

Achievability for Discrete Memoryless Systems

Abbas El Gamal

Stanford University

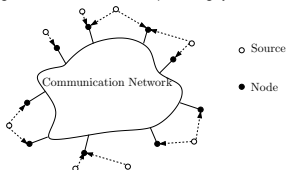
Padovani Lecture 2009

Acknowledgments

- Roberto Padovani for generous gift to the IT society
- IT Membership and Chapters committee
- Summer School of IT general chairs Aylin Yener and Gerhard Kramer and the other organizers
- The lecture is based on Lecture Notes on Network Information Theory jointly developed with Young-Han Kim, UC San Diego

Network Information Flow Problem

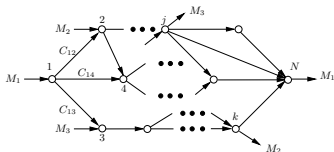
- Consider a general network information processing system:



- ▶ Sources may be data, speech, music, images, video, sensor data, ...
- ▶ Nodes may be handsets, base stations, servers, sensor nodes, ...
- ▶ The communication network may be wired, wireless, or a hybrid

- Each node observes a subset of the sources and wishes to obtain **descriptions** of some or all the sources, or to **compute a function/make a decision** based on these sources
- To achieve these goals, the nodes communicate over the network
- Information flow questions:
 - ▶ What are the **necessary and sufficient** conditions on information flow in the network under which the desired information processing goals can be achieved?
 - ▶ What are the **optimal coding/computing techniques/protocols** needed?
- The difficulty of answering these questions depends on the information processing goals, the source and communication network models, and the computational and storage capabilities of the nodes

- For example, if the sources are commodities with demands (rates in bits/sec), the nodes are connected by noiseless rate-constrained links, each intermediate node simply forwards the bits it receives, and the goal is to send the commodities to desired destination nodes, the problem reduces to the well-studied multi-commodity flow with known necessary and sufficient conditions on optimal flow [1]
- In the case of a single commodity, these necessary and sufficient conditions reduce to the celebrated max-flow min-cut theorem [2, 3]



- This simple information processing system model, however, does not capture many important aspects of real-world systems:
 - Real world information sources have redundancies, time and space correlations, and time variations
 - Real world networks may suffer from noise, interference, time, node failures, delays, and time variations
 - Real world networks may allow for broadcasting
 - Real world communication nodes may allow for more complex node operations than forwarding
 - The goal in many information processing systems is not to merely communicate source information

Network Information Theory

- Network information theory attempts to answer the above information flow questions while capturing some of these aspects of real world networks in the framework of Shannon information theory
- It involves several new challenges beyond point-to-point communication, including: multi-accessing; broadcasting; interference; relaying; cooperation; competition; distributed source coding; use of side information; multiple descriptions; joint source-channel coding; coding for computing; ...
- First paper on multiple user information theory: Claude E. Shannon, "Two-way communication channels" *Proc. 4th Berkeley Symp. Math. Stat. Prob.*, pp. 611-644, 1961
 - He didn't find the optimal rates
 - Problem remains open
- Lots of research activities in the 70s and early 80s with many new results and techniques developed, but
 - Many basic problems remained open
 - Little interest from information theory community and absolutely no interest from communication practitioners
- Wireless communication and the Internet (and advances in technology) have revived the interest in this area
 - Lots of research going on since the mid 90s
 - Some work on old open problems but some new problems as well

State of the Theory

- Focus has been on compression and communication
- Most work assumes **discrete memoryless** (DM) and **Gaussian** source and channel models
- Most results are for separate source–channel settings
- For such a setting, the goal is to establish a coding theorem that determines the set of *achievable rate tuples*, i.e.,
 - ▶ the **capacity region** when sending independent *messages* over a noisy channel, or
 - ▶ the **rate/rate-distortion region** when sending source descriptions over a noiseless channel
- Establishing a coding theorem involves proving:
 - ▶ **Achievability**: There exists a sequence of codes that achieve any rate tuple in the capacity/rate region
 - ▶ **(Weak) converse**: If a rate tuple is achievable, then it must be in the capacity/rate region

- Capacity/rate regions known only for few settings, including:
 - ▶ Point-to-point communication
 - ▶ Multiple access channel
 - ▶ Classes of broadcast, interference, and relay channels
 - ▶ Classes of channels with state
 - ▶ Classes of multiple descriptions and distributed source coding
 - ▶ Classes of channels with feedback
 - ▶ Classes of deterministic networks
 - ▶ Classes of wiretap channels and distributed key generation settings
- Several of the coding techniques developed, such as superposition coding, successive cancellation decoding, Slepian–Wolf, Wyner–Ziv, successive refinement, writing on dirty paper, network coding, decode–forward, compress–forward, and interference alignment, are beginning to have an impact on real world networks
- But many basic problems remain open and a complete theory is yet to be developed

The Lecture

- Will present a simple (yet rigorous) and unified approach to achievability for DM systems through examples of basic channel and source models:
 - ▶ Strong typicality
 - ▶ Small number of coding techniques and performance analysis tools
- The approach has its origin in Shannon’s original work, which was later made rigorous by Forney [7] and Cover [8] among others
- Why focus on achievability?
 - ▶ Achievability is useful—it often leads to practical coding techniques
 - ▶ We have recently developed a more unified approach based on a small set of techniques and tools
- Why discrete instead of Gaussian?
 - ▶ “End-to-end” model for many real world networks
 - ▶ Many real world sources are discrete
 - ▶ Achievability for most Gaussian models follows from their discrete counterparts via scalar quantizations and standard limit theorems

Outline

- 1 Typical Sequences
- 2 DMC
- 3 DM-MAC
- 4 DM-BC with Degraded Message Sets
- 5 DM-IC
- 6 DM-RC: Decode–Forward
- 7 DMS
- 8 DMS with Side Information at Decoder
- 9 DM-RC: Compress–Forward
- 10 DMC with State Available at Encoder
- 11 DM-BC with Private Messages
- 12 DM-WTC
- 13 References and Appendices

Typical Sequences

- Let $x^n \in \mathcal{X}^n$. Define the empirical pmf (or type) of x^n as

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

- Let X_1, X_2, \dots be a sequence of i.i.d. random variables each drawn according to $p(x)$. By the **law of large numbers (LLN)**, for every $x \in \mathcal{X}$, $\pi(x|x^n) \rightarrow p(x)$ in probability
- For $X \sim p(x)$, the set of ϵ -typical n -sequences [4] is defined as

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

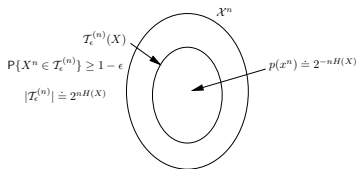
Typical Average Lemma

Let $x^n \in \mathcal{T}_\epsilon^{(n)}(X)$ and $g(x)$ be a nonnegative function, then

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

Properties of Typical Sequences

- Let $X^n \sim \prod_{i=1}^n p_X(x_i)$



- $H(X) := -\mathbf{E}(\log p(X))$
- $p(x^n) \doteq 2^{-nH(X)}$ means $2^{-n(H(X)+\delta(\epsilon))} \leq p(x^n) \leq 2^{-n(H(X)-\delta(\epsilon))}$, where $\delta(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$

Jointly Typical Sequences

- The joint type of $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ is defined as

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

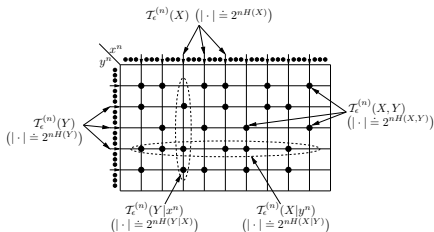
- For $(X, Y) \sim p(x, y)$, the set of *jointly ϵ -typical n -sequences* (x^n, y^n) is defined as before

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

- If $(x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y)$, then $x^n \in \mathcal{T}_\epsilon^{(n)}(X)$, $y^n \in \mathcal{T}_\epsilon^{(n)}(Y)$
- For $x^n \in \mathcal{T}_\epsilon^{(n)}$, the set of *conditionally ϵ -typical y^n sequences* is

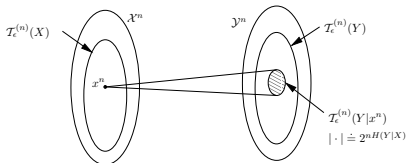
$$\mathcal{T}_\epsilon^{(n)}(Y|x^n) := \{y^n : |\pi(x, y|x^n, y^n) - p(x, y)| \leq \epsilon \cdot p(x, y) \text{ for all } (x, y) \in \mathcal{X} \times \mathcal{Y}\}$$

Properties of Jointly Typical Sequences



- $H(X|Y) := -\mathbf{E}_{X, Y}(\log p(X|Y))$

Another Useful Picture



Other Properties

- Let $x^n \in \mathcal{T}_\epsilon^{(n)}(X)$ and $Y^n | \{X^n = x^n\} \sim \prod_{i=1}^n p_{Y|X}(y_i|x_i)$. Then by the LLN, for every $\epsilon > \epsilon'$,

$$P\{(x^n, Y^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y)\} \rightarrow 1 \text{ as } n \rightarrow \infty$$

Joint Typicality Lemma

Given $(X, Y) \sim p(x, y)$, let $x^n \in \mathcal{T}_\epsilon^{(n)}(X)$ and $\tilde{Y}^n \sim \prod_{i=1}^n p_{Y|\tilde{X}}(\tilde{y}_i|x_i)$ (instead of $\prod_{i=1}^n p_{Y|X}(\tilde{y}_i|x_i)$). Then,

$$P\{(x^n, \tilde{Y}^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y)\} \doteq 2^{-nI(X;Y)}$$

- $I(X;Y) := H(X) + H(Y) - H(X,Y)$

Multivariate Typical Sequences

- The set of ϵ -typical n -sequences (x_1^n, x_2^n, x_3^n) is defined as

$$\{(x_1^n, x_2^n, x_3^n) : |\pi(x_1, x_2, x_3|x_1^n, x_2^n, x_3^n) - p(x_1, x_2, x_3)| \leq \epsilon \cdot p(x_1, x_2, x_3) \text{ for all } (x_1, x_2, x_3) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3\}$$

- Properties of jointly typical sets continue to hold by considering subsets of the random variables as single random variables

Conditional Joint Typicality Lemma

Let $(X_1, X_2, X_3) \sim p(x_1, x_2, x_3)$. Given $(x_1^n, x_2^n) \in \mathcal{T}_\epsilon^{(n)}(X_1, X_2)$, let \tilde{X}_3^n be drawn according to $\prod_{i=1}^n p_{X_3|X_1}(\tilde{x}_{3i}|x_{1i})$ (instead of $p_{X_3|X_1, X_2}(\tilde{x}_{3i}|x_{1i}, x_{2i})$). Then,

$$P\{(x_1^n, x_2^n, \tilde{X}_3^n) \in \mathcal{T}_\epsilon^{(n)}(X_1, X_2, X_3)\} \doteq 2^{-nI(X_2; X_3|X_1)}$$

Markov Lemma [5]

Suppose $X \rightarrow Y \rightarrow Z$ form a Markov chain. Let $(x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y)$. If $Z^n | \{Y^n = y^n\} \sim \prod_{i=1}^n p_{Z|Y}(z_i|y_i)$, then for every $\epsilon > \epsilon'$,

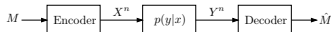
$$P\{(x^n, y^n, Z^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y, Z)\} \rightarrow 1 \text{ as } n \rightarrow \infty$$

- The proof follows by the LLN

Discrete Memoryless Channel

- Consider a DMC $(\mathcal{X}, p(y|x), \mathcal{Y})$ consisting of two finite sets \mathcal{X}, \mathcal{Y} , and a collection of conditional pmfs $p(y|x)$ over \mathcal{Y}

- The sender X wishes to send a message reliably to the receiver Y



- A $(2^{nR}, n)$ code for the DMC consists of:

- A message set $[1 : 2^{nR}] := \{1, 2, \dots, 2^{nR}\}$
- An encoder that assigns a *codeword* $x^n(m)$ to each message $m \in [1 : 2^{nR}]$
- A decoder that assigns to each received sequence y^n a message estimate $\hat{m}(y^n) \in [1 : 2^{nR}]$ or an error e

- Assume that the message $M \sim \text{Unif}[1 : 2^{nR}]$
- The *average probability of error* for a $(2^{nR}, n)$ code is

$$P_e^{(n)} = \mathbb{P}\{\hat{M} \neq M\}$$

- A rate R is *achievable* if there exists a sequence of $(2^{nR}, n)$ codes with probability of error $P_e^{(n)} \rightarrow 0$ as $n \rightarrow \infty$
- The *capacity* of the DMC is the supremum of all achievable rates

Theorem (Shannon [6])

The *capacity* of the DMC $(\mathcal{X}, p(y|x), \mathcal{Y})$ is

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

- We prove achievability using **random codebook generation, joint typicality decoding, and properties of jointly typical sequences**

Proof of Achievability

- Codebook generation:** Fix $p(x)$ that achieves C . Randomly and independently generate $x^n(m)$, $m \in [1 : 2^{nR}]$, sequences each according to $p(x^n) = \prod_{i=1}^n p_X(x_i)$. The generated set of sequences constitutes the *codebook* \mathcal{C}
- The codebook is revealed to both the sender and receiver before transmission commences
- Encoding:** To send a message $m \in [1 : 2^{nR}]$, transmit $x^n(m)$
- Decoding:** We use *joint-typicality decoding*. Upon receiving y^n , the decoder declares that $\hat{m} \in [1 : 2^{nR}]$ is sent if it is the unique message such that $(x^n(\hat{m}), y^n) \in \mathcal{T}_\epsilon^{(n)}$; otherwise an error e is declared

- Analysis of the probability of error:** We bound the probability of error $\mathbb{P}(\mathcal{E})$ averaged over codebooks and M . By symmetry of codebook generation, $\mathbb{P}(\mathcal{E}) = \mathbb{P}(\mathcal{E} | M = 1)$. Thus, we assume wlog that $M = 1$ is sent

A decoding error \mathcal{E} occurs iff

$$\mathcal{E}_1 := \{(X^n(1), Y^n) \notin \mathcal{T}_\epsilon^{(n)}\}, \text{ or}$$

$$\mathcal{E}_2 := \{(X^n(m), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m \in [2 : 2^{nR}]\}$$

Then, $\mathbb{P}(\mathcal{E}) = \mathbb{P}(\mathcal{E}_1 \cup \mathcal{E}_2) \leq \mathbb{P}(\mathcal{E}_1) + \mathbb{P}(\mathcal{E}_2)$

- Since, $(X^n(1), Y^n) \sim \prod_{i=1}^n p_{X,Y}(x_i, y_i)$, $\mathbb{P}(\mathcal{E}_1) \rightarrow 0$ as $n \rightarrow \infty$ by the LLN
- Next, consider $\mathbb{P}(\mathcal{E}_2)$. Since for $m \neq 1$, $(X^n(m), X^n(1), Y^n) \sim \prod_{i=1}^n p_X(x_i(m)) p_{X,Y}(x_i(1), y_i)$, $(X^n(m), Y^n) \sim \prod_{i=1}^n p_X(x_i) p_Y(y_i)$. Thus, the joint typicality lemma,

$$\mathbb{P}\{(X^n(m), Y^n) \in \mathcal{T}_\epsilon^{(n)}\} \leq 2^{-n(I(X;Y) - \delta(\epsilon))} = 2^{-n(C - \delta(\epsilon))},$$

By the union of events bound

$$\begin{aligned} P(\mathcal{E}_2) &\leq \sum_{m=2}^{2^{nR}} P\{(X^n(m), Y^n) \in \mathcal{T}_\epsilon^n\} \\ &\leq \sum_{m=2}^{2^{nR}} 2^{-n(C-\delta(\epsilon))} \leq 2^{-n(C-R-\delta(\epsilon))}, \end{aligned}$$

which $\rightarrow 0$ as $n \rightarrow \infty$ if $R < C - \delta(\epsilon)$

- To complete the proof, note that since the probability of error averaged over the codebooks $\rightarrow 0$, there must exist a sequence of $(2^{nR}, n)$ codes with $P_e^{(n)} \rightarrow 0$ as $n \rightarrow \infty$
- Note that to bound $P(\mathcal{E})$, we divided \mathcal{E} into events with same (X^n, Y^n) pmf. This observation will prove useful when we analyze more complex error events
- We generalize the bound on $P(\mathcal{E}_2)$ in the following *packing lemma* for use in multiple user settings

Packing Lemma

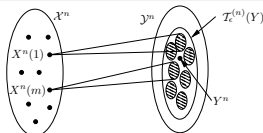
Let $(U, X, Y) \sim p(u, x, y)$, $\epsilon > 0$, and $U^n \sim \prod_{i=1}^n p_U(u_i)$

Let $X^n(m), m \in \mathcal{A}$, where $|\mathcal{A}| \leq 2^{nR}$, be random sequences each drawn according to $\prod_{i=1}^n p_{X|U}(x_i|u_i)$

Let Y^n be a random sequence drawn according to an arbitrary pmf $p(y^n|u^n)$, conditionally independent of each $X^n(m), m \in \mathcal{A}$, given U^n

There exists $\delta(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$ such that if $R < I(X; Y|U) - \delta(\epsilon)$, then

$$P\{(U^n, X^n(m), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m \in \mathcal{A}\} \rightarrow 0 \text{ as } n \rightarrow \infty$$



Packing Lemma

Let $(U, X, Y) \sim p(u, x, y)$, $\epsilon > 0$, and $U^n \sim \prod_{i=1}^n p_U(u_i)$

Let $X^n(m), m \in \mathcal{A}$, where $|\mathcal{A}| \leq 2^{nR}$, be random sequences each drawn according to $\prod_{i=1}^n p_{X|U}(x_i|u_i)$

Let Y^n be a random sequence drawn according to an arbitrary pmf $p(y^n|u^n)$, conditionally independent of each $X^n(m), m \in \mathcal{A}$, given U^n

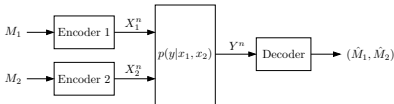
There exists $\delta(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$ such that if $R < I(X; Y|U) - \delta(\epsilon)$, then

$$P\{(U^n, X^n(m), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m \in \mathcal{A}\} \rightarrow 0 \text{ as } n \rightarrow \infty$$

- The proof is in Appendix I
- For the bound on $P(\mathcal{E}_2)$: $|\mathcal{A}| = 2^{nR} - 1$, $U = \emptyset$, for $m \neq 1$, the $X^n(m)$ sequences are i.i.d. each generated according to $\prod_{i=1}^n p_X(x_i)$ and independent of Y^n , which is generated according to $\prod_{i=1}^n p_Y(y_i)$
- We will encounter settings where $U \neq \emptyset$, the $X^n(m)$ sequences are not independent, and Y^n is not generated i.i.d.

