding

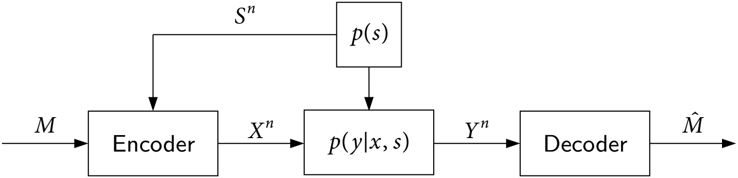
DMC with State Information Available at the Encoder

- Point-to-point communication system with state



- Assume a DMC with DM state model $(\mathcal{X} \times \mathcal{S}, p(y|x,s)p(s), \mathcal{Y})$
  - DMC: $p(y^n|x^n, s^n, m) = \prod_{i=1}^n p_{Y|X,S}(y_i|x_i, s_i)$
  - DM state: $(S_1, S_2, \ldots)$ i.i.d. with $S_i \sim p_S(s_i)$

- A $(2^{nR}, n)$ code for the DMC with state information available at the encoder:
  - Message set: $[1:2^{nR}]$
  - Encoder: $x^n(m, s^n)$
  - Decoder: $\hat{m}(y^n)$



- Expected average cost constraint:

$$\sum_{i=1}^{n} \mathsf{E}[b(x_i(m,S^n))] \le nB \quad \text{for every } m \in [1:2^{nR}]$$

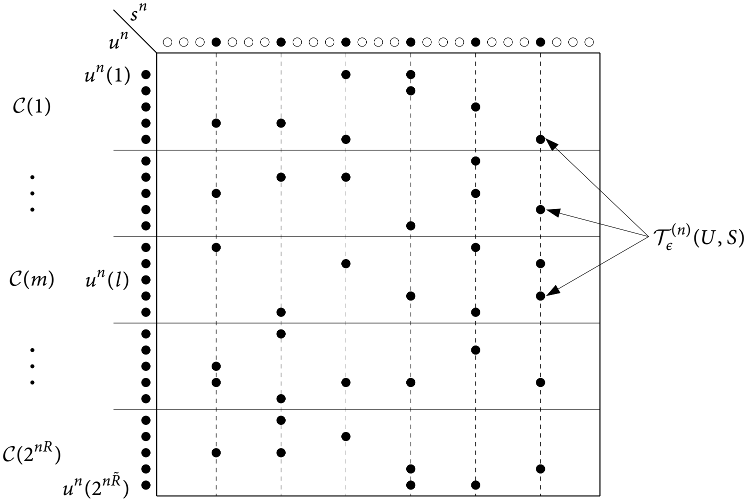
- Probability of error, achievability, capacity–cost function: defined as for DMC

Theorem (Gelfand–Pinsker 1980)

$$C_{\text{SI-E}}(B) = \max_{p(u|s),\, x(u,s):\mathsf{E}(b(X))\le B} \big(I(U;Y) - I(U;S)\big)$$

Proof of Achievability (Heegard–El Gamal 1983)

- We use multicoding



- Codebook generation:

Analysis of the Probability of Error

- Assume $M = 1$
- Let L denote the index of the chosen U^n sequence for $M = 1$ and S^n
- The decoder makes an error only if one or more of the following events occur:

$$\begin{aligned} \mathcal{E}_1 &= \{(U^n(l), S^n) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } U^n(l) \in \mathcal{C}(1)\} \\ \mathcal{E}_2 &= \{(U^n(L), Y^n) \notin \mathcal{T}_{\epsilon}^{(n)}\} \\ \mathcal{E}_3 &= \{(U^n(l), Y^n) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } U^n(l) \notin \mathcal{C}(1)\} \end{aligned}$$

Thus, the probability of error is bounded as

$$\mathsf{P}(\mathcal{E}) \le \mathsf{P}(\mathcal{E}_1) + \mathsf{P}(\mathcal{E}_1^c \cap \mathcal{E}_2) + \mathsf{P}(\mathcal{E}_3)$$

$$\mathcal{E}_1 = \{(U^n(l), S^n) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } U^n(l) \in \mathcal{C}(1)\}$$
$$\mathcal{E}_2 = \{(U^n(L), Y^n) \notin \mathcal{T}_{\epsilon}^{(n)}\}$$
$$\mathcal{E}_3 = \{(U^n(l), Y^n) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } U^n(l) \notin \mathcal{C}(1)\}$$

$$\mathsf{P}(\mathcal{E}) \le \mathsf{P}(\mathcal{E}_1) + \mathsf{P}(\mathcal{E}_1^c \cap \mathcal{E}_2) + \mathsf{P}(\mathcal{E}_3)$$

- By the covering lemma, $\mathsf{P}(\mathcal{E}_1) \to 0$ as $n \to \infty$ if $\tilde{R} - R > I(U;S) + \delta(\epsilon')$
- Since $\epsilon > \epsilon'$, $\mathcal{E}_1^c = \{(U^n(L), S^n) \in \mathcal{T}_{\epsilon'}^{(n)}\} = \{(U^n(L), X^n, S^n) \in \mathcal{T}_{\epsilon'}^{(n)}\}$, and

  $$Y^n | \{U^n(L) = u^n, X^n = x^n, S^n = s^n\} \sim \prod_{i=1}^{n} p_{Y|U,X,S}(y_i | u_i, x_i, s_i) = \prod_{i=1}^{n} p_{Y|X,S}(y_i | x_i, s_i),$$

  by the conditional typicality lemma, $\mathsf{P}(\mathcal{E}_1^c \cap \mathcal{E}_2) \to 0$ as $n \to \infty$
- Since $U^n(l) \notin \mathcal{C}(1)$ is distributed according to $\prod_{i=1}^{n} p(u_i)$, independent of Y^n, by the packing lemma, $\mathsf{P}(\mathcal{E}_3) \to 0$ as $n \to \infty$ if $\tilde{R} < I(U;Y) - \delta(\epsilon)$

  Remark: Y^n is not i.i.d.
- Combining the bounds, we have shown that $\mathsf{P}(\mathcal{E}) \to 0$ as $n \to \infty$ if

  $R < I(U;Y) - I(U;S) - \delta(\epsilon) - \delta(\epsilon') = C_{\text{SI-E}}(B/(1+\epsilon')) - \delta(\epsilon) - \delta(\epsilon')$

Multicoding versus Binning

Multicoding

Channel coding technique

- Given a set of messages
- Generate many codewords for each message
- To communicate a message, send a codeword from its subcodebook

Binning

Source coding technique

- Given a set of indices (sequences)
- Map indices into a smaller number of bins
- To communicate an index, send its bin index

Wyner–Ziv versus Gelfand–Pinsker

- Wyner–Ziv theorem: rate–distortion function for a DMS X with side information Y available at the decoder:

  $$R_{\text{SI-D}}(D) = \min\bigl(I(U;X) - I(U;Y)\bigr)$$

  We proved achievability using binning, covering, and packing

- Gelfand–Pinsker theorem: capacity–cost function of a DMC with state information S available at the encoder:

  $$C_{\text{SI-E}}(B) = \max\bigl(I(U;Y) - I(U;S)\bigr)$$

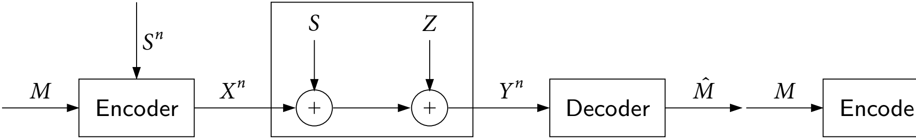
  We proved achievability using multicoding, covering, and packing

- Dualities:

  min ⇔ max
  binning ⇔ multicoding
  covering rate – packing rate ⇔ packing rate – covering rate

Writing on Dirty Paper

- Gaussian channel with additive Gaussian state available at the encoder



  - Noise $Z \sim \mathrm{N}(0, N)$
  - State $S \sim \mathrm{N}(0, Q)$, independent of Z

- Assume expected average power constraint: $\sum_{i=1}^{n} \mathsf{E}(x_i^2(m, S^n)) \le nP$ for every m
- $C = \frac{1}{2} \log\left(1 + \frac{P}{N+Q}\right)$
- $C_{\text{SI-ED}} = \frac{1}{2} \log\left(1 + \frac{P}{N}\right) = C_{\text{SI-D}}$

Writing on Dirty Paper (Costa 1983)

$$C_{\text{SI-E}} = \frac{1}{2} \log\left(1 + \frac{P}{N}\right)$$

Proof of Achievability

- Proof involves a clever choice of $F(u|s)$, $x(u,s)$ and discretization procedure
- Let $X \sim \mathrm{N}(0,P)$ independent of S and $U = X + \alpha S$, where $\alpha = P/(P+N)$. Then

$$I(U;Y) - I(U;S) = \frac{1}{2}\log\left(1+\frac{P}{N}\right)$$

- Let $[U]_j$ and $[S]_{j'}$ be finite quantizations of U and S
- Let $[X]_{jj'} = [U]_j - \alpha[S]_{j'}$ and $[Y_{jj'}]_k$ be a finite quantization of the corresponding channel output $Y_{jj'} = [U]_j - \alpha[S]_{j'} + S + Z$
- We use Gelfand–Pinsker coding for the DMC with DM state $p([y_{jj'}]_k|[x]_{jj'},[s]_{j'})p([s]_{j'})$
    - Joint typicality encoding: $\tilde{R} - R > I(U;S) \geq I([U]_j;[S]_{j'})$
    - Joint typicality decoding: $\tilde{R} < I([U]_j;[Y_{jj'}]_k)$
    - Thus $R < I([U]_j;[Y_{jj'}]_k) - I(U;S)$ is achievable for any j, j', k
- Following similar arguments to the discretization procedure for Gaussian channel coding,

$$\lim_{j\to\infty}\lim_{j'\to\infty}\lim_{k\to\infty} I([U]_j;[Y_{jj'}]_l) = I(U;Y)$$

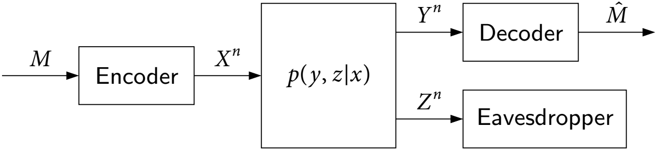
Summary

1. Typical Sequences
2. Point-to-Point Communication
3. Multiple Access Channel
4. Broadcast Channel
5. Lossy Source Coding
6. Wyner–Ziv Coding
7. Gelfand–Pinsker Coding
8. Wiretap Channel
9. Relay Channel
10. Multicast Network

- Multicoding
- Packing lemma with non i.i.d. Y^n
- Writing on dirty paper

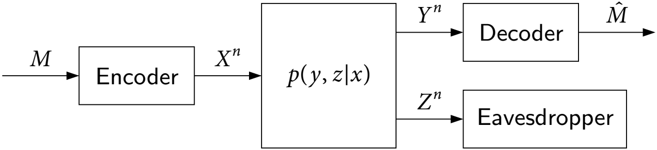
DM Wiretap Channel (WTC)

