

## Recap

We have seen a measure for how much *information* we could ever hope to send through a channel.

Channel Capacity:

$$C = \max_{\mathcal{A}} I(\mathcal{A}, \mathcal{B})$$

$I(\mathcal{A}, \mathcal{B})$  represents how much information in  $\mathcal{A}$  emerges in  $\mathcal{B}$ , for a given  $\Gamma$ .

We wish to *apply*  $\Gamma$  *efficiently*; get our *transfer rate* as close to  $C$  as possible.

Two factors play in on how efficiently we can use  $\Gamma$ :

- (i) How we *improve reliability* (how we transfer redundancy),
- (ii) How we *decide* what was sent from what we receive.

We shall look at (ii) first; then (i).

# Decision Rules, and Quality Thereof

Recall, that for given  $\Gamma$  and distribution of  $\mathcal{A}$ :

$P_{ij} = \Pr(Y = b_j \mid X = a_i)$  ; entry in *forward* probability matrix

$Q_{ij} = \Pr(X = a_i \mid Y = b_j)$  ; entry in *backward* probability matrix

$R_{ij} = \Pr(X = a_i, Y = b_j)$  ; entry in *joint* probability matrix

**Decision Rule  $\Delta$ :** A function for “guessing” input to  $\Gamma$ .

$$\Delta: B \rightarrow A \quad ; \quad \Delta(b_j) = a_{j^*}$$

$a_{j^*}$  denotes what receiver *decides*  $\mathcal{A}$  sent, on receiving  $b_i$ .

**Hopefully**,  $a_{j^*}$  was sent<sup>1</sup>. Chance being correct:

$$\Pr(X = a_{j^*} \mid Y = b_j) = Q_{j^*j}$$

---

<sup>1</sup>making the decision valid

## Decision Rules, and Quality Thereof

Recall, that for given  $\Gamma$  and distribution of  $\mathcal{A}$ :

$P_{ij} = \Pr(Y = b_j \mid X = a_i)$  ; entry in *forward* probability matrix

$Q_{ij} = \Pr(X = a_i \mid Y = b_j)$  ; entry in *backward* probability matrix

$R_{ij} = \Pr(X = a_i, Y = b_j)$  ; entry in *joint* probability matrix

**Decision Rule  $\Delta$ :** A function for “guessing” input to  $\Gamma$ .

$$\Delta: B \rightarrow A \quad ; \quad \Delta(b_j) = a_{j^*}$$

$a_{j^*}$  denotes what receiver *decides*  $\mathcal{A}$  sent, on receiving  $b_i$ .

**Hopefully**,  $a_{j^*}$  was sent<sup>1</sup>. Chance being correct:

$$\Pr(X = a_{j^*} \mid Y = b_j) = Q_{j^*j}$$

**Pr<sub>C</sub>:** Average probability of *correct* decision when using  $\Delta$ .

$$\Pr_C = \sum_j q_j Q_{j^*j} = \sum_j R_{j^*j} \text{ , as } q_j Q_{ij} = R_{ij}$$

**Pr<sub>E</sub>:** Average probability of *incorrect* decision when using  $\Delta$ .

$$\Pr_E = 1 - \Pr_C = 1 - \sum_j R_{j^*j} = \sum_j \sum_{i \neq j^*} R_{ij}$$

---

<sup>1</sup>making the decision valid

Decision Rules

Definitions

Example

Improving  
ReliabilityMajority Decoding  
(Repetition code)

Generalisation

Transmission Rate

Peek Into Linear  
Block Codes

Example

Epilogue

We wish to **maximise**  $\Pr_C$  (minimise  $\Pr_E$ ).  $\longleftarrow$  *ideal observer rule.*

$p_i$  are known: For each  $j$ , choose  $i = j^*$  such that  $Q_{ij} = \Pr(a_i | b_j)$  is *maximal*. Largest entry in column  $j$  in  $(Q_{ij})$ .

- Equivalent to maximising  $R_{ij} = q_j Q_{ij}$ , that is,  $R_{j^*j} \geq R_{ij}$ ,  $\forall i$ .  $R_{j^*j}$  is then the largest entry in column  $j$  in  $(R_{ij})$ .

