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
**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
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
5. Lossy Source Coding
6. Wyner–Ziv Coding
7. Gelfand–Pinsker Coding
8. Wiretap Channel
9. Relay Channel
10. Multicast Network

10-minute break

10-minute break
Typical Sequences

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

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

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

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

**Typical Average Lemma**

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

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

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

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, $\lim_{n \to \infty} P\{X^n \in \mathcal{T}_{\epsilon}^{(n)}\} = 1$
Jointly Typical Sequences

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

  \[
  \pi(x, y|x^n, y^n) = \frac{|\{i: (x_i, y_i) = (x, y)\}|}{n} \quad \text{for} \ (x, y) \in \mathcal{X} \times \mathcal{Y}
  \]

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

  \[
  \mathcal{T}_\epsilon^{(n)}(X, Y) = \{(x^n, y^n): |\pi(x, y|x^n, y^n) - p(x, y)| \leq \epsilon \cdot p(x, y) \ \text{for all} \ (x, y)\}
  \]

  \[
  = \mathcal{T}_\epsilon^{(n)}((X, Y))
  \]

- 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 $y^n \in \mathcal{T}_\epsilon^{(n)}(Y)$
  - $p(x^n) \approx 2^{-nH(X)}$, $p(y^n) \approx 2^{-nH(Y)}$, and $p(x^n, y^n) \approx 2^{-nH(X,Y)}$
  - $p(x^n|y^n) \approx 2^{-nH(X|Y)}$ and $p(y^n|x^n) \approx 2^{-nH(Y|X)}$
### Conditionally Typical Sequences

- **Conditionally typical set**: For $x^n \in \mathcal{X}^n$,
  $\mathcal{T}_\varepsilon(n)(Y|x^n) = \{y^n: (x^n, y^n) \in \mathcal{T}_\varepsilon(n)(X, Y)\}$

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

### Conditional Typicality Lemma

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

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

- If $x^n \in \mathcal{T}_\varepsilon'(n)(X)$ and $\varepsilon > \varepsilon'$, then for $n$ sufficiently large,
  $$|\mathcal{T}_\varepsilon(n)(Y|x^n)| \geq 2^{n(H(Y|X) - \delta(\varepsilon))}$$

- Let $X \sim p(x)$, $Y = g(X)$, and $x^n \in \mathcal{T}_\varepsilon(n)(X)$. Then
  $$y^n \in \mathcal{T}_\varepsilon(n)(Y|x^n) \text{ iff } y_i = g(x_i), \ i \in [1:n]$$
Illustration of Joint Typicality

\[ T_{\epsilon}^{(n)}(X) \left( | \cdot | \leq 2^{nH(X)} \right) \]

\[ T_{\epsilon}^{(n)}(Y) \left( | \cdot | \leq 2^{nH(Y)} \right) \]

\[ T_{\epsilon}^{(n)}(X, Y) \left( | \cdot | \leq 2^{nH(X,Y)} \right) \]

\[ T_{\epsilon}^{(n)}(Y|X^n) \]

\[ T_{\epsilon}^{(n)}(X|Y^n) \]
Another Illustration of Joint Typicality

\[ T_e^n(X) \]
\[ T_e^n(Y) \]
\[ T_e^n(Y|X^n) \]

\[ \left| \cdot \right| = 2^{nH(Y|X)} \]
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
  \[
  P\{(\tilde{x}^n, \tilde{y}^n, \tilde{Z}^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y, Z)\} \leq 2^{-n(I(Y;Z|X)-\delta(\epsilon))}
  \]

- 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,
  \[
  P\{(x^n, y^n, \tilde{Z}^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y, Z)\} \geq 2^{-n(I(Y;Z|X)+\delta(\epsilon))}
  \]
Summary

1. Typical Sequences
   - Typical average lemma
   - Conditional typicality lemma
   - Joint typicality lemma
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
**Discrete Memoryless Channel (DMC)**

- **Point-to-point communication system**

  \[
  M \xrightarrow{\text{Encoder}} X^n \xrightarrow{p(y|x)} Y^n \xrightarrow{\text{Decoder}} \hat{M}
  \]

- **Assume a discrete memoryless channel (DMC) model \((X, p(y|x), Y)\)**
  - **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** \([1 : 2^{nR}] = \{1, 2, \ldots, 2^{[nR]}\}\)
  - **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 $M \sim \text{Unif}[1 : 2^{nR}]$

- **Average probability of error**: $P_e^{(n)} = P\{\hat{M} \neq M\}$
- **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)**

\[
C(B) = \max_{p(x): E(b(X)) \leq B} I(X; Y)
\]
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 $x^n(m) \in T_{\epsilon}^{(n)}$
  (by the typical average lemma, $\sum_{i=1}^{n} b(x_i(m)) \leq nB$)
- Otherwise it transmits $(x_0, \ldots, x_0)$

**Decoding:**
- Decoder declares that $\hat{m}$ is sent if it is unique message such that $(x^n(\hat{m}), y^n) \in T_{\epsilon}^{(n)}$
- Otherwise it declares an error
Analysis of the Probability of Error

- Consider the probability of error $P(\mathcal{E})$ 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:

  $$\mathcal{E}_1 = \{(X^n(1), Y^n) \notin \mathcal{T}_{e}^{(n)}\}$$

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

Thus, by the union of events bound

$$P(\mathcal{E}) = P(\mathcal{E} | M = 1)$$

$$= P(\mathcal{E}_1 \cup \mathcal{E}_2)$$

$$\leq P(\mathcal{E}_1) + P(\mathcal{E}_2)$$
Consider the first term
\[ P(\mathcal{E}_1) = P\{ (X^n(1), Y^n) \notin \mathcal{T}^{(n)}_{\varepsilon} \} \]
\[ = P\{ X^n(1) \in \mathcal{T}^{(n)}_{\varepsilon}, (X^n(1), Y^n) \notin \mathcal{T}^{(n)}_{\varepsilon} \} + P\{ X^n(1) \notin \mathcal{T}^{(n)}_{\varepsilon}, (X^n(1), Y^n) \notin \mathcal{T}^{(n)}_{\varepsilon} \} \]
\[ \leq \sum_{x^n \in \mathcal{T}^{(n)}_{\varepsilon}} \prod_{i=1}^{n} p_X(x_i) \sum_{y^n \notin \mathcal{T}^{(n)}_{\varepsilon}(Y|x^n)} \prod_{i=1}^{n} p_{Y|X}(y_i|x_i) + P\{ X^n(1) \notin \mathcal{T}^{(n)}_{\varepsilon} \} \]
\[ \leq \sum_{(x^n, y^n) \notin \mathcal{T}^{(n)}_{\varepsilon}} \prod_{i=1}^{n} p_X(x_i) p_{Y|X}(y_i|x_i) + P\{ X^n(1) \notin \mathcal{T}^{(n)}_{\varepsilon} \} \]

By the LLN, each term \( \to 0 \) as \( n \to \infty \)
Analysis of the Probability of Error

Consider the second term

\[
P(\mathcal{E}_2) = P\{(X^n(m), Y^n) \in \mathcal{T}_e^{(n)} \text{ for some } m \neq 1\}
\]

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 \( P(\mathcal{E}_2) \), 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 $\tilde{U}^n$

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

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

if $R < I(X; Y|U) - \delta(\epsilon)$
Analysis of the Probability of Error

- Consider the second term

\[ P(\mathcal{E}_2) = P\{(X^n(m), Y^n) \in \mathcal{T}_e^{(n)} \text{ for some } m \neq 1\} \]

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 \( A = [2 : 2^{nR}] \) and \( U = \emptyset \), \( P(\mathcal{E}_2) \to 0 \) if

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

- Since \( P(\mathcal{E}) \to 0 \) 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**

  \[ Y = gX + Z \]

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

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

**Theorem (Shannon 1948)**:

\[
C = \max_{F(x):E(X^2)\leq P} I(X;Y) = \frac{1}{2} \log(1 + S)
\]
Proof of Achievability

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

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

- By the weak convergence and the dominated convergence theorem,

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

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

\[
\begin{align*}
M_1 & \rightarrow \text{Encoder 1} \rightarrow X_1^n \rightarrow p(y|x_1, x_2) \rightarrow \text{Decoder} \rightarrow \hat{M}_1, \hat{M}_2 \\
M_2 & \rightarrow \text{Encoder 2} \rightarrow X_2^n \rightarrow p(y|x_1, x_2) \rightarrow \text{Decoder} \rightarrow \hat{M}_1, \hat{M}_2
\end{align*}
\]