- Point-to-point communication system with an eavesdropper



- Assume a DM-WTC model $(\mathcal{X}, p(y, z|x), \mathcal{Y} \times \mathcal{Z})$

- A $(2^{nR}, n)$ secrecy code for the DM-WTC:
  - Message set: $[1 : 2^{nR}]$
  - Randomized encoder: $X^n(m) \sim p(x^n|m)$ for each $m \in [1 : 2^{nR}]$
  - Decoder: $\hat{m}(y^n)$



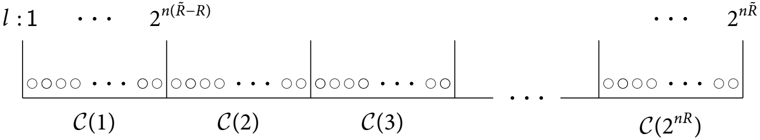
- Assume $M \sim \mathrm{Unif}[1:2^{nR}]$
- Average probability of error: $P_e^{(n)} = \mathsf{P}\{\hat{M} \neq M\}$
- Information leakage rate: $R_{\mathrm{L}}^{(n)} = (1/n) I(M; Z^n)$
- (R, R_{L}) achievable if $\exists$ $(2^{nR}, n)$ codes with $\lim_{n\to\infty} P_e^{(n)} = 0$, $\limsup_{n\to\infty} R_{\mathrm{L}}^{(n)} \le R_{\mathrm{L}}$
- Rate–leakage region $\mathscr{R}^*$: closure of the set of achievable (R, R_{L})
- Secrecy capacity: $C_{\mathrm{S}} = \max\{R: (R, 0) \in \mathscr{R}^*\}$

Theorem (Wyner 1975, Csiszár–Körner 1978)

$$C_{\mathrm{S}} = \max_{p(u,x)} \bigl(I(U; Y) - I(U; Z) \bigr)$$

Proof of Achievability

- We use multicoding and two-step randomized encoding
- Codebook generation:
  - Assume $C_S > 0$ and fix $p(u, x)$ that attains it $(I(U;Y) - I(U;Z) > 0)$
  - For each $m \in [1:2^{nR}]$, generate a subcodebook $\mathcal{C}(m)$ consisting of $2^{n(\tilde{R}-R)}$ randomly and independently generated sequences $u^n(l) \sim \prod_{i=1}^n p_U(u_i)$, $l \in [(m-1)2^{n(\tilde{R}-R)} + 1 : m2^{n(\tilde{R}-R)}]$

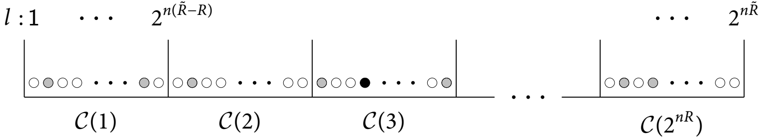


- Encoding:
  - To send m, choose an index $L \in [(m-1)2^{n(\tilde{R}-R)} + 1 : m2^{n(\tilde{R}-R)}]$ uniformly at random
  - Then generate $X^n \sim \prod_{i=1}^n p_{X|U}(x_i | u_i(L))$ and transmit it
- Decoding:
  - Find the unique $\hat{m}$ such that $(u^n(\hat{l}), y^n) \in \mathcal{T}_\epsilon^{(n)}$ for some $u^n(\hat{l}) \in \mathcal{C}(\hat{m})$

    By the LLN and the packing lemma, $\mathrm{P}(\mathcal{E}) \to 0$ as $n \to \infty$ if $\tilde{R} < I(U;Y) - \delta(\epsilon)$

Analysis of the Information Leakage Rate

- For each $\mathcal{C}(m)$, the eavesdropper has $\doteq 2^{n(\tilde{R}-R-I(U;Z))}$ $u^n(l)$ jointly typical with z^n



- If $\tilde{R} - R > I(U;Z)$, the eavesdropper has roughly same number of sequences in each subcodebook, providing it with no information about the message

- Let M be the message sent and L be the randomly selected index

- Every codebook $\mathcal{C}$ induces a pmf of the form

$$p(m, l, u^n, z^n | \mathcal{C}) = 2^{-nR} 2^{-n(\tilde{R}-R)} p(u^n | l, \mathcal{C}) \prod_{i=1}^{n} p_{Z|U}(z_i | u_i)$$

In particular, $p(u^n, z^n) = \prod_{i=1}^{n} p_{U,Z}(u_i, z_i)$

Analysis of the Information Leakage Rate

- Consider the amount of information leakage averaged over codebooks:

$$\begin{aligned} I(M;Z^n|\mathcal{C}) &= H(M|\mathcal{C}) - H(M|Z^n,\mathcal{C}) \\ &= nR - H(M,L|Z^n,\mathcal{C}) + H(L|Z^n,M,\mathcal{C}) \\ &= nR - H(L|Z^n,\mathcal{C}) + H(L|Z^n,M,\mathcal{C}) \end{aligned}$$

- The first equivocation term

$$\begin{aligned} H(L|Z^n,\mathcal{C}) &= H(L|\mathcal{C}) - I(L;Z^n|\mathcal{C}) \\ &= n\tilde{R} - I(L;Z^n|\mathcal{C}) \\ &= n\tilde{R} - I(U^n,L;Z^n|\mathcal{C}) \\ &\geq n\tilde{R} - I(U^n,L,\mathcal{C};Z^n) \\ &\overset{(a)}{=} n\tilde{R} - I(U^n;Z^n) \\ &= n\tilde{R} - nI(U;Z) \end{aligned}$$

(a) $(L,\mathcal{C}) \to U^n \to Z^n$ form a Markov chain

Analysis of the Information Leakage Rate

- Consider the amount of information leakage averaged over codebooks:

$$I(M;Z^n|\mathcal{C}) \le nR - n\tilde{R} + nI(U;Z) + H(L|Z^n,M,\mathcal{C})$$

- The remaining equivocation term can be upper bounded as follows

Lemma

If $\tilde{R} - R \ge I(U;Z)$, then

$$\limsup_{n\to\infty} \frac{1}{n} H(L|Z^n,M,\mathcal{C}) \le \tilde{R} - R - I(U;Z) + \delta(\epsilon)$$

- Substituting (recall that $\tilde{R} < I(U;Y) - \delta(\epsilon)$ for decoding), we have shown that

$$\limsup_{n\to\infty} \frac{1}{n} I(M;Z^n|\mathcal{C}) \le \delta(\epsilon)$$

  if $R < I(U;Y) - I(U;Z) - \delta(\epsilon)$

- Thus, there must exist a sequence of $(2^{nR},n)$ codes such that $P_e^{(n)} \to 0$ and $R_{\mathrm{L}}^{(n)} \le \delta(\epsilon)$ as $n \to \infty$

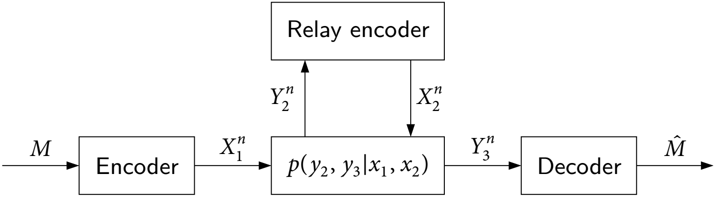
Summary

1. Typical Sequences
2. Point-to-Point Communication
3. Multiple Access Channel
4. Broadcast Channel
5. Lossy Source Coding
6. Wyner–Ziv Coding
7. Gelfand–Pinsker Coding
8. Wiretap Channel
9. Relay Channel
10. Multicast Network

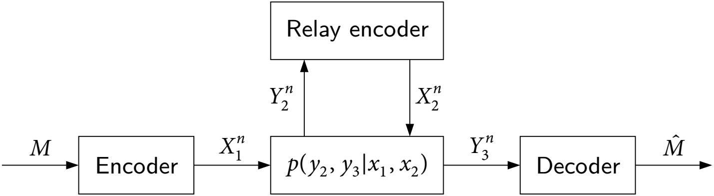
- Randomized encoding
- Bound on equivocation (list size)

DM Relay Channel (RC)

- Point-to-point communication system with a relay



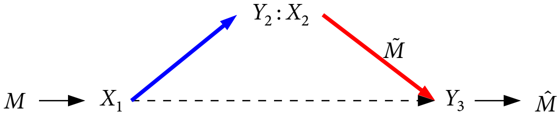
- Assume a DM-RC model $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_2, y_3|x_1, x_2), \mathcal{Y}_2 \times \mathcal{Y}_3)$
- A $(2^{nR}, n)$ code for the DM-RC:
    - Message set: $[1:2^{nR}]$
    - Encoder: $x_1^n(m)$
    - Relay encoder: $x_{2i}(y_2^{i-1})$, $i \in [1:n]$
    - Decoder: $\hat{m}(y_3^n)$
- Probability of error, achievability, capacity: defined as for the DMC



- Capacity of the DM-RC is not known in general
- There are upper and lower bounds that are tight in some cases
- We discuss two lower bounds: decode–forward and compress–forward

Multihop Lower Bound

- The relay recovers the message received from the sender in each block and retransmits it in the following block



Multihop Lower Bound

$$C \ge \max_{p(x_1)p(x_2)} \min\{I(X_2;Y_3),\ I(X_1;Y_2|X_2)\}$$

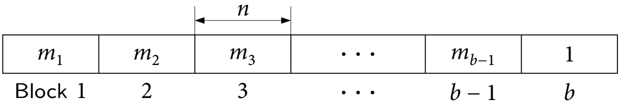
- Tight for a cascade of two DMCs, i.e., $p(y_2, y_3|x_1, x_2) = p(y_2|x_1)p(y_3|x_2)$:

$$C = \min\left\{\max_{p(x_2)} I(X_2;Y_3),\ \max_{p(x_1)} I(X_1;Y_2)\right\}$$

- The scheme uses block Markov coding, where codewords in a block can depend on the message sent in the previous block

Proof of Achievability

- Send $b-1$ messages in b blocks using independently generated codebooks



m_1	m_2	m_3	$\cdots$	m_{b-1}	1
Block 1	2	3	$\cdots$	$b-1$	b

- Codebook generation:
  - Fix $p(x_1)p(x_2)$ that attains the lower bound
  - For each $j \in [1:b]$, randomly and independently generate 2^{nR} sequences $x_1^n(m_j) \sim \prod_{i=1}^n p_{X_1}(x_{1i})$, $m_j \in [1:2^{nR}]$
  - Similarly, generate 2^{nR} sequences $x_2^n(m_{j-1}) \sim \prod_{i=1}^n p_{X_2}(x_{2i})$, $m_{j-1} \in [1:2^{nR}]$
  - Codebooks: $\mathcal{C}_j = \{(x_1^n(m_j), x_2^n(m_{j-1})) : m_{j-1}, m_j \in [1:2^{nR}]\}$, $j \in [1:b]$
- Encoding:
  - To send m_j in block j, transmit $x_1^n(m_j)$ from $\mathcal{C}_j$
- Relay encoding:
  - At the end of block j, find the unique $\tilde{m}_j$ such that $(x_1^n(\tilde{m}_j), x_2^n(\tilde{m}_{j-1}), y_2^n(j)) \in \mathcal{T}_\epsilon^{(n)}$
  - In block $j+1$, transmit $x_2^n(\tilde{m}_j)$ from $\mathcal{C}_{j+1}$
- Decoding:
  - At the end of block $j+1$, find the unique $\hat{m}_j$ such that $(x_2^n(\hat{m}_j), y_3^n(j+1)) \in \mathcal{T}_\epsilon^{(n)}$