DM Multiple Access Channel

- Consider a 2-sender DM-MAC $(\mathcal{X}_1 \times \mathcal{X}_2, p(y|x_1, x_2), \mathcal{Y})$
- Senders X_1, X_2 wish to send independent messages to the receiver Y



- A $(2^{nR_1}, 2^{nR_2}, n)$ code for a DM-MAC consists of:
 - Two message sets $[1 : 2^{nR_1}]$ and $[1 : 2^{nR_2}]$
 - Two encoders: Encoder $j = 1, 2$ assigns a codeword $x_j^n(m_j)$ to each message $m_j \in [1 : 2^{nR_j}]$
 - A decoder that assigns to each received sequence y^n an estimate $(\hat{m}_1, \hat{m}_2) \in [1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$ or an error e

- Assume that $(M_1, M_2) \sim \text{Unif}([1 : 2^{nR_1}] \times [1 : 2^{nR_2}])$. Thus, $X_1^n(M_1)$ and $X_2^n(M_2)$ are independent
- The average probability of error is

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

- A rate pair (R_1, R_2) is *achievable* for the DM-MAC if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes with $P_e^{(n)} \rightarrow 0$ as $n \rightarrow \infty$
- The *capacity region* of the DM-MAC is the closure of the set of achievable rate pairs (R_1, R_2)

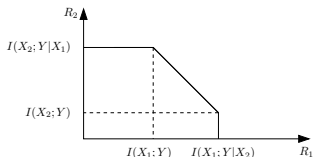
Theorem (Ahlswede [9], Liao [10])

Let $(X_1, X_2) \sim p(x_1)p(x_2)$ and $\mathcal{R}(X_1, X_2)$ be the set of rate pairs (R_1, R_2) such that

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

The capacity region for the DM-MAC is the convex closure of the union of the regions $\mathcal{R}(X_1, X_2)$ over all $p(x_1)p(x_2)$

- $\mathcal{R}(X_1, X_2)$ is in general a pentagon with a 45° side (why?)



- The new achievability ideas are **successive cancellation decoding**, **simultaneous decoding** and **time sharing**, in addition to applying the packing lemma

Achievability

- We first fix $(X_1, X_2) \sim p(x_1)p(x_2)$ and show that every rate pair (R_1, R_2) in the interior of $\mathcal{R}(X_1, X_2)$ is achievable. The rest of the capacity region is achieved using **time sharing** between points in different $\mathcal{R}(X_1, X_2)$ regions: If rate pairs (R_1, R_2) and (R'_1, R'_2) are achievable, then using the codes of first pair α of the time and the codes of second pair the rest of the time, can achieve $(\alpha R_1 + \bar{\alpha} R'_1, \alpha R_2 + \bar{\alpha} R'_2)$, $\bar{\alpha} := (1 - \alpha)$
- Note that time sharing is more general than time/frequency division, where we time-share between $(R_1, 0)$ and $(0, R_2)$
- Codebook generation:** Fix $p(x_1)p(x_2)$. Randomly and independently generate $x_1^n(m_1)$, $m_1 \in [1 : 2^{nR_1}]$, sequences each according to $\prod_{i=1}^n p_{X_1}(x_{1i})$. Similarly generate $x_2^n(m_2)$, $m_2 \in [1 : 2^{nR_2}]$, sequences each according to $\prod_{i=1}^n p_{X_2}(x_{2i})$
- Encoding:** To send the message m_1 , encoder 1 transmits $x_1^n(m_1)$. Similarly, to send m_2 , encoder 2 transmits $x_2^n(m_2)$

Successive Cancellation Decoding

- This decoding rule aims to achieve the two *corner points* of the pentagon $\mathcal{R}(X_1, X_2)$, e.g., $R_1 < I(X_1; Y)$, $R_2 < I(X_2; Y|X_1)$
- The decoder first declares that \hat{m}_1 is sent if it is the unique message such that $(x_1^n(\hat{m}_1), y^n) \in \mathcal{T}_\epsilon^{(n)}$; otherwise an error is declared. If such \hat{m}_1 is found, the receiver finds the unique \hat{m}_2 such that $(x_1^n(\hat{m}_1), x_2^n(\hat{m}_2), y^n) \in \mathcal{T}_\epsilon^{(n)}$; otherwise an error is declared
- Analysis of the probability of error:** Assume wlog that message pair $(1, 1)$ is sent. We first bound the average probability of error for the first decoding step. An error \mathcal{E}_1 occurs iff

$$\mathcal{E}_{11} := \{(X_1^n(1), Y^n) \notin \mathcal{T}_\epsilon^{(n)}\}, \text{ or}$$

$$\mathcal{E}_{12} := \{(X_1^n(m_1), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_1 \neq 1\}$$

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

- Since $(X_1^n(1), Y^n) \sim \prod_{i=1}^n p_{X_1, Y}(x_{1i}, y_i)$, by the LLN, $P(\mathcal{E}_{11}) \rightarrow 0$ as $n \rightarrow \infty$

- Since for $m \neq 1$, $(X_1^n(m_1), Y^n) \sim \prod_{i=1}^n p_{X_1}(x_{1i})p_Y(y_i)$, by the packing lemma (with $|\mathcal{A}| = 2^{nR_1} - 1$, $U = \emptyset$), $P(\mathcal{E}_{12}) \rightarrow 0$ as $n \rightarrow \infty$ if $R_1 < I(X_1; Y) - \delta(\epsilon)$
- Next, we bound the average probability that decoding step 1 succeeds but decoding step 2 fails. An error \mathcal{E}_2 occurs iff

$$\mathcal{E}_{21} := \{(X_1^n(1), X_2^n(1), Y^n) \notin \mathcal{T}_\epsilon^{(n)}\}, \text{ or}$$

$$\mathcal{E}_{22} := \{(X_1^n(1), X_2^n(m_2), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_2 \neq 1\}$$

Then, $P(\mathcal{E}_2) \leq P(\mathcal{E}_{21}) + P(\mathcal{E}_{22})$

- By the LLN, $P(\mathcal{E}_{21}) \rightarrow 0$ as $n \rightarrow \infty$
- Since for $m_2 \neq 1$, $(X_2^n(m_2), X_1^n(1), Y^n) \sim \prod_{i=1}^n p_{X_2}(x_{2i})p_{X_1, Y}(x_{1i}, y_i)$, by the packing lemma (with $|\mathcal{A}| = 2^{nR_2} - 1$, $Y \rightarrow (X_1, Y)$, $U = \emptyset$), $P(\mathcal{E}_{22}) \rightarrow 0$ as $n \rightarrow \infty$ if $R_2 < I(X_2; Y, X_1) - \delta(\epsilon) = I(X_2; Y|X_1) - \delta(\epsilon)$, since X_1, X_2 are independent

- Thus the total average probability of decoding error $P(\mathcal{E}) \leq P(\mathcal{E}_1) + P(\mathcal{E}_2) \rightarrow 0$ as $n \rightarrow \infty$ if $R_1 < I(X_1; Y) - \delta(\epsilon)$ and $R_2 < I(X_2; Y|X_1) - \delta(\epsilon)$
- Achievability of the other corner point follows by changing the decoding order
- To show achievability of other points in $\mathcal{R}(X_1, X_2)$, we use time sharing between corner points and points on the axes
- Finally, to show achievability of points not in any $\mathcal{R}(X_1, X_2)$, we use time sharing between points in different $\mathcal{R}(X_1, X_2)$ regions

Simultaneous Joint Typicality Decoding

- We can prove achievability of *any* rate pair in the interior of $\mathcal{R}(X_1, X_2)$ *without* time sharing
- The decoder declares that (\hat{m}_1, \hat{m}_2) is sent if it is the unique message pair such that $(x_1^n(\hat{m}_1), x_2^n(\hat{m}_2), y^n) \in \mathcal{T}_\epsilon^{(n)}$; otherwise an error is declared
- Analysis of the probability of error:** Assume wlog that message pair $(1, 1)$ is sent.
- To divide the error event, let's look at the pmfs for the triple $(X_1^n(m_1), X_2^n(m_2), Y^n)$

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

- So, an error \mathcal{E} occurs iff

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

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

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

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

$$\text{Thus, } P(\mathcal{E}) \leq \sum_{j=1}^4 P(\mathcal{E}_j)$$

- By the LLN, $P(\mathcal{E}_1) \rightarrow 0$ as $n \rightarrow \infty$
- By the packing lemma, $P(\mathcal{E}_2) \rightarrow 0$ as $n \rightarrow \infty$ if $R_1 < I(X_1; Y|X_2) - \delta(\epsilon)$
- Similarly, $P(\mathcal{E}_3) \rightarrow 0$ as $n \rightarrow \infty$ if $R_2 < I(X_2; Y|X_1) - \delta(\epsilon)$
- And $P(\mathcal{E}_4) \rightarrow 0$ as $n \rightarrow \infty$ if $R_1 + R_2 < I(X_1, X_2; Y) - \delta(\epsilon)$
- So, simultaneous decoding is more powerful than successive cancellation decoding because we don't need time sharing to achieve points in $\mathcal{R}(X_1, X_2)$

Alternative Characterization

- The proof of the weak converse gives the alternative characterization

Theorem

The capacity region of the DM-MAC $(\mathcal{X}_1 \times \mathcal{X}_2, p(y|x_1, x_2), \mathcal{Y})$ is the set of (R_1, R_2) pairs satisfying

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

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

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

for some joint pmf $p(q)p(x_1|q)p(x_2|q)$

- Q is an auxiliary (time sharing) random variable
- Clearly the above region contains the first characterization region
- The first characterization region contains the above region (why?)
- Can we achieve the above region directly?

Coded Time-Sharing

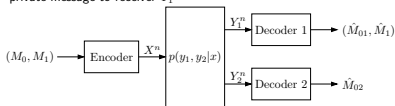
- Fix $p(q)p(x_1|q)p(x_2|q)$
- Codebook generation:** Randomly generate a *time sharing* sequence $q^n \sim \prod_{i=1}^n p_Q(q_i)$
Randomly and conditionally independently generate $x_1^n(m_1)$, $m_1 \in [1 : 2^{nR_1}]$, sequences each according to $\prod_{i=1}^n p_{X_1|Q}(x_{1i}|q_i)$
Similarly generate $x_2^n(m_2)$, $m_2 \in [1 : 2^{nR_2}]$, sequences each according to $\prod_{i=1}^n p_{X_2|Q}(x_{2i}|q_i)$
- The chosen codebook, including q^n , is revealed to the encoders and decoder
- Encoding:** To send (m_1, m_2) transmit $x_1^n(m_1)$ and $x_2^n(m_2)$
- Decoding:** We find the unique (\hat{m}_1, \hat{m}_2) such that $(q^n, x_1^n(\hat{m}_1), x_2^n(\hat{m}_2), y^n) \in \mathcal{T}_\epsilon^{(n)}$
- The rest of the proof follows as in simultaneous decoding with U in the packing lemma replaced by Q

Summary of Achievability Ideas

- Coding techniques:**
 - Random codebook generation: Makes analysis of $P_e^{(n)}$ tractable; has become practical with LDPC codes/Fountain codes
 - Joint typicality decoding (packing): Not optimal but easy to analyze and achieves optimal rates; does not provide good error bounds
 - Successive cancellation decoding
 - Simultaneous joint typicality decoding: More powerful than successive cancellation decoding
 - Time sharing and coded time sharing
- Performance analysis tools:**
 - Division of error event
 - LLN
 - Properties of typicality
 - Packing lemma

DM Broadcast Channel with Degraded Message Sets

- Consider a 2-receiver DM-BC $(\mathcal{X}, p(y_1, y_2|x), \mathcal{Y}_1 \times \mathcal{Y}_2)$
- Sender X wishes to send a common message to both receivers and a private message to receiver Y_1



- A $(2^{nR_0}, 2^{nR_1}, n)$ code for the DM-BC consists of:
 - Two message sets $[1 : 2^{nR_0}]$ and $[1 : 2^{nR_1}]$
 - An encoder that assigns a codeword $x^n(m_0, m_1)$ to each message pair $(m_0, m_1) \in [1 : 2^{nR_0}] \times [1 : 2^{nR_1}]$
 - Two decoders: Decoder 1 assigns to each $y_1^n \in \mathcal{Y}_1^n$ an estimate $(\hat{m}_{01}, \hat{m}_1)(y_1^n) \in [1 : 2^{nR_0}] \times [1 : 2^{nR_1}]$ or an error e . Decoder 2 assigns to each $y_2^n \in \mathcal{Y}_2^n$ an estimate $\hat{m}_{02}(y_2^n) \in [1 : 2^{nR_0}]$ or an error e

- Assume that $(M_0, M_1) \sim \text{Unif}([1 : 2^{nR_0}] \times [1 : 2^{nR_1}])$
- The average probability of error is

$$P_e^{(n)} = \mathbb{P}\{(\hat{M}_{01}, \hat{M}_1) \neq (M_0, M_1) \text{ or } \hat{M}_{02} \neq M_0\}$$

- A rate pair (R_0, R_1) is *achievable* for the DM-BC if there exists a sequence of $(2^{nR_0}, 2^{nR_1}, n)$ codes with $P_e^{(n)} \rightarrow 0$ as $n \rightarrow \infty$
- The *capacity region* of the DM-BC with degraded message sets is the closure of the set of achievable rate pairs

Theorem (Körner–Marton [11])

The capacity region of the DM-BC with degraded message sets is the set of rate pairs (R_0, R_1) such that

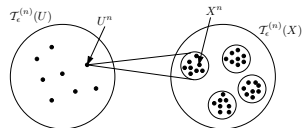
$$\begin{aligned} R_0 &\leq I(U; Y_2), \\ R_1 &\leq I(X; Y_1|U), \\ R_0 + R_1 &\leq I(X; Y_1) \end{aligned}$$

for some $p(u, x)$

- U is an auxiliary random variable
- The new achievability idea is **superposition coding**

Outline of Achievability

- Fix $p(u)p(x|u)$. Generate 2^{nR_0} independent $u^n(m_0)$ sequences (cloud centers). For each $u^n(m_0)$, generate 2^{nR_1} conditionally independent $x^n(m_0, m_1)$ sequences (satellite codewords)



- Receiver Y_2 decodes the cloud center $u^n(m_0)$
- Receiver Y_1 decodes the satellite codeword $x^n(m_0, m_1)$

Proof of Achievability

- We show that any rate pair that satisfies the conditions for any given $p(u)p(x|u)$ is achievable
- **Codebook generation:** Fix $p(u)p(x|u)$. Randomly and independently generate $u^n(m_0)$, $m_0 \in [1 : 2^{nR_0}]$, sequences each according to $\prod_{i=1}^n p_U(u_i)$. For each sequence $u^n(m_0)$, randomly and conditionally independently generate $x^n(m_0, m_1)$, $m_1 \in [1 : 2^{nR_1}]$, sequences each according to $\prod_{i=1}^n p_{X|U}(x_i|u_i(m_0))$
- **Encoding:** To send the message pair (m_0, m_1) , transmit $x^n(m_0, m_1)$
- **Decoding:** Decoder 2 declares that a message \hat{m}_{02} is sent if it is the unique message such that $(u^n(\hat{m}_{02}), y_2^n) \in \mathcal{T}_\epsilon^{(n)}$; otherwise an error is declared. Decoder 1 uses simultaneous decoding. It declares that the message pair $(\hat{m}_{01}, \hat{m}_1)$ is sent if it is the unique message pair such that $(u^n(\hat{m}_{01}), x^n(\hat{m}_{01}, \hat{m}_1), y_1^n) \in \mathcal{T}_\epsilon^{(n)}$; otherwise it declares an error

- Next consider the average probability of error for decoder 1. Let's look at the pmfs for the triple $(U^n(m_0), X^n(m_0, m_1), Y_1^n)$

| m_0 | m_1 | Joint pmf |
|----------|----------|---------------------------|
| 1 | 1 | $p(u^n, x^n)p(y_1^n x^n)$ |
| 1 | \times | $p(u^n, x^n)p(y_1^n u^n)$ |
| \times | \times | $p(u^n, x^n)p(y_1^n)$ |
| \times | 1 | $p(u^n, x^n)p(y_1^n)$ |

The last two cases have the same joint pmf
So we can divide the error event into 3 events

$$\begin{aligned} \mathcal{E}_{11} &:= \{(U^n(1), X^n(1, 1), Y_1^n) \notin \mathcal{T}_\epsilon^{(n)}\}, \\ \mathcal{E}_{12} &:= \{(U^n(1), X^n(1, m_1), Y_1^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_1 \neq 1\}, \\ \mathcal{E}_{13} &:= \{(U^n(m_0), X^n(m_0, m_1), Y_1^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_0 \neq 1, m_1\} \end{aligned}$$

Analysis of the Probability of Error

- Assume wlog that message pair $(1, 1)$ is sent
- First consider the average probability of error for decoder 2. Define the events

$$\begin{aligned} \mathcal{E}_{21} &:= \{(U^n(1), Y_2^n) \notin \mathcal{T}_\epsilon^{(n)}\}, \\ \mathcal{E}_{22} &:= \{(U^n(m_0), Y_2^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_0 \neq 1\} \end{aligned}$$

The probability of error for decoder 2 is then bounded by

$$P(\mathcal{E}_2) \leq P(\mathcal{E}_{21}) + P(\mathcal{E}_{22})$$

- By the LLN, $P(\mathcal{E}_{21}) \rightarrow 0$ as $n \rightarrow \infty$
- Since for $m_0 \neq 1$, $U^n(m_0)$ is independent of Y_2^n , by the packing lemma, $P(\mathcal{E}_{22}) \rightarrow 0$ as $n \rightarrow \infty$ if $R_2 < I(U; Y_2) - \delta(\epsilon)$

- The probability of error for decoder 1 is then bounded by

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

- By the LLN, $P(\mathcal{E}_{11}) \rightarrow 0$ as $n \rightarrow \infty$
- Since for $m_1 \neq 1$, $(U^n(1), X^n(1, m_1), Y_1^n) \sim \prod_{i=1}^n p_U(u_i)p_{X|U}(x_i|u_i)p_{Y|U}(y_i|u_i)$, by the packing lemma, $P(\mathcal{E}_{12}) \rightarrow 0$ as $n \rightarrow \infty$ if $R_1 < I(X; Y_1|U) - \delta(\epsilon)$
- For $m_0 \neq 1$, $(U^n(m_0), X^n(m_0, m_1))$ is independent of Y_1^n . Hence, by the packing lemma, $P(\mathcal{E}_{13}) \rightarrow 0$ as $n \rightarrow \infty$ if $R_0 + R_1 < I(U, X; Y_1) - \delta(\epsilon) = I(X; Y_1) - \delta(\epsilon)$, by the Markovity of $U \rightarrow X \rightarrow Y_1$

Note that here the $(U^n(m_0), X^n(m_0, m_1))$ sequences are not independent of each other

Alternative Characterization

- We can prove the converse for the region consisting of rate pairs (R_0, R_1) satisfying

$$\begin{aligned} R_0 &\leq I(U; Y_2), \\ R_0 + R_1 &\leq I(U; Y_2) + I(X; Y_1|U), \\ R_0 + R_1 &\leq I(X; Y_1) \end{aligned}$$

for some $p(u, x)$

- Clearly this region must include the first characterization region. We can also show that they have the same boundary points
- Can we show that the above region is achievable directly?
- The answer is yes, and the proof involves (unnecessary) **rate splitting**:
- Divide M_1 into two independent messages: M_{10} at rate R_{10} and M_{11} at rate R_{11} . Represent (M_0, M_{10}) by U and (M_0, M_{10}, M_{11}) by X

- Following similar steps to the previous proof of achievability, we can show that (R_0, R_{10}, R_{11}) is achievable if

$$\begin{aligned} R_0 + R_{10} &< I(U; Y_2) - \delta(\epsilon), \\ R_{11} &< I(X; Y_1|U) - \delta(\epsilon), \\ R_0 + R_1 &< I(X; Y_1) - \delta(\epsilon) \end{aligned}$$

for some $p(u, x)$

- Substituting $R_{11} = R_1 - R_{10}$ (and dropping $\delta(\epsilon)$), we have the conditions

$$\begin{aligned} R_{10} &< I(U; Y_2) - R_0, \\ R_{10} &\geq 0, \\ R_{10} &> R_1 - I(X; Y_1|U), \\ R_0 + R_1 &< I(X; Y_1) \end{aligned}$$

for some $p(u, x)$

- Now, we use the Fourier–Motzkin procedure (e.g., see [12], Appendix II) to find the projection of the above (R_0, R_{10}, R_{11}) region on the (R_0, R_1) plane by eliminating R_{10} as follows: The first and second inequalities give

$$R_0 < I(U; Y_2)$$

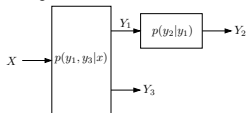
The first and third inequalities give

$$R_0 + R_1 < I(U; Y_2) + I(X; Y_1|U)$$

- This establishes the achievability of the outer bound
- The above rate splitting idea turns out to be crucial for 3-receiver DM-BC

Multi-level DM-BC with Degraded Message Sets

- A 3-receiver *multi-level* DM-BC [13] $(\mathcal{X}, p(y_1, y_3|x)p(y_2|y_1), \mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3)$ is a 3-receiver DM-BC where receiver \mathcal{Y}_2 is a degraded version of receiver \mathcal{Y}_1



- Consider the two degraded message sets scenario: Common message $M_0 \in [1 : 2^{nR_0}]$ is to be reliably sent to all receivers, Private message $M_1 \in [1 : 2^{nR_1}]$ is to be sent only to receiver \mathcal{Y}_1
- What is the capacity region?