Assume a 2-sender DM-MAC model \((X_1 \times X_2, p(y|x_1, x_2), \mathcal{Y})\)

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: \(P_e^{(n)} = P\{ (\hat{M}_1, \hat{M}_2) \neq (M_1, M_2) \}\)

\((R_1, R_2)\) 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 \((R_1, R_2)\)

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

$$R_1 \leq I(X_1; Y|X_2, Q),$$
$$R_2 \leq I(X_2; Y|X_1, Q),$$
$$R_1 + R_2 \leq I(X_1, X_2; Y|Q)$$

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 \( x_1^n(m_1) \sim \prod_{i=1}^{n} p_{X_1|Q}(x_{1i}|q_i), \quad m_1 \in [1 : 2^{nR_1}] \)
  - Similarly generate \( 2^{nR_2} \) sequences \( x_2^n(m_2) \sim \prod_{i=1}^{n} p_{X_2|Q}(x_{2i}|q_i), \quad m_2 \in [1 : 2^{nR_2}] \)

- 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 \( (q^n, x_1^n(\hat{m}_1), x_2^n(\hat{m}_2), y^n) \in \mathcal{T}_e^{(n)} \)
Analysis of the Probability of Error

- Assume \((M_1, M_2) = (1, 1)\)

- Joint pmfs induced by different \((m_1, m_2)\)

<table>
<thead>
<tr>
<th>(m_1)</th>
<th>(m_2)</th>
<th>Joint pmf</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>(p(q^n)p(x_1^n</td>
</tr>
<tr>
<td>*</td>
<td>1</td>
<td>(p(q^n)p(x_1^n</td>
</tr>
<tr>
<td>1</td>
<td>*</td>
<td>(p(q^n)p(x_1^n</td>
</tr>
<tr>
<td>*</td>
<td>*</td>
<td>(p(q^n)p(x_1^n</td>
</tr>
</tbody>
</table>

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

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

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

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

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

- Then \(P(\mathcal{E}) \leq P(\mathcal{E}_1) + P(\mathcal{E}_2) + P(\mathcal{E}_3) + P(\mathcal{E}_4)\)
<table>
<thead>
<tr>
<th>$m_1$</th>
<th>$m_2$</th>
<th>Joint pmf</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>$p(q^n)p(x_1^n</td>
</tr>
<tr>
<td>*</td>
<td>1</td>
<td>$p(q^n)p(x_1^n</td>
</tr>
<tr>
<td>1</td>
<td>*</td>
<td>$p(q^n)p(x_1^n</td>
</tr>
<tr>
<td>*</td>
<td>*</td>
<td>$p(q^n)p(x_1^n</td>
</tr>
</tbody>
</table>

$E_1 = \{(Q^n, X_1^n(1), X_2^n(1), Y^n) \notin T_\epsilon^{(n)}\}$

$E_2 = \{(Q^n, X_1^n(m_1), X_2^n(1), Y^n) \in T_\epsilon^{(n)} \text{ for some } m_1 \neq 1\}$

$E_3 = \{(Q^n, X_1^n(1), X_2^n(m_2), Y^n) \in T_\epsilon^{(n)} \text{ for some } m_2 \neq 1\}$

$E_4 = \{(Q^n, X_1^n(m_1), X_2^n(m_2), Y^n) \in T_\epsilon^{(n)} \text{ for some } m_1 \neq 1, m_2 \neq 1\}$

- By the LLN, $P(E_1) \to 0$ as $n \to \infty$
- By the packing lemma ($A = [2 : 2^nR_1]$, $U \leftarrow Q$, $X \leftarrow X_1$, $Y \leftarrow (X_2, Y)$), $P(E_2) \to 0$ 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, $P(E_3) \to 0$ 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 \(|\mathcal{A}| \leq 2^{nR}\), 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

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

if \(R < I(X; Y|U) - \delta(\epsilon)\)
<table>
<thead>
<tr>
<th>$m_1$</th>
<th>$m_2$</th>
<th>Joint pmf</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>$p(q^n)p(x_1^n</td>
</tr>
<tr>
<td>*</td>
<td>1</td>
<td>$p(q^n)p(x_1^n</td>
</tr>
<tr>
<td>1</td>
<td>*</td>
<td>$p(q^n)p(x_1^n</td>
</tr>
<tr>
<td>*</td>
<td>*</td>
<td>$p(q^n)p(x_1^n</td>
</tr>
</tbody>
</table>

$E_1 = \{(Q^n, X_1^n(1), X_2^n(1), Y^n) \notin T_{e}^{(n)}\}$

$E_2 = \{(Q^n, X_1^n(m_1), X_2^n(1), Y^n) \in T_{e}^{(n)} \text{ for some } m_1 \neq 1\}$

$E_3 = \{(Q^n, X_1^n(1), X_2^n(m_2), Y^n) \in T_{e}^{(n)} \text{ for some } m_2 \neq 1\}$

$E_4 = \{(Q^n, X_1^n(m_1), X_2^n(m_2), Y^n) \in T_{e}^{(n)} \text{ for some } m_1 \neq 1, m_2 \neq 1\}$

- By the LLN, $P(E_1) \to 0$ as $n \to \infty$
- By the packing lemma ($A = [2 : 2^{nR_1}]$, $U \leftarrow Q$, $X \leftarrow X_1$, $Y \leftarrow (X_2, Y)$), $P(E_2) \to 0$ 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, $P(E_3) \to 0$ as $n \to \infty$ if $R_2 < I(X_2; Y|X_1, Q) - \delta(\epsilon)$
<table>
<thead>
<tr>
<th>$m_1$</th>
<th>$m_2$</th>
<th>Joint pmf</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>$p(q^n)p(x_1^n</td>
</tr>
<tr>
<td>*</td>
<td>1</td>
<td>$p(q^n)p(x_1^n</td>
</tr>
<tr>
<td>1</td>
<td>*</td>
<td>$p(q^n)p(x_1^n</td>
</tr>
<tr>
<td>*</td>
<td>*</td>
<td>$p(q^n)p(x_1^n</td>
</tr>
</tbody>
</table>

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

- By the packing lemma ($A = [2 : 2^{nR_1}] \times [2 : 2^{nR_2}]$, $U \leftarrow Q$, $X \leftarrow (X_1, X_2)$), $P(\mathcal{E}_4) \to 0$ as $n \to \infty$ if $R_1 + R_2 < I(X_1, X_2; Y|Q) - \delta(\varepsilon)$
- **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: $P_e^{(n)} = P\{(\hat{M}_1, \hat{M}_2) \neq (M_1, M_2)\}$

$(R_1, R_2)$ 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

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

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

- This bound is tight for several special cases, including
  - Degraded: \(X \rightarrow Y_1 \rightarrow Y_2\) 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 ⇒ Less noisy ⇒ 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
    $$(u^n(\hat{m}_2), y^n_2) \in \mathcal{T}_e^{(n)}$$
    (by the packing lemma, $P(\mathcal{E}_2) \to 0$ as $n \to \infty$ if $R_2 < I(U; Y_2) - \delta(\epsilon)$)
  - Decoder 1 finds the unique message $\hat{m}_1$ such that
    $$(u^n(m_2), x^n(\hat{m}_1, m_2), y^n_1) \in \mathcal{T}_e^{(n)} \text{ for some } m_2$$
Analysis of the Probability of Error for Decoder 1

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

<table>
<thead>
<tr>
<th>(m_1)</th>
<th>(m_2)</th>
<th>Joint pmf</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>(p(u^n, x^n)p(y_1^n</td>
</tr>
<tr>
<td>*</td>
<td>1</td>
<td>(p(u^n, x^n)p(y_1^n</td>
</tr>
<tr>
<td>*</td>
<td>*</td>
<td>(p(u^n, x^n)p(y_1^n))</td>
</tr>
<tr>
<td>1</td>
<td>*</td>
<td>(p(u^n, x^n)p(y_1^n))</td>
</tr>
</tbody>
</table>

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 \(P(\mathcal{E}_1) \leq P(\mathcal{E}_{11}) + P(\mathcal{E}_{12}) + P(\mathcal{E}_{13})\)
### Joint pmf

<table>
<thead>
<tr>
<th>$m_1$</th>
<th>$m_2$</th>
<th>Joint pmf</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>$p(u^n, x^n) p(y_1^n</td>
</tr>
<tr>
<td>*</td>
<td>1</td>
<td>$p(u^n, x^n) p(y_1^n</td>
</tr>
<tr>
<td>*</td>
<td>*</td>
<td>$p(u^n, x^n) p(y_1^n)$</td>
</tr>
<tr>
<td>1</td>
<td>*</td>
<td>$p(u^n, x^n) p(y_1^n)$</td>
</tr>
</tbody>
</table>

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

- By the packing lemma ($A = [2 : 2^{nR_1}]$), $P(\mathcal{E}_{12}) \to 0$ as $n \to \infty$ if $R_1 < I(X; Y_1 | U) - \delta(\epsilon)$
- By the packing lemma ($A = [2 : 2^{nR_1}] \times [2 : 2^{nR_2}]$, $U \leftarrow \emptyset$, $X \leftarrow (U, X)$), $P(\mathcal{E}_{13}) \to 0$ as $n \to \infty$ if $R_1 + R_2 < I(U, X; Y_1) - \delta(\epsilon) = I(X; Y_1) - \delta(\epsilon)$
- **Remark:** $P(\mathcal{E}_{14}) = P\{(U^n(m_2), X^n(1, m_2), Y_1^n) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } m_2 \neq 1\} \to 0$ 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 $M_2$. 