Analysis of the Probability of Error

- We analyze the probability of decoding error for M_j averaged over codebooks
- Assume $M_j = 1$
- Let $\tilde{M}_j$ be the relay's decoded message at the end of block j
- Since $\{\hat{M}_j \neq 1\} \subseteq \{\tilde{M}_j \neq 1\} \cup \{\hat{M}_j \neq \tilde{M}_j\}$, the decoder makes an error only if one of the following events occur:

$$\begin{aligned}
\tilde{\mathcal{E}}_1(j) &= \left\{(X_1^n(1), X_2^n(\tilde{M}_{j-1}), Y_2^n(j)) \notin \mathcal{T}_\epsilon^{(n)}\right\} \\
\tilde{\mathcal{E}}_2(j) &= \left\{(X_1^n(m_j), X_2^n(\tilde{M}_{j-1}), Y_2^n(j)) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_j \neq 1\right\} \\
\mathcal{E}_1(j) &= \left\{(X_2^n(\tilde{M}_j), Y_3^n(j+1)) \notin \mathcal{T}_\epsilon^{(n)}\right\} \\
\mathcal{E}_2(j) &= \left\{(X_2^n(m_j), Y_3^n(j+1)) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_j \neq \tilde{M}_j\right\}
\end{aligned}$$

Thus, the probability of error is upper bounded as

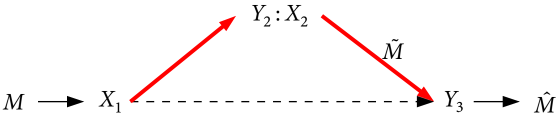
$$\mathsf{P}(\mathcal{E}(j)) = \mathsf{P}\{\hat{M}_j \neq 1\} \leq \mathsf{P}(\tilde{\mathcal{E}}_1(j)) + \mathsf{P}(\tilde{\mathcal{E}}_2(j)) + \mathsf{P}(\mathcal{E}_1(j)) + \mathsf{P}(\mathcal{E}_2(j))$$

$$\tilde{\mathcal{E}}_1(j) = \{(X_1^n(1), X_2^n(\tilde{M}_{j-1}), Y_2^n(j)) \notin \mathcal{T}_\epsilon^{(n)}\}$$
$$\tilde{\mathcal{E}}_2(j) = \{(X_1^n(m_j), X_2^n(\tilde{M}_{j-1}), Y_2^n(j)) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_j \neq 1\}$$
$$\mathcal{E}_1(j) = \{(X_2^n(\tilde{M}_j), Y_3^n(j+1)) \notin \mathcal{T}_\epsilon^{(n)}\}$$
$$\mathcal{E}_2(j) = \{(X_2^n(m_j), Y_3^n(j+1)) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_j \neq \tilde{M}_j\}$$

- By the independence of the codebooks, $\tilde{M}_{j-1}$, which is a function of $Y_2^n(j-1)$ and codebook $\mathcal{C}_{j-1}$, is independent of the codewords $X_1^n(1), X_2^n(\tilde{M}_{j-1})$ in $\mathcal{C}_j$
Thus by the LLN, $\mathsf{P}(\tilde{\mathcal{E}}_1(j)) \to 0$ as $n \to \infty$
- By the packing lemma, $\mathsf{P}(\tilde{\mathcal{E}}_2(j)) \to 0$ as $n \to \infty$ if $R < I(X_1; Y_2|X_2) - \delta(\epsilon)$
- By the independence of the codebooks and the LLN, $\mathsf{P}(\mathcal{E}_1(j)) \to 0$ as $n \to \infty$
- By the same independence and the packing lemma, $\mathsf{P}(\mathcal{E}_2(j)) \to 0$ as $n \to \infty$ if $R < I(X_2; Y_3) - \delta(\epsilon)$
- Thus we have shown that under the given constraints on the rate, $\mathsf{P}\{\hat{M}_j \neq M_j\} \to 0$ as $n \to \infty$ for each $j \in [1:b-1]$

Coherent Multihop Lower Bound

- In the multihop coding scheme, the sender knows what the relay transmits in each block



- Hence, the multihop coding scheme can be improved via coherent cooperation between the sender and the relay

Coherent Multihop Lower Bound

$$C \geq \max_{p(x_1,x_2)} \min\{I(X_2;Y_3),\ I(X_1;Y_2|X_2)\}$$

Proof of Achievability

- We again use a block Markov coding scheme
  - Send $b-1$ messages in b blocks using independently generated codebooks
- Codebook generation:
  - Fix $p(x_1, x_2)$ that attains the lower bound
  - For $j \in [1:b]$, randomly and independently generate 2^{nR} sequences $x_2^n(m_{j-1}) \sim \prod_{i=1}^n p_{X_2}(x_{2i})$, $m_{j-1} \in [1:2^{nR}]$
  - For each $m_{j-1} \in [1:2^{nR}]$, randomly and conditionally independently generate 2^{nR} sequences $x_1^n(m_j|m_{j-1}) \sim \prod_{i=1}^n p_{X_1|X_2}(x_{1i}|x_{2i}(m_{j-1}))$, $m_j \in [1:2^{nR}]$
  - Codebooks: $\mathcal{C}_j = \{(x_1^n(m_j|m_{j-1}), x_2^n(m_{j-1})) : m_{j-1}, m_j \in [1:2^{nR}]\}$, $j \in [1:b]$

Block	1	2	3	…	$b-1$	b
X_1	$x_1^n(m_1\|1)$	$x_1^n(m_2\|m_1)$	$x_1^n(m_3\|m_2)$	…	$x_1^n(m_{b-1}\|m_{b-2})$	$x_1^n(1\|m_{b-1})$
Y_2	$\tilde{m}_1$	$\tilde{m}_2$	$\tilde{m}_3$	…	$\tilde{m}_{b-1}$	$\emptyset$
X_2	$x_2^n(1)$	$x_2^n(\tilde{m}_1)$	$x_2^n(\tilde{m}_2)$	…	$x_2^n(\tilde{m}_{b-2})$	$x_2^n(\tilde{m}_{b-1})$
Y_3	$\emptyset$	$\hat{m}_1$	$\hat{m}_2$	…	$\hat{m}_{b-2}$	$\hat{m}_{b-1}$

- Encoding:
  - In block j, transmit $x_1^n(m_j|m_{j-1})$ from codebook $\mathcal{C}_j$
- Relay encoding:
  - At the end of block j, find the unique $\tilde{m}_j$ such that $(x_1^n(\tilde{m}_j|\tilde{m}_{j-1}), x_2^n(\tilde{m}_{j-1}), y_2^n(j)) \in \mathcal{T}_\epsilon^{(n)}$
  - In block $j+1$, transmit $x_2^n(\tilde{m}_j)$ from codebook $\mathcal{C}_{j+1}$
- Decoding:
  - At the end of block $j+1$, find unique message $\hat{m}_j$ such that $(x_2^n(\hat{m}_j), y_3^n(j+1)) \in \mathcal{T}_\epsilon^{(n)}$

Analysis of the Probability of Error

- We analyze the probability of decoding error for M_j averaged over codebooks
- Assume $M_{j-1} = M_j = 1$
- Let $\tilde{M}_j$ be the relay's decoded message at the end of block j
- The decoder makes an error only if one of the following events occur:

$$\begin{aligned} \tilde{\mathcal{E}}(j) &= \{\tilde{M}_j \neq 1\} \\ \mathcal{E}_1(j) &= \left\{(X_2^n(\tilde{M}_j), Y_3^n(j+1)) \notin \mathcal{T}_\epsilon^{(n)}\right\} \\ \mathcal{E}_2(j) &= \left\{(X_2^n(m_j), Y_3^n(j+1)) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_j \neq \tilde{M}_j\right\} \end{aligned}$$

Thus, the probability of error is upper bounded as

$$\mathrm{P}(\mathcal{E}(j)) = \mathrm{P}\{\hat{M}_j \neq 1\} \leq \mathrm{P}(\tilde{\mathcal{E}}(j)) + \mathrm{P}(\mathcal{E}_1(j)) + \mathrm{P}(\mathcal{E}_2(j))$$

- Following the same steps as in the multihop coding scheme, the last two terms $\to 0$ as $n \to \infty$ if $R < I(X_2; Y_3) - \delta(\epsilon)$

Analysis of the Probability of Error

- To upper bound $\mathsf{P}(\tilde{\mathcal{E}}(j)) = \mathsf{P}\{\tilde{M}_j \neq 1\}$, define

$$\tilde{\mathcal{E}}_1(j) = \{(X_1^n(1|\tilde{M}_{j-1}), X_2^n(\tilde{M}_{j-1}), Y_2^n(j)) \notin \mathcal{T}_\epsilon^{(n)}\}$$
$$\tilde{\mathcal{E}}_2(j) = \{(X_1^n(m_j|\tilde{M}_{j-1}), X_2^n(\tilde{M}_{j-1}), Y_2^n(j)) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_j \neq 1\}$$

- Then

$$\mathsf{P}(\tilde{\mathcal{E}}(j)) \leq \mathsf{P}(\tilde{\mathcal{E}}(j-1)) + \mathsf{P}(\tilde{\mathcal{E}}_1(j) \cap \tilde{\mathcal{E}}^c(j-1)) + \mathsf{P}(\tilde{\mathcal{E}}_2(j))$$

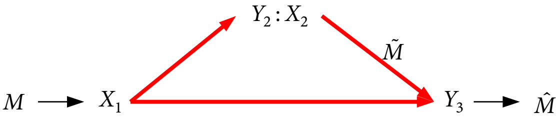
- Consider the second term

$$\begin{aligned}\mathsf{P}(\tilde{\mathcal{E}}_1(j) \cap \tilde{\mathcal{E}}^c(j-1)) &= \mathsf{P}\{(X_1^n(1|\tilde{M}_{j-1}), X_2^n(\tilde{M}_{j-1}), Y_2^n(j)) \notin \mathcal{T}_\epsilon^{(n)}, \tilde{M}_{j-1} = 1\} \\ &\leq \mathsf{P}\{(X_1^n(1|1), X_2^n(1), Y_2^n(j)) \notin \mathcal{T}_\epsilon^{(n)} \,|\, \tilde{M}_{j-1} = 1\},\end{aligned}$$

  which, by the independence of the codebooks and the LLN, $\to 0$ as $n \to \infty$
- By the packing lemma, $\mathsf{P}(\tilde{\mathcal{E}}_2(j)) \to 0$ as $n \to \infty$ if $R < I(X_1; Y_2|X_2) - \delta(\epsilon)$
- Since $\tilde{M}_0 = 1$, by induction, $\mathsf{P}(\tilde{\mathcal{E}}(j)) \to 0$ as $n \to \infty$ for every $j \in [1:b-1]$
- Thus we have shown that under the given constraints on the rate, $\mathsf{P}\{\hat{M}_j \neq M_j\} \to 0$ as $n \to \infty$ for every $j \in [1:b-1]$