- A straightforward extension of Körner–Marton capacity region for 2 receivers gives the inner bound consisting of (R_0, R_1) such that

$$R_0 < \min\{I(U; Y_2), I(U; Y_3)\}, \\ R_1 < I(X; Y_1|U)$$

for some $p(u, x)$

Note the bound $R_0 + R_1 < I(X; Y_1)$ drops out since $X \rightarrow Y_1 \rightarrow Y_2$

- This region turned out not to be optimal in general

Theorem (Nair–El Gamal [14])

The capacity region of the 3-receiver multi-level DM-BC is the set of rate pairs (R_0, R_1) such that

$$R_0 \leq \min\{I(U; Y_2), I(V; Y_3)\}, \\ R_1 \leq I(X; Y_1|U), \\ R_0 + R_1 \leq I(V; Y_3) + I(X; Y_1|V)$$

for some $p(u)p(v|u)p(x|v)$

Proof of Achievability

- The new achievability ideas are **rate splitting** and **indirect decoding**
- Rate splitting:** Split the private message M_1 into two independent parts M_{10}, M_{11} with rates R_{10}, R_{11} , respectively. Thus $R_1 = R_{10} + R_{11}$
- Codebook generation:** Fix $p(u)p(v|u)p(x|v)$. Randomly and independently generate $u^n(m_0)$, $m_0 \in [1 : 2^{nR_{10}}]$, sequences each according to $\prod_{i=1}^n p_U(u_i)$
For each $u^n(m_0)$, randomly and conditionally independently generate $v^n(m_0, m_{10})$, $m_{10} \in [1 : 2^{nR_{10}}]$, sequences each according to $\prod_{i=1}^n p_{V|U}(v_i|u_i(m_0))$
For each $v^n(m_0, m_{10})$, randomly and conditionally independently generate $x^n(m_0, m_{10}, m_{11})$, $m_{11} \in [1 : 2^{nR_{11}}]$, sequences each according to $\prod_{i=1}^n p_{X|V}(x_i|v_i(m_0, m_{10}))$
- Encoding:** To send $(m_0, m_1) \in [1 : 2^{nR_0}] \times [1 : 2^{nR_1}]$, where m_1 is represented by $(m_{10}, m_{11}) \in [1 : 2^{nR_{10}}] \times [1 : 2^{nR_{11}}]$, the sender transmits $x^n(m_0, m_{10}, m_{11})$

Decoding and Analysis of the Probability of Error

- Receiver Y_2 declares that $\hat{m}_{02} \in [1 : 2^{nR_{10}}]$ is sent if it is the unique message such that $(u^n(\hat{m}_{02}), y_2^n) \in \mathcal{T}_\epsilon^{(n)}$. By the LLN and the packing lemma, the probability of error $\rightarrow 0$ as $n \rightarrow \infty$ if

$$R_0 < I(U; Y_2) - \delta(\epsilon)$$

- Receiver Y_1 declares that $(\hat{m}_{01}, \hat{m}_{10}, \hat{m}_{11})$ is sent if it is the unique triple such that $(u^n(\hat{m}_{01}), v^n(\hat{m}_{01}, \hat{m}_{10}), x^n(\hat{m}_{01}, \hat{m}_{10}, \hat{m}_{11}), y_1^n) \in \mathcal{T}_\epsilon^{(n)}$. By the LLN and the packing lemma, the probability of error $\rightarrow 0$ as $n \rightarrow \infty$ if

$$R_{11} < I(X; Y_1|V) - \delta(\epsilon), \\ R_{10} + R_{11} < I(X; Y_1|U) - \delta(\epsilon), \\ R_0 + R_{10} + R_{11} < I(X; Y_1) - \delta(\epsilon)$$

- If receiver Y_3 decodes m_0 directly by finding the unique \hat{m}_{03} such that $(u^n(\hat{m}_{03}), y_3^n) \in \mathcal{T}_\epsilon^{(n)}$, we obtain $R_0 < I(U; Y_3)$, which together with previous conditions gives the extended Körner–Marton region
- To achieve the larger region, receiver Y_3 decodes m_0 *indirectly*. It declares that \hat{m}_{03} is sent if it is the unique index such that $(u^n(\hat{m}_{03}), v^n(\hat{m}_{03}, m_{10}), y_3^n) \in \mathcal{T}_\epsilon^{(n)}$ for some $m_{10} \in [1 : 2^{nR_{10}}]$
- Analysis of the probability of error:** Assume $(m_0, m_{10}) = (1, 1)$ is sent
- Consider the pmfs for the triple $(U^n(m_0), V^n(m_0, m_{10}), Y_3^n)$

| m_0 | m_{10} | Joint pmf |
|-------|----------|---------------------------|
| 1 | 1 | $p(u^n, v^n)p(y_3^n v^n)$ |
| 1 | × | $p(u^n, v^n)p(y_3^n u^n)$ |
| × | × | $p(u^n, v^n)p(y_3^n)$ |
| × | 1 | $p(u^n, v^n)p(y_3^n)$ |

- The second case does not result in an error, and the last 2 cases have the same pmf

- Thus, we are left with only two error events

$$\mathcal{E}_1 := \{(U^n(1), V^n(1, 1), Y_3^n) \notin \mathcal{T}_\epsilon^{(n)}\},$$

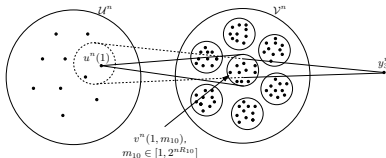
$$\mathcal{E}_2 := \{(U^n(m_0), V^n(m_0, m_{10}), Y_3^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_0 \neq 1, m_{10}\}$$

Then, the probability of error for decoder 3 averaged over codebooks $P(\mathcal{E}_3) \leq P(\mathcal{E}_{31}) + P(\mathcal{E}_{32})$

- By the LLN, $P(\mathcal{E}_{31}) \rightarrow 0$ as $n \rightarrow \infty$
- By the packing lemma (with $|\mathcal{A}| = 2^{n(R_0 + R_{10})} - 1$, $X \rightarrow (U, V)$, $U = \emptyset$), $P(\mathcal{E}_{32}) \rightarrow 0$ as $n \rightarrow \infty$ if $R_0 + R_{10} < I(U, V; Y_3) - \delta(\epsilon) = I(V; Y_3) - \delta(\epsilon)$
- Combining the bounds, substituting $R_{10} + R_{11} = R_1$, and using the Fourier-Motzkin procedure to eliminate R_{10} and R_{11} complete the proof of achievability

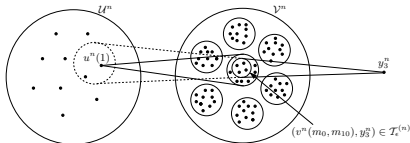
Indirect Decoding Interpretation

- Suppose that $R_0 > I(U; Y_3)$
- Y_3 cannot directly decode the cloud center $u^n(1)$



Indirect Decoding Interpretation

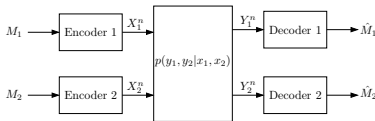
- Assume $R_0 > I(U; Y_3)$ but $R_0 + R_{10} < I(V; Y_3)$
- Y_3 decodes cloud center *indirectly*



- The above condition suffices in general (even when $R_0 < I(U; Y_3)$)

DM Interference Channel

- Consider a 2-user pair DM-IC ($\mathcal{X}_1 \times \mathcal{X}_2, p(y_1, y_2|x_1, x_2), \mathcal{Y}_1 \times \mathcal{Y}_2$)
- Sender $j = 1, 2$ wishes to send an independent message to receiver Y_j



- A $(2^{nR_1}, 2^{nR_2}, n)$ code for the interference channel consists of:
 - Two message sets $[1 : 2^{nR_1}]$ and $[1 : 2^{nR_2}]$
 - Two encoders: Encoder $j = 1, 2$ assigns a codeword $x_j^n(m_j)$ to each message $m_j \in [1 : 2^{nR_j}]$
 - Two decoders: Decoder $j = 1, 2$ assigns an estimate $\hat{m}_j(y_j^n)$ or an error e to each received sequence y_j^n

- Assume that $(M_1, M_2) \sim \text{Unif}([1 : 2^{nR_1}] \times [1 : 2^{nR_2}])$
- The average probability of error is

$$P_e^{(n)} = P\{\hat{M}_1 \neq M_1 \text{ or } \hat{M}_2 \neq M_2\}$$

- A rate pair (R_1, R_2) is *achievable* for the DM-IC if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes with $P_e^{(n)} \rightarrow 0$
- The *capacity region* of the DM-IC is the closure of the set of achievable rate pairs (R_1, R_2)
- The capacity region of the DM-IC is not known in general

Han-Kobayashi Inner Bound

- This is the best known inner bound on the capacity region of the 2-user pair DM-IC
- The following is a recent characterization [15] of this inner bound

Theorem (Han-Kobayashi [16])

A rate pair (R_1, R_2) is achievable for a DM-IC if it satisfies the inequalities

$$R_1 < I(X_1; Y_1 | U_2, Q),$$

$$R_2 < I(X_2; Y_2 | U_1, Q),$$

$$R_1 + R_2 < I(X_1, U_2; Y_1 | Q) + I(X_2; Y_2 | U_1, U_2, Q),$$

$$R_1 + R_2 < I(X_1; Y_1 | U_1, U_2, Q) + I(X_2; U_1; Y_2 | Q),$$

$$R_1 + R_2 < I(X_1, U_2; Y_1 | U_1, Q) + I(X_2; U_1; Y_2 | U_2, Q),$$

$$2R_1 + R_2 < I(X_1, U_2; Y_1 | Q) + I(X_1; Y_1 | U_1, U_2, Q) + I(X_2; U_1; Y_2 | U_2, Q),$$

$$R_1 + 2R_2 < I(X_2; U_1; Y_2 | Q) + I(X_2; Y_2 | U_1, U_2, Q) + I(X_1; U_2; Y_1 | U_1, Q)$$

for some $p(q)p(u_1, x_1 | q)p(u_2, x_2 | q)$

Proof of Achievability

- Split message M_j , $j = 1, 2$, into independent "public" part of rate R_{j0} and "private" part of rate R_{jj} . Thus, $R_j = R_{j0} + R_{jj}$
- We first show that $(R_{10}, R_{20}, R_{11}, R_{22})$ is achievable if

$$R_{11} < I(X_1; Y_1 | U_1, U_2, Q),$$

$$R_{11} + R_{10} < I(X_1; Y_1 | U_2, Q),$$

$$R_{11} + R_{20} < I(X_1, U_2; Y_1 | U_1, Q),$$

$$R_{11} + R_{10} + R_{20} < I(X_1, U_2; Y_1 | Q),$$

$$R_{22} < I(X_2; Y_2 | U_1, U_2, Q),$$

$$R_{22} + R_{20} < I(X_2; Y_2 | U_1, Q),$$

$$R_{22} + R_{10} < I(X_2; U_1; Y_2 | U_2, Q),$$

$$R_{22} + R_{20} + R_{10} < I(X_2; U_1; Y_2 | Q)$$

for some $p(q)p(u_1, x_1 | q)p(u_2, x_2 | q)$

- Codebook generation:** Fix $p(q)p(u_1, x_1 | q)p(u_2, x_2 | q)$. Generate a sequence $q^n \sim \prod_{i=1}^n p_Q(q_i)$. For $j = 1, 2$, randomly and conditionally independently generate $u_j^n(m_{j0}), m_{j0} \in [1 : 2^{nR_{j0}}]$, sequences each according to $\prod_{i=1}^n p_{U_j | Q}(u_{ji} | q_i)$. For each $u_j^n(m_{j0})$, randomly and conditionally independently generate $x_j^n(m_{j0}, m_{jj}), m_{jj} \in [1 : 2^{nR_{jj}}]$, sequences each according to $\prod_{i=1}^n p_{X_j | U_j, Q}(x_{ji} | u_{ji}(m_{j0}), q_i)$
- Encoding:** To send the message $m_j = (m_{j0}, m_{jj})$, encoder $j = 1, 2$ sends the codeword $x_j^n(m_{j0}, m_{jj})$
- Decoding:** Upon receiving y_1^n , decoder 1 finds the unique message pair $(\hat{m}_{10}, \hat{m}_{11})$ such that $(q^n, u_1^n(\hat{m}_{10}), u_2^n(m_{20}), x_1^n(\hat{m}_{10}, \hat{m}_{11}), y_1^n) \in \mathcal{T}_\epsilon^n$ for some $m_{20} \in [1 : 2^{nR_{20}}]$. If no such unique pair exists, the decoder declares an error. Decoder 2 decodes the message pair $(\hat{m}_{20}, \hat{m}_{22})$ similarly

Analysis of the Probability of Error

- Assume message pair $((1,1), (1,1))$ is sent. We bound the average probability of error for each decoder. First consider decoder 1
- We have 8 cases to consider (conditioning on q^n suppressed)

| | m_{10} | m_{20} | m_{11} | Joint pmf |
|---|----------|----------|----------|--|
| 1 | 1 | 1 | 1 | $p(u_1^n, x_1^n)p(u_2^n)p(y_1^n x_1^n, u_2^n)$ |
| 2 | 1 | 1 | × | $p(u_1^n, x_1^n)p(u_2^n)p(y_1^n u_1^n, u_2^n)$ |
| 3 | × | 1 | × | $p(u_1^n, x_1^n)p(u_2^n)p(y_1^n u_2^n)$ |
| 4 | × | 1 | 1 | $p(u_1^n, x_1^n)p(u_2^n)p(y_1^n u_2^n)$ |
| 5 | 1 | × | × | $p(u_1^n, x_1^n)p(u_2^n)p(y_1^n u_1^n)$ |
| 6 | × | × | 1 | $p(u_1^n, x_1^n)p(u_2^n)p(y_1^n)$ |
| 7 | × | × | × | $p(u_1^n, x_1^n)p(u_2^n)p(y_1^n)$ |
| 8 | 1 | × | 1 | $p(u_1^n, x_1^n)p(u_2^n)p(y_1^n x_1^n)$ |

- Cases 3,4 and 6,7 share same pmf, and case 8 does not cause an error

- We are left with only 5 error events:

$$\mathcal{E}_{10} := \{(Q^n, U_1^n(1), U_2^n(1), X_1^n(1,1), Y_1^n) \notin \mathcal{T}_\epsilon^{(n)}\},$$

$$\mathcal{E}_{11} := \{(Q^n, U_1^n(1), U_2^n(1), X_1^n(1, m_{11}), Y_1^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_{11} \neq 1\},$$

$$\mathcal{E}_{12} := \{(Q^n, U_1^n(m_{10}), U_2^n(1), X_1^n(m_{10}, m_{11}), Y_1^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_{10} \neq 1, m_{11}\},$$

$$\mathcal{E}_{13} := \{(Q^n, U_1^n(1), U_2^n(m_{20}), X_1^n(1, m_{11}), Y_1^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_{20} \neq 1, m_{11} \neq 1\},$$

$$\mathcal{E}_{14} := \{(Q^n, U_1^n(m_{10}), U_2^n(m_{20}), X_1^n(m_{10}, m_{11}), Y_1^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } m_{10} \neq 1, m_{20} \neq 1, m_{11}\}$$

Then, the average probability of error for decoder 1 is

$$P(\mathcal{E}_1) \leq \sum_{j=0}^4 P(\mathcal{E}_{1j})$$

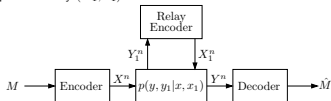
- Now, we bound each probability of error term
- By the LLN, $P(\mathcal{E}_{10}) \rightarrow 0$ as $n \rightarrow \infty$
- By the packing lemma, $P(\mathcal{E}_{11}) \rightarrow 0$ as $n \rightarrow \infty$ if $R_{11} < I(X_1; Y_1|U_1, U_2, Q) - \delta(\epsilon)$
- By the packing lemma, $P(\mathcal{E}_{12}) \rightarrow 0$ as $n \rightarrow \infty$ if $R_{11} + R_{10} < I(X_1; Y_1|U_2, Q) - \delta(\epsilon)$
- By the packing lemma, $P(\mathcal{E}_{13}) \rightarrow 0$ as $n \rightarrow \infty$ if $R_{11} + R_{20} < I(X_1, U_2; Y_1|U_1, Q) - \delta(\epsilon)$
- By the packing lemma, $P(\mathcal{E}_{14}) \rightarrow 0$ as $n \rightarrow \infty$ if $R_{11} + R_{10} + R_{20} < I(X_1, U_2; Y_1|Q) - \delta(\epsilon)$
- The average probability of error for decoder 2 can be bounded similarly
- Finally, we use the Fourier-Motzkin procedure with the constraints $R_{j0} = R_j - R_{jj}, 0 \leq R_{jj} \leq R_j$ for $j = 1, 2$, to eliminate R_{11}, R_{22} and obtain the region given in the theorem (see Appendix II)