By the packing lemma ($A = [2 : 2^{nR_1}]$), $P(\mathcal{E}_{12}) \to 0$ as $n \to \infty$ if $R_1 < I(X; Y_1 | U) - \delta(\epsilon)$

By the packing lemma ($A = [2 : 2^{nR_1}] \times [2 : 2^{nR_2}]$, $U \leftarrow \emptyset$, $X \leftarrow (U, X)$), $P(\mathcal{E}_{13}) \to 0$ as $n \to \infty$ if $R_1 + R_2 < I(U, X; Y_1) - \delta(\epsilon) = I(X; Y_1) - \delta(\epsilon)$

**Remark:** $P(\mathcal{E}_{14}) = P\{(U^n(m_2), X^n(1, m_2), Y_1^n) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } m_2 \neq 1\} \to 0$ 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 $M_2$. 

By the packing lemma ($A = [2 : 2^{nR_1}]$), $P(\mathcal{E}_{12}) \to 0$ as $n \to \infty$ if $R_1 < I(X; Y_1 | U) - \delta(\epsilon)$

By the packing lemma ($A = [2 : 2^{nR_1}] \times [2 : 2^{nR_2}]$, $U \leftarrow \emptyset$, $X \leftarrow (U, X)$), $P(\mathcal{E}_{13}) \to 0$ as $n \to \infty$ if $R_1 + R_2 < I(U, X; Y_1) - \delta(\epsilon) = I(X; Y_1) - \delta(\epsilon)$

**Remark:** $P(\mathcal{E}_{14}) = P\{(U^n(m_2), X^n(1, m_2), Y_1^n) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } m_2 \neq 1\} \to 0$ 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 $M_2$. 

### Elements of NIT Tutorial, ISIT 2011 42 / 118
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:

\[ D = E(d(X^n, \hat{X}^n)) = \sum_{x^n} p(x^n) d(x^n, \hat{x}^n(m(x^n))) \]

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

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

\[ R(D) = \min_{p(\hat{x}|x):E(d(x,\hat{x})) \leq D} I(X; \hat{X}) \]

for \(D \geq D_{\text{min}} = E[\min_{\hat{x}(x)} d(X, \hat{x}(X))]\)
Proof of Achievability

- We use random coding and joint typicality encoding

- **Codebook generation:**
  - Fix $p(\hat{x}|x)$ that attains $R(D/(1 + \varepsilon))$ and compute $p(\hat{x}) = \sum_x p(x)p(\hat{x}|x)$
  - Randomly and independently generate sequences $\hat{x}^n(m) \sim \prod_{i=1}^n p(\hat{x}_i), \ m \in [1 : 2^{nR}]$

- **Encoding:**
  - Find an index $m$ such that $(x^n, \hat{x}^n(m)) \in T^{(n)}_\varepsilon$
  - 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

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

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

- To bound \( P(\mathcal{E}) \), 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
\[
\lim_{n \to \infty} P\{(U^n, X^n) \in \mathcal{T}^{(n)}_{\epsilon'}(U, X)\} = 1
\]

Let \(\hat{X}^n(m) \sim \prod_{i=1}^{n} p_{\hat{X}|U}(\hat{x}_i|u_i), m \in \mathcal{A},\) where \(|\mathcal{A}| \geq 2^{nR}\), 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
\[
\lim_{n \to \infty} P\{(U^n, X^n, \hat{X}^n(m)) \notin \mathcal{T}^{(n)}_{\epsilon} \text{ for all } m \in \mathcal{A}\} = 0,
\]
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:

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

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

- By the covering lemma \((U = \emptyset)\), \( P(\mathcal{E}) \to 0 \) 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,

  \[ E[d(X^n, \hat{X}^n)] = P(\mathcal{E}) E[d(X^n, \hat{X}^n)|\mathcal{E}] + P(\mathcal{E}^c) E[d(X^n, \hat{X}^n)|\mathcal{E}^c] \leq P(\mathcal{E}) d_{\text{max}} + P(\mathcal{E}^c)(1 + \epsilon) E(d(X, \hat{X})) \]

- Hence, \( \limsup_{n \to \infty} E[d(X^n, \hat{X}^n)] \leq D \) 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 $\lim_{n \to \infty} P\{\hat{X}^n \neq X^n\} = 0$
- 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 $X$, $\hat{x} = x$, 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
    $$
    \lim_{n \to \infty} E(d(X^n, \hat{X}^n)) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} P\{\hat{X}_i \neq X_i\} \leq \lim_{n \to \infty} P\{\hat{X}^n \neq X^n\}
    $$
  - Hence, any sequence of $(2^{nR}, n)$ codes with $\lim_{n \to \infty} P\{\hat{X}^n \neq X^n\} = 0$ 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 ($p(\hat{x}) = p_X(\hat{x})$)
  - As before, generate a random code $\hat{x}^n(m), m \in [1:2^{nR}]$
  - Then $P(\mathcal{E}) = P\{(X^n, \hat{X}^n) \notin \mathcal{T}_\varepsilon^{(n)}\} \to 0$ as $n \to \infty$ if $R > I(X; \hat{X}) + \delta(\varepsilon) = R(0) + \delta(\varepsilon)$
  - Now recall that if $(x^n, \hat{x}^n) \notin \mathcal{T}_\varepsilon^{(n)}$, then $\hat{x}^n = x^n$ (or if $\hat{x}^n \neq x^n$, then $(x^n, \hat{x}^n) \notin \mathcal{T}_\varepsilon^{(n)}$)
  - Hence, $P\{\hat{X}^n \neq X^n\} \to 0$ as $n \to \infty$ if $R > R(0) + \delta(\varepsilon)$
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

\[ X^n \xrightarrow{Encoder} M \xrightarrow{Decoder} (\hat{X}^n, D) \]

- 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)**

\[
R_{SI-D}(D) = \min(I(X; U) - I(Y; U)) = \min I(X; U|Y),
\]

where the minimum is over all \(p(u|x)\) and \(\hat{x}(u, y)\) such that \(E(d(X, \hat{X})) \leq D\)
Proof of Achievability

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

\[ y^n \]
\[ x^n \]
\[ u^n(1) \]
\[ B(1) \]
\[ \cdots \]
\[ B(m) \]
\[ u^n(1) \]
\[ \cdots \]
\[ B(2^{nR}) \]
\[ u^n(2^{nR}) \]
\[ T^{(n)}_e(U,Y) \]
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

\[
E = \{(U^n(\hat{L}), X^n, Y^n) \notin T_{\epsilon}^{(n)}\}
\]

and consider

\[
E_1 = \{(U^n(l), X^n) \notin T_{\epsilon}^{(n)} \text{ for all } l \in [1 : 2^{n\tilde{R}}]\}
\]

\[
E_2 = \{(U^n(L), X^n, Y^n) \notin T_{\epsilon}^{(n)}\}
\]

\[
E_3 = \{(U^n(\tilde{l}), Y^n) \in T_{\epsilon}^{(n)} \text{ for some } \tilde{l} \in \mathcal{B}(M), \tilde{l} \neq L\}
\]

- The probability of “error” is bounded as

\[
P(E) \leq P(E_1) + P(E_1^c \cap E_2) + P(E_3)
\]
\( \mathcal{E}_1 = \{(U^n(l), X^n) \notin T^{(n)}_{\epsilon l} \text{ for all } l \in [1 : 2^n\tilde{R}]\} \)

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

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

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

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

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

by the conditional typicality lemma, \( P(\mathcal{E}_1^c \cap \mathcal{E}_2) \to 0 \text{ as } n \to \infty \)

- To bound \( P(\mathcal{E}_3) \), it can be shown that

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

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

by the packing lemma, \( P(\mathcal{E}_3) \to 0 \text{ as } n \to \infty \text{ if } \tilde{R} - R < I(Y; U) - \delta(\epsilon) \)

- Combining the bounds, we have shown that \( P(\mathcal{E}) \to 0 \text{ as } n \to \infty \text{ if } R > I(X; U) - I(Y; U) + \delta(\epsilon) + \delta(\epsilon') = R_{SI-D}(D/(1 + \epsilon)) + \delta(\epsilon) + \delta(\epsilon') \)
What is the minimum rate $R_{\text{SI-D}}^*$ needed to recover $X$ \textit{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 $(1/n) \sum_{i=1}^{n} P\{\hat{X}_i \neq X_i\} \leq P\{\hat{X}^n \neq X^n\}$.
  - $R_{\text{SI-D}}^* \leq R_{\text{SI-D}}(0)$ by Wyner–Ziv coding with $\hat{X} = U = X$. 