Decode–Forward Lower Bound

- Coherent multihop can be further improved by combining the information through the direct path with the information from the relay



Decode–Forward Lower Bound (Cover–El Gamal 1979)

$$C \geq \max_{p(x_1,x_2)} \min\{I(X_1,X_2;Y_3),\ I(X_1;Y_2|X_2)\}$$

- Tight for a physically degraded DM-RC, i.e.,

$$p(y_2,y_3|x_1,x_2) = p(y_2|x_1,x_2)p(y_3|y_2,x_2)$$

Proof of Achievability (Zeng–Kuhlmann–Buzo 1989)

- We use backward decoding (Willems–van der Meulen 1985)
- Codebook generation, encoding, relay encoding:
  - Same as coherent multihop
  - Codebooks: $\mathcal{C}_j = \{(x_1^n(m_j|m_{j-1}), x_2^n(m_{j-1})) : m_{j-1}, m_j \in [1:2^{nR}]\}$, $j \in [1:b]$

Block	1	2	3	$\ldots$	$b-1$	b
X_1	$x_1^n(m_1\|1)$	$x_1^n(m_2\|m_1)$	$x_1^n(m_3\|m_2)$	$\ldots$	$x_1^n(m_{b-1}\|m_{b-2})$	$x_1^n(1\|m_{b-1})$
Y_2	$\tilde{m}_1 \rightarrow$	$\tilde{m}_2 \rightarrow$	$\tilde{m}_3 \rightarrow$	$\ldots$	$\tilde{m}_{b-1}$	$\emptyset$
X_2	$x_2^n(1)$	$x_2^n(\tilde{m}_1)$	$x_2^n(\tilde{m}_2)$	$\ldots$	$x_2^n(\tilde{m}_{b-2})$	$x_2^n(\tilde{m}_{b-1})$
Y_3	$\emptyset$	$\hat{m}_1$	$\leftarrow \hat{m}_2$	$\ldots$	$\leftarrow \hat{m}_{b-2}$	$\leftarrow \hat{m}_{b-1}$

- Decoding:
  - Decoding at the receiver is done backwards after all b blocks are received
  - For $j = b-1, \ldots, 1$, the receiver finds the unique message $\hat{m}_j$ such that $(x_1^n(\hat{m}_{j+1}|\hat{m}_j), x_2^n(\hat{m}_j), y_3^n(j+1)) \in \mathcal{T}_\epsilon^{(n)}$, successively with the initial condition $\hat{m}_b = 1$

Analysis of the Probability of Error

- We analyze the probability of decoding error for M_j averaged over codebooks
- Assume $M_j = M_{j+1} = 1$
- The decoder makes an error only if one or more of the following events occur:

$$\begin{aligned}\tilde{\mathcal{E}}(j) &= \{\tilde{M}_j \neq 1\}\\ \mathcal{E}(j+1) &= \{\hat{M}_{j+1} \neq 1\}\\ \mathcal{E}_1(j) &= \{(X_1^n(\hat{M}_{j+1}|\tilde{M}_j), X_2^n(\tilde{M}_j), Y_3^n(j+1)) \notin \mathcal{T}_\epsilon^{(n)}\}\\ \mathcal{E}_2(j) &= \{(X_1^n(\hat{M}_{j+1}|m_j), X_2^n(m_j), Y_3^n(j+1)) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_j \neq \tilde{M}_j\}\end{aligned}$$

Thus, the probability of error is upper bounded as

$$\begin{aligned}\mathrm{P}(\mathcal{E}(j)) &= \mathrm{P}\{\hat{M}_j \neq 1\}\\ &\le \mathrm{P}(\tilde{\mathcal{E}}(j) \cup \mathcal{E}(j+1) \cup \mathcal{E}_1(j) \cup \mathcal{E}_2(j))\\ &\le \mathrm{P}(\tilde{\mathcal{E}}(j)) + \mathrm{P}(\mathcal{E}(j+1)) + \mathrm{P}(\mathcal{E}_1(j) \cap \tilde{\mathcal{E}}^c(j) \cap \mathcal{E}^c(j+1)) + \mathrm{P}(\mathcal{E}_2(j))\end{aligned}$$

- As in the coherent multihop scheme, the first term → 0 as $n \to \infty$ if $R < I(X_1; Y_2|X_2) - \delta(\epsilon)$

$$\begin{aligned}
\tilde{\mathcal{E}}(j) &= \{\tilde{M}_j \neq 1\} \\
\mathcal{E}(j+1) &= \{\hat{M}_{j+1} \neq 1\} \\
\mathcal{E}_1(j) &= \{(X_1^n(\hat{M}_{j+1} | \tilde{M}_j), X_2^n(\tilde{M}_j), Y_3^n(j+1)) \notin \mathcal{T}_\epsilon^{(n)}\} \\
\mathcal{E}_2(j) &= \{(X_1^n(\hat{M}_{j+1} | m_j), X_2^n(m_j), Y_3^n(j+1)) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_j \neq \tilde{M}_j\}
\end{aligned}$$

$$\mathsf{P}(\mathcal{E}(j)) \le \mathsf{P}(\tilde{\mathcal{E}}(j)) + \mathsf{P}(\mathcal{E}(j+1)) + \mathsf{P}(\mathcal{E}_1(j) \cap \tilde{\mathcal{E}}^c(j) \cap \mathcal{E}^c(j+1)) + \mathsf{P}(\mathcal{E}_2(j))$$

- The third term is upper bounded as

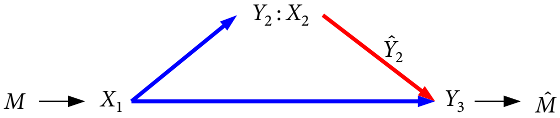
$$\begin{aligned}
&\mathsf{P}\big(\mathcal{E}_1(j) \cap \{\hat{M}_{j+1} = 1\} \cap \{\tilde{M}_j = 1\}\big) \\
&\quad = \mathsf{P}\big\{(X_1^n(1|1), X_2^n(1), Y_3^n(j+1)) \notin \mathcal{T}_\epsilon^{(n)},\ \hat{M}_{j+1} = 1,\ \tilde{M}_j = 1\big\} \\
&\quad \le \mathsf{P}\big\{(X_1^n(1|1), X_2^n(1), Y_3^n(j+1)) \notin \mathcal{T}_\epsilon^{(n)} \,|\, \tilde{M}_j = 1\big\},
\end{aligned}$$

  which, by the independence of the codebooks and the LLN, $\to 0$ as $n \to \infty$

- By the same independence and the packing lemma, the fourth term $\mathsf{P}(\mathcal{E}_2(j)) \to 0$ as $n \to \infty$ if $R < I(X_1, X_2; Y_3) - \delta(\epsilon)$
- Finally for the second term, since $\hat{M}_b = M_b = 1$, by induction, $\mathsf{P}\{\hat{M}_j = M_j\} \to 0$ as $n \to \infty$ for every $j \in [1 : b-1]$ if the given constraints on the rate are satisfied

Compress–Forward Lower Bound

- In the decode–forward coding scheme, the relay recovers the entire message



- If channel from sender to relay is worse than direct channel to receiver, this requirement can reduce rate below that of direct transmission (relay is not used)

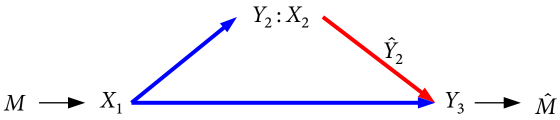
- In the compress–forward coding scheme, the relay helps communication by sending a description of its received sequence to the receiver

Compress–Forward Lower Bound (Cover–El Gamal 1979, El Gamal–Mohseni–Zahedi 2006)

$$C \geq \max_{p(x_1)p(x_2)p(\hat{y}_2|y_2,x_2)} \min\{I(X_1, X_2; Y_3) - I(Y_2; \hat{Y}_2 | X_1, X_2, Y_3),\ I(X_1; \hat{Y}_2, Y_3 | X_2)\}$$

Proof of Achievability

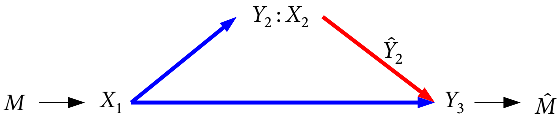
- We use block Markov coding, joint typicality encoding, binning, and simultaneous nonunique decoding



  - At the end of block j, the relay chooses a reconstruction sequence $\hat{y}_2^n(j)$ of the received sequence $y_2^n(j)$
  - Since the receiver has side information $y_3^n(j)$, we use binning to reduce the rate
  - The bin index is sent to the receiver in block $j+1$ via $x_2^n(j+1)$
  - At the end of block $j+1$, the receiver recovers the bin index and then m_j and the compression index simultaneously

Proof of Achievability

- We use block Markov coding, joint typicality encoding, binning, and simultaneous nonunique decoding



- Codebook generation:
  - Fix $p(x_1)p(x_2)p(\hat{y}_2|y_2,x_2)$ that attains the lower bound
  - For $j \in [1:b]$, randomly and independently generate 2^{nR} sequences $x_1^n(m_j) \sim \prod_{i=1}^n p_{X_1}(x_{1i})$, $m_j \in [1:2^{nR}]$
  - Similarly generate 2^{nR_2} sequences $x_2^n(l_{j-1}) \sim \prod_{i=1}^n p_{X_2}(x_{2i})$, $l_{j-1} \in [1:2^{nR_2}]$
  - For each $l_{j-1} \in [1:2^{nR_2}]$, randomly and conditionally independently generate $2^{n\tilde{R}_2}$ sequences $\hat{y}_2^n(k_j|l_{j-1}) \sim \prod_{i=1}^n p_{\hat{Y}_2|X_2}(\hat{y}_{2i}|x_{2i}(l_{j-1}))$, $k_j \in [1:2^{n\tilde{R}_2}]$
  - Codebooks: $\mathcal{C}_j = \{(x_1^n(m_j), x_2^n(l_{j-1})) : m_j \in [1:2^{nR}],\ l_{j-1} \in [1:2^{nR_2}]\}$, $j \in [1:b]$
  - Partition the set $[1:2^{n\tilde{R}_2}]$ into 2^{nR_2} equal-size bins $\mathcal{B}(l_j)$, $l_j \in [1:2^{nR_2}]$

