Bounds on Permutation Channel Capacity

Anuran Makur

Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology

IEEE International Symposium on Information Theory 2020

Outline

Three Motivations

- **Coding theory:** [DG01], [Mit06], [Met09], [KV15], [KT18], ...

Three Motivations

- **Coding theory:** [DG01], [Mit06], [Met09], [KV15], [KT18], ...
  - *Random deletion channel*: LDPC codes nearly achieve capacity for large alphabets
  - Codes correct for transpositions of symbols

Three Motivations

- **Coding theory:** [DG01], [Mit06], [Met09], [KV15], [KT18], ...
  - *Random deletion channel*: LDPC codes nearly achieve capacity for large alphabets
  - Codes correct for transpositions of symbols
  - Permutation channels with insertions, deletions, substitutions, or erasures
  - Construction and analysis of *multiset codes*

Three Motivations

- **Coding theory:** [DG01], [Mit06], [Met09], [KV15], [KT18], ...
  - *Random deletion channel*: LDPC codes nearly achieve capacity for large alphabets
  - Codes correct for transpositions of symbols
  - Permutation channels with insertions, deletions, substitutions, or erasures
  - Construction and analysis of *multiset codes*
- **Communication networks:** [XZ02], [WWM09], [GG10], [KV13], ...
  - Mobile ad hoc networks, multipath routed networks, etc.

Three Motivations

- **Coding theory:** [DG01], [Mit06], [Met09], [KV15], [KT18], . . .
  - *Random deletion channel*: LDPC codes nearly achieve capacity for large alphabets
  - Codes correct for transpositions of symbols
  - Permutation channels with insertions, deletions, substitutions, or erasures
  - Construction and analysis of *multiset codes*
- **Communication networks:** [XZ02], [WWM09], [GG10], [KV13], . . .
  - Mobile ad hoc networks, multipath routed networks, etc.
  - *Out-of-order delivery* of packets
  - Correct for packet errors/losses when packets *do not have sequence numbers*

Three Motivations

- **Coding theory:** [DG01], [Mit06], [Met09], [KV15], [KT18], ...
  - *Random deletion channel*: LDPC codes nearly achieve capacity for large alphabets
  - Codes correct for transpositions of symbols
  - Permutation channels with insertions, deletions, substitutions, or erasures
  - Construction and analysis of *multiset codes*
- **Communication networks:** [XZ02], [WWM09], [GG10], [KV13], ...
  - Mobile ad hoc networks, multipath routed networks, etc.
  - *Out-of-order delivery* of packets
  - Correct for packet errors/losses when packets *do not have sequence numbers*
- **Molecular/Biological Communications:** [YKGR$^+$15], [KPM16], [HSRT17], [SH19], ...

Three Motivations

- **Coding theory:** [DG01], [Mit06], [Met09], [KV15], [KT18], ...
  - *Random deletion channel*: LDPC codes nearly achieve capacity for large alphabets
  - Codes correct for transpositions of symbols
  - Permutation channels with insertions, deletions, substitutions, or erasures
  - Construction and analysis of *multiset codes*
- **Communication networks:** [XZ02], [WWM09], [GG10], [KV13], ...
  - Mobile ad hoc networks, multipath routed networks, etc.
  - *Out-of-order delivery* of packets
  - Correct for packet errors/losses when packets *do not have sequence numbers*
- **Molecular/Biological Communications:** [YKGR$^+$15], [KPM16], [HSRT17], [SH19], ...
  - *DNA based storage systems*
  - Source data encoded into DNA molecules

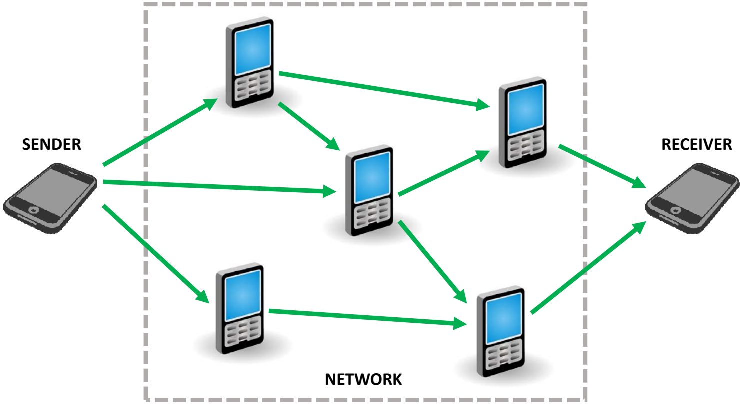
Three Motivations

- **Coding theory:** [DG01], [Mit06], [Met09], [KV15], [KT18], ...
  - *Random deletion channel*: LDPC codes nearly achieve capacity for large alphabets
  - Codes correct for transpositions of symbols
  - Permutation channels with insertions, deletions, substitutions, or erasures
  - Construction and analysis of *multiset codes*
- **Communication networks:** [XZ02], [WWM09], [GG10], [KV13], ...
  - Mobile ad hoc networks, multipath routed networks, etc.
  - *Out-of-order delivery* of packets
  - Correct for packet errors/losses when packets *do not have sequence numbers*
- **Molecular/Biological Communications:** [YKGR$^+$15], [KPM16], [HSRT17], [SH19], ...
  - *DNA based storage systems*
  - Source data encoded into DNA molecules
  - Fragments of DNA molecules cached
  - Receiver reads encoded data by *shotgun sequencing* (i.e., random sampling)

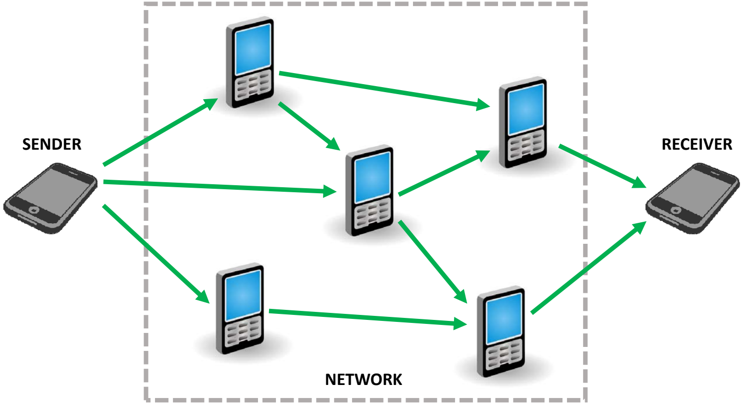
Three Motivations

- **Coding theory:** [DG01], [Mit06], [Met09], [KV15], [KT18], ...
  - *Random deletion channel*: LDPC codes nearly achieve capacity for large alphabets
  - Codes correct for transpositions of symbols
  - Permutation channels with insertions, deletions, substitutions, or erasures
  - Construction and analysis of *multiset codes*
- **Communication networks:** [XZ02], [WWM09], [GG10], [KV13], ...
  - Mobile ad hoc networks, multipath routed networks, etc.
  - *Out-of-order delivery* of packets
  - Correct for packet errors/losses when packets *do not have sequence numbers*
- **Molecular/Biological Communications:** [YKGR$^+$15], [KPM16], [HSRT17], [SH19], ...
  - *DNA based storage systems*
  - Source data encoded into DNA molecules
  - Fragments of DNA molecules cached
  - Receiver reads encoded data by *shotgun sequencing* (i.e., random sampling)