Summary of Achievability Ideas

- Coding techniques:
 - Random codebook generation
 - Joint typicality decoding (packing)
 - Successive cancellation decoding
 - Simultaneous joint typicality decoding
 - Time sharing and coded time sharing
 - Superposition coding
 - Rate splitting
 - Indirect decoding
- Performance analysis tools:
 - Division of error event
 - LLN
 - Properties of typicality
 - Packing lemma
 - Fourier-Motzkin elimination

DM Relay Channel

- Consider a DM-RC $(\mathcal{X} \times \mathcal{X}_1, p(y, y_1|x, x_1), \mathcal{Y} \times \mathcal{Y}_1)$
- The sender X wishes to send a message to the receiver Y with the help of the relay (X_1, Y_1)



- A $(2^{nR}, n)$ code for a DM-RC consists of:
 - A message set $[1 : 2^{nR}]$
 - An encoder that assigns a codeword $x^n(m)$ to each message $m \in [1 : 2^{nR}]$
 - A relay encoder that assigns a symbol $x_{1i}(y_i^{i-1})$ to every received sequence y_1^{i-1} for $i \in [1 : n]$
 - A decoder that assigns to each received sequence y^n an estimate $\hat{m}(y^n) \in [1 : 2^{nR}]$ or an error e

- Assume that $M \sim \text{Unif}[1 : 2^{nR}]$
- The average probability of error is $P_e^{(n)} = \mathbb{P}\{\hat{M} \neq M\}$
- A rate R is achievable for the DM-RC if there exists a sequence of $(2^{nR}, n)$ codes with $P_e^{(n)} \rightarrow 0$
- The *capacity* C of the DM-RC is the supremum of all achievable rates
- The capacity of the DM-RC is not known in general
- We know an upper bound (cutset bound) and several lower bounds on capacity that are tight only for some special cases
- We present the *decode-forward* coding scheme, which involves the new ideas of [block Markov coding](#) and [backward decoding](#)
- For reference

Cutset Upper Bound [29]

$$C \leq \max_{p(x_1, x_2)} \min\{I(X, X_1; Y), I(X; Y, Y_1|X_1)\}$$

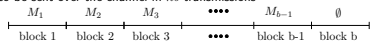
Multi-hop Lower Bound

- In this scheme, the relay decodes the message received from the sender in each block and retransmits it in the following block

Proposition: Multi-hop Lower Bound

$$C \geq \max_{p(x)p(x_1)} \min\{I(X_1; Y), I(X; Y_1|X_1)\}$$

- Achievability uses the following multi-block scheme: Consider b blocks, each consisting of n transmissions. A sequence of $b-1$ i.i.d. messages $m_j \in [1 : 2^{nR}]$, $j \in [1 : b-1]$, is to be sent over the channel in nb transmissions



- So the average rate is $R(b-1)/b \rightarrow R$ as $b \rightarrow \infty$

Proof of Achievability

- Codebook generation:** Fix $p(x)p(x_1)$. Randomly and independently generate $x^n(m)$, $m \in [1 : 2^{nR}]$, each according to $\prod_{i=1}^n p_X(x_i)$, and $x_1^n(m)$, $m \in [1 : 2^{nR}]$, each according to $\prod_{i=1}^n p_{X_1}(x_{1i})$
- Encoding:** To send m_j in block j , the encoder transmits $x^n(m_j)$. The relay has an estimate \hat{m}_{j-1} of m_{j-1} . It transmits $x_1^n(\hat{m}_{j-1})$
- Decoding and analysis of the probability of error:** Assume that at the end of block $j-1$ the relay knows $(m_1, m_2, \dots, m_{j-1})$. The decoding procedures at the end of block j are as follows:
 - Upon receiving $y_1^n(j)$, the relay declares that \hat{m}_j is sent if it is the unique message such that $(x^n(\hat{m}_j), x_1^n(m_{j-1}), y_1^n(j)) \in \mathcal{T}_\epsilon^{(n)}$. By the LLN and the packing lemma, the probability of error $\rightarrow 0$ as $n \rightarrow \infty$ if $R < I(X; Y_1|X_1) - \delta(\epsilon)$
 - The receiver declares that \hat{m}_{j-1} is sent if it is the unique message such that $(x_1^n(\hat{m}_{j-1}), y^n(j)) \in \mathcal{T}_\epsilon^{(n)}$. By the LLN and the packing lemma, the probability of error $\rightarrow 0$ as $n \rightarrow \infty$ if $R < I(X_1; Y) - \delta(\epsilon)$

Bounding the Probability of Error for the b Blocks

- Consider the probability of error averaged over codebooks:

$$P(\mathcal{E}) = P\{(M_1, M_2, \dots, M_{b-1}) \neq (\tilde{M}_1, \tilde{M}_2, \dots, \tilde{M}_{b-1})\}$$

- For $j \in [1 : b-1]$, define $\mathcal{E}_j := \{\tilde{M}_j \neq M_j \text{ or } \hat{M}_{j-1} \neq \tilde{M}_{j-1}\}$
- Now consider

$$\begin{aligned} P(\mathcal{E}) &\leq P\left(\bigcup_{j=1}^{b-1} \mathcal{E}_j\right) \\ &= \sum_{j=1}^{b-1} P(\mathcal{E}_j \cap \mathcal{E}_1^c \cap \mathcal{E}_2^c \cap \dots \cap \mathcal{E}_{j-1}^c) \\ &\leq \sum_{j=1}^{b-1} P(\mathcal{E}_j | \mathcal{E}_1^c \cap \mathcal{E}_2^c \cap \dots \cap \mathcal{E}_{j-1}^c) \end{aligned}$$

- Thus $P(\mathcal{E}) \rightarrow 0$ if $P(\mathcal{E}_j | \mathcal{E}_1^c \cap \mathcal{E}_2^c \cap \dots \cap \mathcal{E}_{j-1}^c) \rightarrow 0$ for all $j \in [1 : b-1]$

Cooperative Multi-hop Lower Bound

- The multi-hop scheme can be improved by having the sender and relay *coherently* cooperate in transmitting their codewords

Proposition: Cooperative Multi-hop Lower Bound

$$C \geq \max_{p(x, x_1)} \min\{I(X_1; Y), I(X; Y_1 | X_1)\}$$

- Again consider b blocks each consisting of n transmissions; a sequence of $b-1$ i.i.d. messages m_j , $j \in [1 : b-1]$, is to be sent in nb transmissions. We use a **block Markov coding** scheme
- Codebook generation:** Fix $p(x_1)p(x|x_1)$ that achieve the lower bound. Randomly and independently generate $x_1^n(m')$, $m' \in [1 : 2^{nR}]$, sequences each according to $\prod_{i=1}^n p_{X_1}(x_{1i})$. For each $x_1^n(m')$, randomly and conditionally independently generate $x^n(m|m')$, $m \in [1 : 2^{nR}]$, sequences each according to $\prod_{i=1}^n p_{X|X_1}(x_i|x_{1i}(m'))$

- Encoding and decoding are described in the following table:

| Block | 1 | 2 | 3 | ... | $b-1$ | b |
|-------|---------------|----------------------|----------------------|-----|--------------------------|--------------------------|
| X | $x^n(m_1 1)$ | $x^n(m_2 m_1)$ | $x^n(m_3 m_2)$ | ... | $x^n(m_{b-1} m_{b-2})$ | $x^n(1 m_{b-1})$ |
| Y_1 | \tilde{m}_1 | \tilde{m}_2 | \tilde{m}_3 | ... | \tilde{m}_{b-1} | \emptyset |
| X_1 | $x_1^n(1)$ | $x_1^n(\tilde{m}_1)$ | $x_1^n(\tilde{m}_2)$ | ... | $x_1^n(\tilde{m}_{b-2})$ | $x_1^n(\tilde{m}_{b-1})$ |
| Y | \emptyset | \hat{m}_1 | \hat{m}_2 | ... | \hat{m}_{b-2} | \hat{m}_{b-1} |

- Encoding:** To send m_j in block j , the encoder sends $x^n(m_j|m_{j-1})$. The relay has an estimate \tilde{m}_{j-1} of the previous message m_{j-1} . It sends $x_1^n(\tilde{m}_{j-1})$ in block j .

Decoding and Analysis of the Probability of Error

- Assume that at the end of block $j-1$ the relay knows $(m_1, m_2, \dots, m_{j-1})$. The decoding procedures at the end of block j are as follows:
- Upon receiving $y_1^n(j)$, the relay receiver declares that \tilde{m}_j is sent if it is the unique message such that $(x^n(\tilde{m}_j|m_{j-1}), x_1^n(m_{j-1}), y_1^n(j)) \in \mathcal{T}_\epsilon^n$; otherwise an error is declared. By the LLN and the packing lemma, the probability of error $\rightarrow 0$ as $n \rightarrow \infty$ if $R < I(X; Y_1 | X_1) - \delta(\epsilon)$.
- The receiver declares that \hat{m}_{j-1} is sent if it is the unique message such that $(x_1^n(\hat{m}_{j-1}), y_1^n(j)) \in \mathcal{T}_\epsilon^n$; otherwise an error is declared. By the LLN and the packing lemma, the probability of error $\rightarrow 0$ as $n \rightarrow \infty$ if $R < I(X_1; Y) - \delta(\epsilon)$.
- The probability of error for the $b-1$ messages can be upper bounded as before

Decode-Forward Lower Bound

- The cooperative multi-hop scheme can be improved by using both the information received through the direct path and from the relay. This leads to the following *decode-forward* lower bound on the capacity of the DM-RC

Theorem (Cover-El Gamal [29])

$$C \geq \max_{p(x,x_1)} \min \{I(X, X_1; Y), I(X; Y_1|X_1)\}$$

- Remark: The decode-forward lower bound is tight when the DM-RC is *physically degraded*, i.e.,

$$p(y, y_1|x, x_1) = p(y_1|x, x_1)p(y|y_1, x_1)$$

Proof of Achievability

- Again we consider b transmission blocks, each consisting of n transmissions and use a block Markov coding scheme
A sequence of $b-1$ i.i.d. messages, $m_j \in [1 : 2^{nR}]$, $j \in [1 : b-1]$, is to be sent over the channel in the nb transmissions
- Codebook generation:** We use the same codebook generation as in cooperative multi-hop. Fix $p(x_1)p(x|x_1)$ that achieves lower bound Randomly and independently generate $x_1^n(m')$, $m' \in [1, 2^{nR}]$, sequences each according to $\prod_{i=1}^n p_{X_1}(x_{1i})$. For each $x_1^n(m')$, randomly and conditionally independently generate $x^n(m|x')$, $m \in [1, 2^{nR}]$, sequences each according to $\prod_{i=1}^n p_{X|X_1}(x_i|x_{1i}(m'))$
- Encoding:** Encoding is again the same as in cooperative multi-hop To send m_j in block j , the encoder sends $x^n(m_j|m_{j-1})$
The relay has an estimate \hat{m}_{j-1} of the previous message m_{j-1}
It sends $x_1^n(\hat{m}_{j-1})$ in block j

Backward Decoding

- Decoding at the receiver is done backwards after all bn transmissions are received [30]
- Encoding and decoding are described in the following table:

| Block | 1 | 2 | 3 | ... | $b-1$ | b |
|-------|-------------------------|-------------------------|-------------------------|-----|----------------------------|----------------------------|
| X | $x^n(m_1 1)$ | $x^n(m_2 m_1)$ | $x^n(m_3 m_2)$ | ... | $x^n(m_{b-1} m_{b-2})$ | $x^n(1 m_{b-1})$ |
| Y_1 | $\hat{m}_1 \rightarrow$ | $\hat{m}_2 \rightarrow$ | $\hat{m}_3 \rightarrow$ | ... | \hat{m}_{b-1} | \emptyset |
| X_1 | $x_1^n(1)$ | $x_1^n(\hat{m}_1)$ | $x_1^n(\hat{m}_2)$ | ... | $x_1^n(\hat{m}_{b-2})$ | $x_1^n(\hat{m}_{b-1})$ |
| Y | \emptyset | \hat{m}_1 | $\leftarrow \hat{m}_2$ | ... | $\leftarrow \hat{m}_{b-2}$ | $\leftarrow \hat{m}_{b-1}$ |

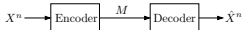
- Decoding at the relay is done as in cooperative multi-hop
Suppose the relay knows $(m_1, m_2, \dots, m_{j-1})$ at the end of block $j-1$
Upon receiving $y_1^n(j)$, the relay receiver declares \hat{m}_j is sent if it is the unique message such that $(x^n(\hat{m}_j|m_{j-1}), x_1^n(m_{j-1}), y_1^n(j)) \in \mathcal{T}_\epsilon^{(n)}$; otherwise an error is declared
By the LLN and the packing lemma, the probability of error $\rightarrow 0$ as $n \rightarrow \infty$ if $R < I(X; Y_1|X_1) - \delta(\epsilon)$
- Decoding at the receiver is done successively backwards
Suppose the receiver has successfully decoded $(m_{b-1}, \dots, m_{j+1})$
Based on the received signal $y^n(j+1)$, the receiver finds the unique message \hat{m}_j such that $(x^n(m_{j+1}|\hat{m}_j), x_1^n(\hat{m}_j), y^n(j+1)) \in \mathcal{T}_\epsilon^{(n)}$; otherwise an error is declared
By the LLN and the packing lemma, the probability of error $\rightarrow 0$ as $n \rightarrow \infty$ if $R < I(X, X_1; Y) - \delta(\epsilon)$
- The probability of error for the b blocks can be bounded as before
- Remark: The excessive delay of backward decoding can be alleviated via *binning* [29] or *sliding window decoding* [31]

Summary of Achievability Ideas

- **Coding techniques:**
 - ▶ Random codebook generation
 - ▶ Joint typicality decoding (packing)
 - ▶ Successive cancellation decoding
 - ▶ Simultaneous joint typicality decoding
 - ▶ Time sharing and coded time sharing
 - ▶ Superposition coding
 - ▶ Rate splitting
 - ▶ Indirect decoding
 - ▶ **Block Markov coding**
 - ▶ **Backward decoding**
- **Performance analysis tools:**
 - ▶ Division of error event
 - ▶ LLN
 - ▶ Properties of typicality
 - ▶ Packing lemma
 - ▶ Fourier–Motzkin elimination

Discrete Memoryless Source

- Let X be a discrete memoryless source (DMS) (i.e., i.i.d. finite alphabet source) $(\mathcal{X}, p(x))$ and $d(x, \hat{x})$, $\hat{x} \in \mathcal{X}$, be a distortion measure
- The decoder wishes to obtain a description of X with distortion D



- The average per-letter distortion between x^n and \hat{x}^n is

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

- A $(2^{nR}, n)$ *lossy source code* consists of:
 - An encoder that assigns to each sequence $x^n \in \mathcal{X}^n$ an index $m(x^n) \in [1 : 2^{nR}] := \{1, 2, \dots, 2^{[nR]}\}$, and
 - A decoder that assigns to each index $m \in [1 : 2^{nR}]$ an estimate $\hat{x}^n(m) \in \mathcal{X}^n$

- The distortion associated with the $(2^{nR}, n)$ code is defined as

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

- A rate distortion pair (R, D) is *achievable* if there exists a sequence of $(2^{nR}, n)$ codes with $\limsup_{n \rightarrow \infty} \mathbb{E}(d(X^n, \hat{X}^n)) \leq D$
- The *rate distortion function* $R(D)$ is the infimum of rates R such that (R, D) is achievable
- Note that $R(D)$ is nonincreasing and convex (thus continuous)

Theorem (Shannon [17])

The rate distortion function for a DMS $(\mathcal{X}, p(x))$ and distortion measure $d(x, \hat{x})$ for $D \geq D_{\min} := \mathbb{E}(\min_{\hat{x}} d(X, \hat{x}))$ is

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

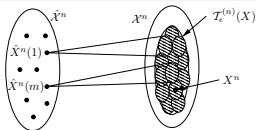
- The new achievability ideas are **joint typicality encoding** (covering) and the **covering lemma**

Proof of Achievability

- **Codebook generation:** Fix $p(\hat{x}|x)$ that achieves $R(D)/(1 + \epsilon)$, and compute $p(\hat{x}) = \sum_x p(x)p(\hat{x}|x)$. Randomly and independently generate $\hat{x}^n(m)$, $m \in [1 : 2^{nR}]$, sequences each according to $\prod_{i=1}^n p_{\hat{X}}(\hat{x}_i)$. The codebook is revealed to both the encoder and decoder
- **Encoding:** We use *joint typicality encoding*. Given a sequence x^n , choose an index m such that $(x^n, \hat{x}^n(m)) \in \mathcal{T}_\epsilon^{(n)}$. If there is more than one such m , choose the smallest index. If there is no such m , choose $m = 1$.
- **Decoding:** Upon receiving the index m , the decoder chooses the reproduction sequence $\hat{x}^n(m)$.
- The following lemma establishes the condition for successful joint typicality encoding

Covering Lemma

Let $(U, X, \hat{X}) \sim p(u, x, \hat{x})$, $\epsilon > 0$, and $(U^n, X^n) \in \mathcal{T}_\epsilon^{(n)}$. Let $\hat{X}^n(m)$, $m \in \mathcal{A}$, $|\mathcal{A}| \geq 2^{nR}$, be random sequences conditionally independent of each other and X^n given U^n , each distributed according to $\prod_{i=1}^n p_{\hat{X}|U}(\hat{x}_i|u_i)$. There exists $\delta(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$ such that if $R > I(X; \hat{X}|U) + \delta(\epsilon)$, then

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


- The proof is in Appendix III

Analysis of Average Distortion

- We bound the distortion averaged over X^n and the codebook
- Define the "error" event

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

and partition it into the events

$$\mathcal{E}_1 := \{X^n \notin \mathcal{T}_\epsilon^{(n)}\},$$

$$\mathcal{E}_2 := \{X^n \in \mathcal{T}_\epsilon^{(n)}, (X^n, \hat{X}^n(m)) \notin \mathcal{T}_\epsilon^{(n)} \text{ for all } m \in [1 : 2^{nR}]\}$$

Then, $P(\mathcal{E}) = P(\mathcal{E}_1) + P(\mathcal{E}_2)$

- By the LLN, $P(\mathcal{E}_1) \rightarrow 0$ as $n \rightarrow \infty$
- By the covering lemma (with $U = \emptyset$), $P(\mathcal{E}_2) \rightarrow 0$ as $n \rightarrow \infty$ if $R > I(X; \hat{X}) + \delta(\epsilon)$

- Now, denote the reproduction sequence for X^n by \hat{X}^n , then by the law of total expectation and the typical average lemma

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

where $d_{\max} := \max_{(x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}} d(x, \hat{x})$

- Since by assumption $E(d(X, \hat{X})) \leq D/(1 + \epsilon)$,

$$\limsup_{n \rightarrow \infty} E(d(X^n, \hat{X}^n)) \leq D,$$

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

- Since the average per-letter distortion is asymptotically $\leq D$, there must exist a sequence of codes with asymptotic distortion $\leq D$, which proves the achievability of $(R(D)/(1 + \epsilon)) + \delta(\epsilon)$, D
- Finally, since $R(D)$ is continuous in D , $R(D/(1 + \epsilon)) + \delta(\epsilon) \rightarrow R(D)$ as $\epsilon \rightarrow 0$

Lossless Source Coding

- In lossless source coding, we wish to find the minimum rate R^* such that if $R > R^*$, then there exists a sequence of $(2^{nR}, n)$ codes with $P_e^{(n)} = P\{\hat{X}^n \neq X^n\} \rightarrow 0$ as $n \rightarrow \infty$

Theorem (Shannon [6])

The optimal lossless source coding rate for DMS X is $R^* = H(X)$

- We show that this lossless source coding theorem is a special case of the lossy source coding theorem
- Consider the lossy source coding problem for a DMS X , reproduction alphabet $\hat{\mathcal{X}} = \mathcal{X}$, and Hamming distortion measure $(d(x, \hat{x}) = 0$ if $x = \hat{x}$, and $d(x, \hat{x}) = 1$ otherwise)
- If we let $D = 0$, we can show that $R(0) = H(X)$ (why?)
- We want to show that $R^* = R(0)$

- The converse for the lossy source coding theorem under the above conditions implies that for any sequence of $(2^{nR}, n)$ codes if the average per-symbol error probability $(1/n) \sum_{i=1}^n \mathbb{P}\{\hat{X}_i \neq X_i\} \rightarrow 0$, then $R \geq H(X)$

Since the average per-symbol error probability is \leq the block error probability $P_e^{(n)}$, then $R^* \geq H(X)$

- As for the achievability, we can still use random coding and covering!