Block	1	2	3	…	$b-1$	b
X_1	$x_1^n(m_1)$	$x_1^n(m_2)$	$x_1^n(m_3)$	…	$x_1^n(m_{b-1})$	$x_1^n(1)$
Y_2	$\hat{y}_2^n(k_1\|1), l_1$	$\hat{y}_2^n(k_2\|l_1), l_2$	$\hat{y}_2^n(k_3\|l_2), l_3$	…	$\hat{y}_2^n(k_{b-1}\|l_{b-2}), l_{b-1}$	$\emptyset$
X_2	$x_2^n(1)$	$x_2^n(l_1)$	$x_2^n(l_2)$	…	$x_2^n(l_{b-2})$	$x_2^n(l_{b-1})$
Y_3	$\emptyset$	$\hat{l}_1, \hat{k}_1, \hat{m}_1$	$\hat{l}_2, \hat{k}_2, \hat{m}_2$	…	$\hat{l}_{b-2}, \hat{k}_{b-2}, \hat{m}_{b-2}$	$\hat{l}_{b-1}, \hat{k}_{b-1}, \hat{m}_{b-1}$

- Encoding:
  - Transmit $x_1^n(m_j)$ from codebook $\mathcal{C}_j$
- Relay encoding:
  - At the end of block j, find an index k_j such that $(y_2^n(j), \hat{y}_2^n(k_j|l_{j-1}), x_2^n(l_{j-1})) \in \mathcal{T}_{\epsilon'}^{(n)}$
  - In block $j+1$, transmit $x_2^n(l_j)$, where l_j is the bin index of k_j
- Decoding:
  - At the end of block $j+1$, find the unique $\hat{l}_j$ such that $(x_2^n(\hat{l}_j), y_3^n(j+1)) \in \mathcal{T}_\epsilon^{(n)}$
  - Find the unique $\hat{m}_j$ such that $(x_1^n(\hat{m}_j), x_2^n(\hat{l}_{j-1}), \hat{y}_2^n(\hat{k}_j|\hat{l}_{j-1}), y_3^n(j)) \in \mathcal{T}_\epsilon^{(n)}$ for some $\hat{k}_j \in \mathcal{B}(\hat{l}_j)$

Analysis of the Probability of Error

- Assume $M_j = 1$ and let L_{j-1}, L_j, K_j denote the indices chosen by the relay
- The decoder makes an error only if one or more of the following events occur:

$$\begin{aligned}
\tilde{\mathcal{E}}(j) &= \left\{(X_2^n(L_{j-1}), \hat{Y}_2^n(k_j|L_{j-1}), Y_2^n(j)) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } k_j \in [1:2^{n\tilde{R}_2}]\right\} \\
\mathcal{E}_1(j-1) &= \{\hat{L}_{j-1} \neq L_{j-1}\} \\
\mathcal{E}_1(j) &= \{\hat{L}_j \neq L_j\} \\
\mathcal{E}_2(j) &= \left\{(X_1^n(1), X_2^n(\hat{L}_{j-1}), \hat{Y}_2^n(K_j|\hat{L}_{j-1}), Y_3^n(j)) \notin \mathcal{T}_{\epsilon}^{(n)}\right\} \\
\mathcal{E}_3(j) &= \left\{(X_1^n(m_j), X_2^n(\hat{L}_{j-1}), \hat{Y}_2^n(K_j|\hat{L}_{j-1}), Y_3^n(j)) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } m_j \neq 1\right\} \\
\mathcal{E}_4(j) &= \Big\{(X_1^n(m_j), X_2^n(\hat{L}_{j-1}), \hat{Y}_2^n(\hat{k}_j|\hat{L}_{j-1}), Y_3^n(j)) \in \mathcal{T}_{\epsilon}^{(n)} \\
&\qquad\qquad \text{for some } \hat{k}_j \in \mathcal{B}(\hat{L}_j), \hat{k}_j \neq K_j, m_j \neq 1\Big\}
\end{aligned}$$

Thus, the probability of error is bounded as

$$\begin{aligned}
\mathsf{P}(\mathcal{E}(j)) &= \mathsf{P}\{\hat{M}_j \neq 1\} \\
&\leq \mathsf{P}(\tilde{\mathcal{E}}(j)) + \mathsf{P}(\mathcal{E}_1(j-1)) + \mathsf{P}(\mathcal{E}_1(j)) + \mathsf{P}(\mathcal{E}_2(j) \cap \tilde{\mathcal{E}}^c(j) \cap \mathcal{E}_1^c(j-1)) \\
&\qquad + \mathsf{P}(\mathcal{E}_3(j)) + \mathsf{P}(\mathcal{E}_4(j) \cap \mathcal{E}_1^c(j-1) \cap \mathcal{E}_1^c(j))
\end{aligned}$$

$$\begin{aligned}
\tilde{\mathcal{E}}(j) &= \{(X_2^n(L_{j-1}), \hat{Y}_2^n(k_j|L_{j-1}), Y_2^n(j)) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } k_j \in [1:2^{n\tilde{R}_2}]\} \\
\mathcal{E}_1(j-1) &= \{\hat{L}_{j-1} \neq L_{j-1}\} \\
\mathcal{E}_1(j) &= \{\hat{L}_j \neq L_j\} \\
\mathcal{E}_2(j) &= \{(X_1^n(1), X_2^n(\hat{L}_{j-1}), \hat{Y}_2^n(K_j|\hat{L}_{j-1}), Y_3^n(j)) \notin \mathcal{T}_{\epsilon}^{(n)}\} \\
\mathcal{E}_3(j) &= \{(X_1^n(m_j), X_2^n(\hat{L}_{j-1}), \hat{Y}_2^n(K_j|\hat{L}_{j-1}), Y_3^n(j)) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } m_j \neq 1\} \\
\mathcal{E}_4(j) &= \{(X_1^n(m_j), X_2^n(\hat{L}_{j-1}), \hat{Y}_2^n(\hat{k}_j|\hat{L}_{j-1}), Y_3^n(j)) \in \mathcal{T}_{\epsilon}^{(n)} \\
&\qquad \text{for some } \hat{k}_j \in \mathcal{B}(\hat{L}_j), \hat{k}_j \neq K_j, m_j \neq 1\}
\end{aligned}$$

$$\begin{aligned}
\mathsf{P}(\mathcal{E}(j)) \le\ & \mathsf{P}(\tilde{\mathcal{E}}(j)) + \mathsf{P}(\mathcal{E}_1(j-1)) + \mathsf{P}(\mathcal{E}_1(j)) + \mathsf{P}(\mathcal{E}_2(j) \cap \tilde{\mathcal{E}}^c(j) \cap \mathcal{E}_1^c(j-1)) \\
& + \mathsf{P}(\mathcal{E}_3(j)) + \mathsf{P}(\mathcal{E}_4(j) \cap \mathcal{E}_1^c(j-1) \cap \mathcal{E}_1^c(j))
\end{aligned}$$

- By the independence of codebooks and the covering lemma ($U \leftarrow X_2$, $X \leftarrow Y_2$, $\hat{X} \leftarrow \hat{Y}_2$), the first term → 0 as $n \to \infty$ if $\tilde{R}_2 > I(\hat{Y}_2; Y_2|X_2) + \delta(\epsilon')$
- As in the multihop coding scheme, the next two terms $\mathsf{P}\{\hat{L}_{j-1} \neq L_{j-1}\} \to 0$ and $\mathsf{P}\{\hat{L}_j \neq L_j\} \to 0$ as $n \to \infty$ if $R_2 < I(X_2; Y_3) - \delta(\epsilon)$
- The fourth term $\le \mathsf{P}\{(X_1^n(1), X_2^n(L_{j-1}), \hat{Y}_2^n(K_j|L_{j-1}), Y_3^n(j)) \notin \mathcal{T}_{\epsilon}^{(n)} | \tilde{\mathcal{E}}^c(j)\} \to 0$ by the independence of codebooks and the conditional typicality lemma

Covering Lemma

- Let $(U, X, \hat{X}) \sim p(u, x, \hat{x})$ and $\epsilon' < \epsilon$
- Let $(U^n, X^n) \sim p(u^n, x^n)$ be arbitrarily distributed such that

$$\lim_{n\to\infty} \mathrm{P}\{(U^n, X^n) \in \mathcal{T}_{\epsilon'}^{(n)}(U, X)\} = 1$$

- Let $\hat{X}^n(m) \sim \prod_{i=1}^n p_{\hat{X}|U}(\hat{x}_i|u_i)$, $m \in \mathcal{A}$, where $|\mathcal{A}| \ge 2^{nR}$, be conditionally independent of each other and of X^n given U^n

Covering Lemma

There exists $\delta(\epsilon) \to 0$ as $\epsilon \to 0$ such that

$$\lim_{n\to\infty} \mathrm{P}\{(U^n, X^n, \hat{X}^n(m)) \notin \mathcal{T}_{\epsilon}^{(n)} \text{ for all } m \in \mathcal{A}\} = 0,$$

if $R > I(X; \hat{X}|U) + \delta(\epsilon)$

$$\begin{aligned}
\tilde{\mathcal{E}}(j) &= \{(X_2^n(L_{j-1}), \hat{Y}_2^n(k_j|L_{j-1}), Y_2^n(j)) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } k_j \in [1:2^{n\tilde{R}_2}]\} \\
\mathcal{E}_1(j-1) &= \{\hat{L}_{j-1} \neq L_{j-1}\} \\
\mathcal{E}_1(j) &= \{\hat{L}_j \neq L_j\} \\
\mathcal{E}_2(j) &= \{(X_1^n(1), X_2^n(\hat{L}_{j-1}), \hat{Y}_2^n(K_j|\hat{L}_{j-1}), Y_3^n(j)) \notin \mathcal{T}_{\epsilon}^{(n)}\} \\
\mathcal{E}_3(j) &= \{(X_1^n(m_j), X_2^n(\hat{L}_{j-1}), \hat{Y}_2^n(K_j|\hat{L}_{j-1}), Y_3^n(j)) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } m_j \neq 1\} \\
\mathcal{E}_4(j) &= \{(X_1^n(m_j), X_2^n(\hat{L}_{j-1}), \hat{Y}_2^n(\hat{k}_j|\hat{L}_{j-1}), Y_3^n(j)) \in \mathcal{T}_{\epsilon}^{(n)} \\
&\qquad \text{for some } \hat{k}_j \in \mathcal{B}(\hat{L}_j), \hat{k}_j \neq K_j, m_j \neq 1\}
\end{aligned}$$

$$\begin{aligned}
\mathsf{P}(\mathcal{E}(j)) \leq\ & \mathsf{P}(\tilde{\mathcal{E}}(j)) + \mathsf{P}(\mathcal{E}_1(j-1)) + \mathsf{P}(\mathcal{E}_1(j)) + \mathsf{P}(\mathcal{E}_2(j) \cap \tilde{\mathcal{E}}^c(j) \cap \mathcal{E}_1^c(j-1)) \\
& + \mathsf{P}(\mathcal{E}_3(j)) + \mathsf{P}(\mathcal{E}_4(j) \cap \mathcal{E}_1^c(j-1) \cap \mathcal{E}_1^c(j))
\end{aligned}$$