Motivation: Point-to-point Communication in Packet Networks

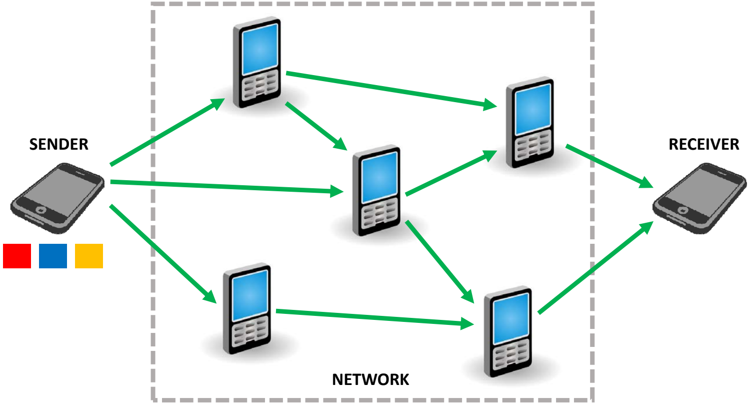


Motivation: Point-to-point Communication in Packet Networks



Model communication network as a channel

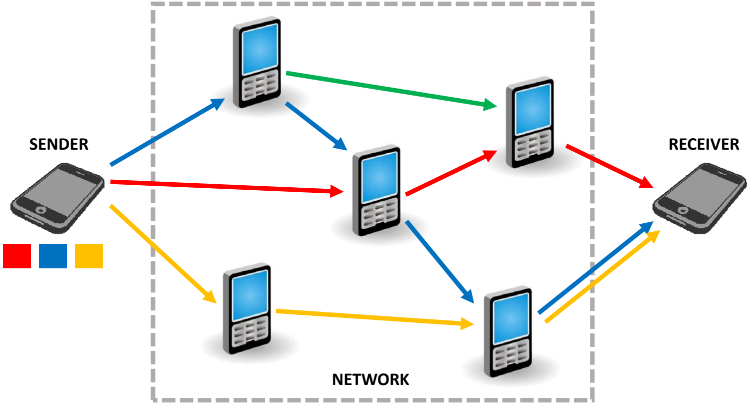
Motivation: Point-to-point Communication in Packet Networks



Model communication network as a channel:

- Alphabet symbols = all possible b-bit packets $\Rightarrow$ 2^b input symbols

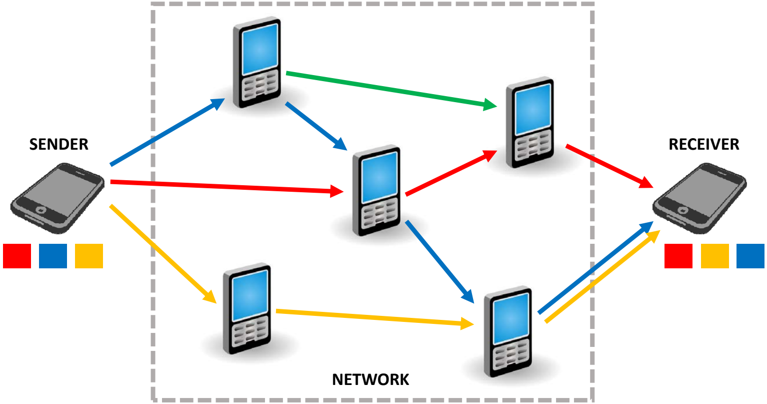
Motivation: Point-to-point Communication in Packet Networks



Model communication network as a channel:

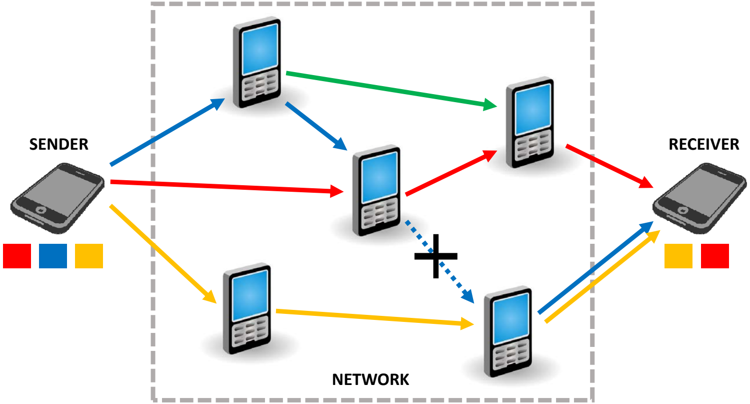
- Alphabet symbols = all possible b-bit packets
- Multipath routed network or evolving network topology

Motivation: Point-to-point Communication in Packet Networks



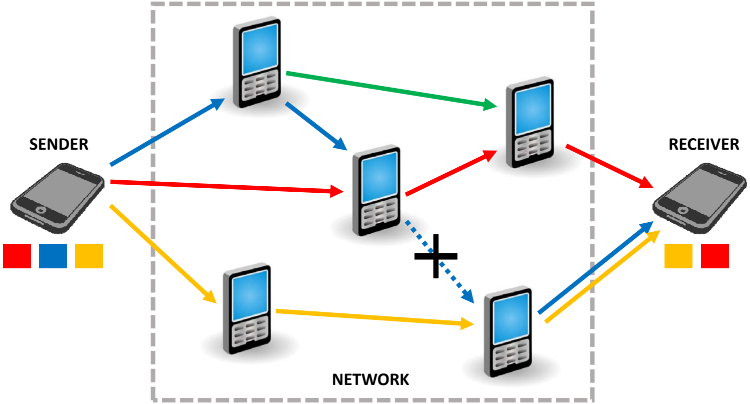
Model communication network as a channel:

- Alphabet symbols = all possible b-bit packets
- Multipath routed network $\Rightarrow$ packets received with transpositions



Model communication network as a channel:

- Alphabet symbols = all possible b-bit packets
- Multipath routed network $\Rightarrow$ packets received with transpositions
- Packets are impaired (e.g., deletions, substitutions, etc.)

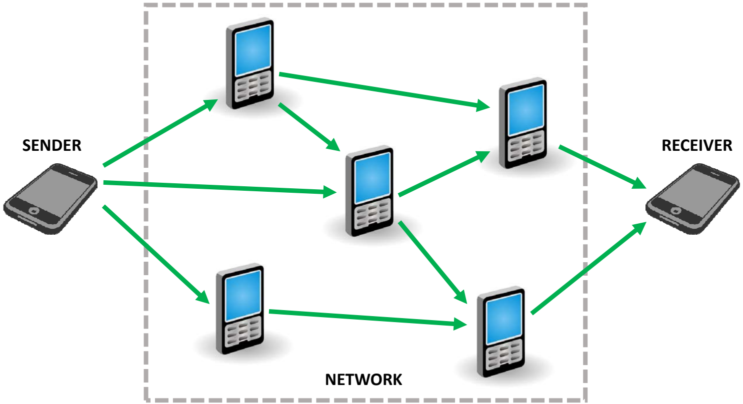


Model communication network as a channel:

- Alphabet symbols = all possible b-bit packets
- Multipath routed network $\Rightarrow$ packets received with transpositions
- Packets are impaired $\Rightarrow$ model using channel probabilities

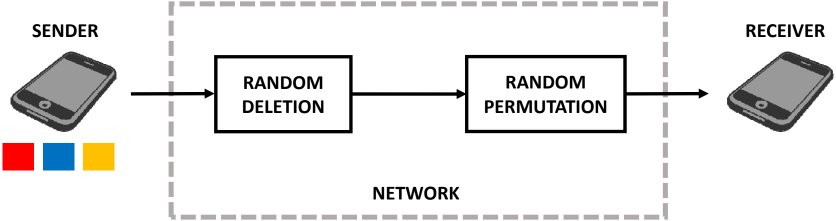
Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:



Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

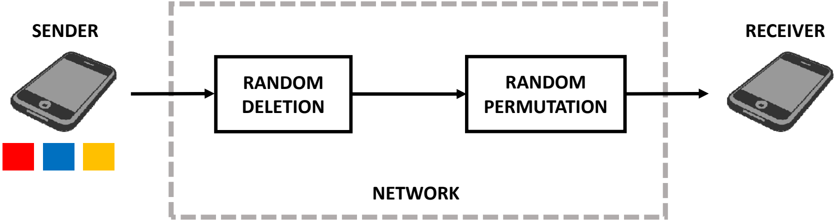


Abstraction:

- n-length codeword = sequence of n packets

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

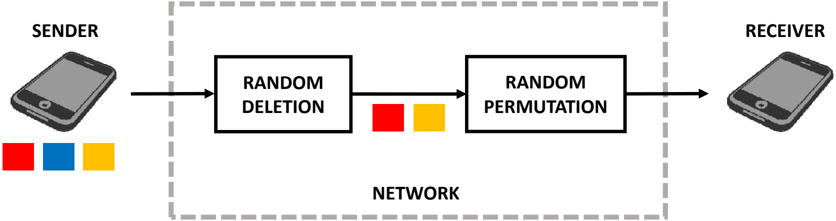


Abstraction:

- n-length codeword = sequence of n packets
- Random deletion channel: Delete each symbol/packet independently with prob $p \in (0, 1)$

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

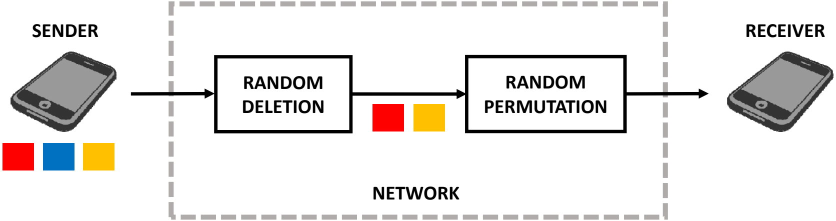


Abstraction:

- n-length codeword $=$ sequence of n packets
- Random deletion channel: Delete each symbol/packet independently with prob $p \in (0, 1)$

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

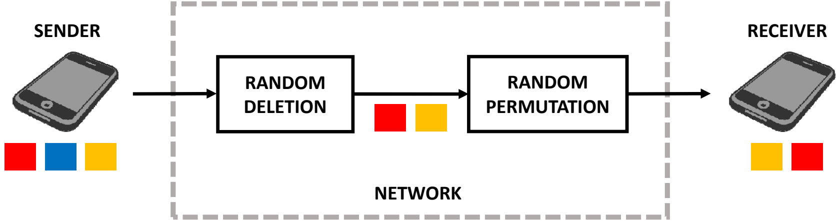


Abstraction:

- n-length codeword = sequence of n packets
- Random deletion channel: Delete each symbol/packet independently with prob $p \in (0, 1)$
- Random permutation block: Randomly permute packets of codeword

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

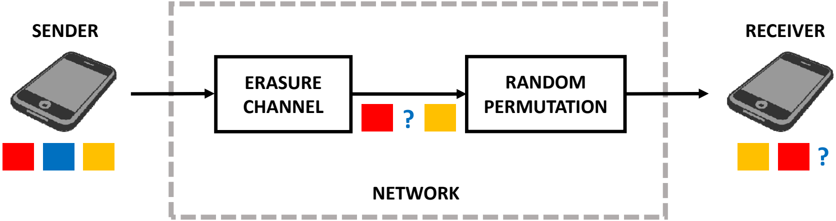


Abstraction:

- n-length codeword = sequence of n packets
- Random deletion channel: Delete each symbol/packet independently with prob $p \in (0, 1)$
- Random permutation block: Randomly permute packets of codeword

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

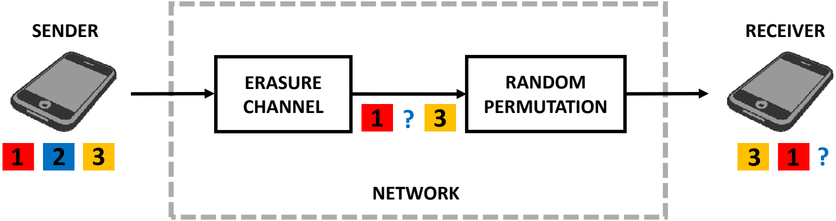


Abstraction:

- n-length codeword = sequence of n packets
- Equivalent Erasure channel: Erase each symbol/packet independently with prob $p \in (0,1)$
- Random permutation block: Randomly permute packets of codeword

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

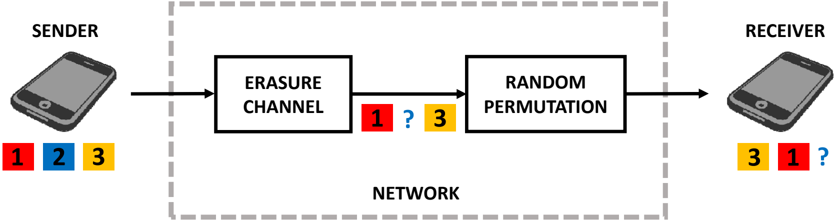


Abstraction:

- n-length codeword = sequence of n packets
- Erasure channel: Erase each symbol/packet independently with prob $p \in (0, 1)$
- Random permutation block: Randomly permute packets of codeword
- Coding: Add sequence numbers (packet size $= b + \log(n)$ bits, alphabet size $= n\,2^b$)

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

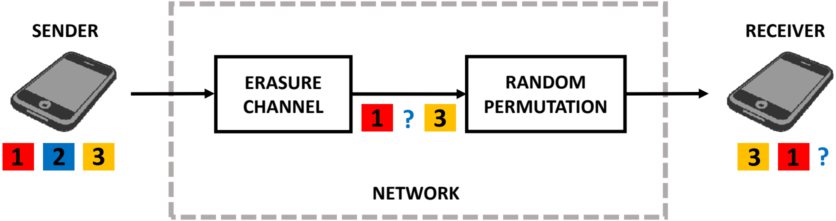


Abstraction:

- n-length codeword = sequence of n packets
- Erasure channel: Erase each symbol/packet independently with prob $p \in (0, 1)$
- Random permutation block: Randomly permute packets of codeword
- Coding: Add sequence numbers and use standard coding techniques

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

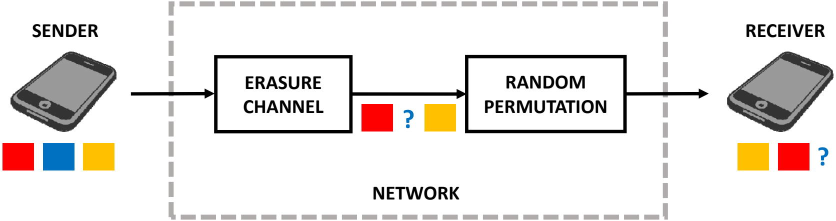


Abstraction:

- n-length codeword $=$ sequence of n packets
- Erasure channel: Erase each symbol/packet independently with prob $p \in (0, 1)$
- Random permutation block: Randomly permute packets of codeword
- Coding: Add sequence numbers and use standard coding techniques
- More refined coding techniques *simulate* sequence numbers [Mit06], [Met09]

Example: Coding for Random Deletion Network

Consider a communication network where packets can be dropped:

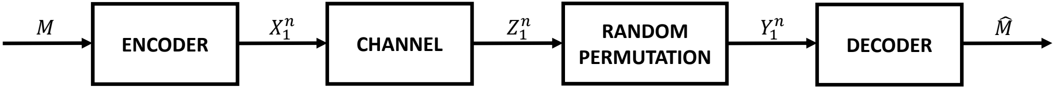


Abstraction:

- n-length codeword = sequence of n packets
- Erasure channel: Erase each symbol/packet independently with prob $p \in (0, 1)$
- Random permutation block: Randomly permute packets of codeword

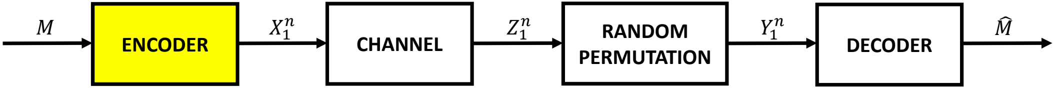
How do you code in such channels without increasing alphabet size?