El Gamal & Kim (Stanford & UCSD)
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

![Diagram of the communication system with state](image)

- Assume a DMC with DM state model \( (\mathcal{X} \times S, p(y|x, s)p(s), \mathcal{Y}) \)
  - 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:

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

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

Theorem (Gelfand–Pinsker 1980)

\[ C_{SI-E}(B) = \max_{p(u|s), x(u,s):\mathbb{E}(b(X)) \leq B} \left( I(U; Y) - I(U; S) \right) \]
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:

  $$\mathcal{E}_1 = \{(U^n(l), S^n) \notin T_e^{(n)} \text{ for all } U^n(l) \in \mathcal{C}(1)\}$$

  $$\mathcal{E}_2 = \{(U^n(L), Y^n) \notin T_e^{(n)}\}$$

  $$\mathcal{E}_3 = \{(U^n(l), Y^n) \in T_e^{(n)} \text{ for some } U^n(l) \notin \mathcal{C}(1)\}$$

Thus, the probability of error is bounded as

$$P(\mathcal{E}) \leq P(\mathcal{E}_1) + P(\mathcal{E}_1^c \cap \mathcal{E}_2) + P(\mathcal{E}_3)$$
\[ \mathcal{E}_1 = \{(U^n(l), S^n) \notin \mathcal{T}_{\epsilon'}(n) \text{ for all } U^n(l) \in C(1)\} \]
\[ \mathcal{E}_2 = \{(U^n(l), Y^n) \notin \mathcal{T}_{\epsilon}(n)\} \]
\[ \mathcal{E}_3 = \{(U^n(l), Y^n) \in \mathcal{T}_{\epsilon}(n) \text{ for some } U^n(l) \notin C(1)\} \]

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

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

Remark: \( Y^n \) is not i.i.d.
- Combining the bounds, we have shown that \( P(\mathcal{E}) \to 0 \) as \( n \to \infty \) if \( R < I(U; Y) - I(U; S) - \delta(\epsilon) - \delta(\epsilon') = C_{SI-E}(B/(1 + \epsilon')) - \delta(\epsilon) - \delta(\epsilon') \)
## Multicoding versus Binning

<table>
<thead>
<tr>
<th><strong>Multicoding</strong></th>
<th><strong>Binning</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Channel coding technique</strong></td>
<td><strong>Source coding technique</strong></td>
</tr>
<tr>
<td>- Given a set of messages</td>
<td>- Given a set of indices (sequences)</td>
</tr>
<tr>
<td>- Generate many codewords for each message</td>
<td>- Map indices into a smaller number of bins</td>
</tr>
<tr>
<td>- To communicate a message, send a codeword from its subcodebook</td>
<td>- To communicate an index, send its bin index</td>
</tr>
</tbody>
</table>
Wyner–Ziv versus Gelfand–Pinsker

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

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

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:

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

We proved achievability using multicoding, covering, and packing.

**Dualities**:

- $\min \leftrightarrow \max$
- $\text{binning} \leftrightarrow \text{multicoding}$
- $\text{covering rate} - \text{packing rate} \leftrightarrow \text{packing rate} - \text{covering rate}$
Gelfand–Pinsker Coding

Writing on Dirty Paper

- Gaussian channel with additive Gaussian state available at the encoder

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

- Assume expected average power constraint: $\sum_{i=1}^{n} E(x_i^2(m, S^n)) \leq nP$ for every $m$

- $C = \frac{1}{2} \log \left( 1 + \frac{P}{N+Q} \right)$

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

Writing on Dirty Paper (Costa 1983)

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

- Proof involves a clever choice of $F(u|s)$, $x(u, s)$ and discretization procedure.
- Let $X \sim N(0, P)$ 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 $M \sim \text{Unif}[1:2^{nR}]$

Average probability of error: $P_e^{(n)} = P\{\hat{M} \neq M\}$

Information leakage rate: $R_L^{(n)} = (1/n)I(M; Z^n)$

$(R, R_L)$ achievable if $\exists (2^{nR}, n)$ codes with $\lim_{n \to \infty} P_e^{(n)} = 0$, $\limsup_{n \to \infty} R_L^{(n)} \leq R_L$

Rate–leakage region $\mathcal{R}^*$: closure of the set of achievable $(R, R_L)$

Secrecy capacity: $C_S = \max\{R: (R, 0) \in \mathcal{R}^*\}$

**Theorem (Wyner 1975, Csiszár–Körner 1978)**

$$C_S = \max_{p(u,x)} \left( I(U; Y) - I(U; Z) \right)$$
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 $C(m)$ consisting of $2^{n(\bar{R} - R)}$ randomly and independently generated sequences $u^n(l) \sim \prod_{i=1}^{n} p_{U}(u_i),
  l \in [(m - 1)2^{n(\bar{R} - R)} + 1 : m2^{n(\bar{R} - R)}]$

\[
\begin{array}{cccc}
  l : 1 & \cdots & 2^{n(\bar{R} - R)} & \cdots & 2^{n\bar{R}} \\
  \circ \circ \circ \circ \cdots \circ & \circ \circ \circ \cdots \circ & \circ \circ \circ \cdots \circ & \cdots & \circ \circ \circ \cdots \circ \\
  C(1) & C(2) & C(3) & \cdots & C(2^nR)
\end{array}
\]

**Encoding:**

- To send $m$, choose an index $L \in [(m - 1)2^{n(\bar{R} - R)} + 1 : m2^{n(\bar{R} - R)}]$ 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 $(u^n(\hat{l}), y^n) \in T_\varepsilon^{(n)}$ for some $u^n(\hat{l}) \in C(\hat{m})$
  - By the LLN and the packing lemma, $P(\mathcal{E}) \to 0$ as $n \to \infty$ if $\bar{R} < I(U; Y) - \delta(\varepsilon)$
Analysis of the Information Leakage Rate

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

$$l : 1 \cdots 2^n(\tilde{R} - R)$$

\[
\begin{array}{cccc}
\circ \circ \circ \circ \circ \circ & \cdots & \circ \circ \circ \circ \circ \circ & \circ \circ \circ \circ \circ \circ \\
C(1) & C(2) & C(3) & \cdots & C(2^nR)
\end{array}
\]

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 $C$ induces a pmf of the form

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

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:

\[ I(M; Z^n | C) = H(M|C) - H(M|Z^n, C) \]
\[ = nR - H(M, L|Z^n, C) + H(L|Z^n, M, C) \]
\[ = nR - H(L|Z^n, C) + H(L|Z^n, M, C) \]

The first equivocation term

\[ H(L|Z^n, C) = H(L|C) - I(L; Z^n | C) \]
\[ = n\tilde{R} - I(L; Z^n | C) \]
\[ = n\tilde{R} - I(U^n, L; Z^n | C) \]
\[ \geq n\tilde{R} - I(U^n, L, C; Z^n) \]
\[ (a) \quad n\tilde{R} - I(U^n; Z^n) \]
\[ = n\tilde{R} - nI(U; Z) \]

\((a) \quad (L, C) \rightarrow U^n \rightarrow Z^n\) form a Markov chain
Consider the amount of information leakage averaged over codebooks:

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

The remaining equivocation term can be upper bounded as follows:

**Lemma**

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

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

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

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

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 $$R_L^{(n)} \leq \delta(\epsilon)$$ 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
   - Randomized encoding
   - Bound on equivocation (list size)
9. Relay Channel
10. Multicast Network
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**
The relay recovers the message received from the sender in each block and retransmits it in the following block.

\[ M \rightarrow X_1 \rightarrow Y_2 : X_2 \rightarrow \hat{M} \]

\[ M \rightarrow X_1 \rightarrow Y_3 \rightarrow \hat{M} \]

Multihop Lower Bound

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

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

<table>
<thead>
<tr>
<th>$m_1$</th>
<th>$m_2$</th>
<th>$m_3$</th>
<th>⋯</th>
<th>$m_{b-1}$</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1</td>
<td>2</td>
<td>3</td>
<td>⋯</td>
<td>$b - 1$</td>
<td>$b$</td>
</tr>
</tbody>
</table>

**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: $C_j = \{(x_1^n(m_j), x_2^n(m_{j-1})) : m_{j-1}, m_j \in [1:2^{nR}]\}, j \in [1:b]$

**Encoding:**

- To send $m_j$ in block $j$, transmit $x_1^n(m_j)$ from $C_j$

**Relay encoding:**

- At the end of block $j$, find the unique $\tilde{m}_j$ such that $(x_1^n(\tilde{m}_j), x_2^n(\tilde{m}_{j-1}), y_2^n(j)) \in T_{e}^{(n)}$
- In block $j + 1$, transmit $x_2^n(\tilde{m}_j)$ from $C_{j+1}$

**Decoding:**

- At the end of block $j + 1$, find the unique $\hat{m}_j$ such that $(x_2^n(\hat{m}_j), y_3^n(j + 1)) \in T_{e}^{(n)}$
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:

$$\tilde{E}_1(j) = \{(X_1^n(1), X_2^n(\tilde{M}_{j-1}), Y_2^n(j)) \not\in T_e^{(n)}\}$$

$$\tilde{E}_2(j) = \{(X_1^n(m_j), X_2^n(\tilde{M}_{j-1}), Y_2^n(j)) \in T_e^{(n)} \text{ for some } m_j \neq 1\}$$

$$E_1(j) = \{(X_2^n(\tilde{M}_j), Y_3^n(j+1)) \not\in T_e^{(n)}\}$$

$$E_2(j) = \{(X_2^n(m_j), Y_3^n(j+1)) \in T_e^{(n)} \text{ for some } m_j \neq \tilde{M}_j\}$$

Thus, the probability of error is upper bounded as

$$P(\mathcal{E}(j)) = P\{\hat{M}_j \neq 1\} \leq P(\tilde{E}_1(j)) + P(\tilde{E}_2(j)) + P(E_1(j)) + P(E_2(j))$$
\[ \tilde{E}_1(j) = \{(X_1^n(1), X_2^n(\tilde{M}_{j-1}), Y_2^n(j)) \notin \mathcal{T}_\varepsilon^{(n)}\} \]
\[ \tilde{E}_2(j) = \{(X_1^n(m_j), X_2^n(\tilde{M}_{j-1}), Y_2^n(j)) \in \mathcal{T}_\varepsilon^{(n)} \text{ for some } m_j \neq 1\} \]
\[ E_1(j) = \{(X_2^n(\tilde{M}_j), Y_3^n(j + 1)) \notin \mathcal{T}_\varepsilon^{(n)}\} \]
\[ E_2(j) = \{(X_2^n(m_j), Y_3^n(j + 1)) \in \mathcal{T}_\varepsilon^{(n)} \text{ for some } m_j \neq \tilde{M}_j\} \]

- By the independence of the codebooks, \( \tilde{M}_{j-1} \), which is a function of \( Y_2^n(j-1) \) and codebook \( C_{j-1} \), is independent of the codewords \( X_1^n(1), X_2^n(\tilde{M}_{j-1}) \) in \( C_j \).
  Thus by the LLN, \( P(\tilde{E}_1(j)) \rightarrow 0 \) as \( n \rightarrow \infty \).
- By the packing lemma, \( P(\tilde{E}_2(j)) \rightarrow 0 \) as \( n \rightarrow \infty \) if \( R < I(X_1; Y_2|X_2) - \delta(\varepsilon) \).
- By the independence of the codebooks and the LLN, \( P(E_1(j)) \rightarrow 0 \) as \( n \rightarrow \infty \).
- By the same independence and the packing lemma, \( P(E_2(j)) \rightarrow 0 \) as \( n \rightarrow \infty \) if \( R < I(X_2; Y_3) - \delta(\varepsilon) \).
- Thus we have shown that under the given constraints on the rate, \( P\{\tilde{M}_j \neq M_j\} \rightarrow 0 \) as \( n \rightarrow \infty \) 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.

\[ Y_2 : X_2 \]

\[ M \rightarrow X_1 \rightarrow Y_3 \rightarrow \hat{M} \]

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

\[ C \geq \max_{p(x_1, x_2)} \min\{I(X_2; Y_3), I(X_1; Y_2|X_2)\} \]
Relay Channel  
Coherent Multihop

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
  \[ x_2^n(m_{j-1}) \sim \prod_{i=1}^{n} p_{X_2}(x_{2i}), \quad m_{j-1} \in [1:2^{nR}] \]
- For each \( m_{j-1} \in [1:2^{nR}] \), randomly and conditionally independently generate \( 2^{nR} \) sequences
  \[ x_1^n(m_j|m_{j-1}) \sim \prod_{i=1}^{n} p_{X_1|X_2}(x_{1i}|x_{2i}(m_{j-1})), \quad m_j \in [1:2^{nR}] \]
- Codebooks: \( C_j = \{(x_1^n(m_j|m_{j-1}), x_2^n(m_{j-1})) : m_{j-1}, m_j \in [1:2^{nR}]\}, \quad j \in [1:b] \)
### Encoding:
- In block $j$, transmit $x_1^n(m_j|m_{j-1})$ from codebook $C_j$

### Relay encoding:
- At the end of block $j$, find the unique $\tilde{m}_j$ such that 
\[(x_1^n(\tilde{m}_j|\tilde{m}_{j-1}), x_2^n(\tilde{m}_{j-1}), y_2^n(j)) \in T(e(n))\]
- In block $j + 1$, transmit $x_2^n(\tilde{m}_j)$ from codebook $C_{j+1}$

### Decoding:
- At the end of block $j + 1$, find unique message $\hat{m}_j$ such that 
\[(x_2^n(\hat{m}_j), y_3^n(j + 1)) \in T(e(n))\]
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:

$$\tilde{E}(j) = \{\tilde{M}_j \neq 1\}$$
$$E_1(j) = \{(X_2^n(\tilde{M}_j), Y_3^n(j + 1)) \notin T_e^{(n)}\}$$
$$E_2(j) = \{(X_2^n(m_j), Y_3^n(j + 1)) \in T_e^{(n)} \text{ for some } m_j \neq \tilde{M}_j\}$$

Thus, the probability of error is upper bounded as

$$P(E(j)) = P(\tilde{M}_j \neq 1) \leq P(\tilde{E}(j)) + P(E_1(j)) + P(E_2(j))$$

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

- To upper bound $P(\tilde{E}(j)) = P\{\tilde{M}_j \neq 1\}$, define
  \[
  \tilde{E}_1(j) = \{(X_1^n(1|\tilde{M}_{j-1}), X_2^n(\tilde{M}_{j-1}), Y_2^n(j)) \notin T^{(n)}_e\}
  \]
  \[
  \tilde{E}_2(j) = \{(X_1^n(m_{j}|\tilde{M}_{j-1}), X_2^n(\tilde{M}_{j-1}), Y_2^n(j)) \in T^{(n)}_e \text{ for some } m_{j} \neq 1\}
  \]

- Then
  \[
P(\tilde{E}(j)) \leq P(\tilde{E}(j - 1)) + P(\tilde{E}_1(j) \cap \tilde{E}^c(j - 1)) + P(\tilde{E}_2(j))
  \]

- Consider the second term
  \[
P(\tilde{E}_1(j) \cap \tilde{E}^c(j - 1)) = P\{(X_1^n(1|\tilde{M}_{j-1}), X_2^n(\tilde{M}_{j-1}), Y_2^n(j)) \notin T^{(n)}_e, \tilde{M}_{j-1} = 1\}
  \leq P\{(X_1^n(1|1), X_2^n(1), Y_2^n(j)) \notin T^{(n)}_e | \tilde{M}_{j-1} = 1\},
  \]

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

- By the packing lemma, $P(\tilde{E}_2(j)) \to 0$ as $n \to \infty$ if $R < I(X_1; Y_2|X_2) - \delta(e)$

- Since $\tilde{M}_0 = 1$, by induction, $P(\tilde{E}(j)) \to 0$ as $n \to \infty$ for every $j \in [1:b - 1]$

- Thus we have shown that under the given constraints on the rate, $P\{\hat{M}_j \neq M_j\} \to 0$ as $n \to \infty$ for every $j \in [1:b - 1]$
Coherent multihop can be further improved by combining the information through the direct path with the information from the relay.