- By the independence of codebooks and the covering lemma ($U \leftarrow X_2$, $X \leftarrow Y_2$, $\hat{X} \leftarrow \hat{Y}_2$), the first term $\to 0$ as $n \to \infty$ if $\tilde{R}_2 > I(\hat{Y}_2; Y_2|X_2) + \delta(\epsilon')$
- As in the multihop coding scheme, the next two terms $\mathsf{P}\{\hat{L}_{j-1} \neq L_{j-1}\} \to 0$ and $\mathsf{P}\{\hat{L}_j \neq L_j\} \to 0$ as $n \to \infty$ if $R_2 < I(X_2; Y_3) - \delta(\epsilon)$
- The fourth term $\leq \mathsf{P}\{(X_1^n(1), X_2^n(L_{j-1}), \hat{Y}_2^n(K_j|L_{j-1}), Y_3^n(j)) \notin \mathcal{T}_{\epsilon}^{(n)} | \tilde{\mathcal{E}}^c(j)\} \to 0$ by the independence of codebooks and the conditional typicality lemma

$$\tilde{\mathcal{E}}(j) = \{(X_2^n(L_{j-1}), \hat{Y}_2^n(k_j|L_{j-1}), Y_2^n(j)) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } k_j \in [1:2^{n\tilde{R}_2}]\}$$

$$\mathcal{E}_1(j-1) = \{\hat{L}_{j-1} \neq L_{j-1}\}$$

$$\mathcal{E}_1(j) = \{\hat{L}_j \neq L_j\}$$

$$\mathcal{E}_2(j) = \{(X_1^n(1), X_2^n(\hat{L}_{j-1}), \hat{Y}_2^n(K_j|\hat{L}_{j-1}), Y_3^n(j)) \notin \mathcal{T}_{\epsilon}^{(n)}\}$$

$$\mathcal{E}_3(j) = \{(X_1^n(m_j), X_2^n(\hat{L}_{j-1}), \hat{Y}_2^n(K_j|\hat{L}_{j-1}), Y_3^n(j)) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } m_j \neq 1\}$$

$$\mathcal{E}_4(j) = \{(X_1^n(m_j), X_2^n(\hat{L}_{j-1}), \hat{Y}_2^n(\hat{k}_j|\hat{L}_{j-1}), Y_3^n(j)) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } \hat{k}_j \in \mathcal{B}(\hat{L}_j), \hat{k}_j \neq K_j, m_j \neq 1\}$$

$$\mathsf{P}(\mathcal{E}(j)) \le \mathsf{P}(\tilde{\mathcal{E}}(j)) + \mathsf{P}(\mathcal{E}_1(j-1)) + \mathsf{P}(\mathcal{E}_1(j)) + \mathsf{P}(\mathcal{E}_2(j) \cap \tilde{\mathcal{E}}^c(j) \cap \mathcal{E}_1^c(j-1)) + \mathsf{P}(\mathcal{E}_3(j)) + \mathsf{P}(\mathcal{E}_4(j) \cap \mathcal{E}_1^c(j-1) \cap \mathcal{E}_1^c(j))$$

- By the same independence and the packing lemma, $\mathsf{P}(\mathcal{E}_3(j)) \to 0$ as $n \to \infty$ if $R < I(X_1; X_2, \hat{Y}_2, Y_3) + \delta(\epsilon) = I(X_1; \hat{Y}_2, Y_3|X_2) + \delta(\epsilon)$
- As in Wyner–Ziv coding, the last term $\le \mathsf{P}\{(X_1^n(m_j), X_2^n(L_{j-1}), \hat{Y}_2^n(\hat{k}_j|L_{j-1}), Y_3^n(j)) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } \hat{k}_j \in \mathcal{B}(1), m_j \neq 1\}$, which, by the independence of codebooks, joint typicality lemma, and union bound, $\to 0$ as $n \to \infty$ if $R + \tilde{R}_2 - R_2 < I(X_1; Y_3|X_2) + I(\hat{Y}_2; X_1, Y_3|X_2) - \delta(\epsilon)$

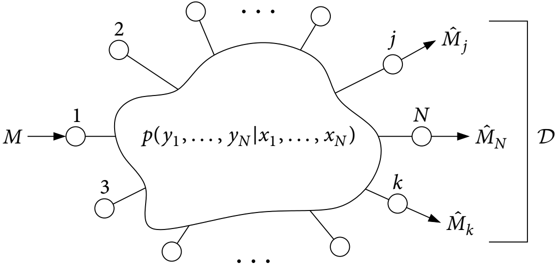
Summary

1. Typical Sequences
2. Point-to-Point Communication
3. Multiple Access Channel
4. Broadcast Channel
5. Lossy Source Coding
6. Wyner–Ziv Coding
7. Gelfand–Pinsker Coding
8. Wiretap Channel
9. Relay Channel
10. Multicast Network

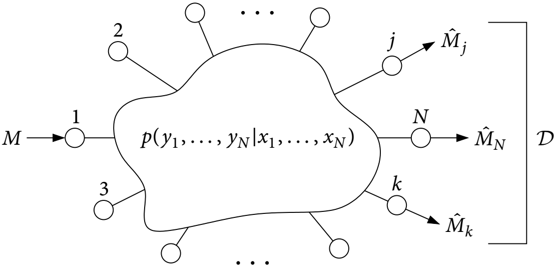
- Block Markov coding
- Coherent cooperation
- Decode–forward
- Backward decoding
- Compress–forward

DM Multicast Network (MN)

- Multicast communication network



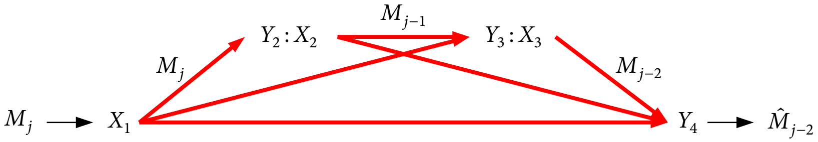
- Assume an N-node DM-MN model $(\bigtimes_{j=1}^N \mathcal{X}_j, p(y^N|x^N), \bigtimes_{j=1}^N \mathcal{Y}_j)$
- Topology of the network is defined through $p(y^N|x^N)$
- A $(2^{nR}, n)$ code for the DM-MN:
  - Message set: $[1:2^{nR}]$
  - Source encoder: $x_{1i}(m, y_1^{i-1})$, $i \in [1:n]$
  - Relay encoder $j \in [2:N]$: $x_{ji}(y_j^{i-1})$, $i \in [1:n]$
  - Decoder $k \in \mathcal{D}$: $\hat{m}_k(y_k^n)$



- Assume $M \sim \text{Unif}[1:2^{nR}]$
- Average probability of error: $P_e^{(n)} = \mathsf{P}\{\hat{M}_k \neq M \text{ for some } k \in \mathcal{D}\}$
- R achievable if there exists a sequence of $(2^{nR}, n)$ codes with $\lim_{n\to\infty} P_e^{(n)} = 0$
- Capacity C: supremum of achievable R
- Special cases:
  - DMC with feedback ($N = 2,\ Y_1 = Y_2,\ X_2 = \emptyset$, and $\mathcal{D} = \{2\}$)
  - DM-RC ($N = 3,\ X_3 = Y_1 = \emptyset$, and $\mathcal{D} = \{3\}$)
  - Common-message DM-BC ($X_2 = \cdots = X_N = Y_1 = \emptyset$ and $\mathcal{D} = [2:N]$)
  - DM unicast network ($\mathcal{D} = \{N\}$)

Network Decode–Forward

- Decode–forward for RC can be extended to MN



Network Decode–Forward Lower Bound (Xie–Kumar 2005, Kramer–Gastpar–Gupta 2005)

$$C \geq \max_{p(x^N)} \min_{k \in [1:N-1]} I(X^k; Y_{k+1} | X_{k+1}^N)$$

- For $N = 3$ and $X_3 = \emptyset$, reduces to the decode–forward lower bound for DM-RC
- Tight for a degraded DM-MN, i.e., $p(y_{k+2}^N | x^N, y^{k+1}) = p(y_{k+2}^N | x_{k+1}^N, y_{k+1})$
- Holds for any $\mathcal{D} \subseteq [2:N]$
- Can be improved by removing some relay nodes and relabeling the nodes

Proof of Achievability

- We use block Markov coding and sliding window decoding (Carleial 1982)
- We illustrate this scheme for DM-RC
- Codebook generation, encoding, and relay encoding: same as before

Block	1	2	3	$\cdots$	$b-1$	b
X_1	$x_1^n(m_1\|1)$	$x_1^n(m_2\|m_1)$	$x_1^n(m_3\|m_2)$	$\cdots$	$x_1^n(m_{b-1}\|m_{b-2})$	$x_1^n(1\|m_{b-1})$
Y_2	$\tilde{m}_1$	$\tilde{m}_2$	$\tilde{m}_3$	$\cdots$	$\tilde{m}_{b-1}$	$\emptyset$
X_2	$x_2^n(1)$	$x_2^n(\tilde{m}_1)$	$x_2^n(\tilde{m}_2)$	$\cdots$	$x_2^n(\tilde{m}_{b-2})$	$x_2^n(\tilde{m}_{b-1})$
Y_3	$\emptyset$	$\hat{m}_1$	$\hat{m}_2$	$\cdots$	$\hat{m}_{b-2}$	$\hat{m}_{b-1}$

- Decoding:
  - At the end of block $j+1$, find the unique $\hat{m}_j$ such that $(x_1^n(\hat{m}_j|\hat{m}_{j-1}), x_2^n(\hat{m}_{j-1}), y_3^n(j)) \in \mathcal{T}_\epsilon^{(n)}$ and $(x_2^n(\hat{m}_j), y_3^n(j+1)) \in \mathcal{T}_\epsilon^{(n)}$ simultaneously

Analysis of the Probability of Error

- Assume that $M_{j-1} = M_j = 1$
- The decoder makes an error only if one or more of the following events occur:

$$\tilde{\mathcal{E}}(j-1) = \{\tilde{M}_{j-1} \neq 1\}$$

$$\tilde{\mathcal{E}}(j) = \{\tilde{M}_j \neq 1\}$$

$$\mathcal{E}(j-1) = \{\hat{M}_{j-1} \neq 1\}$$

$$\mathcal{E}_1(j) = \left\{(X_1^n(\tilde{M}_j|\hat{M}_{j-1}), X_2^n(\hat{M}_{j-1}), Y_3^n(j)) \notin \mathcal{T}_\epsilon^{(n)} \text{ or } (X_2^n(\tilde{M}_j), Y_3^n(j+1)) \notin \mathcal{T}_\epsilon^{(n)}\right\}$$

$$\mathcal{E}_2(j) = \left\{(X_1^n(m_j|\hat{M}_{j-1}), X_2^n(\hat{M}_{j-1}), Y_3^n(j)) \in \mathcal{T}_\epsilon^{(n)} \text{ and } (X_2^n(m_j), Y_3^n(j+1)) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_j \neq \tilde{M}_j\right\}$$