Permutation Channel Model



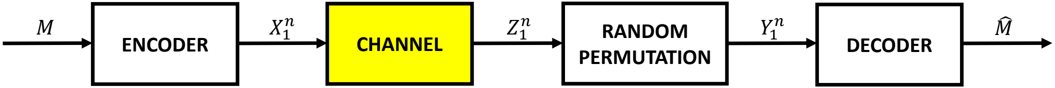
- Sender sends message $M \sim \text{Uniform}(\mathcal{M})$
- $n =$ blocklength

Permutation Channel Model



- Sender sends message $M \sim \text{Uniform}(\mathcal{M})$
- $n =$ blocklength
- Randomized encoder $f_n : \mathcal{M} \to \mathcal{X}^n$ produces codeword $X_1^n = (X_1, \ldots, X_n) = f_n(M)$

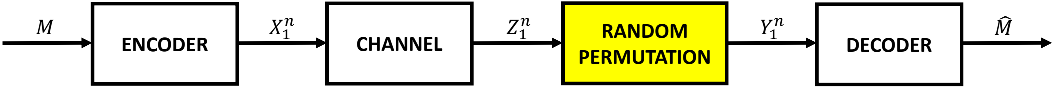
Permutation Channel Model



- Sender sends message $M \sim \text{Uniform}(\mathcal{M})$
- $n =$ blocklength
- Randomized encoder $f_n : \mathcal{M} \to \mathcal{X}^n$ produces codeword $X_1^n = (X_1, \ldots, X_n) = f_n(M)$
- Discrete memoryless channel $P_{Z|X}$ with input & output alphabets $\mathcal{X}$ & $\mathcal{Y}$ produces Z_1^n:

$$P_{Z_1^n|X_1^n}(z_1^n|x_1^n) = \prod_{i=1}^{n} P_{Z|X}(z_i|x_i)$$

Permutation Channel Model

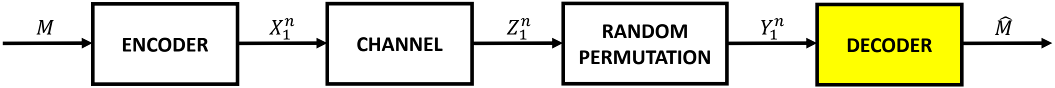


- Sender sends message $M \sim \text{Uniform}(\mathcal{M})$
- $n =$ blocklength
- Randomized encoder $f_n : \mathcal{M} \to \mathcal{X}^n$ produces codeword $X_1^n = (X_1, \dots, X_n) = f_n(M)$
- Discrete memoryless channel $P_{Z|X}$ with input & output alphabets $\mathcal{X}$ & $\mathcal{Y}$ produces Z_1^n:

$$P_{Z_1^n|X_1^n}(z_1^n|x_1^n) = \prod_{i=1}^{n} P_{Z|X}(z_i|x_i)$$

- Random permutation π generates Y_1^n from Z_1^n: $Y_{\pi(i)} = Z_i$ for $i \in \{1, \dots, n\}$

Permutation Channel Model



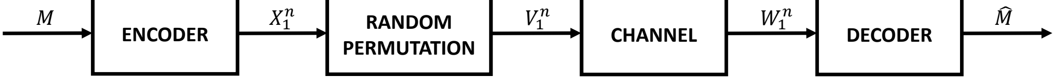
- Sender sends message $M \sim \text{Uniform}(\mathcal{M})$
- $n =$ blocklength
- Randomized encoder $f_n : \mathcal{M} \to \mathcal{X}^n$ produces codeword $X_1^n = (X_1, \dots, X_n) = f_n(M)$
- Discrete memoryless channel $P_{Z|X}$ with input & output alphabets $\mathcal{X}$ & $\mathcal{Y}$ produces Z_1^n:

$$P_{Z_1^n|X_1^n}(z_1^n|x_1^n) = \prod_{i=1}^{n} P_{Z|X}(z_i|x_i)$$

- Random permutation π generates Y_1^n from Z_1^n: $Y_{\pi(i)} = Z_i$ for $i \in \{1, \dots, n\}$
- Randomized decoder $g_n : \mathcal{Y}^n \to \mathcal{M} \cup \{\text{error}\}$ produces estimate $\hat{M} = g_n(Y_1^n)$ at receiver

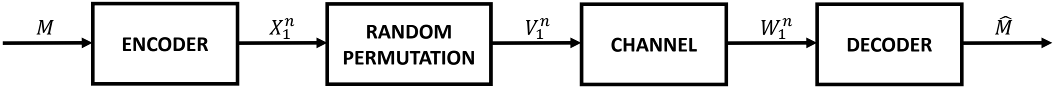
Permutation Channel Model

What if we analyze the "swapped" model?



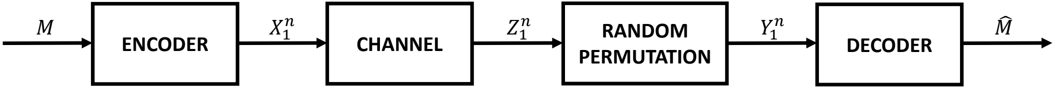
Permutation Channel Model

What if we analyze the "swapped" model?



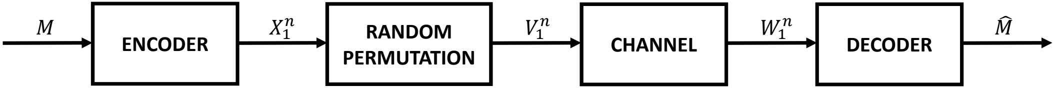
Proposition (Equivalent Models)

If channel $P_{W|V}$ is equal to channel $P_{Z|X}$, then channel $P_{W_1^n|X_1^n}$ is equal to channel $P_{Y_1^n|X_1^n}$.



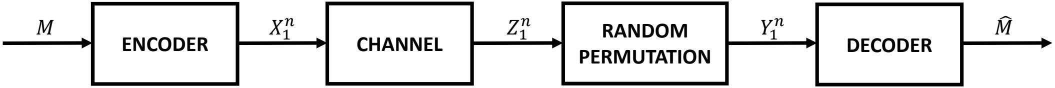
Permutation Channel Model

What if we analyze the "swapped" model?



Proposition (Equivalent Models)

If channel $P_{W|V}$ is equal to channel $P_{Z|X}$, then channel $P_{W_1^n|X_1^n}$ is equal to channel $P_{Y_1^n|X_1^n}$.

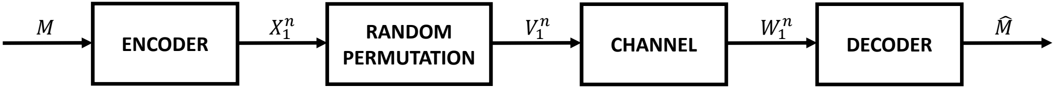


Remarks:

- Proof follows from direct calculation.

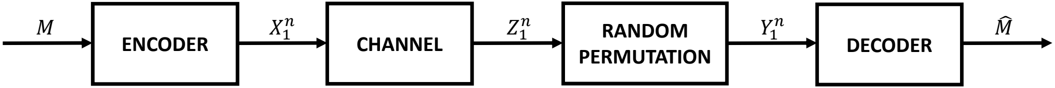
Permutation Channel Model

What if we analyze the "swapped" model?



Proposition (Equivalent Models)

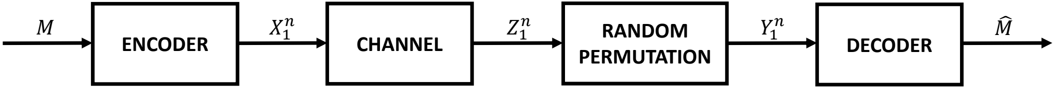
If channel $P_{W|V}$ is equal to channel $P_{Z|X}$, then channel $P_{W_1^n|X_1^n}$ is equal to channel $P_{Y_1^n|X_1^n}$.