Fix a test channel $p(\hat{x}|x) = 1$ if $x = \hat{x}$ and 0, otherwise

Note that if $(x^n, \hat{x}^n) \in \mathcal{T}_\epsilon^{(n)}$, then $x^n = \hat{x}^n$

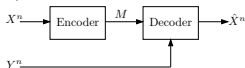
As before, randomly generate a codebook $\hat{x}^n(m)$, $m \in [1 : 2^{nR}]$

By the covering lemma if $R > I(X; \hat{X}) + \delta(\epsilon) = H(X) + \delta(\epsilon)$, the average probability of error $\rightarrow 0$ as $n \rightarrow \infty$

Thus there exists a sequence of $(2^{nR}, n)$ lossless source codes with $P_e^{(n)} \rightarrow 0$ as $n \rightarrow \infty$

DMS with Side Information at Decoder

- Let (X, Y) be a 2-DMS $(\mathcal{X}, \mathcal{Y}, p(x, y))$ and $d(x, \hat{x})$ be a distortion measure
- The decoder observes the DMS Y (side information) and wishes to obtain a description of the DMS X with distortion D



- A $(2^{nR}, n)$ rate distortion code with side information available at the decoder consists of:
 - An encoder that assigns to each $x^n \in \mathcal{X}^n$ an index $m(x^n) \in [1 : 2^{nR}]$
 - A decoder that assigns an estimate $\hat{x}^n(m, y^n)$ to each received index m and side information sequence y^n

- The *rate distortion function with side information available at the decoder*, $R_{\text{SI-D}}(D)$, is the infimum of rates R such that there exists a sequence of $(2^{nR}, n)$ codes with $\limsup_{n \rightarrow \infty} E(d(X^n, \hat{X}^n)) \leq D$
- Note that $R_{\text{SI-D}}(D)$ is nonincreasing and convex (thus continuous)

Theorem (Wyner-Ziv [18])

Let (X, Y) be a 2-DMS and $d(x, \hat{x})$ be a distortion measure. The rate distortion function for X with side information Y available only at the decoder is

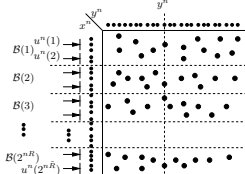
$$R_{\text{SI-D}}(D) = \min(I(X; U) - I(Y; U)) = \min I(X; U|Y),$$

where the minimization is over all functions $\hat{x} : \mathcal{Y} \times \mathcal{U} \rightarrow \hat{\mathcal{X}}$ and conditional pmfs $p(u|x)$ such that $\mathbb{E}_{X,Y,U}(d(X, \hat{X})) \leq D$

- The new achievability ideas are [binning](#) and the [Markov lemma](#)

Outline of Achievability

- Fix $p(u|x)$, $\hat{x}(u, y)$. Generate $2^{nI(U;X)}$ $u^n(l)$ sequences and partition their indices into 2^{nR} bins
- Given x^n find a jointly typical u^n sequence. Send the bin index of u^n



- The decoder finds unique $l \in \mathcal{B}(m)$ such that $(u^n(l), y^n) \in \mathcal{T}_\epsilon^{(n)}$
- The decoder then constructs the estimate $(\hat{x}(u_1, y_1), \dots, \hat{x}(u_n, y_n))$

Proof of Achievability

- **Codebook generation:** Fix $p(u|x)$, $\hat{x}(u, y)$ that achieve the rate distortion function $R_{\text{SI-D}}(D/(1+\epsilon))$. Randomly and independently generate $u^n(l)$, $l \in [1 : 2^{n\tilde{R}}]$, sequences each according to $\prod_{i=1}^n p(u_i)$. Partition the set of indices $l \in [1 : 2^{n\tilde{R}}]$ into equal-size bins $\mathcal{B}(m) := [(m-1)2^{n(\tilde{R}-R)} + 1 : m2^{n(\tilde{R}-R)}]$, $m \in [1 : 2^{nR}]$.
- **Encoding:** Given x^n , the encoder finds an l such that $(x^n, u^n(l)) \in \mathcal{T}_\epsilon^{(n)}$. If there is more than one such index, it selects one uniformly at random. If there is no such index, it selects an arbitrary l . The encoder sends the index m such that $l \in \mathcal{B}(m)$.
- **Decoding:** Let $\epsilon > \epsilon'$. The decoder finds the unique $\tilde{l} \in \mathcal{B}(m)$ such that $(u^n(\tilde{l}), y^n) \in \mathcal{T}_{\epsilon'}^{(n)}$. If there is such an index \tilde{l} , the reproduction \hat{x}^n is computed as $\hat{x}_i = \hat{x}(u_i(\tilde{l}), y_i)$, $i \in [1 : n]$; otherwise an arbitrary \hat{x}^n is chosen.

Analysis of Average Distortion

- Let L be the index of the chosen U^n sequence and M be the bin index of L . Define the "error" events:

$$\mathcal{E}_0 := \{(X^n, Y^n) \notin \mathcal{T}_\epsilon^{(n)}\}$$

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

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

$$\mathcal{E}_3 := \{(U^n(\tilde{l}), Y^n) \in \mathcal{T}_{\epsilon'}^{(n)} \text{ for some } \tilde{l} \in \mathcal{B}(M), \tilde{l} \neq L\}$$

The total probability of "error" is bounded by

$$P(\mathcal{E}) \leq P(\mathcal{E}_0) + P(\mathcal{E}_0^c \cap \mathcal{E}_1) + P(\mathcal{E}_1^c \cap \mathcal{E}_2) + P(\mathcal{E}_3)$$

- By the LLN, $P(\mathcal{E}_0) \rightarrow 0$ as $n \rightarrow \infty$
- By the covering lemma, $P(\mathcal{E}_0^c \cap \mathcal{E}_1) \rightarrow 0$ as $n \rightarrow \infty$ if $\tilde{R} > I(X; U) + \delta(\epsilon')$

- Since $\mathcal{E}_1^c = \{(U^n(L), X^n) \in \mathcal{T}_{\epsilon'}^{(n)}\}$, $Y^n | \{X^n = x^n\} \sim \prod_{i=1}^n p_{Y_i|X}(y_i|x_i)$, and $\epsilon > \epsilon'$, by the Markov lemma, $P(\mathcal{E}_1^c \cap \mathcal{E}_2) \rightarrow 0$ as $n \rightarrow \infty$
- For $P(\mathcal{E}_3)$, we can show via a sequence of careful conditioning, using the symmetry of code generation, and simple probability bounds (see Appendix IV) that

$$P(\mathcal{E}_3) \leq P\{(U^n(\tilde{l}), Y^n) \in \mathcal{T}_{\epsilon'}^{(n)} \text{ for some } \tilde{l} \in \mathcal{B}(1)\}$$

- Since each $U^n \sim \prod_{i=1}^n p(u_i)$ and is independent of Y^n , by the packing lemma, $P(\mathcal{E}_3) \rightarrow 0$ as $n \rightarrow \infty$ if $\tilde{R} - R < I(Y; U) - \delta(\epsilon)$
- Combining the bounds, we have shown that $P(\mathcal{E}) \rightarrow 0$ as $n \rightarrow \infty$ if $R > I(X; U) - I(Y; U) + \delta(\epsilon) + \delta(\epsilon')$

- When there is no error, $(U^n(L), X^n, Y^n) \in \mathcal{T}_\epsilon^{(n)}$. Thus by the typical average lemma, the asymptotic distortion averaged over the random codebook is bounded as

$$E(d(X^n, \hat{X}^n)) \leq d_{\max} P(\mathcal{E}) + (1 + \epsilon) E(d(X, \hat{X})) P(\mathcal{E}^c) \leq D$$

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

- Finally, by the continuity of $R_{\text{SI-D}}(D)$ in D , taking $\epsilon \rightarrow 0$ shows that any (R, D) pair with $R > R_{\text{SI-D}}(D)$ is achievable, which completes the proof

Lossless Source Coding with Side Information

- In lossless source coding with side information at the decoder, we wish to find the minimum rate $R_{\text{SI-D}}^*$ needed to recover X with $P_e^{(n)} \rightarrow 0$

Theorem (Slepian–Wolf [19])

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

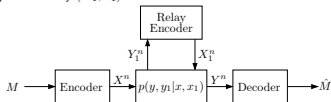
- We show that Slepian–Wolf is a special of Wyner–Ziv
- Let d be the Hamming distortion measure and consider the case of $D = 0$. It is not difficult to show that in this case

$$R_{\text{SI-D}}(0) = H(X|Y),$$

- Can show that $R_{\text{SI-D}}^* = R_{\text{SI-D}}(0) = H(X|Y)$
- The converse follows by the converse for the Wyner–Ziv theorem
- Achievability follows by covering, binning, and packing with $U = X$

DM Relay Channel

- Again, consider a DM-RC $(\mathcal{X} \times \mathcal{X}_1, p(y, y_1|x, x_1), \mathcal{Y} \times \mathcal{Y}_1)$
- The sender X wishes to send a message to the receiver Y with the help of the relay (X_1, Y_1)



- The code and achievability are defined as before
- Here, we present the *compress-forward* coding scheme
- For reference

Cutset Upper Bound [29]

$$C \leq \max_{p(x_1, x_2)} \min \{I(X, X_1; Y), I(X; Y, Y_1|X_1)\}$$

Compress-Forward Lower Bound

- In the decode-forward scheme, the relay decodes the entire message. If the channel from the sender to the relay is worse than the direct channel to the receiver, this requirement can make the transmission rate lower than that when the relay is not used at all
- In the *compress-forward* scheme, the relay helps communication by sending a compressed version of its received signal to the receiver. Because this compressed version is correlated with the received sequence, Wyner–Ziv coding is used to reduce its rate to the receiver. This scheme achieves the following lower bound

Theorem (Cover–El Gamal [29], El Gamal–Mohseni–Zahedi [32])

$$C \geq \max_{p(x)p(x_1)p(\hat{y}_1|y_1, x_1)} \min \{I(X, X_1; Y) - I(Y_1; \hat{Y}_1|X, X_1, Y), I(X; Y, \hat{Y}_1|X_1)\}$$

- This characterization is equivalent to that in Cover–El Gamal [29]

Outline of Achievability

- A block Markov coding scheme is used to send $b - 1$ i.i.d. messages in b blocks
- In block j , a “compressed” version $\hat{y}_1^n(j - 1)$ of $y_1^n(j - 1)$ conditioned on $x_1^n(j - 1)$, which is known to both the relay and receiver, is constructed by the relay
- As in Wyner–Ziv coding, $\hat{y}_1^n(j - 1)$ is encoded with side information $y^n(j - 1)$ at the receiver and sent to the receiver via $x_1^n(j)$
- At the end of block j , the receiver decodes $x_1^n(j)$ and finds $\hat{y}_1^n(j - 1)$
- The receiver then combines $\hat{y}_1^n(j - 1)$ with $y^n(j - 1)$ to decode m_{j-1}

Proof of Achievability

- **Codebook generation:** Fix $p(x)p(x_1)p(\hat{y}_1|y_1, x_1)$ that achieves the lower bound
Randomly and independently generate $x^n(m)$, $m \in [1 : 2^{nR}]$, sequences each according to $\prod_{i=1}^n p_X(x_i)$
Randomly and independently generate $x_1^n(l)$, $l \in [1 : 2^{nR_1}]$, sequences each according to $\prod_{i=1}^n p_{X_1}(x_{1i})$
For each $x_1^n(l)$, $l \in [1 : 2^{nR_1}]$, randomly and conditionally independently generate 2^{nR_1} sequences $\hat{y}_1^n(k|l)$, $k \in [1, 2^{nR_1}]$, each according to $\prod_{i=1}^n p_{Y_1|X_1}(\hat{y}_{1i}|x_{1i}(l))$
Partition the set $[1 : 2^{nR}]$ into 2^{nR_1} equal size bins $\mathcal{B}(l)$, $l \in [1 : 2^{nR_1}]$

- **Encoding:** Let m_j be the message to be sent in block j , and assume that $(\hat{y}_1^n(k_{j-1}|l_{j-2}), y_1^n(j-1), x_1^n(l_{j-2})) \in \mathcal{T}_\epsilon^{(n)}$ and $k_{j-1} \in \mathcal{B}(l_{j-1})$. Then the codeword pair $(x^n(m_j), x_1^n(l_{j-1}))$ is transmitted in block j
- **Decoding and analysis of the probability of error:** Assume that at the end of block $j-1$ the receiver knows (m_1, \dots, m_{j-2}) and (l_1, \dots, l_{j-2}) , the decoding procedure at the end of block j is as follows:
 - The relay, upon receiving $y_1^n(j)$, finds an index k such that $(\hat{y}_1^n(k|l_{j-1}), y_1^n(j), x_1^n(l_{j-1})) \in \mathcal{T}_\epsilon^{(n)}$
By the LLN and the covering lemma, the probability of error $\rightarrow 0$ as $n \rightarrow \infty$ if $R_1 > I(Y_1; Y_1|X_1) + \delta(\epsilon')$
 - Upon receiving $y^n(j)$, the receiver finds the unique \hat{l}_{j-1} such that $(x_1^n(\hat{l}_{j-1}), y^n(j)) \in \mathcal{T}_\epsilon^{(n)}$
By the LLN and the packing lemma, the probability of error $\rightarrow 0$ as $n \rightarrow \infty$ if $R_1 < I(X_1; Y) - \delta(\epsilon)$

- Encoding and decoding are described in the following table:

| Block | 1 | 2 | 3 | ... | $b-1$ | b |
|-------|------------------------|------------------------|------------------------|-----|--------------------------------|--------------------------------|
| X | $x^n(m_1)$ | $x^n(m_2)$ | $x^n(m_3)$ | ... | $x^n(m_{b-1})$ | $x^n(1)$ |
| Y_1 | $\hat{y}_1^n(k_1 l_1)$ | $\hat{y}_1^n(k_2 l_1)$ | $\hat{y}_1^n(k_3 l_2)$ | ... | $\hat{y}_1^n(k_{b-1} l_{b-2})$ | \emptyset |
| X_1 | $x_1^n(1)$ | $x_1^n(l_1)$ | $x_1^n(l_2)$ | ... | $x_1^n(l_{b-2})$ | $x_1^n(l_{b-1})$ |
| Y | \emptyset | \hat{l}_1, \hat{k}_1 | \hat{l}_2, \hat{k}_2 | ... | $\hat{l}_{b-2}, \hat{k}_{b-2}$ | $\hat{l}_{b-1}, \hat{k}_{b-1}$ |

- Let $\epsilon > \epsilon'$. Assuming the receiver correctly decodes l_{j-1} , it finds the unique message \hat{m}_{j-1} such that $(x^n(\hat{m}_{j-1}), x_1^n(l_{j-2}), \hat{y}_1^n(\hat{k}_{j-1}|l_{j-2}), y^n(j-1)) \in \mathcal{T}_\epsilon^{(n)}$ for some $\hat{k}_{j-1} \in \mathcal{B}(l_{j-1})$
- To bound the probability of the error, assume that $(M_{j-1}, L_{j-2}) = (1, 1)$ and let K_{j-1} be the index of the covering \hat{Y}_1^n and L_{j-1} be the bin index of K_{j-1} . Define the events
 - $\mathcal{E}_1 := \{(X^n(1), X_1^n(1), \hat{Y}_1^n(K_{j-1}|1), Y^n(j-1)) \notin \mathcal{T}_\epsilon^{(n)}\}$,
 - $\mathcal{E}_2 := \{(X^n(m), X_1^n(1), \hat{Y}_1^n(K_{j-1}|1), Y^n(j-1)) \in \mathcal{T}_\epsilon^{(n)}$
for some $m \neq 1\}$,
 - $\mathcal{E}_3 := \{(X^n(m), X_1^n(1), \hat{Y}_1^n(\hat{k}|1), Y^n(j-1)) \in \mathcal{T}_\epsilon^{(n)}$
for some $\hat{k} \in \mathcal{B}(L_{j-1}), \hat{k} \neq K_{j-1}, m \neq 1\}$

Then the probability of error is bounded by
 $P(\mathcal{E}) \leq P(\mathcal{E}_1) + P(\mathcal{E}_2) + P(\mathcal{E}_3)$