Thus, the probability of error is upper bounded as

$$\begin{aligned} \mathsf{P}(\mathcal{E}(j)) &\le \mathsf{P}(\tilde{\mathcal{E}}(j-1) \cup \tilde{\mathcal{E}}(j) \cup \mathcal{E}(j-1) \cup \mathcal{E}_1(j) \cup \mathcal{E}_2(j)) \\ &\le \mathsf{P}(\tilde{\mathcal{E}}(j-1)) + \mathsf{P}(\tilde{\mathcal{E}}(j)) + \mathsf{P}(\mathcal{E}(j-1)) \\ &\quad + \mathsf{P}(\mathcal{E}_1(j) \cap \tilde{\mathcal{E}}^c(j-1) \cap \tilde{\mathcal{E}}^c(j) \cap \mathcal{E}^c(j-1)) + \mathsf{P}(\mathcal{E}_2(j) \cap \tilde{\mathcal{E}}^c(j)) \end{aligned}$$

- By independence of the codebooks, the LLN, the packing lemma, and induction, the first four terms tend to zero as $n \to \infty$ if $R < I(X_1; Y_2|X_2) - \delta(\epsilon)$

- For the last term, consider

$$
\begin{aligned}
\mathsf{P}(\mathcal{E}_2(j) \cap \tilde{\mathcal{E}}^c(j)) = {} & \mathsf{P}\{(X_1^n(m_j|\hat{M}_{j-1}), X_2^n(\hat{M}_{j-1}), Y_3^n(j)) \in \mathcal{T}_\epsilon^{(n)}, \\
& \quad (X_2^n(m_j), Y_3^n(j+1)) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_j \neq 1, \text{ and } \tilde{M}_j = 1\} \\
\le {} & \sum_{m_j \neq 1} \mathsf{P}\{(X_1^n(m_j|\hat{M}_{j-1}), X_2^n(\hat{M}_{j-1}), Y_3^n(j)) \in \mathcal{T}_\epsilon^{(n)}, \\
& \qquad (X_2^n(m_j), Y_3^n(j+1)) \in \mathcal{T}_\epsilon^{(n)}, \text{ and } \tilde{M}_j = 1\} \\
\overset{(a)}{=} {} & \sum_{m_j \neq 1} \mathsf{P}\{(X_1^n(m_j|\hat{M}_{j-1}), X_2^n(\hat{M}_{j-1}), Y_3^n(j)) \in \mathcal{T}_\epsilon^{(n)} \text{ and } \tilde{M}_j = 1\} \\
& \qquad \cdot \mathsf{P}\{(X_2^n(m_j), Y_3^n(j+1)) \in \mathcal{T}_\epsilon^{(n)} \,|\, \tilde{M}_j = 1\} \\
\le {} & \sum_{m_j \neq 1} \mathsf{P}\{(X_1^n(m_j|\hat{M}_{j-1}), X_2^n(\hat{M}_{j-1}), Y_3^n(j)) \in \mathcal{T}_\epsilon^{(n)}\} \\
& \qquad \cdot \mathsf{P}\{(X_2^n(m_j), Y_3^n(j+1)) \in \mathcal{T}_\epsilon^{(n)} \,|\, \tilde{M}_j = 1\} \\
\overset{(b)}{\le} {} & 2^{nR} 2^{-n(I(X_1;Y_3|X_2)-\delta(\epsilon))} 2^{-n(I(X_2;Y_3)-\delta(\epsilon))}
\end{aligned}
$$

$\to 0$ as $n \to \infty$ if $R < I(X_1;Y_3|X_2) + I(X_2;Y_3) - 2\delta(\epsilon) = I(X_1,X_2;Y_3) - 2\delta(\epsilon)$

(a) $\{(X_1^n(m_j|\hat{M}_{j-1}), X_2^n(\hat{M}_{j-1}), Y_3^n(j)) \in \mathcal{T}_\epsilon^{(n)}\}$ and $\{(X_2^n(m_j), Y_3^n(j+1)) \in \mathcal{T}_\epsilon^{(n)}\}$ are conditionally independent given $\tilde{M}_j = 1$ for $m_j \neq 1$

(b) independence of the codebooks and the joint typicality lemma

Noisy Network Coding

- Compress–forward for DM-RC can be extended to DM-MN

Theorem (Noisy Network Coding Lower Bound)

$$C \geq \max_{k \in \mathcal{D}} \min_{\mathcal{S}: 1 \in \mathcal{S}, k \in \mathcal{S}^c} \left(I(X(\mathcal{S}); \hat{Y}(\mathcal{S}^c), Y_k | X(\mathcal{S}^c)) - I(Y(\mathcal{S}); \hat{Y}(\mathcal{S}) | X^N, \hat{Y}(\mathcal{S}^c), Y_k) \right),$$

where the maximum is over all $\prod_{k=1}^N p(x_k) p(\hat{y}_k | y_k, x_k)$, $\hat{Y}_1 = \emptyset$ by convention, $X(\mathcal{S})$ denotes inputs in $\mathcal{S}$, and $Y(\mathcal{S}^c)$ denotes outputs in $\mathcal{S}^c$

- Special cases:
  - Compress–forward lower bound for DM-RC ($N = 3$ and $X_3 = \emptyset$)
  - Network coding theorem for graphical MN (Ahlswede–Cai–Li–Yeung 2000)
  - Capacity of deterministic MN with no interference (Ratnakar–Kramer 2006)
  - Capacity of wireless erasure MN (Dana–Gowaikar–Palanki–Hassibi–Effros 2006)
  - Lower bound for general deterministic MN (Avestimehr–Diggavi–Tse 2011)
- Can be extended to Gaussian networks (giving best known gap result) and to multiple messages (Lim–Kim–El Gamal–Chung 2011)

Proof of Achievability

- We use several new ideas beyond compress–forward for DM-RC
  - The source node sends the same message $m \in [1:2^{nbR}]$ over b blocks
  - Relay node j sends the index of the compressed version $\hat{Y}_j^n$ of Y_j^n without binning
  - Each receiver node performs simultaneous nonunique decoding of the message and compression indices from all b blocks

- We illustrate this scheme for DM-RC

- Codebook generation:
  - Fix $p(x_1)p(x_2)p(\hat{y}_2|y_2,x_2)$ that attains the lower bound
  - For each $j \in [1:b]$, randomly and independently generate 2^{nbR} sequences $x_1^n(j,m) \sim \prod_{i=1}^n p_{X_1}(x_{1i})$, $m \in [1:2^{nbR}]$
  - Randomly and independently generate 2^{nR_2} sequences $x_2^n(l_{j-1}) \sim \prod_{i=1}^n p_{X_2}(x_{2i})$, $l_{j-1} \in [1:2^{nR_2}]$
  - For each $l_{j-1} \in [1:2^{nR_2}]$, randomly and conditionally independently generate 2^{nR_2} sequences $\hat{y}_2^n(l_j|l_{j-1}) \sim \prod_{i=1}^n p_{\hat{Y}_2|X_2}(\hat{y}_{2i}|x_{2i}(l_{j-1}))$, $l_j \in [1:2^{nR_2}]$
  - $\mathcal{C}_j = \{(x_1^n(j,m), x_2^n(l_{j-1}), \hat{y}_2^n(l_j|l_{j-1})) : m \in [1:2^{nbR}], l_j, l_{j-1} \in [1:2^{nR_2}]\}$, $j \in [1:b]$

Block	1	2	3	$\cdots$	$b-1$	b
X_1	$x_1^n(1,m)$	$x_1^n(2,m)$	$x_1^n(3,m)$	$\cdots$	$x_1^n(b-1,m)$	$x_1^n(b,m)$
Y_2	$\hat{y}_2^n(l_1\|1), l_1$	$\hat{y}_2^n(l_2\|l_1), l_2$	$\hat{y}_2^n(l_3\|l_2), l_3$	$\cdots$	$\hat{y}_2^n(l_{b-1}\|l_{b-2}), l_{b-1}$	$\hat{y}_2^n(l_b\|l_{b-1}), l_b$
X_2	$x_2^n(1)$	$x_2^n(l_1)$	$x_2^n(l_2)$	$\cdots$	$x_2^n(l_{b-2})$	$x_2^n(l_{b-1})$
Y_3	$\emptyset$	$\emptyset$	$\emptyset$	$\cdots$	$\emptyset$	$\hat{m}$

- Encoding:
  - To send $m \in [1:2^{nbR}]$, transmit $x_1^n(j,m)$ in block j
- Relay encoding:
  - At the end of block j, find an index l_j such that $(y_2^n(j), \hat{y}_2^n(l_j|l_{j-1}), x_2^n(l_{j-1})) \in \mathcal{T}_{\epsilon'}^{(n)}$
  - In block $j+1$, transmit $x_2^n(l_j)$
- Decoding:
  - At the end of block b, find the unique $\hat{m}$ such that $(x_1^n(j,\hat{m}), x_2^n(l_{j-1}), \hat{y}_2^n(l_j|l_{j-1}), y_3^n(j)) \in \mathcal{T}_{\epsilon}^{(n)}$ for all $j \in [1:b]$ for some $l_1, l_2, \ldots, l_b$

Analysis of the Probability of Error

- Assume $M = 1$ and $L_1 = L_2 = \cdots = L_b = 1$
- The decoder makes an error only if one or more of the following events occur:

$$\mathcal{E}_1 = \{(Y_2^n(j), \hat{Y}_2^n(l_j|1), X_2^n(1)) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } l_j \text{ for some } j \in [1:b]\}$$

$$\mathcal{E}_2 = \{(X_1^n(j,1), X_2^n(1), \hat{Y}_2^n(1|1), Y_3^n(j)) \notin \mathcal{T}_{\epsilon}^{(n)} \text{ for some } j \in [1:b]\}$$

$$\mathcal{E}_3 = \{(X_1^n(j,m), X_2^n(l_{j-1}), \hat{Y}_2^n(l_j|l_{j-1}), Y_3^n(j)) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for all } j \text{ for some } l^b,\ m \neq 1\}$$

Thus, the probability of error is upper bounded as

$$\mathsf{P}(\mathcal{E}) \leq \mathsf{P}(\mathcal{E}_1) + \mathsf{P}(\mathcal{E}_2 \cap \mathcal{E}_1^c) + \mathsf{P}(\mathcal{E}_3)$$

- By the covering lemma and the union of events bound (over b blocks), $\mathsf{P}(\mathcal{E}_1) \to 0$ as $n \to \infty$ if $R_2 > I(\hat{Y}_2; Y_2|X_2) + \delta(\epsilon')$
- By the conditional typicality lemma and the union of events bound, $\mathsf{P}(\mathcal{E}_2 \cap \mathcal{E}_1^c) \to 0$ as $n \to \infty$