Remarks:

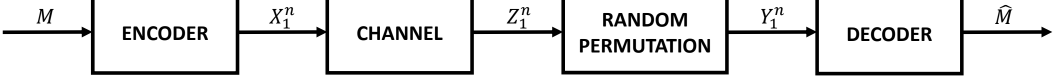
- Proof follows from direct calculation.
- Can analyze *either* model!

Coding for the Permutation Channel



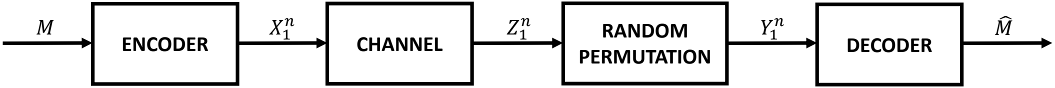
- **General Principle:**
  "Encode the information in an object that is invariant under the [permutation] transformation." [KV13]

Coding for the Permutation Channel



- **General Principle:**
  "Encode the information in an object that is invariant under the [permutation] transformation." [KV13]
- Multiset codes are studied in [KV13], [KV15], and [KT18].

Coding for the Permutation Channel

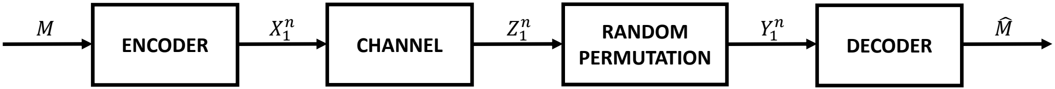


- **General Principle:**
  "Encode the information in an object that is invariant under the [permutation] transformation." [KV13]
- Multiset codes are studied in [KV13], [KV15], and [KT18].

In contrast, in [Mak18], we asked:

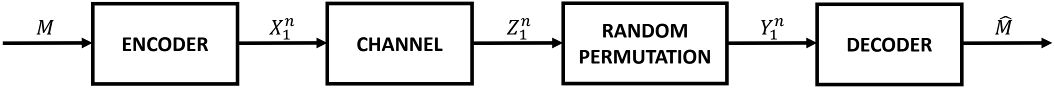
What are the fundamental information theoretic limits?

Information Capacity of the Permutation Channel



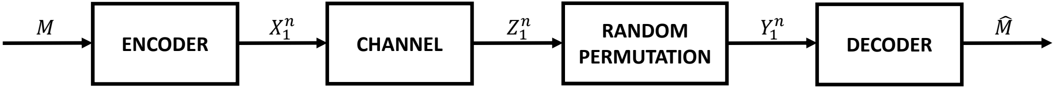
- Average probability of error $P_{\text{error}}^n \triangleq \mathbb{P}(M \neq \hat{M})$

Information Capacity of the Permutation Channel

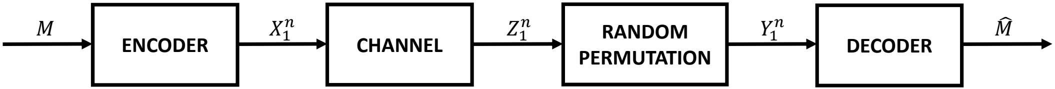


- Average probability of error $P^n_{\text{error}} \triangleq \mathbb{P}(M \neq \hat{M})$
- "Rate" of coding scheme (f_n, g_n) is $R \triangleq \dfrac{\log(|\mathcal{M}|)}{\log(n)}$

Information Capacity of the Permutation Channel

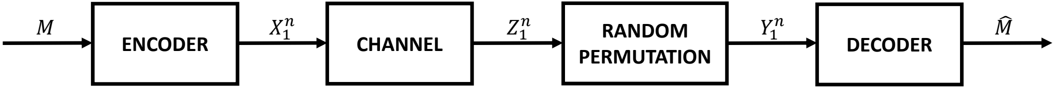


- Average probability of error $P_{\text{error}}^n \triangleq \mathbb{P}(M \neq \hat{M})$
- "Rate" of coding scheme (f_n, g_n) is $R \triangleq \dfrac{\log(|\mathcal{M}|)}{\log(n)}$
- $|\mathcal{M}| = n^R$



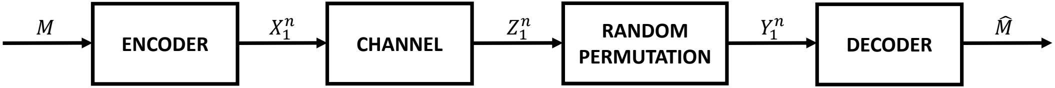
- Average probability of error $P_{\text{error}}^n \triangleq \mathbb{P}(M \neq \hat{M})$
- "Rate" of coding scheme (f_n, g_n) is $R \triangleq \dfrac{\log(|\mathcal{M}|)}{\log(n)}$
- $|\mathcal{M}| = n^R$ because number of empirical distributions of Y_1^n is $poly(n)$

Information Capacity of the Permutation Channel



- Average probability of error $P_{\text{error}}^n \triangleq \mathbb{P}(M \neq \hat{M})$
- "Rate" of coding scheme (f_n, g_n) is $R \triangleq \dfrac{\log(|\mathcal{M}|)}{\log(n)}$
- $|\mathcal{M}| = n^R$
- Rate $R \geq 0$ is achievable $\Leftrightarrow \exists \{(f_n, g_n)\}_{n \in \mathbb{N}}$ such that $\lim_{n \to \infty} P_{\text{error}}^n = 0$

Information Capacity of the Permutation Channel

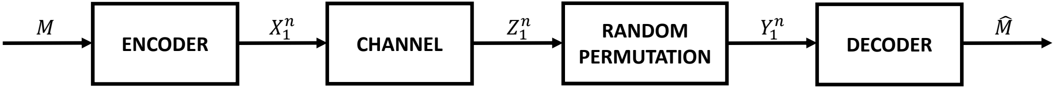


- Average probability of error $P_{\text{error}}^n \triangleq \mathbb{P}(M \neq \hat{M})$
- "Rate" of coding scheme (f_n, g_n) is $R \triangleq \dfrac{\log(|\mathcal{M}|)}{\log(n)}$
- $|\mathcal{M}| = n^R$
- Rate $R \geq 0$ is achievable $\Leftrightarrow \exists \{(f_n, g_n)\}_{n \in \mathbb{N}}$ such that $\lim_{n \to \infty} P_{\text{error}}^n = 0$

Definition (Permutation Channel Capacity [Mak18])

$$C_{\text{perm}}(P_{Z|X}) \triangleq \sup\{R \geq 0 : R \text{ is achievable}\}$$

Information Capacity of the Permutation Channel



- Average probability of error $P_{\text{error}}^n \triangleq \mathbb{P}(M \neq \hat{M})$
- "Rate" of coding scheme (f_n, g_n) is $R \triangleq \dfrac{\log(|\mathcal{M}|)}{\log(n)}$
- $|\mathcal{M}| = n^R$
- Rate $R \geq 0$ is achievable $\Leftrightarrow \exists \{(f_n, g_n)\}_{n \in \mathbb{N}}$ such that $\lim_{n \to \infty} P_{\text{error}}^n = 0$

Definition (Permutation Channel Capacity [Mak18])

$$C_{\text{perm}}(P_{Z|X}) \triangleq \sup\{R \geq 0 : R \text{ is achievable}\}$$

Main Question

What is the permutation channel capacity of a general $P_{Z|X}$?