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

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:  
    \[ C_j = \{(x_1^n(m_j|m_{j-1}), x_2^n(m_{j-1})) : m_{j-1}, m_j \in [1:2^{nR}]\}, \ j \in [1:b] \]

<table>
<thead>
<tr>
<th>Block</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>...</th>
<th>b − 1</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X_1)</td>
<td>(x_1^n(m_1</td>
<td>1))</td>
<td>(x_1^n(m_2</td>
<td>m_1))</td>
<td>(x_1^n(m_3</td>
<td>m_2))</td>
</tr>
<tr>
<td>(Y_2)</td>
<td>(\hat{m}_1 \rightarrow \hat{m}_2 \rightarrow \hat{m}<em>3 \rightarrow \ldots \rightarrow \hat{m}</em>{b-1} \rightarrow 0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(X_2)</td>
<td>(x_2^n(1))</td>
<td>(x_2^n(\hat{m}_1))</td>
<td>(x_2^n(\hat{m}_2))</td>
<td>...</td>
<td>(x_2^n(\hat{m}_{b-2}))</td>
<td>(x_2^n(\hat{m}_{b-1}))</td>
</tr>
<tr>
<td>(Y_3)</td>
<td>(\emptyset)</td>
<td>(\hat{m}_1)</td>
<td>(\leftarrow \hat{m}_2)</td>
<td>...</td>
<td>(\leftarrow \hat{m}_{b-2})</td>
<td>(\leftarrow \hat{m}_{b-1})</td>
</tr>
</tbody>
</table>

**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 
  \[(x_1^n(\hat{m}_{j+1}|\hat{m}_j), x_2^n(\hat{m}_j), y_3^n(j + 1)) \in \mathcal{T}_\epsilon^{(n)}, \text{ successively with the initial condition } \hat{m}_b = 1\]
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:

  \[ \tilde{E}(j) = \{ \tilde{M}_j \neq 1 \} \]
  \[ E(j + 1) = \{ \hat{M}_{j+1} \neq 1 \} \]
  \[ E_1(j) = \{ (X_1^n(\hat{M}_{j+1} | \tilde{M}_j), X_2^n(\tilde{M}_j), Y_3^n(j + 1)) \notin T^{(n)}_\epsilon \} \]
  \[ E_2(j) = \{ (X_1^n(\hat{M}_{j+1} | m_j), X_2^n(m_j), Y_3^n(j + 1)) \in T^{(n)}_\epsilon \text{ for some } m_j \neq \tilde{M}_j \} \]

  Thus, the probability of error is upper bounded as

  \[
P(\mathcal{E}(j)) = P(\hat{M}_j \neq 1) \leq P(\tilde{E}(j) \cup E(j + 1) \cup E_1(j) \cup E_2(j)) \leq P(\tilde{E}(j)) + P(E(j + 1)) + P(E_1(j) \cap \tilde{E}^c(j) \cap E^c(j + 1)) + P(E_2(j))\]

- As in the coherent multihop scheme, the first term $\to 0$ as $n \to \infty$ if $R < I(X_1; Y_2 | X_2) - \delta(\epsilon)$.
\( \mathcal{E}(j) = \{ \tilde{M}_j \neq 1 \} \)

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

\( \mathcal{E}_1(j) = \{(X_1^n(\hat{M}_{j+1}|\tilde{M}_j), X_2^n(\tilde{M}_j), Y_3^n(j+1)) \notin \mathcal{T}_e^{(n)} \} \)

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

\[
P(\mathcal{E}(j)) \leq P(\tilde{\mathcal{E}}(j)) + P(\mathcal{E}(j+1)) + P(\mathcal{E}_1(j) \cap \tilde{\mathcal{E}}^c(j) \cap \mathcal{E}^c(j+1)) + P(\mathcal{E}_2(j))
\]

- The third term is upper bounded as

\[
P(\mathcal{E}_1(j) \cap \{ \hat{M}_{j+1} = 1 \} \cap \{ \tilde{M}_j = 1 \}) = P\{(X_1^n(1|1), X_2^n(1), Y_3^n(j+1)) \notin \mathcal{T}_e^{(n)}, \hat{M}_{j+1} = 1, \tilde{M}_j = 1 \} \leq P\{(X_1^n(1|1), X_2^n(1), Y_3^n(j+1)) \notin \mathcal{T}_e^{(n)} | \tilde{M}_j = 1 \},
\]

which, by the independence of the codebooks and the LLN, \( \rightarrow 0 \) as \( n \rightarrow \infty \)

- By the same independence and the packing lemma, the fourth term \( P(\mathcal{E}_2(j)) \rightarrow 0 \) as \( n \rightarrow \infty \) 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, \( P\{\hat{M}_j = M_j\} \rightarrow 0 \) as \( n \rightarrow \infty \) 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

\[
Y_2 : X_2 \quad \quad \hat{Y}_2
\]

\[
M \rightarrow X_1 \rightarrow Y_3 \rightarrow \hat{M}
\]

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

\[
C \geq \max_{p(x_1)p(x_2)p(\hat{y}_2|y_2,x_2)} \min\{I(X_1, X_2; Y_3) - I(Y_2; \hat{Y}_2|X_1, X_2, Y_3), I(X_1; \hat{Y}_2, Y_3|X_2)\}
\]
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: $C_j = \{(x_1^n(m_j), x_2^n(l_{j-1})): \ m_j \in [1:2^{nR}], \ l_{j-1} \in [1:2^{nR_2}]\}, \ j \in [1:b]$
  - Partition the set $[1:2^{n\tilde{R}_2}]$ into $2^{nR_2}$ equal-size bins $B(l_j), \ l_j \in [1:2^{nR_2}]$
### Encoding:

- Transmit $x_1^n(m_j)$ from codebook $C_j$

### Relay Encoding:

- At the end of block $j$, find an index $k_j$ such that $(y_2^n(j), \hat{y}_2^n(k_j|l_{j-1}), x_2^n(l_{j-1})) \in \mathcal{T}_e^{(n)}$
- 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 $(x_2^n(\hat{l}_j), y_3^n(j + 1)) \in \mathcal{T}_e^{(n)}$
- Find the unique $\hat{m}_j$ such that $(x_1^n(\hat{m}_j), x_2^n(\hat{l}_{j-1}), \hat{y}_2^n(\hat{k}_j|\hat{l}_{j-1}), y_3^n(j)) \in \mathcal{T}_e^{(n)}$ for some $\hat{k}_j \in \mathcal{B}(\hat{l}_j)$

<table>
<thead>
<tr>
<th>Block</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>...</th>
<th>$b - 1$</th>
<th>$b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>$x_1^n(m_1)$</td>
<td>$x_1^n(m_2)$</td>
<td>$x_1^n(m_3)$</td>
<td>...</td>
<td>$x_1^n(m_{b-1})$</td>
<td>$x_1^n(1)$</td>
</tr>
<tr>
<td>$Y_2$</td>
<td>$\hat{y}_2^n(k_1</td>
<td>1), l_1$</td>
<td>$\hat{y}_2^n(k_2</td>
<td>l_1), l_2$</td>
<td>$\hat{y}_2^n(k_3</td>
<td>l_2), l_3$</td>
</tr>
<tr>
<td>$X_2$</td>
<td>$x_2^n(1)$</td>
<td>$x_2^n(l_1)$</td>
<td>$x_2^n(l_2)$</td>
<td>...</td>
<td>$x_2^n(l_{b-2})$</td>
<td>$x_2^n(l_{b-1})$</td>
</tr>
<tr>
<td>$Y_3$</td>
<td>$\emptyset$</td>
<td>$\hat{l}_1, \hat{k}_1, \hat{m}_1$</td>
<td>$\hat{l}_2, \hat{k}_2, \hat{m}_2$</td>
<td>...</td>
<td>$\hat{l}<em>{b-2}, \hat{k}</em>{b-2}, \hat{m}_{b-2}$</td>
<td>$\hat{l}<em>{b-1}, \hat{k}</em>{b-1}, \hat{m}_{b-1}$</td>
</tr>
</tbody>
</table>
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:

  $\tilde{E}(j) = \{(X_2^n(L_{j-1}), \hat{Y}_2^n(k_j|L_{j-1}), Y_2^n(j)) \notin T_e^{(n)} \text{ for all } k_j \in [1:2^{n\tilde{R}_2}]\}$

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

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

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

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

  $\mathcal{E}_4(j) = \{(X_1^n(m_j), X_2^n(\hat{L}_{j-1}), \hat{Y}_2^n(\hat{k}_j|\hat{L}_{j-1}), Y_3^n(j)) \in T_e^{(n)}$

  for some $\hat{k}_j \in \mathcal{B}(\hat{L}_j), \hat{k}_j \neq K_j, m_j \neq 1\}$

Thus, the probability of error is bounded as

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

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

$$+ P(\mathcal{E}_3(j)) + P(\mathcal{E}_4(j) \cap \mathcal{E}_1^c(j-1) \cap \mathcal{E}_1^c(j))$$
\[
\tilde{E}(j) = \{(X_2^n(L_{j-1}), \hat{Y}_2^n(k_j|L_{j-1}), Y_2^n(j)) \notin \mathcal{T}_e^{(n)} \text{ for all } k_j \in [1 : 2^{n\tilde{R}_2}]\}
\]

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

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

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

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

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

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

\[
+ P(\mathcal{E}_3(j)) + P(\mathcal{E}_4(j) \cap \mathcal{E}_1^c(j-1) \cap \mathcal{E}_1^c(j))
\]

- 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 \(\rightarrow 0\) as \(n \rightarrow \infty\) if \(\tilde{R}_2 > I(\hat{Y}_2; Y_2|X_2) + \delta(e')\)

- As in the multihop coding scheme, the next two terms \(P\{\hat{L}_{j-1} \neq L_{j-1}\} \rightarrow 0\) and \(P\{\hat{L}_j \neq L_j\} \rightarrow 0\) as \(n \rightarrow \infty\) if \(R_2 < I(X_2; Y_3) - \delta(e)\)

- The fourth term \(\leq P\{(X_1^n(1), X_2^n(L_{j-1}), \hat{Y}_2^n(K_j|L_{j-1}), Y_3^n(j)) \notin \mathcal{T}_e^{(n)} | \hat{E}^c(j)\} \rightarrow 0\) 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

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

Let \(\hat{X}^n(m) \sim \prod_{i=1}^{n} p_{\hat{X}|U}(\hat{x}_i|u_i), m \in A\), where \(|A| \geq 2^{nR}\), 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

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

if \(R > I(X; \hat{X}|U) + \delta(\epsilon)\)
\[ \tilde{E}(j) = \{(X_2^n(L_{j-1}), \hat{Y}_2^n(k_j|L_{j-1}), Y_2^n(j)) \notin \mathcal{T}_\varepsilon^{(n)} \text{ for all } k_j \in [1:2^n\tilde{R}_2]\} \]

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

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

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

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

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

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

\[ + P(\mathcal{E}_3(j)) + P(\mathcal{E}_4(j) \cap \mathcal{E}_1^c(j - 1) \cap \mathcal{E}_1^c(j)) \]

- 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 \(\rightarrow 0\) as \(n \rightarrow \infty\) if \(\tilde{R}_2 > I(\hat{Y}_2; Y_2|X_2) + \delta(\varepsilon')\)

- As in the multihop coding scheme, the next two terms \(P\{\hat{L}_{j-1} \neq L_{j-1}\} \rightarrow 0\) and \(P\{\hat{L}_j \neq L_j\} \rightarrow 0\) as \(n \rightarrow \infty\) if \(R_2 < I(X_2; Y_3) - \delta(\varepsilon)\)

- The fourth term \(\leq P\{(X_1^n(1), X_2^n(L_{j-1}), \hat{Y}_2^n(K_j|L_{j-1}), Y_3^n(j)) \notin \mathcal{T}_\varepsilon^{(n)} | \tilde{E}^c(j)\} \rightarrow 0\) by the independence of codebooks and the conditional typicality lemma
\[ \mathcal{E}(j) = \{ (X_2^n(L_{j-1}), \hat{Y}_2^n(k_j|L_{j-1}), Y_2^n(j)) \notin \mathcal{T}_e^{(n)} \text{ for all } k_j \in [1:2^n]\tilde{R}_2] \}
\]

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

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

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

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

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

\[
\Pr(\mathcal{E}(j)) \leq \Pr(\mathcal{E}(j)) + \Pr(\mathcal{E}_1(j-1)) + \Pr(\mathcal{E}_1(j)) + \Pr(\mathcal{E}_2(j) \cap \mathcal{E}_c(j) \cap \mathcal{E}_1^c(j-1)) \\
+ \Pr(\mathcal{E}_3(j)) + \Pr(\mathcal{E}_4(j) \cap \mathcal{E}_1^c(j-1) \cap \mathcal{E}_1^c(j))
\]

By the same independence and the packing lemma, \( \Pr(\mathcal{E}_3(j)) \to 0 \) as \( n \to \infty \) if \( R < I(X_1; X_2, \hat{Y}_2, Y_3) + \delta(e) = I(X_1; \hat{Y}_2, Y_3|X_2) + \delta(e) \)

As in Wyner–Ziv coding, the last term
\[
\leq \Pr\{ (X_1^n(m_j), X_2^n(L_{j-1}), \hat{Y}_2^n(\hat{k}_j|L_{j-1}), Y_3^n(j)) \in \mathcal{T}_e^{(n)} \text{ for some } \hat{k}_j \in \mathcal{B}(1), m_j \neq 1 \},
\]
which, by the independence of codebooks, joint typicality lemma, and union bound, \( \to 0 \) as \( n \to \infty \) if \( R + \tilde{R}_2 - R_2 < I(X_1; Y_3|X_2) + I(\hat{Y}_2; X_1, Y_3|X_2) - \delta(e) \)
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 $(\times_{j=1}^{N} X_j, p(y^N|x^N), \times_{j=1}^{N} Y_j)$

- **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: $P_{e}^{(n)} = P\{\hat{M}_k \neq M \text{ for some } k \in D\}$

$R$ 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 $D = \{2\}$)
- DM-RC ($N = 3$, $X_3 = Y_1 = \emptyset$, and $D = \{3\}$)
- Common-message DM-BC ($X_2 = \cdots = X_N = Y_1 = \emptyset$ and $D = [2 : N]$)
- DM unicast network ($D = \{N\}$)
Network Decode–Forward

- Decode–forward for RC can be extended to MN

![Diagram]

$M_j \rightarrow X_1 \rightarrow Y_2 : X_2 \rightarrow M_{j-1} \rightarrow Y_3 : X_3 \rightarrow M_{j-2} \rightarrow Y_4 \rightarrow \hat{M}_{j-2}$

Network Decode–Forward Lower Bound

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

- 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., $p(y^N_{k+2}|x^N, y^{k+1}) = p(y^N_{k+2}|x^N_{k+1}, y_{k+1})$
- 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

<table>
<thead>
<tr>
<th>Block</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>...</th>
<th>b−1</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>$x_1^n(m_1</td>
<td>1)$</td>
<td>$x_1^n(m_2</td>
<td>m_1)$</td>
<td>$x_1^n(m_3</td>
<td>m_2)$</td>
</tr>
<tr>
<td>$Y_2$</td>
<td>$\hat{m}_1$</td>
<td>$\hat{m}_2$</td>
<td>$\hat{m}_3$</td>
<td>...</td>
<td>$\hat{m}_{b−1}$</td>
<td>$\emptyset$</td>
</tr>
<tr>
<td>$X_2$</td>
<td>$x_2^n(1)$</td>
<td>$x_2^n(\hat{m}_1)$</td>
<td>$x_2^n(\hat{m}_2)$</td>
<td>...</td>
<td>$x_2^n(\hat{m}_{b−2})$</td>
<td>$x_2^n(\hat{m}_{b−1})$</td>
</tr>
<tr>
<td>$Y_3$</td>
<td>$\emptyset$</td>
<td>$\hat{m}_1$</td>
<td>$\hat{m}_2$</td>
<td>...</td>
<td>$\hat{m}_{b−2}$</td>
<td>$\hat{m}_{b−1}$</td>
</tr>
</tbody>
</table>