- By the Markov lemma, $P(\mathcal{E}_1) \rightarrow 0$ as $n \rightarrow \infty$
- By the packing lemma, $P(\mathcal{E}_2) \rightarrow 0$ as $n \rightarrow \infty$ if $R < I(X; Y, \hat{Y}_1, X_1) - \delta(\epsilon) = I(X; Y, \hat{Y}_1 | X_1) - \delta(\epsilon)$
- By first bounding $P(\mathcal{E}_3)$ as in Appendix IV, and then using properties of joint typicality and the union of events bound (the packing lemma here is not general enough), we can show $P(\mathcal{E}_3) \rightarrow 0$ as $n \rightarrow \infty$ if

$$R + \tilde{R}_1 - R_1 < I(\hat{Y}_1; X, Y | X_1) + I(X; Y | X_1) - \delta(\epsilon)$$

- Combining with the bounds on \tilde{R}_1 and R_1 , $P(\mathcal{E}) \rightarrow 0$ as $n \rightarrow \infty$ if

$$\begin{aligned} R &< I(X, X_1; Y) + I(\hat{Y}_1; X, Y | X_1) - I(\hat{Y}_1; Y_1 | X_1) - 2\delta(\epsilon) - \delta(\epsilon') \\ &= I(X, X_1; Y) + I(\hat{Y}_1; X, Y | X_1) - I(\hat{Y}_1; Y_1, X, Y | X_1) - \delta'(\epsilon) \\ &= I(X, X_1; Y) - I(\hat{Y}_1; Y_1 | X, X_1, Y) - \delta'(\epsilon) \end{aligned}$$

- This completes the proof of achievability

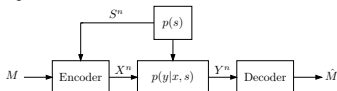
Summary of Achievability Ideas

- Coding techniques:
 - Random codebook generation
 - Joint typicality decoding (packing)
 - Successive cancellation decoding
 - Simultaneous joint typicality decoding
 - Time sharing and coded time sharing
 - Superposition coding
 - Rate splitting
 - Indirect decoding
 - Block Markov coding
 - Backward decoding
 - Joint typicality encoding (covering): dual to packing
 - Binning: Many-to-one indexing

- Performance analysis tools:
 - Division of error event
 - LLN
 - Properties of typicality
 - Packing lemma
 - Fourier-Motzkin elimination
 - Covering lemma

DMC with State Available at Encoder

- Consider a DM channel with state $(\mathcal{X} \times \mathcal{S}, p(y|x, s), \mathcal{Y})$, where \mathcal{S} is a DMS $(\mathcal{S}, p(s))$
- The sender X who knows the state S noncausally wishes to send a message to the receiver Y



- A $(2^{nR}, n)$ code for the channel with state noncausally known at the encoder is defined by an encoder $x^n(m, s^n)$ and a decoder $\hat{m}(y^n)$
- We wish to find the channel capacity $C_{\text{SI-E}}$ for this setting

Theorem (Gelfand-Pinsker [20])

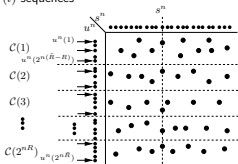
The capacity of a DM channel with state available noncausally at the encoder is

$$C_{\text{SI-E}} = \max_{p(u,x|s)} (I(U; Y) - I(U; S))$$

- The new achievability idea is **subcode generation**, which as we will see is "dual" to binning

Outline of Achievability

- Fix $p(u, x|s)$. For each message $m \in [1 : 2^{nR}]$, generate a **subcode** of $2^{n(\tilde{R}-R)}$ $u^n(l)$ sequences



- To send m given s^n , find $u^n(l) \in C(m)$ that is jointly typical with s^n and transmit $X^n \sim \prod_{i=1}^n P_{X|U,S}(x_i|u_i, s_i)$
- The receiver finds a jointly typical $u^n(l)$ with y^n and declares the subcode index of $u^n(l)$ to be the message sent

Proof of Achievability [21]

- Codebook generation:** Fix $p(u|s)p(x|u, s)$ that achieves capacity. For each message $m \in [1 : 2^{nR}]$ generate a subcode $C(m)$ consisting of $2^{n(\tilde{R}-R)}$ randomly and independently generated $u^n(l)$, $l \in [(m-1)2^{n(\tilde{R}-R)} + 1 : m2^{n(\tilde{R}-R)}]$, sequences each according to $\prod_{i=1}^n p_U(u_i)$
- Encoding:** To send the message $m \in [1 : 2^{nR}]$ with the state sequence s^n observed, the sender chooses a $u^n(l) \in C(m)$ such that $(u^n(l), s^n) \in \mathcal{T}_\epsilon^{(n)}$. If $s^n \notin \mathcal{T}_\epsilon^{(n)}$ or no such $u^n(l)$ exists, then it picks an arbitrary sequence $u^n(l) \in C(m)$. The sender then generates a sequence x^n according to $\prod_{i=1}^n P_{X|U,S}(x_i|u_i(l), s_i)$ and transmits it
- Decoding:** Let $\epsilon > \epsilon'$. Upon receiving y^n , the decoder declares that a message \hat{m} is sent if it is the unique message such that $(u^n(l), y^n) \in \mathcal{T}_\epsilon^{(n)}$ for some $u^n(l) \in C(\hat{m})$; otherwise an error is declared

Analysis of the Probability of Error

- Assume $M = 1$ and let L be the index of the chosen U^n codeword for $M = 1$ and S^n
- Define the events

$$\mathcal{E}_0 := \{S^n \notin \mathcal{T}_{\epsilon'}^{(n)}\}$$

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

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

$$\mathcal{E}_3 := \{(U^n(\tilde{l}), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } U^n(\tilde{l}) \notin C(1)\}$$

The probability of error is bounded by

$$P(\mathcal{E}) \leq P(\mathcal{E}_0) + P(\mathcal{E}_0^c \cap \mathcal{E}_1) + P(\mathcal{E}_1^c \cap \mathcal{E}_2) + P(\mathcal{E}_3)$$

- By the LLN, $P(\mathcal{E}_0) \rightarrow 0$ as $n \rightarrow \infty$
 - By the covering lemma, $P(\mathcal{E}_0^c \cap \mathcal{E}_1) \rightarrow 0$ as $n \rightarrow \infty$ if $\bar{R} - R > I(U; S) + \delta(\epsilon')$
 - By the Markov lemma, since $\epsilon > \epsilon'$, $P(\mathcal{E}_1^c \cap \mathcal{E}_2) \rightarrow 0$ as $n \rightarrow \infty$
 - Since $U^n(\bar{i}) \notin \mathcal{C}(1)$ is distributed according to $\prod_{i=1}^n p(u_i)$ and is independent of Y^n , by the packing lemma, $P(\mathcal{E}_2) \rightarrow 0$ as $n \rightarrow \infty$ if $\bar{R} < I(U; Y) - \delta(\epsilon)$
- Note that here Y^n is not generated i.i.d.
- Thus $P(\mathcal{E}) \rightarrow 0$ as $n \rightarrow \infty$ if $R < I(U; Y) - I(U; S) - \delta(\epsilon) - \delta(\epsilon')$
 - This completes the proof of achievability

Subcode Generation is “Dual” to Binning

- **Subcode generation** is a *channel coding* technique—we have a set of messages and we generate many codewords for each message. To send a message, we send one of the codewords in its subcode
- **Binning** is a *source coding* technique—we have many indices/sequences and we map them into a smaller number of bin indices. To send an index/sequence, we send its bin index

Wyner–Ziv versus Gelfand–Pinsker

- The Wyner–Ziv theorem establishes the rate distortion function for a source X with side information Y noncausally known to the decoder:

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

We proved achievability using binning, covering, packing

- The Gelfand–Pinsker theorem establishes the capacity of DMC with state noncausally known to the encoder:

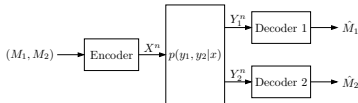
$$C_{\text{SI-E}} = \max(I(U; Y) - I(U; S))$$

We proved achievability using subcode generation, covering, packing

- Note the following “dualities”:
- min is dual to max
- Binning is dual to subcode generation
- $R_{\text{SI-D}}$ is the difference between the *covering rate* $I(U; X)$ and the *packing rate* $I(U; Y)$, while $C_{\text{SI-E}}$ is the difference between the *packing rate* $I(U; Y)$ and the *covering rate* $I(U; S)$

DM-BC with Private Messages

- Consider a 2-receiver DM-BC $(\mathcal{X}, p(y_1, y_2|x), \mathcal{Y}_1 \times \mathcal{Y}_2)$
- Sender X wishes to send a private message to each receiver



- A $(2^{nR_1}, 2^{nR_2}, n)$ code for the DM-BC consists of:
 - Two message sets $[1 : 2^{nR_1}]$ and $[1 : 2^{nR_2}]$
 - An encoder that assigns a codeword $x^n(m_1, m_2)$ to each message pair $(m_1, m_2) \in [1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$
 - Two decoders: Decoder 1 assigns to each $y_1^n \in \mathcal{Y}_1^n$ an estimate $\hat{m}_1(y_1^n) \in [1 : 2^{nR_1}]$ or an error e . Decoder 2 assigns to each $y_2^n \in \mathcal{Y}_2^n$ an estimate $\hat{m}_2(y_2^n) \in [1 : 2^{nR_2}]$ or an error e

- Assume that $(M_1, M_2) \sim \text{Unif}([1 : 2^{nR_1}] \times [1 : 2^{nR_2}])$
- The average probability of error is

$$P_e^{(n)} = \mathbb{P}\{\hat{M}_1 \neq M_1 \text{ or } \hat{M}_2 \neq M_2\}$$

- A rate pair (R_1, R_2) is *achievable* for the DM-BC if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes with $P_e^{(n)} \rightarrow 0$ as $n \rightarrow \infty$
- The *capacity region* of the DM-BC with private messages is the closure of the set of achievable rate pairs
- Capacity region not known in general
- We consider inner bounds

Cover-van der Meulen Inner Bound

Theorem (Cover-van der Meulen [22, 23])

A rate pair (R_1, R_2) is achievable for a DM-BC $(\mathcal{X}, p(y_1, y_2|x), \mathcal{Y}_1 \times \mathcal{Y}_2)$ if

$$R_1 < I(U_1; Y_1), \quad R_2 < I(U_2; Y_2)$$

for some $p(u_1, u_2, x) = p(u_1)p(u_2)p(x|u_1, u_2)$

- Codebook generation:** Fix $p(u_1)p(u_2)p(x|u_1, u_2)$. Randomly and independently generate $u_1^n(m_1)$, $m_1 \in [1 : 2^{nR_1}]$, sequences each according to $\prod_{i=1}^n p_{U_1}(u_{1i})$, and $u_2^n(m_2)$, $m_2 \in [1 : 2^{nR_2}]$, sequences each according to $\prod_{i=1}^n p_{U_2}(u_{2i})$. For each $(u_1^n(m_1), u_2^n(m_2))$ pair, generate the sequence $x^n(m_1, m_2) \sim \prod_{i=1}^n p_{X|U_1, U_2}(x_i|u_{1i}(m_1), u_{2i}(m_2))$

- Encoding:** To send (m_1, m_2) , transmit $x^n(m_1, m_2)$
- Decoding and analysis of the probability of error:** Decoder $j = 1, 2$ declares that message \hat{m}_j is sent if it is the unique message such that $(U_j^n(\hat{m}_j), Y_j^n) \in \mathcal{T}_\epsilon^{(n)}$. By the LLN and packing lemma, the probability of decoding error $\rightarrow 0$ if $R_j < I(U_j; Y_j) - \delta(\epsilon)$ for $j = 1, 2$

Marton Inner Bound

- Marton's inner bound allows U_1, U_2 in the Cover-van der Meulen bound to be dependent (even though the messages are independent). This comes with a penalty term in the sum rate

Theorem (Marton [24, 25])

A rate pair (R_1, R_2) is achievable for a DM-BC $(\mathcal{X}, p(y_1, y_2|x), \mathcal{Y}_1 \times \mathcal{Y}_2)$ if

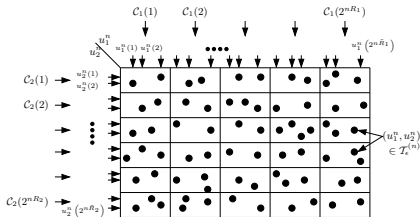
$$\begin{aligned} R_1 &< I(U_1; Y_1), \\ R_2 &< I(U_2; Y_2), \\ R_1 + R_2 &< I(U_1; Y_1) + I(U_2; Y_2) - I(U_1; U_2) \end{aligned}$$

for some $p(u_1, u_2, x) = p(u_1, u_2)p(x|u_1, u_2)$

- Remark: This inner bound is tight for semi-deterministic BC and MIMO Gaussian BC
- The new achievability ideas are **simultaneous joint typicality encoding** and the **mutual covering lemma**

Proof of Achievability

- Codebook generation:** Fix $p(u_1, u_2)p(x|u_1, u_2)$. Let $\tilde{R}_j \geq R_j$, $j = 1, 2$. For each message $m_1 \in [1 : 2^{n\tilde{R}_1}]$ generate a subcode $C_1(m_1)$ consisting of randomly and independently generated $u_1^n(l_1)$, $l_1 \in [(m_1 - 1)2^{n(\tilde{R}_1 - R_1)} + 1 : m_1 2^{n(\tilde{R}_1 - R_1)}]$, sequences each according to $\prod_{i=1}^n p_{U_1}(u_{1i})$. Similarly, for each message $m_2 \in [1 : 2^{n\tilde{R}_2}]$ generate a subcode $C_2(m_2)$ consisting of randomly and independently generated $u_2^n(l_2)$, $l_2 \in [(m_2 - 1)2^{n(\tilde{R}_2 - R_2)} + 1 : m_2 2^{n(\tilde{R}_2 - R_2)}]$, sequences each according to $\prod_{i=1}^n p_{U_2}(u_{2i})$. For $m_1 \in [1 : 2^{n\tilde{R}_1}]$ and $m_2 \in [1 : 2^{n\tilde{R}_2}]$, define $\mathcal{C}(m_1, m_2) := \{(u_1^n(l_1), u_2^n(l_2)) \in C_1(m_1) \times C_2(m_2) : (u_1^n(l_1), u_2^n(l_2)) \in \mathcal{T}_\epsilon^{(n)}\}$



- For each message pair $(m_1, m_2) \in [1 : 2^{n\tilde{R}_1}] \times [1 : 2^{n\tilde{R}_2}]$, pick one index pair (l_1, l_2) such that $(u_1^n(l_1), u_2^n(l_2)) \in \mathcal{C}(m_1, m_2)$. If no such pair exists, pick an arbitrary pair (l_1, l_2) such that $(u_1^n(l_1), u_2^n(l_2)) \in C_1(m_1) \times C_2(m_2)$

- Generate a sequence $x^n(m_1, m_2)$ according to $\prod_{i=1}^n p_{X|U_1, U_2}(x_i | u_{1i}(l_1), u_{2i}(l_2))$
- Encoding:** To send message pair (m_1, m_2) , transmit $x^n(m_1, m_2)$
- Decoding:** Let $\epsilon > \epsilon'$. Decoder 1 declares that \hat{m}_1 is sent if it is the unique index such that $(u_1^n(l_1), y_1^n) \in \mathcal{T}_{\epsilon'}^{(n)}$ for some $u_1^n(l_1) \in C_1(\hat{m}_1)$; otherwise an error is declared. Similarly, decoder 2 finds the unique index \hat{m}_2 such that $(u_2^n(l_2), y_2^n) \in \mathcal{T}_{\epsilon'}^{(n)}$ for some $u_2^n(l_2) \in C_2(\hat{m}_2)$; otherwise an error is declared.
- A crucial requirement for this coding scheme to work is that we have at least one sequence pair $(u_1^n(l_1), u_2^n(l_2)) \in \mathcal{C}(m_1, m_2)$. The constraint on the subcode sizes to guarantee this is provided by the following lemma

Mutual Covering Lemma [25]

Let $(U_1, U_2) \sim p(u_1, u_2)$, $\epsilon > 0$. Let $U_1^n(m_1)$, $m_1 \in [1 : 2^{nR_1}]$, be pairwise independent random sequences, each distributed according to $\prod_{i=1}^n p_{U_1}(u_{1i})$. Similarly, let $U_2^n(m_2)$, $m_2 \in [1 : 2^{nR_2}]$, be pairwise independent random sequences, each distributed according to $\prod_{i=1}^n p_{U_2}(u_{2i})$. Assume $\{U_1^n(m_1) : m_1 \in [1 : 2^{nR_1}]\}$ and $\{U_2^n(m_2) : m_2 \in [1 : 2^{nR_2}]\}$ are independent. There exists $\delta(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$ such that if $R_1 + R_2 > I(U_1; U_2) + \delta(\epsilon)$, then $P\{(U_1^n(m_1), U_2^n(m_2)) \notin \mathcal{T}_\epsilon^{(n)} \text{ for all } m_1 \in [1 : 2^{nR_1}], m_2 \in [1 : 2^{nR_2}]\} \rightarrow 0$

- The proof of the lemma uses Chebychev inequality (see Appendix V)
- This lemma extends the covering lemma in two ways:
- By considering a single U_1^n sequence ($R_2 = 0$), we obtain the same rate requirement $R_2 > I(U_1; U_2) + \delta(\epsilon)$ as for the covering lemma
- We assumed only pairwise independence among the $U_2^n(m_2)$ sequences, which strengthens the covering lemma

Analysis of the Probability of Error

- Assume that the message pair $(1, 1)$ is sent. Let (L_1, L_2) be the index pair of the chosen sequence pair in $\mathcal{C}(1, 1)$
- Consider the average probability of error for decoder 1. Define the error events

$$\begin{aligned}\mathcal{E}_0 &:= \{|\mathcal{C}(1, 1)| = 0\}, \\ \mathcal{E}_{11} &:= \{(U_1^n(L_1), Y_1^n) \notin \mathcal{T}_\epsilon^{(n)}\}, \\ \mathcal{E}_{12} &:= \{(U_1^n(l_1), Y_1^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } l_1 \notin [1 : 2^{n(\tilde{R}_1 - R_1)}]\}\end{aligned}$$

Then the probability of error for decoder 1 is bounded by

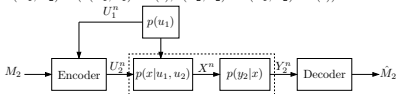
$$P(\mathcal{E}_1) \leq P(\mathcal{E}_0) + P(\mathcal{E}_0^c \cap \mathcal{E}_{11}) + P(\mathcal{E}_{12})$$

- To bound $P(\mathcal{E}_0)$, note that subcode $\mathcal{C}_1(1)$ corresponds to $2^{n(\tilde{R}_1 - R_1)}$ $U_1^n(l_1)$ sequences and subcode $\mathcal{C}_2(1)$ corresponds to $2^{n(\tilde{R}_2 - R_2)}$ $U_2^n(l_2)$ sequences. Hence, by the mutual covering lemma, $P\{|\mathcal{C}(1, 1)| = 0\} \rightarrow 0$ as $n \rightarrow \infty$ if $(\tilde{R}_1 - R_1) + (\tilde{R}_2 - R_2) > I(U_1; U_2) + \delta(\epsilon)$