Outline

1. Introduction
2. Achievability Bound
   - Coding Scheme
   - Rank Bound
3. Converse Bounds
4. Conclusion

Achievability: Coding Scheme

- Let $r = \text{rank}(P_{Z|X})$ and $k = \lfloor \sqrt{n} \rfloor$

Achievability: Coding Scheme

- Let $r = \text{rank}(P_{Z|X})$ and $k = \lfloor \sqrt{n} \rfloor$
- Consider $\mathcal{X}' \subseteq \mathcal{X}$ with $|\mathcal{X}'| = r$ such that $\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}'\}$ are linearly independent

Achievability: Coding Scheme

- Let $r = \text{rank}(P_{Z|X})$ and $k = \lfloor \sqrt{n} \rfloor$
- Consider $\mathcal{X}' \subseteq \mathcal{X}$ with $|\mathcal{X}'| = r$ such that $\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}'\}$ are linearly independent
- **Message set:**

$$\mathcal{M} \triangleq \left\{ p = (p(x) : x \in \mathcal{X}') \in (\mathbb{Z}_+)^{\mathcal{X}'} : \sum_{x \in \mathcal{X}'} p(x) = k \right\}$$

Achievability: Coding Scheme

- Let $r = \text{rank}(P_{Z|X})$ and $k = \lfloor \sqrt{n} \rfloor$
- Consider $\mathcal{X}' \subseteq \mathcal{X}$ with $|\mathcal{X}'| = r$ such that $\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}'\}$ are linearly independent
- **Message set:**

$$\mathcal{M} \triangleq \left\{ p = (p(x) : x \in \mathcal{X}') \in (\mathbb{Z}_+)^{\mathcal{X}'} : \sum_{x \in \mathcal{X}'} p(x) = k \right\}$$

where $|\mathcal{M}| = \binom{k+r-1}{r-1} = \Theta\left(n^{\frac{r-1}{2}}\right)$

Achievability: Coding Scheme

- Let $r = \text{rank}(P_{Z|X})$ and $k = \lfloor \sqrt{n} \rfloor$
- Consider $\mathcal{X}' \subseteq \mathcal{X}$ with $|\mathcal{X}'| = r$ such that $\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}'\}$ are linearly independent
- **Message set:**

$$\mathcal{M} \triangleq \left\{ p = (p(x) : x \in \mathcal{X}') \in (\mathbb{Z}_+)^{\mathcal{X}'} : \sum_{x \in \mathcal{X}'} p(x) = k \right\}$$

  where $|\mathcal{M}| = \binom{k+r-1}{r-1} = \Theta\left(n^{\frac{r-1}{2}}\right)$

- **Randomized Encoder:**

$$\forall p \in \mathcal{M}, \quad f_n(p) = X_1^n \overset{\text{i.i.d.}}{\sim} P_X \quad \text{where} \quad P_X(x) = \begin{cases} \frac{p(x)}{k}, & \text{for } x \in \mathcal{X}' \\ 0, & \text{for } x \in \mathcal{X} \backslash \mathcal{X}' \end{cases}$$

Achievability: Coding Scheme

- Let stochastic matrix $\tilde{P}_{Z|X} \in \mathbb{R}^{r \times |\mathcal{Y}|}$ have rows $\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}'\}$
- Let $\tilde{P}_{Z|X}^{\dagger}$ denote its *Moore-Penrose pseudoinverse*

Achievability: Coding Scheme

- Let stochastic matrix $\tilde{P}_{Z|X} \in \mathbb{R}^{r \times |\mathcal{Y}|}$ have rows $\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}'\}$
- Let $\tilde{P}_{Z|X}^{\dagger}$ denote its *Moore-Penrose pseudoinverse*
- **(Sub-optimal) Thresholding Decoder:** For any $y_1^n \in \mathcal{Y}^n$,
  Step 1: Construct its type/empirical distribution/histogram

$$\forall y \in \mathcal{Y}, \quad \hat{P}_{y_1^n}(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{y_i = y\}$$

Achievability: Coding Scheme

- Let stochastic matrix $\tilde{P}_{Z|X} \in \mathbb{R}^{r \times |\mathcal{Y}|}$ have rows $\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}'\}$
- Let $\tilde{P}^{\dagger}_{Z|X}$ denote its *Moore-Penrose pseudoinverse*
- **(Sub-optimal) Thresholding Decoder:** For any $y_1^n \in \mathcal{Y}^n$,

  Step 1: Construct its type/empirical distribution/histogram

$$\forall y \in \mathcal{Y}, \quad \hat{P}_{y_1^n}(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{y_i = y\}$$

  Step 2: Generate estimate $\hat{p} \in (\mathbb{Z}_+)^{\mathcal{X}'}$ with components

$$\forall x \in \mathcal{X}', \quad \hat{p}(x) = \underset{j \in \{0,\dots,k\}}{\arg\min} \left| \sum_{y \in \mathcal{Y}} \hat{P}_{y_1^n}(y) \left[\tilde{P}^{\dagger}_{Z|X}\right]_{y,x} - \frac{j}{k} \right|$$

Achievability: Coding Scheme

- Let stochastic matrix $\tilde{P}_{Z|X} \in \mathbb{R}^{r \times |\mathcal{Y}|}$ have rows $\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}'\}$
- Let $\tilde{P}^{\dagger}_{Z|X}$ denote its *Moore-Penrose pseudoinverse*
- **(Sub-optimal) Thresholding Decoder:** For any $y_1^n \in \mathcal{Y}^n$,

  Step 1: Construct its type/empirical distribution/histogram

$$\forall y \in \mathcal{Y}, \quad \hat{P}_{y_1^n}(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{y_i = y\}$$

  Step 2: Generate estimate $\hat{p} \in (\mathbb{Z}_+)^{\mathcal{X}'}$ with components

$$\forall x \in \mathcal{X}', \quad \hat{p}(x) = \underset{j \in \{0,\dots,k\}}{\arg\min} \left| \sum_{y \in \mathcal{Y}} \hat{P}_{y_1^n}(y) \left[\tilde{P}^{\dagger}_{Z|X} \right]_{y,x} - \frac{j}{k} \right|$$

  Step 3: Output decoded message

$$g_n(y_1^n) = \begin{cases} \hat{p}, & \text{if } \hat{p} \in \mathcal{M} \\ \text{error}, & \text{otherwise} \end{cases}$$

Achievability: Rank Bound

Theorem (Rank Bound)

For any channel $P_{Z|X}$:

$$C_{\text{perm}}(P_{Z|X}) \geq \frac{\operatorname{rank}(P_{Z|X}) - 1}{2} .$$

Remarks about Coding Scheme:

- Showing $\lim_{n \to \infty} P_{\text{error}}^n = 0$ proves theorem.

Achievability: Rank Bound

Theorem (Rank Bound)

For any channel $P_{Z|X}$:

$$C_{\text{perm}}(P_{Z|X}) \geq \frac{\text{rank}(P_{Z|X}) - 1}{2} .$$

Remarks about Coding Scheme:

- Showing $\lim_{n\to\infty} P_{\text{error}}^n = 0$ proves theorem.
- *Intuition*: Conditioned on $M = p$, $\hat{P}_{Y_1^n} \approx P_Z$ with high probability as $n \to \infty$.

Achievability: Rank Bound

Theorem (Rank Bound)

For any channel $P_{Z|X}$:

$$C_{\text{perm}}(P_{Z|X}) \geq \frac{\text{rank}(P_{Z|X}) - 1}{2} .$$

Remarks about Coding Scheme:

- Showing $\lim_{n\to\infty} P^n_{\text{error}} = 0$ proves theorem.
- *Intuition*: Conditioned on $M = p$, $\hat{P}_{Y_1^n} \approx P_Z$ with high probability as $n \to \infty$. Hence, $\sum_{y\in\mathcal{Y}} \hat{P}_{Y_1^n}(y)\big[\tilde{P}^{\dagger}_{Z|X}\big]_{y,x} \approx P_X(x)$ for all $x \in \mathcal{X}'$ with high probability.