---

<sup>2</sup>“On average”, it is.

We wish to **maximise**  $\Pr_C$  (minimise  $\Pr_E$ ).  $\longleftarrow$  *ideal observer rule.*

$p_i$  are known: For each  $j$ , choose  $i = j^*$  such that  $Q_{ij} = \Pr(a_i | b_j)$  is *maximal*. Largest entry in column  $j$  in  $(Q_{ij})$ .

- Equivalent to maximising  $R_{ij} = q_j Q_{ij}$ , that is,  $R_{j^*j} \geq R_{ij}$ ,  $\forall i$ .  $R_{j^*j}$  is then the largest entry in column  $j$  in  $(R_{ij})$ .
- Why is that interesting:  $(R_{ij})$  is *easily computed*, since  $R_{ij} = p_i P_{ij}$  (Thank you, Bayes!).

$$p_i P_{ij} = \Pr(a_i) \Pr(b_j | a_i) = \underbrace{\Pr(a_i, b_j)}_{=R_{ij}} = \Pr(b_j) \Pr(a_i | b_j) = q_j Q_{ij}$$

Easily, how: Scale each  $P_{ij}$  with  $p_i$ . Equivalent to

$$(R_{ij}) = \begin{pmatrix} p_1 & & \\ & \ddots & \\ & & p_r \end{pmatrix} (P_{ij})$$

<sup>2</sup>"On average", it is.

We wish to **maximise**  $\Pr_C$  (minimise  $\Pr_E$ ).  $\leftarrow$  *ideal observer rule.*

$p_i$  are known: For each  $j$ , choose  $i = j^*$  such that  $Q_{ij} = \Pr(a_i | b_j)$  is *maximal*. Largest entry in column  $j$  in  $(Q_{ij})$ .

- Equivalent to maximising  $R_{ij} = q_j Q_{ij}$ , that is,  $R_{j^*j} \geq R_{ij}$ ,  $\forall i$ .  $R_{j^*j}$  is then the largest entry in column  $j$  in  $(R_{ij})$ .
- Why is that interesting:  $(R_{ij})$  is *easily computed*, since  $R_{ij} = p_i P_{ij}$  (Thank you, Bayes!).

$$p_i P_{ij} = \Pr(a_i) \Pr(b_j | a_i) = \underbrace{\Pr(a_i, b_j)}_{=R_{ij}} = \Pr(b_j) \Pr(a_i | b_j) = q_j Q_{ij}$$

Easily, how: Scale each  $P_{ij}$  with  $p_i$ . Equivalent to

$$(R_{ij}) = \begin{pmatrix} p_1 & & \\ & \ddots & \\ & & p_r \end{pmatrix} (P_{ij})$$

$p_i$  are unknown: Assume  $(p_i)$  is uniformly distributed<sup>2</sup>. Then proceed as above. Note: Then we are essentially picking  $i = j^*$  maximising  $P_{ij}$ .

$\leftarrow$  *maximum likelihood rule*

<sup>2</sup>"On average", it is.

# Binary Symmetric Channel $\Gamma$ , with $P$

Recall: We are given  $(p, \bar{p})$  (distribution of  $\mathcal{A}$ ), and *forward probabilities*

$$(P_{ij}) = \begin{pmatrix} P & \bar{P} \\ \bar{P} & P \end{pmatrix}, \text{ from which } (q, \bar{q}) = (p, \bar{p}) \begin{pmatrix} P & \bar{P} \\ \bar{P} & P \end{pmatrix} = (pP + \bar{p}\bar{P}, p\bar{P} + \bar{p}P).$$

## Lecture 7: Using $\Gamma$ Efficiently

Decision Rules

Definitions

Example

Improving  
ReliabilityMajority Decoding  
(Repetition code)