**Decoding:**

- At the end of block $j + 1$, find the unique $\hat{m}_j$ such that

$$(x_1^n(\hat{m}_j|\hat{m}_{j−1}), x_2^n(\hat{m}_{j−1}), y_3^n(j)) \in \mathcal{T}_e^{(n)} \text{ and } (x_2^n(\hat{m}_j), y_3^n(j + 1)) \in \mathcal{T}_e^{(n)} \text{ simultaneously}$$
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) = \{(X_1^n(\tilde{M}_j | \tilde{M}_{j-1}), X_2^n(\tilde{M}_j - 1), Y_3^n(j)) \notin \mathcal{T}_e^{(n)} \text{ or } (X_2^n(\tilde{M}_j), Y_3^n(j + 1)) \notin \mathcal{T}_e^{(n)}\} \\
  \mathcal{E}_2(j) = \{(X_1^n(m_j | \tilde{M}_{j-1}), X_2^n(\tilde{M}_j - 1), Y_3^n(j)) \in \mathcal{T}_e^{(n)} \text{ and } (X_2^n(m_j), Y_3^n(j + 1)) \in \mathcal{T}_e^{(n)}\}
  
  \text{for some } m_j \neq \tilde{M}_j\} \\

  \text{Thus, the probability of error is upper bounded as}
  \[
  P(\mathcal{E}(j)) \leq P(\tilde{\mathcal{E}}(j - 1) \cup \tilde{\mathcal{E}}(j) \cup \mathcal{E}(j - 1) \cup \mathcal{E}_1(j) \cup \mathcal{E}_2(j)) \\
  \leq P(\tilde{\mathcal{E}}(j - 1)) + P(\tilde{\mathcal{E}}(j)) + P(\mathcal{E}(j - 1)) \\
  + P(\mathcal{E}_1(j) \cap \tilde{\mathcal{E}}^c(j - 1) \cap \tilde{\mathcal{E}}^c(j) \cap \mathcal{E}_1^c(j - 1)) + P(\mathcal{E}_2(j) \cap \tilde{\mathcal{E}}^c(j)) \\
  \]

- 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

\[
P(\mathcal{E}_2(j) \cap \tilde{\mathcal{E}}^c(j)) = P\{ (X_1^n(m_j | \hat{M}_{j-1}), X_2^n(\hat{M}_{j-1}), Y_3^n(j)) \in \mathcal{T}_{\varepsilon}^{(n)},

(X_2^n(m_j), Y_3^n(j + 1)) \in \mathcal{T}_{\varepsilon}^{(n)} \text{ for some } m_j \neq 1, \text{ and } \tilde{M}_j = 1 \}
\]

\[
\leq \sum_{m_j \neq 1} P\{ (X_1^n(m_j | \hat{M}_{j-1}), X_2^n(\hat{M}_{j-1}), Y_3^n(j)) \in \mathcal{T}_{\varepsilon}^{(n)},

(X_2^n(m_j), Y_3^n(j + 1)) \in \mathcal{T}_{\varepsilon}^{(n)}, \text{ and } \tilde{M}_j = 1 \}
\]

\[
= \sum_{m_j \neq 1} P\{ (X_1^n(m_j | \hat{M}_{j-1}), X_2^n(\hat{M}_{j-1}), Y_3^n(j)) \in \mathcal{T}_{\varepsilon}^{(n)} \text{ and } \tilde{M}_j = 1 \}
\]

\[
\cdot P\{ (X_2^n(m_j), Y_3^n(j + 1)) \in \mathcal{T}_{\varepsilon}^{(n)} | \tilde{M}_j = 1 \}
\]

\[
\leq \sum_{m_j \neq 1} P\{ (X_1^n(m_j | \hat{M}_{j-1}), X_2^n(\hat{M}_{j-1}), Y_3^n(j)) \in \mathcal{T}_{\varepsilon}^{(n)} \}
\]

\[
\cdot P\{ (X_2^n(m_j), Y_3^n(j + 1)) \in \mathcal{T}_{\varepsilon}^{(n)} | \tilde{M}_j = 1 \}
\]

\[
\leq 2^{nR} 2^{-n(I(X_1; Y_3 | X_2) - \delta(\varepsilon))} 2^{-n(I(X_2; Y_3) - \delta(\varepsilon))}
\]

\[
\rightarrow 0 \text{ as } n \rightarrow \infty \text{ if } R < I(X_1; Y_3 | X_2) + I(X_2; Y_3) - 2\delta(\varepsilon) = I(X_1, X_2; Y_3) - 2\delta(\varepsilon)
\]

\[(a) \quad \{(X_1^n(m_j | \hat{M}_{j-1}), X_2^n(\hat{M}_{j-1}), Y_3^n(j)) \in \mathcal{T}_{\varepsilon}^{(n)}\} \text{ and } \{(X_2^n(m_j), Y_3^n(j + 1)) \in \mathcal{T}_{\varepsilon}^{(n)}\} \text{ are}
\]

\[
\text{conditionally independent given } \tilde{M}_j = 1 \text{ for } m_j \neq 1
\]

\[(b) \quad \text{independence of the codebooks and the joint typicality lemma}
\]
Noisy Network Coding

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

**Theorem (Noisy Network Coding Lower Bound)**

\[
C \geq \max_{k \in D} \min_{S:1 \in S,k \in S^c} \min_{\prod_{k=1}^{N} p(x_k) p(\hat{y}_k|y_k,x_k)} (I(X(S);\hat{Y}(S^c),Y_k|X(S^c)) - I(Y(S);\hat{Y}(S)|X^N,\hat{Y}(S^c),Y_k)),
\]

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, \(X(S)\) denotes inputs in \(S\), and \(Y(S^c)\) denotes outputs in \(S^c\)

- 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)
  - 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 $\hat{y}_2^n(l_j|l_{j-1}) \sim \prod_{i=1}^n p_{\hat{Y}_2|X_2}(\hat{y}_{2i}|x_{2i}(l_{j-1})), l_j \in [1 : 2^{nR_2}]$
- $\mathcal{C}_j = \{(x_1^n(j, m), x_2^n(l_{j-1}), \hat{y}_2^n(l_j|l_{j-1})): m \in [1 : 2^{nbR}], l_j, l_{j-1} \in [1 : 2^{nR_2}], j \in [1 : b]\}$
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 \( (y_2^n(j), \hat{y}_2^n(l_j|l_{j-1}), x_2^n(l_{j-1})) \in T_e(n) \)
- In block \( j + 1 \), transmit \( x_2^n(l_j) \)

Decoding:
- At the end of block \( b \), find the unique \( \hat{m} \) such that \( (x_1^n(j, \hat{m}), x_2^n(l_{j-1}), \hat{y}_2^n(l_j|l_{j-1}), y_3^n(j)) \in T_e(n) \) 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

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

By the covering lemma and the union of events bound (over $b$ blocks), $P(\mathcal{E}_1) \to 0$ 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, $P(\mathcal{E}_2 \cap \mathcal{E}_1^c) \to 0$ as $n \to \infty$
Define $\tilde{E}_j(m, l_{j-1}, l_j) = \{(X_1^n(j, m), X_2^n(l_{j-1}), \hat{Y}_2^n(l_j|l_{j-1}), Y_3^n(j)) \in \mathcal{T}_\epsilon^{(n)}\}$

Then

$$P(\mathcal{E}_3) = P\left(\bigcup_{m \neq 1} \bigcup_{l^b} \bigcap_{j=1}^{b} \tilde{E}_j(m, l_{j-1}, l_j)\right)$$

$$\leq \sum_{m \neq 1} \sum_{l^b} P\left(\bigcap_{j=1}^{b} \tilde{E}_j(m, l_{j-1}, l_j)\right)$$

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

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

If $l_{j-1} = 1$, then by the joint typicality lemma, $P(\tilde{E}_j) \leq 2^{-n(I(X_1; \hat{Y}_2, Y_3|X_2) - \delta(\epsilon))}$

Similarly, if $l_{j-1} \neq 1$, then $P(\tilde{E}_j) \leq 2^{-n(I(X_1, X_2; Y_3) + I(\hat{Y}_2; X_1, Y_3|X_2) - \delta(\epsilon))}$

Thus, if $l^{b-1}$ has $k$ 1s, then

$$\prod_{j=2}^{b} P(\tilde{E}_j(m, l_{j-1}, l_j)) \leq 2^{-n(kI_1 + (b-1-k)I_2 - (b-1)\delta(\epsilon))}$$
Continuing with the bound,

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

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

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

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

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

which \(\to 0\) 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 \(P(\mathcal{E}) \to 0\) 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


References (cont.)