Achievability: Rank Bound

Theorem (Rank Bound)

For any channel $P_{Z|X}$:

$$C_{\text{perm}}(P_{Z|X}) \geq \frac{\text{rank}(P_{Z|X}) - 1}{2} \, .$$

Remarks about Coding Scheme:

- Showing $\lim_{n\to\infty} P^n_{\text{error}} = 0$ proves theorem.
- *Intuition*: Conditioned on $M = p$, $\hat{P}_{Y_1^n} \approx P_Z$ with high probability as $n \to \infty$. Hence, $\sum_{y\in\mathcal{Y}} \hat{P}_{Y_1^n}(y)\big[\tilde{P}^{\dagger}_{Z|X}\big]_{y,x} \approx P_X(x)$ for all $x \in \mathcal{X}'$ with high probability.
- *Computational complexity*: Decoder has $O(n)$ running time.

Achievability: Rank Bound

Theorem (Rank Bound)

For any channel $P_{Z|X}$:

$$C_{\text{perm}}(P_{Z|X}) \geq \frac{\operatorname{rank}(P_{Z|X}) - 1}{2} .$$

Remarks about Coding Scheme:

- Showing $\lim_{n\to\infty} P_{\text{error}}^n = 0$ proves theorem.
- *Intuition*: Conditioned on $M = p$, $\hat{P}_{Y_1^n} \approx P_Z$ with high probability as $n \to \infty$.
  Hence, $\sum_{y\in\mathcal{Y}} \hat{P}_{Y_1^n}(y) \big[\tilde{P}_{Z|X}^{\dagger}\big]_{y,x} \approx P_X(x)$ for all $x \in \mathcal{X}'$ with high probability.
- *Computational complexity*: Decoder has $O(n)$ running time.
- *Probabilistic method*: Good deterministic codes exist.

Achievability: Rank Bound

Theorem (Rank Bound)

For any channel $P_{Z|X}$:

$$C_{\text{perm}}(P_{Z|X}) \geq \frac{\text{rank}(P_{Z|X}) - 1}{2}.$$

Remarks about Coding Scheme:

- Showing $\lim_{n\to\infty} P^n_{\text{error}} = 0$ proves theorem.
- *Intuition*: Conditioned on $M = p$, $\hat{P}_{Y_1^n} \approx P_Z$ with high probability as $n \to \infty$. Hence, $\sum_{y\in\mathcal{Y}} \hat{P}_{Y_1^n}(y)\big[\tilde{P}^{\dagger}_{Z|X}\big]_{y,x} \approx P_X(x)$ for all $x \in \mathcal{X}'$ with high probability.
- *Computational complexity*: Decoder has $O(n)$ running time.
- *Probabilistic method*: Good deterministic codes exist.
- *Expurgation*: Achievability bound holds under maximal probability of error criterion.

Outline

Converse: Output Alphabet Bound

Theorem (Output Alphabet Bound)

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$C_{\mathsf{perm}}(P_{Z|X}) \leq \frac{|\mathcal{Y}| - 1}{2} .$$

Converse: Output Alphabet Bound

Theorem (Output Alphabet Bound)

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$C_{\text{perm}}(P_{Z|X}) \leq \frac{|\mathcal{Y}| - 1}{2} .$$

Remarks:

- Proof hinges on *Fano's inequality* and CLT-based approximation of *binomial entropy*.

Converse: Output Alphabet Bound

Theorem (Output Alphabet Bound)

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$C_{\text{perm}}(P_{Z|X}) \leq \frac{|\mathcal{Y}| - 1}{2} \,.$$

Remarks:

- Proof hinges on *Fano's inequality* and CLT-based approximation of *binomial entropy*.
- What if $|\mathcal{X}|$ is much smaller than $|\mathcal{Y}|$?

Converse: Output Alphabet Bound

Theorem (Output Alphabet Bound)

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$C_{\text{perm}}(P_{Z|X}) \leq \frac{|\mathcal{Y}| - 1}{2} .$$

Remarks:

- Proof hinges on *Fano's inequality* and CLT-based approximation of *binomial entropy*.
- What if $|\mathcal{X}|$ is much smaller than $|\mathcal{Y}|$?
- **Want:** Converse bound in terms of input alphabet size.

Converse: Effective Input Alphabet Bound

Theorem (Effective Input Alphabet Bound)

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$C_{\mathrm{perm}}(P_{Z|X}) \leq \frac{\mathrm{ext}(P_{Z|X}) - 1}{2}$$

where $\mathrm{ext}(P_{Z|X})$ denotes the number of *extreme points* of $\mathrm{conv}\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}\}$.

Converse: Effective Input Alphabet Bound

Theorem (Effective Input Alphabet Bound)

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$C_{\text{perm}}(P_{Z|X}) \leq \frac{\text{ext}(P_{Z|X}) - 1}{2}$$

where $\text{ext}(P_{Z|X})$ denotes the number of *extreme points* of $\text{conv}\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}\}$.

Remarks:

- *Effective input alphabet size*: $\text{rank}(P_{Z|X}) \leq \text{ext}(P_{Z|X}) \leq |\mathcal{X}|$.

Converse: Effective Input Alphabet Bound

Theorem (Effective Input Alphabet Bound)

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$C_{\text{perm}}(P_{Z|X}) \leq \frac{\text{ext}(P_{Z|X}) - 1}{2}$$

where $\text{ext}(P_{Z|X})$ denotes the number of *extreme points* of $\text{conv}\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}\}$.

Remarks:

- *Effective input alphabet size*: $\text{rank}(P_{Z|X}) \leq \text{ext}(P_{Z|X}) \leq |\mathcal{X}|$.
- For any channel $P_{Z|X} > 0$, $C_{\text{perm}}(P_{Z|X}) \leq (\min\{\text{ext}(P_{Z|X}), |\mathcal{Y}|\} - 1)/2$.

Converse: Effective Input Alphabet Bound

Theorem (Effective Input Alphabet Bound)

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$C_{\text{perm}}(P_{Z|X}) \leq \frac{\text{ext}(P_{Z|X}) - 1}{2}$$

where $\text{ext}(P_{Z|X})$ denotes the number of *extreme points* of $\text{conv}\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}\}$.

Remarks:

- *Effective input alphabet size*: $\text{rank}(P_{Z|X}) \leq \text{ext}(P_{Z|X}) \leq |\mathcal{X}|$.
- For any channel $P_{Z|X} > 0$, $C_{\text{perm}}(P_{Z|X}) \leq \big(\min\{\text{ext}(P_{Z|X}), |\mathcal{Y}|\} - 1\big)/2$.
- For any *general* channel $P_{Z|X}$, $C_{\text{perm}}(P_{Z|X}) \leq \min\{\text{ext}(P_{Z|X}), |\mathcal{Y}|\} - 1$.

Converse: Effective Input Alphabet Bound

Theorem (Effective Input Alphabet Bound)

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$C_{\text{perm}}(P_{Z|X}) \leq \frac{\text{ext}(P_{Z|X}) - 1}{2}$$

where $\text{ext}(P_{Z|X})$ denotes the number of *extreme points* of $\text{conv}\{P_{Z|X}(\cdot|x) : x \in \mathcal{X}\}$.

Remarks:

- *Effective input alphabet size*: $\text{rank}(P_{Z|X}) \leq \text{ext}(P_{Z|X}) \leq |\mathcal{X}|$.
- For any channel $P_{Z|X} > 0$, $C_{\text{perm}}(P_{Z|X}) \leq \left(\min\{\text{ext}(P_{Z|X}), |\mathcal{Y}|\} - 1\right)/2$.
- For any *general* channel $P_{Z|X}$, $C_{\text{perm}}(P_{Z|X}) \leq \min\{\text{ext}(P_{Z|X}), |\mathcal{Y}|\} - 1$.
- **How do we prove above theorem?**

Proof Idea: Degradation by Symmetric Channels

Definition (Degradation/Blackwell Order [Bla51], [She51], [Ste51], [Cov72], [Ber73])

Given channels $P_{Z_1|X}$ and $P_{Z_2|X}$ with common input alphabet $\mathcal{X}$, $P_{Z_2|X}$ is a degraded version of $P_{Z_1|X}$ if $P_{Z_2|X} = P_{Z_1|X}P_{Z_2|Z_1}$ for some channel $P_{Z_2|Z_1}$.