- Define $\tilde{\mathcal{E}}_j(m, l_{j-1}, l_j) = \{(X_1^n(j,m), X_2^n(l_{j-1}), \hat{Y}_2^n(l_j|l_{j-1}), Y_3^n(j)) \in \mathcal{T}_\epsilon^{(n)}\}$
  Then

$$\begin{aligned} \mathrm{P}(\mathcal{E}_3) &= \mathrm{P}\Big(\bigcup_{m\neq 1}\bigcup_{l^b}\bigcap_{j=1}^{b}\tilde{\mathcal{E}}_j(m, l_{j-1}, l_j)\Big) \\ &\le \sum_{m\neq 1}\sum_{l^b}\mathrm{P}\Big(\bigcap_{j=1}^{b}\tilde{\mathcal{E}}_j(m, l_{j-1}, l_j)\Big) \\ &= \sum_{m\neq 1}\sum_{l^b}\prod_{j=1}^{b}\mathrm{P}(\tilde{\mathcal{E}}_j(m, l_{j-1}, l_j)) \\ &\le \sum_{m\neq 1}\sum_{l^b}\prod_{j=2}^{b}\mathrm{P}(\tilde{\mathcal{E}}_j(m, l_{j-1}, l_j)) \end{aligned}$$

- If $l_{j-1} = 1$, then by the joint typicality lemma, $\mathrm{P}(\tilde{\mathcal{E}}_j) \le 2^{-n(\overbrace{I(X_1;\hat{Y}_2, Y_3|X_2)}^{I_1}-\delta(\epsilon))}$
- Similarly, if $l_{j-1} \neq 1$, then $\mathrm{P}(\tilde{\mathcal{E}}_j) \le 2^{-n(\underbrace{I(X_1, X_2; Y_3) + I(\hat{Y}_2; X_1, Y_3|X_2)}_{I_2}-\delta(\epsilon))}$
- Thus, if l^{b-1} has k 1s, then

$$\prod_{j=2}^{b}\mathrm{P}(\tilde{\mathcal{E}}_j(m, l_{j-1}, l_j)) \le 2^{-n\left(kI_1+(b-1-k)I_2-(b-1)\delta(\epsilon)\right)}$$

- Continuing with the bound,

$$\sum_{m\neq 1}\sum_{l^b}\prod_{j=2}^{b}\mathsf{P}(\tilde{\mathcal{E}}_j(m,l_{j-1},l_j)) = \sum_{m\neq 1}\sum_{l_b}\sum_{l^{b-1}}\prod_{j=2}^{b}\mathsf{P}(\tilde{\mathcal{E}}_j(m,l_{j-1},l_j))$$

$$\leq \sum_{m\neq 1}\sum_{l_b}\sum_{j=0}^{b-1}\binom{b-1}{j}2^{n(b-1-j)R_2}\cdot 2^{-n\left(jI_1+(b-1-j)I_2-(b-1)\delta(\epsilon)\right)}$$

$$= \sum_{m\neq 1}\sum_{l_b}\sum_{j=0}^{b-1}\binom{b-1}{j}2^{-n(jI_1+(b-1-j)(I_2-R_2)-(b-1)\delta(\epsilon))}$$

$$\leq \sum_{m\neq 1}\sum_{l_b}\sum_{j=0}^{b-1}\binom{b-1}{j}2^{-n((b-1)(\min\{I_1,I_2-R_2\}-\delta(\epsilon)))}$$

$$\leq 2^{nbR}\cdot 2^{nR_2}\cdot 2^{b}\cdot 2^{-n(b-1)(\min\{I_1,I_2-R_2\}-\delta(\epsilon))},$$

  which → 0 as $n\to\infty$ if $R<((b-1)(\min\{I_1,\ I_2-R_2\}-\delta'(\epsilon))-R_2)/b$
- Finally, by eliminating $R_2>I(\hat{Y}_2;Y_2|X_2)+\delta(\epsilon')$, substituting I_1 and I_2, and taking $b\to\infty$, we have shown that $\mathsf{P}(\mathcal{E})\to 0$ as $n\to\infty$ if

$$R<\min\{I(X_1;\hat{Y}_2,Y_3|X_2),\ I(X_1,X_2;Y_3)-I(\hat{Y}_2;Y_2|X_1,X_2,Y_3)\}-\delta'(\epsilon)-\delta(\epsilon')$$

- This completes the proof of achievability for noisy network coding

Summary

1. Typical Sequences
2. Point-to-Point Communication
3. Multiple Access Channel
4. Broadcast Channel
5. Lossy Source Coding
6. Wyner–Ziv Coding
7. Gelfand–Pinsker Coding
8. Wiretap Channel
9. Relay Channel
10. Multicast Network

- Network decode–forward
- Sliding window decoding
- Noisy network coding
- Sending same message multiple times using independent codebooks
- Beyond packing lemma

Conclusion

- Presented a unified approach to achievability proofs for DM networks:
  - Typicality and elementary lemmas
  - Coding techniques: random coding, joint typicality encoding/decoding, simultaneous (nonunique) decoding, superposition coding, binning, multicoding
- Results can be extended to Gaussian models via discretization procedures
- Lossless source coding is a corollary of lossy source coding
- Network Information Theory book:
  - Comprehensive coverage of this approach
  - More advanced coding techniques and analysis tools
  - Converse techniques (DM and Gaussian)
  - Open problems
- Although the theory is far from complete, we hope that our approach will
  - Make the subject accessible to students, researchers, and communication engineers
  - Help in the quest for a unified theory of information flow in networks

References

Ahlswede, R. (1971). Multiway communication channels. In *Proc. 2nd Int. Symp. Inf. Theory*, Tsahkadsor, Armenian SSR, pp. 23–52.

Ahlswede, R., Cai, N., Li, S.-Y. R., and Yeung, R. W. (2000). Network information flow. *IEEE Trans. Inf. Theory*, 46(4), 1204–1216.

Avestimehr, A. S., Diggavi, S. N., and Tse, D. N. C. (2011). Wireless network information flow: A deterministic approach. *IEEE Trans. Inf. Theory*, 57(4), 1872–1905.

Bergmans, P. P. (1973). Random coding theorem for broadcast channels with degraded components. *IEEE Trans. Inf. Theory*, 19(2), 197–207.

Carleial, A. B. (1982). Multiple-access channels with different generalized feedback signals. *IEEE Trans. Inf. Theory*, 28(6), 841–850.

Costa, M. H. M. (1983). Writing on dirty paper. *IEEE Trans. Inf. Theory*, 29(3), 439–441.

Cover, T. M. (1972). Broadcast channels. *IEEE Trans. Inf. Theory*, 18(1), 2–14.

Cover, T. M. and El Gamal, A. (1979). Capacity theorems for the relay channel. *IEEE Trans. Inf. Theory*, 25(5), 572–584.

Csiszár, I. and Körner, J. (1978). Broadcast channels with confidential messages. *IEEE Trans. Inf. Theory*, 24(3), 339–348.

Dana, A. F., Gowaikar, R., Palanki, R., Hassibi, B., and Effros, M. (2006). Capacity of wireless erasure networks. *IEEE Trans. Inf. Theory*, 52(3), 789–804.

References (cont.)

El Gamal, A., Mohseni, M., and Zahedi, S. (2006). Bounds on capacity and minimum energy-per-bit for AWGN relay channels. *IEEE Trans. Inf. Theory*, 52(4), 1545–1561.

Elias, P., Feinstein, A., and Shannon, C. E. (1956). A note on the maximum flow through a network. *IRE Trans. Inf. Theory*, 2(4), 117–119.

Ford, L. R., Jr. and Fulkerson, D. R. (1956). Maximal flow through a network. *Canad. J. Math.*, 8(3), 399–404.

Gelfand, S. I. and Pinsker, M. S. (1980). Coding for channel with random parameters. *Probl. Control Inf. Theory*, 9(1), 19–31.

Han, T. S. and Kobayashi, K. (1981). A new achievable rate region for the interference channel. *IEEE Trans. Inf. Theory*, 27(1), 49–60.

Heegard, C. and El Gamal, A. (1983). On the capacity of computer memories with defects. *IEEE Trans. Inf. Theory*, 29(5), 731–739.

Kramer, G., Gastpar, M., and Gupta, P. (2005). Cooperative strategies and capacity theorems for relay networks. *IEEE Trans. Inf. Theory*, 51(9), 3037–3063.

Liao, H. H. J. (1972). *Multiple access channels*. Ph.D. thesis, University of Hawaii, Honolulu, HI.

Lim, S. H., Kim, Y.-H., El Gamal, A., and Chung, S.-Y. (2011). Noisy network coding. *IEEE Trans. Inf. Theory*, 57(5), 3132–3152.

McEliece, R. J. (1977). *The Theory of Information and Coding*. Addison-Wesley, Reading, MA.

References (cont.)

Orlitsky, A. and Roche, J. R. (2001). Coding for computing. *IEEE Trans. Inf. Theory*, 47(3), 903–917.

Ratnakar, N. and Kramer, G. (2006). The multicast capacity of deterministic relay networks with no interference. *IEEE Trans. Inf. Theory*, 52(6), 2425–2432.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.*, 27(3), 379–423, 27(4), 623–656.

Shannon, C. E. (1959). Coding theorems for a discrete source with a fidelity criterion. In *IRE Int. Conv. Rec.*, vol. 7, part 4, pp. 142–163. Reprint with changes (1960). In R. E. Machol (ed.) *Information and Decision Processes*, pp. 93–126. McGraw-Hill, New York.

Shannon, C. E. (1961). Two-way communication channels. In *Proc. 4th Berkeley Symp. Math. Statist. Probab.*, vol. I, pp. 611–644. University of California Press, Berkeley.

Slepian, D. and Wolf, J. K. (1973a). Noiseless coding of correlated information sources. *IEEE Trans. Inf. Theory*, 19(4), 471–480.

Slepian, D. and Wolf, J. K. (1973b). A coding theorem for multiple access channels with correlated sources. *Bell Syst. Tech. J.*, 52(7), 1037–1076.

Willems, F. M. J. and van der Meulen, E. C. (1985). The discrete memoryless multiple-access channel with cribbing encoders. *IEEE Trans. Inf. Theory*, 31(3), 313–327.

Wyner, A. D. (1975). The wire-tap channel. *Bell Syst. Tech. J.*, 54(8), 1355–1387.

References (cont.)

Wyner, A. D. and Ziv, J. (1976). The rate–distortion function for source coding with side information at the decoder. *IEEE Trans. Inf. Theory*, 22(1), 1–10.

Xie, L.-L. and Kumar, P. R. (2005). An achievable rate for the multiple-level relay channel. *IEEE Trans. Inf. Theory*, 51(4), 1348–1358.

Zeng, C.-M., Kuhlmann, F., and Buzo, A. (1989). Achievability proof of some multiuser channel coding theorems using backward decoding. *IEEE Trans. Inf. Theory*, 35(6), 1160–1165.