Generalisation

Transmission Rate

Peek Into Linear  
Block Codes

Example

Epilogue

Binary Symmetric Channel  $\Gamma$ , with  $P$ 

Recall: We are given  $(p, \bar{p})$  (distribution of  $\mathcal{A}$ ), and *forward probabilities*

$$(P_{ij}) = \begin{pmatrix} P & \bar{P} \\ \bar{P} & P \end{pmatrix}, \text{ from which } (q, \bar{q}) = (p, \bar{p}) \begin{pmatrix} P & \bar{P} \\ \bar{P} & P \end{pmatrix} = (pP + \bar{p}\bar{P}, p\bar{P} + \bar{p}P).$$

Recall Bayes' formulae.

$$p_i P_{ij} = \Pr(a_i) \Pr(b_j | a_i) = \underbrace{\Pr(a_i, b_j)}_{=R_{ij}} = \Pr(b_j) \Pr(a_i | b_j) = q_j Q_{ij}$$

From this, since  $Q_{ij} = \frac{p_i}{q_j} P_{ij}$ , and since  $R_{ij} = q_j Q_{ij}$ ,  $(p_0, p_1) = (p, \bar{p})$ ,  $(q_0, q_1) = (q, \bar{q})$ ,

$$(Q_{ij}) = \begin{pmatrix} \frac{p}{q} P & \frac{p}{\bar{q}} \bar{P} \\ \frac{\bar{p}}{q} \bar{P} & \frac{\bar{p}}{\bar{q}} P \end{pmatrix}, \text{ and } (R_{ij}) = \begin{pmatrix} pP & p\bar{P} \\ \bar{p}\bar{P} & \bar{p}P \end{pmatrix} = \underline{\begin{pmatrix} p & 0 \\ 0 & \bar{p} \end{pmatrix} (P_{ij})}$$



Binary Symmetric Channel  $\Gamma$ , with  $P$ 

Recall: We are given  $(p, \bar{p})$  (distribution of  $\mathcal{A}$ ), and *forward probabilities*

$$(P_{ij}) = \begin{pmatrix} P & \bar{P} \\ \bar{P} & P \end{pmatrix}, \text{ from which } (q, \bar{q}) = (p, \bar{p}) \begin{pmatrix} P & \bar{P} \\ \bar{P} & P \end{pmatrix} = (pP + \bar{p}\bar{P}, p\bar{P} + \bar{p}P).$$

Recall Bayes' formulae.

$$p_i P_{ij} = \Pr(a_i) \Pr(b_j | a_i) = \underbrace{\Pr(a_i, b_j)}_{=R_{ij}} = \Pr(b_j) \Pr(a_i | b_j) = q_j Q_{ij}$$

From this, since  $Q_{ij} = \frac{p_i}{q_j} P_{ij}$ , and since  $R_{ij} = q_j Q_{ij}$ ,  $(p_0, p_1) = (p, \bar{p})$ ,  $(q_0, q_1) = (q, \bar{q})$ ,

$$(Q_{ij}) = \begin{pmatrix} \frac{p}{q} P & \frac{p}{\bar{q}} \bar{P} \\ \frac{\bar{p}}{q} \bar{P} & \frac{\bar{p}}{\bar{q}} P \end{pmatrix}, \text{ and } (R_{ij}) = \begin{pmatrix} pP & p\bar{P} \\ \bar{p}\bar{P} & \bar{p}P \end{pmatrix} = \underline{\begin{pmatrix} p & 0 \\ 0 & \bar{p} \end{pmatrix} (P_{ij})}$$

Applying *ideal observer rule*, we take

$$\Delta(0) = \begin{cases} 0 & \text{if } pP > \bar{p}\bar{P} \\ 1 & \text{if } pP < \bar{p}\bar{P} \end{cases}, \text{ and } \Delta(1) = \begin{cases} 1 & \text{if } \bar{p}P > p\bar{P} \\ 0 & \text{if } \bar{p}P < p\bar{P} \end{cases}.$$

Example ( $\Gamma$  reliable, high input uncertainty)

Let  $p = 0.6$  and  $P = 0.7$ . Then

$$(R_{ij}) = \begin{pmatrix} p & 0 \\ 0 & \bar{p} \end{pmatrix} (P_{ij}) = \begin{pmatrix} 0.6 & 0 \\ 0 & 0.4 \end{pmatrix} \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix} = \begin{pmatrix} 0.42 & 0.18 \\ 0.12 & 0.28 \end{pmatrix}.$$