Proof Idea: Degradation by Symmetric Channels

Definition (Degradation/Blackwell Order [Bla51], [She51], [Ste51], [Cov72], [Ber73])

Given channels $P_{Z_1|X}$ and $P_{Z_2|X}$ with common input alphabet $\mathcal{X}$, $P_{Z_2|X}$ is a degraded version of $P_{Z_1|X}$ if $P_{Z_2|X} = P_{Z_1|X} P_{Z_2|Z_1}$ for some channel $P_{Z_2|Z_1}$.

Definition (q-ary Symmetric Channel)

A q-ary symmetric channel, denoted q-SC(δ), with total crossover probability $\delta \in [0, 1]$ and alphabet $\mathcal{X}$ where $|\mathcal{X}| = q$, is given by the doubly stochastic matrix:

$$W_\delta \triangleq \begin{bmatrix} 1-\delta & \frac{\delta}{q-1} & \cdots & \frac{\delta}{q-1} \\ \frac{\delta}{q-1} & 1-\delta & \cdots & \frac{\delta}{q-1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\delta}{q-1} & \frac{\delta}{q-1} & \cdots & 1-\delta \end{bmatrix}.$$

Proof Idea: Degradation by Symmetric Channels

Proposition (Degradation by Symmetric Channels)

Given channel $P_{Z|X}$ with $\nu = \min_{x \in \mathcal{X}, y \in \mathcal{Y}} P_{Z|X}(y|x)$, if we have:

$$0 \leq \delta \leq \frac{\nu}{1 - \nu + \frac{\nu}{q-1}},$$

then $P_{Z|X}$ is a degraded version of q-SC(δ).

Proof Idea: Degradation by Symmetric Channels

Proposition (Degradation by Symmetric Channels)

Given channel $P_{Z|X}$ with $\nu = \min_{x \in \mathcal{X},\, y \in \mathcal{Y}} P_{Z|X}(y|x)$, if we have:

$$0 \leq \delta \leq \frac{\nu}{1 - \nu + \frac{\nu}{q-1}} ,$$

then $P_{Z|X}$ is a degraded version of q-SC(δ).

- Prop follows from computing extremal δ such that $W_\delta^{-1} P_{Z|X}$ is row stochastic.
- Many other applications in information theory and statistics [MP18], [MOS13].

Proof Idea: Degradation by Symmetric Channels

Proposition (Degradation by Symmetric Channels)

Given channel $P_{Z|X}$ with $\nu = \min_{x \in \mathcal{X},\, y \in \mathcal{Y}} P_{Z|X}(y|x)$, if we have:

$$0 \leq \delta \leq \frac{\nu}{1 - \nu + \frac{\nu}{q-1}} \, ,$$

then $P_{Z|X}$ is a degraded version of q-SC(δ).

- Prop follows from computing extremal δ such that $W_{\delta}^{-1} P_{Z|X}$ is row stochastic.
- Many other applications in information theory and statistics [MP18], [MOS13].
- Prop + "swapped" model + *tensorization* of degradation $\Rightarrow$ $I(X_1^n; Y_1^n) \leq I(X_1^n; \tilde{Y}_1^n)$, where Y_1^n and $\tilde{Y}_1^n$ are outputs of permutation channels with $P_{Z|X}$ and q-SC(δ).

Proof Idea: Degradation by Symmetric Channels

Proposition (Degradation by Symmetric Channels)

Given channel $P_{Z|X}$ with $\nu = \min_{x \in \mathcal{X}, y \in \mathcal{Y}} P_{Z|X}(y|x)$, if we have:

$$0 \leq \delta \leq \frac{\nu}{1 - \nu + \frac{\nu}{q-1}} ,$$

then $P_{Z|X}$ is a degraded version of q-SC(δ).

- Prop follows from computing extremal δ such that $W_\delta^{-1} P_{Z|X}$ is row stochastic.
- Many other applications in information theory and statistics [MP18], [MOS13].
- Prop + "swapped" model + *tensorization* of degradation $\Rightarrow$ $I(X_1^n; Y_1^n) \leq I(X_1^n; \tilde{Y}_1^n)$, where Y_1^n and $\tilde{Y}_1^n$ are outputs of permutation channels with $P_{Z|X}$ and q-SC(δ).
- *Convexity* of KL divergence $\Rightarrow$ Reduce $|\mathcal{X}|$ to ext$(P_{Z|X})$.

Proof Idea: Degradation by Symmetric Channels

Proposition (Degradation by Symmetric Channels)

Given channel $P_{Z|X}$ with $\nu = \min_{x \in \mathcal{X},\, y \in \mathcal{Y}} P_{Z|X}(y|x)$, if we have:

$$0 \leq \delta \leq \frac{\nu}{1 - \nu + \frac{\nu}{q-1}},$$

then $P_{Z|X}$ is a degraded version of q-SC(δ).

- Prop follows from computing extremal δ such that $W_\delta^{-1} P_{Z|X}$ is row stochastic.
- Many other applications in information theory and statistics [MP18], [MOS13].
- Prop + "swapped" model + *tensorization* of degradation $\Rightarrow$ $I(X_1^n; Y_1^n) \leq I(X_1^n; \tilde{Y}_1^n)$, where Y_1^n and $\tilde{Y}_1^n$ are outputs of permutation channels with $P_{Z|X}$ and q-SC(δ).
- *Convexity* of KL divergence $\Rightarrow$ Reduce $|\mathcal{X}|$ to ext$(P_{Z|X})$.
- *Fano argument* of output alphabet bound $\Rightarrow$ effective input alphabet bound.

Outline

Strictly Positive and "Full Rank" Channels

Achievability and converse bounds yield:

Theorem (Strictly Positive and "Full Rank" Channels)

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$ that is *"full rank"* in the sense that $r \triangleq \text{rank}(P_{Z|X}) = \min\{\text{ext}(P_{Z|X}), |\mathcal{Y}|\}$:

$$C_{\text{perm}}(P_{Z|X}) = \frac{r-1}{2} .$$

Strictly Positive and "Full Rank" Channels

Achievability and converse bounds yield:

Theorem (Strictly Positive and "Full Rank" Channels)

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$ that is *"full rank"* in the sense that $r \triangleq \text{rank}(P_{Z|X}) = \min\{\text{ext}(P_{Z|X}), |\mathcal{Y}|\}$:

$$C_{\text{perm}}(P_{Z|X}) = \frac{r-1}{2} .$$

Example [Mak18]**:** C_{perm} of non-trivial binary symmetric channel is $\frac{1}{2}$.

Conclusion

Main Result:

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$\frac{\text{rank}(P_{Z|X}) - 1}{2} \leq C_{\text{perm}}(P_{Z|X}) \leq \frac{\min\{\text{ext}(P_{Z|X}), |\mathcal{Y}|\} - 1}{2}.$$

Conclusion

Main Result:

For any entry-wise *strictly positive* channel $P_{Z|X} > 0$:

$$\frac{\text{rank}(P_{Z|X}) - 1}{2} \leq C_{\text{perm}}(P_{Z|X}) \leq \frac{\min\{\text{ext}(P_{Z|X}), |\mathcal{Y}|\} - 1}{2}.$$

Future Direction:

Characterize C_{perm} of all entry-wise strictly positive channels, and more generally, all channels.

Thank You!