- Given that $(U_1^n(L_1), U_2^n(L_2)) \in \mathcal{T}'_\epsilon^{(n)}$ and $\epsilon > \epsilon'$, $P\{U_1^n(L_1), U_2^n(L_2), X^n, Y_1^n \notin \mathcal{T}_\epsilon^{(n)}\} \rightarrow 0$ as $n \rightarrow \infty$ by the LLN. Hence, $P(\mathcal{E}_0^c \cap \mathcal{E}_{11}) \rightarrow 0$ as $n \rightarrow \infty$
- Since Y_1^n is independent of every $U_1^n(l_1) \notin \mathcal{C}(1)$, by the packing lemma $P(\mathcal{E}_{12}) \rightarrow 0$ as $n \rightarrow \infty$ if $\tilde{R}_1 < I(U_1; Y_1) - \delta(\epsilon)$
- Similarly, the average probability of error for decoder 2 $\rightarrow 0$ as $n \rightarrow \infty$ if $\tilde{R}_2 < I(U_2; Y_2) - \delta(\epsilon)$ and $(\tilde{R}_1 - R_1) + (\tilde{R}_2 - R_2) > I(U_1; U_2) + \delta(\epsilon)$
- Eliminating \tilde{R}_1, \tilde{R}_2 gives the inequalities in the theorem
- Remark: Simultaneous encoding is "dual" to simultaneous decoding. Alternatively, we can achieve each corner point of the region using covering and then use time sharing to achieve the rest of the region (as we did in successive cancellation decoding). Simultaneous encoding, however, is more powerful because it can achieve every point in the region

Relationship to Gelfand–Pinsker

- Fix $p(u_1, u_2)p(x|u_1, u_2)$ and consider the achievable rate pair $(R_1, R_2) = (I(U_1; Y_1) - \delta(\epsilon), I(U_2; Y_2) - I(U_1; U_2) - \delta(\epsilon))$



- So the coding scheme for sending M_2 to Y_2 is identical to that of sending M_2 over a channel $p(y_2|u_2, u_1) = \sum_x p(u_1|u_2)p(x|u_1, u_2)p(y_2|x)$ with "state" U_1 available noncausally at the encoder
- Since we do not know if the Marton inner bound is optimal in general, this relationship may or may not be fundamental
- This relationship, however, turned out to be crucial in establishing the capacity of MIMO Gaussian BC

Multivariate Mutual Covering Lemma

- Mutual covering lemma can be generalized to more than two variables
- Let $(U_1, U_2, U_3) \sim p(u_1, u_2, u_3)$, $\epsilon > 0$. For each $j = 1, 2, 3$, let $U_j^n(m_j), m_j \in [1 : 2^{nR_j}]$, be pairwise independent random sequences, each distributed according to $\prod_{i=1}^n p_{U_j}(u_{ji})$. Assume that $\{U_1^n(m_1) : m_1 \in [1 : 2^{nR_1}]\}$, $\{U_2^n(m_2) : m_2 \in [1 : 2^{nR_2}]\}$, and $\{U_3^n(m_3) : m_3 \in [1 : 2^{nR_3}]\}$ are independent. There exists $\delta(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$ such that if

$$R_1 + R_2 > I(U_1; U_2) + \delta(\epsilon),$$

$$R_1 + R_3 > I(U_1; U_3) + \delta(\epsilon),$$

$$R_2 + R_3 > I(U_2; U_3) + \delta(\epsilon),$$

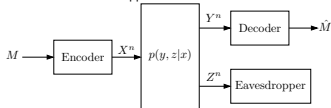
$$R_1 + R_2 + R_3 > I(U_1; U_2) + I(U_1, U_2; U_3) + \delta(\epsilon), \text{ then}$$

$$P\{(\{U_1^n(m_1), U_2^n(m_2), U_3^n(m_3)\} \notin \mathcal{T}_\epsilon^{(n)} \text{ for all } (m_1, m_2, m_3))\} \rightarrow 0 \text{ as } n \rightarrow \infty$$

- This extension is used to extend the Marton region to > 2 receivers and in the proof of achievability for multiple description coding [26]

DM Wiretap Channel

- Consider a DM-WTC, which is a 2-receiver DM-BC $(\mathcal{X}, p(y, z|x), \mathcal{Y} \times \mathcal{Z})$
- We wish to send a message M to the legitimate receiver Y and keep it secret from the eavesdropper Z



- A $(2^{nR}, n)$ secrecy code for the DM-WTC consists of:
 - A message set $[1 : 2^{nR}]$
 - An encoder that generates a codeword $X^n(m)$, $m \in [1 : 2^{nR}]$, according to $p(x^n|m)$, i.e., a random assignment that depends on m
 - A decoder that assigns a message estimate $\hat{m}(y^n)$ to each received sequence $y^n \in \mathcal{Y}^n$

- Assume that $M \sim \text{Unif}[1 : 2^{nR}]$
- The average probability of error for the secrecy code is defined as

$$P_e^{(n)} = \mathbb{P}\{M \neq \hat{M}\}$$

- The *information leakage rate* is defined as $E^{(n)} := (1/n)I(M; Z^n)$
- A rate R is achievable if there exists a sequence of $(2^{nR}, n)$ codes with $P_e^{(n)} \rightarrow 0$ and $E^{(n)} \rightarrow 0$ as $n \rightarrow \infty$
- The *secrecy capacity* C_S of the DM-WTC is the supremum of all achievable rates

Theorem (Wyner [27], Csiszár-Körner [28])

The secrecy capacity of the DM-WTC $(\mathcal{X}, p(y, z|x), \mathcal{Y} \times \mathcal{Z})$ is

$$C_S = \max_{p(u)p(x|u)} (I(U; Y) - I(U; Z))$$

- The wiretap channel was first introduced by Wyner [27], who assumed that Z is a degraded version of Y , i.e., $p(y, z|x) = p(y|x)p(z|y)$. Under this assumption,

$$\begin{aligned} I(U; Y) - I(U; Z) &= I(U; Y|Z) \\ &\leq I(X; Y|Z) = I(X; Y) - I(X; Z) \end{aligned}$$

Thus, the secrecy capacity is

$$C_S = \max_{p(x)} (I(X; Y) - I(X; Z))$$

- The above result also holds if Y is *more capable* than Z , i.e., $I(X; Y) \geq I(X; Z)$ for all $p(x)$

Proof of Achievability

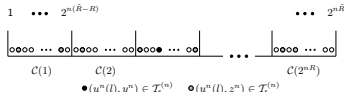
- Codebook generation:** Assume $C_S > 0$ and fix $p(u)p(x|u)$ that achieves it. Thus $I(U; Y) - I(U; Z) > 0$. For each message $m \in [1 : 2^{nR}]$, generate a subcode $\mathcal{C}(m)$ consisting of $2^{n(\hat{R}-R)}$ randomly and independently generated $u^n(l)$, $l \in [(m-1)2^{n(\hat{R}-R)} + 1 : m2^{n(\hat{R}-R)}]$, sequences each according to $\prod_{i=1}^n p(u_i)$. The chosen set of subcodes are revealed to the encoder, decoder, and eavesdropper
- Encoding:** To send message $m \in [1 : 2^{nR}]$, the encoder picks an index $l \in [(m-1)2^{n(\hat{R}-R)} + 1 : m2^{n(\hat{R}-R)}]$ uniformly at random. It then generates $X^n \sim \prod_{i=1}^n p_{X|U}(x_i|u_i(l))$ and transmits it

- Decoding and analysis of the probability of error: The decoder declares that \hat{l} is sent if it is the unique index such that $(u^n(\hat{l}), y^n) \in \mathcal{T}_\epsilon^{(n)}$ and declares that the message sent is the subcode index \hat{m} of $u^n(\hat{l})$

By the LLN and the packing lemma, this decoding procedure succeeds with average probability of error $P(\mathcal{E}) \rightarrow 0$ as $n \rightarrow \infty$ if $\tilde{R} < I(U; Y) - \delta(\epsilon)$

Analysis of the Information Leakage Rate

- Outline: For each subcode $\mathcal{C}(m)$, the eavesdropper (on average) has $\doteq 2^{n(\tilde{R}-R-I(U;Z))}$ $u^n(l)$ sequences that are jointly typical with z^n
- Thus if we take $\tilde{R} - R > I(U; Z)$, the eavesdropper would have roughly equal number of indices (in the exponent) in each subcode, providing it with almost no information about the message sent



- For each message m , let $L(m)$ be the randomly selected index by the encoder. Every codebook \mathcal{C} induces a joint pmf on (M, L, U^n, Z^n) of the form $p(m, l, u^n, z^n | \mathcal{C}) = 2^{-nR} 2^{-n(\tilde{R}-R)} p(u^n | l, \mathcal{C}) \prod_{i=1}^n p_{Z|U}(z_i | u_i)$

- Now consider the eavesdropper message rate averaged over the randomly chosen codebook \mathcal{C}

$$\begin{aligned} I(M; Z^n | \mathcal{C}) &= I(M, L; Z^n | \mathcal{C}) - I(L; Z^n | M, \mathcal{C}) \\ &= I(L; Z^n | \mathcal{C}) - H(L | M, \mathcal{C}) + H(L | Z^n, M, \mathcal{C}) \\ &= I(L; Z^n | \mathcal{C}) - n(\tilde{R} - R) + H(L | Z^n, M, \mathcal{C}) \\ &\leq I(U^n; Z^n | \mathcal{C}) - n(\tilde{R} - R) + H(L | Z^n, M, \mathcal{C}) \\ &= H(Z^n | \mathcal{C}) - \sum_{i=1}^n H(Z_i | U^i, Z^{i-1}, \mathcal{C}) - n(\tilde{R} - R) \\ &\quad + H(L | Z^n, M, \mathcal{C}) \\ &\leq \sum_{i=1}^n H(Z_i | \mathcal{C}) - \sum_{i=1}^n H(Z_i | U_i, \mathcal{C}) - n(\tilde{R} - R) \\ &\quad + H(L | Z^n, M, \mathcal{C}) \\ &\leq nH(Z; \mathcal{C}) - nH(Z | U) - n(\tilde{R} - R) + H(L | Z^n, M, \mathcal{C}) \\ &= nI(U; Z) - n(\tilde{R} - R) + H(L | Z^n, M, \mathcal{C}) \end{aligned}$$

- The equivocation term in the last step is bounded as follows

Equivocation Bound Lemma

If $R < I(U; Y) - I(U; Z) - 4\delta(\epsilon)$, then

$$\lim_{n \rightarrow \infty} H(L | Z^n, M, \mathcal{C}) / n \leq \tilde{R} - R - I(U; Z) + \delta(\epsilon)$$

The proof of the lemma uses Chebychev inequality (see Appendix VI)

- Substituting, we have shown that if $R < I(U; Y) - I(U; Z) - 4\delta(\epsilon)$, then $\lim_{n \rightarrow \infty} I(M; Z^n | \mathcal{C}) / n \leq \delta(\epsilon)$
- From the above, it follows that there must exist a sequence of codes with $P_e^{(n)} \rightarrow 0$ and $E^{(n)} \rightarrow 0$, which completes the proof of achievability
- Remark: We can use subcode generation instead of randomly generating x^n

For each l , randomly and conditionally independently generate $> 2^{nI(X; Z|U)}$ x^n sequences each according to $\prod_{i=1}^n p_{X|U}(x_i | u_i(l))$. The encoder randomly selects one of them for transmission

Summary of Achievability Ideas

- Coding techniques:
 - ▶ Random codebook generation
 - ▶ Joint typicality decoding (packing)
 - ▶ Successive cancellation decoding
 - ▶ Simultaneous joint typicality decoding
 - ▶ Time sharing and coded time sharing
 - ▶ Superposition coding
 - ▶ Rate splitting
 - ▶ Indirect decoding
 - ▶ Block Markov coding
 - ▶ Backward decoding
 - ▶ Joint typicality encoding (covering)
 - ▶ Binning
 - ▶ Subcode generation: One-to-many coding; dual to binning
 - ▶ Simultaneous joint typicality encoding (mutual covering): generalization of covering, dual to simultaneous decoding
 - ▶ Random encoding (for secrecy)

- Performance analysis tools:
 - ▶ Division of error event
 - ▶ LLN
 - ▶ Properties of typicality
 - ▶ Packing lemma
 - ▶ Fourier–Motzkin elimination
 - ▶ Covering lemma
 - ▶ Mutual covering lemma
 - ▶ Chebychev inequality
 - ▶ Entropy and mutual information bounds
- There are other coding techniques that we didn't discuss, e.g., linear coding, network coding, iterative refinement for channels with feedback, and corresponding analysis tools

Conclusion

- Presented a simple and unified approach to achievability for DM systems
 - ▶ Strong typicality
 - ▶ Small number of coding techniques and performance analysis tools
- We observed some interesting "dualities"
 - Covering \leftrightarrow Packing
 - Binning \leftrightarrow Subcode generation
 - Simultaneous covering \leftrightarrow Simultaneous packing
- Proofs can be extended to Gaussian sources/channels via appropriate scalar quantizations and standard limit theorems
- As we have seen, the same region can often be achieved using different coding schemes, and there can be several equivalent characterizations of the same region. It would be interesting to develop a deeper understanding of these intriguing observations

- [1] T. C. Hu.
Multi-commodity network flows.
Operations Research, 11(3):344–360, 1963.
- [2] L. R. Ford, Jr. and D. R. Fulkerson.
Maximal flow through a network.
Canad. J. Math., 8:399–404, 1956.
- [3] Peter Elias, Amiel Feinstein, and Claude E. Shannon.
A note on the maximum flow through a network.
IRE Trans. Inf. Theory, 2(4):117–119, December 1956.
- [4] Alon Orlitsky and James R. Roche.
Coding for computing.
IEEE Trans. Inf. Theory, 47(3):903–917, 2001.
- [5] Toby Berger.
Multiterminal source coding.
In Giuseppe Longo, editor, *The Information Theory Approach to Communications*. Springer-Verlag, New York, 1978.
- [6] Claude E. Shannon.

- A mathematical theory of communication.
Bell System Tech. J., 27:379–423, 623–656, 1948.
- [7] G. David Forney, Jr.
Information theory.
Unpublished course notes, Stanford University, 1972.
- [8] Thomas M. Cover and Joy A. Thomas.
Elements of Information Theory.
Wiley, New York, second edition, 2006.
- [9] Rudolf Ahlswede.
Multiway communication channels.
In Proc. 2nd Int. Symp. Inf. Theory, pages 23–52, Tsahkadsor, Armenian S.S.R., 1971.
- [10] Henry H. J. Liao.
Multiple Access Channels.
Ph.d. thesis, University of Hawaii, Honolulu, September 1972.
- [11] János Körner and Katalin Marton.
General broadcast channels with degraded message sets.
IEEE Trans. Inf. Theory, 23(1):60–64, 1977.

- [12] Günter M. Ziegler.
Lectures on Polytopes.
Springer-Verlag, New York, 1995.
- [13] Shashi Borade, Lizhong Zheng, and Mitchell Trott.
Multilevel broadcast networks.
In Proc. IEEE International Symposium on Information Theory, pages 1151–1155, Nice, France, June 2007.
- [14] Chandra Nair and Abbas El Gamal.
The capacity region of a class of 3-receiver broadcast channels with degraded message sets.
2008.
- [15] Hon-Fah Chong, M. Motani, H. K. Garg, and H. El Gamal.
On the Han–Kobayashi region for the interference channel.
IEEE Trans. Inf. Theory, 54(7):3188–3195, July 2008.
- [16] Te Sun Han and Kingo Kobayashi.
A new achievable rate region for the interference channel.
IEEE Trans. Inf. Theory, 27(1):49–60, 1981.
- [17] Claude E. Shannon.

- Coding theorems for a discrete source with a fidelity criterion.
In IRE Int. Conv. Rec., part 4, volume 7, pages 142–163, 1959.
Reprinted with changes in *Information and Decision Processes*, R. E. Machol, Ed. New York: McGraw-Hill, 1960, pp. 93–126.
- [18] Aaron D. Wyner and Jacob Ziv.
The rate-distortion function for source coding with side information at the decoder.
IEEE Trans. Inf. Theory, 22(1):1–10, 1976.
- [19] David Slepian and Jack Keil Wolf.
A coding theorem for multiple access channels with correlated sources.
Bell System Tech. J., 52:1037–1076, 1973.
- [20] S. I. Gelfand and M. S. Pinsker.
Coding for channel with random parameters.
Probl. Control Inf. Theory, 9(1):19–31, 1980.
- [21] Chris Heegard and Abbas El Gamal.
On the capacity of computer memories with defects.
IEEE Trans. Inf. Theory, 29(5):731–739, 1983.
- [22] Thomas M. Cover.

- An achievable rate region for the broadcast channel.
IEEE Trans. Inf. Theory, 21:399–404, 1975.
- [23] Edward C. van der Meulen.
Random coding theorems for the general discrete memoryless broadcast channel.
IEEE Trans. Inf. Theory, 21:180–190, 1975.
- [24] Katalin Marton.
A coding theorem for the discrete memoryless broadcast channel.
IEEE Trans. Inf. Theory, 25(3):306–311, 1979.
- [25] Abbas El Gamal and Edward C. van der Meulen.
A proof of Marton’s coding theorem for the discrete memoryless broadcast channel.
IEEE Trans. Inf. Theory, 27(1):120–122, January 1981.
- [26] Abbas El Gamal and Thomas M. Cover.
Achievable rates for multiple descriptions.
IEEE Trans. Inf. Theory, 28(6):851–857, 1982.
- [27] A. D. Wyner.
The wire-tap channel.
Bell System Tech. J., 54(8):1355–1387, 1975.

- [28] Imre Csiszár and János Körner.

Broadcast channels with confidential messages.
IEEE Trans. Inf. Theory, 24(3):339–348, 1978.

- [29] Thomas M. Cover and Abbas El Gamal.

Capacity theorems for the relay channel.
IEEE Trans. Inf. Theory, 25(5):572–584, September 1979.

- [30] Frans M. J. Willems and Edward C. van der Meulen.

The discrete memoryless multiple-access channel with cribbing encoders.
IEEE Trans. Inf. Theory, 31(3):313–327, 1985.

- [31] Liang-Liang Xie and P. R. Kumar.

An achievable rate for the multiple-level relay channel.
IEEE Trans. Inf. Theory, 51(4):1348–1358, 2005.

- [32] Abbas El Gamal, Mehdi Mohseni, and Sina Zahedi.

Bounds on capacity and minimum energy-per-bit for AWGN relay channels.
IEEE Trans. Inf. Theory, 52(4):1545–1561, 2006.

Appendix I: Proof of the Packing lemma

- Define the events $\mathcal{E}_m := \{(U^n, X^n(m), Y^n) \in \mathcal{T}_\epsilon^{(n)}\}$, $m \in \mathcal{A}$
- The probability of the event of interest can be bounded as

$$P\left(\bigcup_{m \in \mathcal{A}} \mathcal{E}_m\right) \leq \sum_{m \in \mathcal{A}} P(\mathcal{E}_m)$$

Now, consider

$$\begin{aligned} P(\mathcal{E}_m) &= P\{(U^n, X^n(m), Y^n) \in \mathcal{T}_\epsilon^{(n)}(U, X, Y)\} \\ &= \sum_{(u^n, y^n) \in \mathcal{T}_\epsilon^{(n)}} p(u^n, y^n) P\{(u^n, X^n(m), y^n) \in \mathcal{T}_\epsilon^{(n)}(U, X, Y) | U^n = u^n\} \\ &\stackrel{(a)}{\leq} \sum_{(u^n, y^n) \in \mathcal{T}_\epsilon^{(n)}} p(u^n, y^n) 2^{-n(I(X; Y|U) - \delta(\epsilon))} \leq 2^{-n(I(X; Y|U) - \delta(\epsilon))}, \end{aligned}$$

where (a) follows by the conditional joint typicality lemma, since $(u^n, y^n) \in \mathcal{T}_\epsilon^{(n)}$, $X^n(m) \sim \prod_{i=1}^n p_{X|U}(x_i | u_i)$, and $\delta(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$. Hence

$$\sum_{m \in \mathcal{A}} P(\mathcal{E}_m) \leq |\mathcal{A}| 2^{-n(I(X; Y|U) - \delta(\epsilon))} \leq 2^{-n(I(X; Y|U) - R - \delta(\epsilon))},$$

which tends to 0 as $n \rightarrow \infty$ if $R < I(X; Y|U) - \delta(\epsilon)$

Appendix II: Fourier–Motzkin for Han–Kobayashi Bound

- Substituting $R_{3j} = R_j - R_{j0}$ for $j = 1, 2$, we obtain the system of inequalities:

$$\begin{aligned} R_1 - R_{10} &\leq I_1, \\ R_1 &\leq I_2, \\ R_1 - R_{10} + R_{20} &\leq I_3, \\ R_1 + R_{20} &\leq I_4, \\ R_2 - R_{20} &\leq I_5, \\ R_2 &\leq I_6, \\ R_2 - R_{20} + R_{10} &\leq I_7, \\ R_2 + R_{10} &\leq I_8 \end{aligned}$$

- Next we eliminate R_{j0} for $j = 1, 2$

- Elimination of R_{10} :

We have two upper bounds on R_{10} :

$$\begin{aligned} R_{10} &\leq I_7 - R_2 + R_{20}, \\ R_{10} &\leq I_8 - R_2 \end{aligned}$$

- and two lower bounds on R_{10} :

$$\begin{aligned} R_{10} &\geq R_1 - I_1, \\ R_{10} &\geq R_1 + R_{20} - I_3 \end{aligned}$$

Comparing upper and lower bounds, and copying four inequalities in the original system that do not involve R_{10} , we obtain a new system of inequalities in (R_{20}, R_1, R_2) given by

$$\begin{aligned} R_1 &\leq I_2, \\ R_1 + R_{20} &\leq I_4, \\ R_2 - R_{20} &\leq I_5, \\ R_2 &\leq I_6, \\ R_1 + R_2 - R_{20} &\leq I_1 + I_7, \\ R_1 + R_2 &\leq I_1 + I_8, \\ R_1 + R_2 &\leq I_3 + I_7, \\ R_1 + R_2 + R_{20} &\leq I_3 + I_8 \end{aligned}$$

Noting that none of these 8 inequalities is redundant, we move on to eliminate R_{20} from them

- Elimination of R_{20} :

Comparing upper bounds on R_{20}

$$R_{20} \leq I_4 - R_1, \quad R_{20} \leq I_3 + I_8 - R_1 - R_2$$

with lower bounds

$$R_{20} \geq R_2 - I_5, \quad R_{20} \geq R_1 + R_2 - I_1 - I_7,$$

and copying inequalities that do not involve R_{20} , we obtain

$$\begin{aligned} R_1 &\leq I_2, \\ R_2 &\leq I_6, \\ R_1 + R_2 &\leq I_1 + I_8, \\ R_1 + R_2 &\leq I_3 + I_7, \\ R_1 + R_2 &\leq I_4 + I_5, \\ 2R_1 + R_2 &\leq I_1 + I_4 + I_7, \\ R_1 + 2R_2 &\leq I_3 + I_5 + I_8, \\ 2R_1 + 2R_2 &\leq I_1 + I_3 + I_7 + I_8 \end{aligned}$$

- Finally we note that the last inequality is redundant since it is implied by the third and fourth inequalities; thus we have 7 inequalities on (R_1, R_2)

Appendix III: Proof of the Covering Lemma

- From the joint typicality lemma,

$$\mathbb{P}\{(u^n, x^n, \hat{X}^n(m)) \in \mathcal{T}_\epsilon^{(n)}\} \geq 2^{-n(I(X; \hat{X}|U) + \delta(\epsilon))}$$

for each $m \in \mathcal{A}$, where $\delta(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$

- Hence, the probability of the event of interest is bounded by

$$\begin{aligned} \mathbb{P}\{(u^n, x^n, \hat{X}^n(m)) \notin \mathcal{T}_\epsilon^{(n)} \text{ for all } m\} &= \prod_{m \in \mathcal{A}} \mathbb{P}\{(u^n, x^n, \hat{X}^n(m)) \notin \mathcal{T}_\epsilon^{(n)}\} \\ &\leq \left[1 - 2^{-n(I(X; \hat{X}|U) + \delta(\epsilon))}\right]^{|\mathcal{A}|} \\ &\leq e^{-\left(|\mathcal{A}| \cdot 2^{-n(I(X; \hat{X}|U) + \delta(\epsilon))}\right)} \\ &\leq e^{-\left(2^n(n - I(X; \hat{X}|U) - \delta(\epsilon))\right)}, \end{aligned}$$

which goes to zero as $n \rightarrow \infty$, provided that $R > I(X; \hat{X}|U) + \delta(\epsilon)$

Taking expectation over (U^n, X^n) completes the proof

Appendix IV: Bounding $\mathbb{P}(\mathcal{E}_3)$ in Wyner–Ziv Achievability

- First we show that

$$\begin{aligned} \mathbb{P}\{(U^n(\tilde{l}), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } \tilde{l} \in \mathcal{B}(m), \tilde{l} \neq l | M = m\} \\ \leq \mathbb{P}\{(U^n(\tilde{l}), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } \tilde{l} \in \mathcal{B}(1) | M = m\} \end{aligned}$$

- The claim holds trivially for $m = 1$

- For $m \neq 1$, consider

$$\begin{aligned} \mathbb{P}\{(U^n(\tilde{l}), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } \tilde{l} \in \mathcal{B}(m), \tilde{l} \neq l | M = m\} \\ = \sum_{\tilde{l} \in \mathcal{B}(m)} p(l|m) \mathbb{P}\{(U^n(\tilde{l}), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } \tilde{l} \in \mathcal{B}(m), \tilde{l} \neq l | L = l, M = m\} \\ \stackrel{(a)}{=} \sum_{\tilde{l} \in \mathcal{B}(m)} p(l|m) \mathbb{P}\{(U^n(\tilde{l}), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } \tilde{l} \in \mathcal{B}(m), \tilde{l} \neq l | L = l\} \\ \stackrel{(b)}{=} \sum_{\tilde{l} \in \mathcal{B}(m)} p(l|m) \mathbb{P}\{(U^n(\tilde{l}), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } \tilde{l} \in [1 : 2^{n(\tilde{R}-R)} - 1] | L = l\} \\ \leq \sum_{\tilde{l} \in \mathcal{B}(m)} p(l|m) \mathbb{P}\{(U^n(\tilde{l}), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } \tilde{l} \in \mathcal{B}(1) | L = l\} \\ \stackrel{(c)}{=} \sum_{\tilde{l} \in \mathcal{B}(m)} p(l|m) \mathbb{P}\{(U^n(\tilde{l}), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } \tilde{l} \in \mathcal{B}(1) | L = l, M = m\} \end{aligned}$$

$$= \mathbb{P}\{(U^n(\tilde{l}), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } \tilde{l} \in \mathcal{B}(1) | M = m\}$$

where (a) and (c) follow since M is a function of L and (b) follows since by the code generation, given $L = l$, any collection of $2^{n(\tilde{R}-R)} - 1$ $U^n(\tilde{l})$ sequences with $\tilde{l} \neq l$ has the same distribution

- Hence

$$\begin{aligned} \mathbb{P}(\mathcal{E}_3) &= \sum_m p(m) \mathbb{P}\{(U^n(\tilde{l}), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } U^n(\tilde{l}) \in \mathcal{B}(m), \tilde{l} \neq l | M = m\} \\ &\leq \sum_m p(m) \mathbb{P}\{(U^n(\tilde{l}), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } U^n(\tilde{l}) \in \mathcal{B}(1) | M = m\} \\ &= \mathbb{P}\{(U^n(\tilde{l}), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } U^n(\tilde{l}) \in \mathcal{B}(1)\} \end{aligned}$$

Appendix V: Proof of Mutual Covering Lemma

• Let

$\mathcal{B} = \{(m_1, m_2) \in [1 : 2^{nR_1}] \times [1 : 2^{nR_2}] : (U_1^n(m_1), U_2^n(m_2)) \in \mathcal{T}_\epsilon^{(n)}(U_1, U_2)\}$.
Then the probability of the event of interest can be bounded as

$$\mathbb{P}\{|\mathcal{B}| = 0\} \leq \mathbb{P}\{(|\mathcal{B}| - \mathbb{E}\{|\mathcal{B}|\})^2 \geq (\mathbb{E}\{|\mathcal{B}|\})^2\} \leq \frac{\text{Var}\{|\mathcal{B}|\}}{(\mathbb{E}\{|\mathcal{B}|\})^2}$$

by Chebychev's inequality

• Using indicator random variables, we can express $|\mathcal{B}|$ as

$$|\mathcal{B}| = \sum_{m_1=1}^{2^{nR_1}} \sum_{m_2=1}^{2^{nR_2}} E(m_1, m_2),$$

where

$$E(m_1, m_2) := \begin{cases} 1 & \text{if } (U_1^n(m_1), U_2^n(m_2)) \in \mathcal{T}_\epsilon^{(n)}, \\ 0 & \text{otherwise} \end{cases}$$

for each $(m_1, m_2) \in [1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$

• Let

$$p_1 := \mathbb{P}\{(U_1^n(1), U_2^n(1)) \in \mathcal{T}_\epsilon^{(n)}\},$$

$$p_2 := \mathbb{P}\{(U_1^n(1), U_2^n(1)) \in \mathcal{T}_\epsilon^{(n)}, (U_1^n(1), U_2^n(2)) \in \mathcal{T}_\epsilon^{(n)}\},$$

$$p_3 := \mathbb{P}\{(U_1^n(1), U_2^n(1)) \in \mathcal{T}_\epsilon^{(n)}, (U_1^n(2), U_2^n(1)) \in \mathcal{T}_\epsilon^{(n)}\},$$

$$p_4 := \mathbb{P}\{(U_1^n(1), U_2^n(1)) \in \mathcal{T}_\epsilon^{(n)}, (U_1^n(2), U_2^n(2)) \in \mathcal{T}_\epsilon^{(n)}\} = p_1^2$$

Then

$$\mathbb{E}\{|\mathcal{B}|\} = \sum_{m_1, m_2} \mathbb{P}\{(U_1^n(m_1), U_2^n(m_2)) \in \mathcal{T}_\epsilon^{(n)}\} = 2^{n(R_1+R_2)} p_1$$

and

$$\begin{aligned} \mathbb{E}\{|\mathcal{B}|^2\} &= \sum_{m_1, m_2} \mathbb{P}\{(U_1^n(m_1), U_2^n(m_2)) \in \mathcal{T}_\epsilon^{(n)}\} \\ &+ \sum_{\substack{m_1, m_2, m_1' \neq m_2 \\ m_1, m_2, m_1' \neq m_1}} \mathbb{P}\{(U_1^n(m_1), U_2^n(m_2)) \in \mathcal{T}_\epsilon^{(n)}, (U_1^n(m_1), U_2^n(m_2')) \in \mathcal{T}_\epsilon^{(n)}\} \\ &+ \sum_{\substack{m_1, m_2, m_1' \neq m_1 \\ m_1, m_2, m_1' \neq m_2}} \mathbb{P}\{(U_1^n(m_1), U_2^n(m_2)) \in \mathcal{T}_\epsilon^{(n)}, (U_1^n(m_1'), U_2^n(m_2)) \in \mathcal{T}_\epsilon^{(n)}\} \\ &+ \sum_{\substack{m_1, m_2, m_1' \neq m_1, m_2' \neq m_2}} \mathbb{P}\{(U_1^n(m_1), U_2^n(m_2)) \in \mathcal{T}_\epsilon^{(n)}, (U_1^n(m_1'), U_2^n(m_2')) \in \mathcal{T}_\epsilon^{(n)}\} \\ &\leq 2^{n(R_1+R_2)} p_1 + 2^{n(R_1+2R_2)} p_2 + 2^{n(2R_1+R_2)} p_3 + 2^{n(R_1+R_2)} p_4 \end{aligned}$$

• Hence

$$\text{Var}\{|\mathcal{B}|\} \leq 2^{n(R_1+R_2)} p_1 + 2^{n(R_1+2R_2)} p_2 + 2^{n(2R_1+R_2)} p_3$$

• Now by the joint typicality lemma, we have

$$p_1 \geq 2^{-n(I(U_1; U_2) + \delta(\epsilon))},$$

$$p_2 \leq 2^{-2n(I(U_1; U_2) - \delta(\epsilon))},$$

$$p_3 \leq 2^{-2n(I(U_1; U_2) - \delta(\epsilon))},$$

hence

$$p_2/p_1^2 \leq 2^{4n\delta(\epsilon)}, \quad p_3/p_1^2 \leq 2^{4n\delta(\epsilon)}$$

• Therefore,

$$\frac{\text{Var}\{|\mathcal{B}|\}}{(\mathbb{E}\{|\mathcal{B}|\})^2} \leq 2^{-n(R_1+R_2 - I(U_1; U_2) - \delta(\epsilon))} + 2^{-n(R_1 - 4\delta(\epsilon))} + 2^{-n(R_2 - 4\delta(\epsilon))},$$

which $\rightarrow 0$ as $n \rightarrow \infty$, provided that $R_1 > 4\delta(\epsilon)$, $R_2 > 4\delta(\epsilon)$,
 $R_1 + R_2 > I(U_1; U_2) + \delta(\epsilon)$

• Similarly, $\mathbb{P}\{|\mathcal{B}| = 0\} \rightarrow 0$ as $n \rightarrow \infty$ if $R_1 = 0$ and $R_2 > I(U_1; U_2) + \delta(\epsilon)$, or if $R_1 > I(U_1; U_2) + \delta(\epsilon)$ and $R_2 = 0$ • Combining three sets of inequalities, we have shown that $\mathbb{P}\{|\mathcal{B}| = 0\} \rightarrow 0$ as $n \rightarrow \infty$ if $R_1 + R_2 > I(U_1; U_2) + 5\delta(\epsilon)$

Appendix VI: Proof of the Equivocation Bound Lemma

• We bound $H(L|Z^n, m, \mathcal{C})/n$ for every m

• We first bound the average probability for the following events:

For every m and codebook \mathcal{C} , let $N(m, \mathcal{C}) := \{l \in \mathcal{C}(m) : (U^n(l), Z^n) \in \mathcal{T}_\epsilon^{(n)}\}$.
It is not difficult to show that

$$\mathbb{E}\{N(m, \mathcal{C})\} \doteq 2^{n(\bar{R} - R - I(U; Z) + \delta(\epsilon))},$$

$$\text{Var}\{N(m, \mathcal{C})\} \leq 2^{n(\bar{R} - R - I(U; Z) + \delta(\epsilon))}$$

Define the random event $\mathcal{E}_2(m, \mathcal{C}) := \{N(m, \mathcal{C}) \geq 2\mathbb{E}\{N(m, \mathcal{C})\}\}$. Using the Chebyshev inequality, the probability of \mathcal{E}_2 averaged over codebooks is

$$\mathbb{P}\{\mathcal{E}_2(m, \mathcal{C})\} \leq \frac{\text{Var}\{N(m, \mathcal{C})\}}{(\mathbb{E}\{N(m, \mathcal{C})\})^2} \leq 2^{-n(\bar{R} - R - I(U; Z) - 3\delta(\epsilon))}$$

Thus if $\bar{R} - R - I(U; Z) - 3\delta(\epsilon) > 0$, i.e., $R < I(U; Y) - I(U; Z) - 4\delta(\epsilon)$, $\mathbb{P}\{\mathcal{E}_2(m, \mathcal{C})\} \rightarrow 0$ as $n \rightarrow \infty$ for every m

• For each code c and message m , define the indicator random variable $I(m, c) := 0$ if $(U^n(L), Z^n) \in \mathcal{T}_\epsilon^{(n)}$ and the event \mathcal{E}_2 occurs, and $I(m, c) := 1$, otherwise
Clearly, $\mathbb{E}\{I(m, \mathcal{C})\} = \mathbb{P}\{I(m, \mathcal{C}) = 1\} \leq \mathbb{P}\{(U^n(L), Z^n) \notin \mathcal{T}_\epsilon^{(n)}\} + \mathbb{P}\{\mathcal{E}_2(m, \mathcal{C})\}$
The first term $\rightarrow 0$ as $n \rightarrow \infty$ by the LLN and the second term $\rightarrow 0$ as $n \rightarrow \infty$ if $R < I(U; Y) - I(U; Z) - 4\delta(\epsilon)$

- We are now ready to bound the last eavesdropper message rate term. Consider

$$\begin{aligned}
 H(L|Z^n, m, C) &\leq 1 + \sum_C p(c) \mathbb{P}\{I(m, c) = 1 | C = c\} H(L|Z^n, m, I(m, c) = 1, c) \\
 &\quad + H(L|Z^n, m, I(m, C) = 0, C) \\
 &\leq 1 + n(\tilde{R} - R) \mathbb{P}\{I(m, C) = 1\} + H(L|Z^n, m, I(m, C) = 0, C) \\
 &\leq 1 + n(\tilde{R} - R) \mathbb{P}\{I(m, C) = 1\} + n(\tilde{R} - R - I(U; Z) + \delta(\epsilon))
 \end{aligned}$$

Now since $\mathbb{P}\{I(m, C) = 1\} \rightarrow 0$ as $n \rightarrow \infty$ if $R < I(U; Y) - I(U; Z) - 4\delta(\epsilon)$, then for every m ,

$$\lim_{n \rightarrow \infty} H(L|Z^n, m, C)/n \leq \tilde{R} - R - I(U; Z) + \delta(\epsilon)$$

- This completes the proof of the lemma