*Ideal observer rule:*  $\Delta(0) = 0$ , and,  $\Delta(1) = 1$ .

$$\Pr_C = \sum_j R_{j*j} = pP + \bar{p}P = 0.42 + 0.28 = 0.7 = P$$

Lecture 7: Using  $\Gamma$  Efficiently

Decision Rules

Definitions

**Example**Improving  
ReliabilityMajority Decoding  
(Repetition code)

Generalisation

Transmission Rate

Peek Into Linear  
Block Codes

Example

Epilogue

Example ( $\Gamma$  reliable, high input uncertainty)

Let  $p = 0.6$  and  $P = 0.7$ . Then

$$(R_{ij}) = \begin{pmatrix} p & 0 \\ 0 & \bar{p} \end{pmatrix} (P_{ij}) = \begin{pmatrix} 0.6 & 0 \\ 0 & 0.4 \end{pmatrix} \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix} = \begin{pmatrix} 0.42 & 0.18 \\ 0.12 & 0.28 \end{pmatrix}.$$

*Ideal observer rule:*  $\Delta(0) = 0$ , and,  $\Delta(1) = 1$ .

$$\text{Pr}_C = \sum_j R_{j*j} = pP + \bar{p}P = 0.42 + 0.28 = 0.7 = P$$

Example ( $\Gamma$  unreliable, high input uncertainty)

Let  $p = 0.6$  and  $P = 0.2$ . Then

$$(R_{ij}) = \begin{pmatrix} p & 0 \\ 0 & \bar{p} \end{pmatrix} (P_{ij}) = \begin{pmatrix} 0.6 & 0 \\ 0 & 0.4 \end{pmatrix} \begin{pmatrix} 0.2 & 0.8 \\ 0.8 & 0.2 \end{pmatrix} = \begin{pmatrix} 0.12 & 0.48 \\ 0.32 & 0.08 \end{pmatrix}.$$

*Ideal observer rule:*  $\Delta(0) = 1$ , and,  $\Delta(1) = 0$ .

$$\text{Pr}_C = \sum_j R_{j*j} = \bar{p}\bar{P} + p\bar{P} = 0.32 + 0.48 = 0.8 = \bar{P}$$

**Note:** The *closer*  $P$  is to  $\frac{1}{2}$ , the more often we *decide incorrectly* (how much greater than other elements in column vector is the maximum).

## Example (low input uncertainty)

Let  $p = 0.9$  and  $P = 0.8$ . Then

$$(R_{ij}) = \begin{pmatrix} p & 0 \\ 0 & \bar{p} \end{pmatrix} (P_{ij}) = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.1 \end{pmatrix} \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix} = \begin{pmatrix} 0.72 & 0.18 \\ 0.02 & 0.08 \end{pmatrix}.$$

*Ideal observer rule:*  $\Delta(0) = 0$ , and,  $\Delta(1) = 0$ . **Wait, what?**

$$\Pr_C = \sum_j R_{j*j} = pP + p\bar{P} = 0.72 + 0.18 = 0.9 = p$$

Lecture 7: Using  
Γ Efficiently

Decision Rules

Definitions

**Example**Improving  
ReliabilityMajority Decoding  
(Repetition code)

Generalisation

Transmission Rate

Peek Into Linear  
Block Codes

Example

Epilogue

## Example (low input uncertainty)

Let  $p = 0.9$  and  $P = 0.8$ . Then

$$(R_{ij}) = \begin{pmatrix} p & 0 \\ 0 & \bar{p} \end{pmatrix} (P_{ij}) = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.1 \end{pmatrix} \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix} = \begin{pmatrix} 0.72 & 0.18 \\ 0.02 & 0.08 \end{pmatrix}.$$

*Ideal observer rule:*  $\Delta(0) = 0$ , and,  $\Delta(1) = 0$ . **Wait, what?**

$$\Pr_C = \sum_j R_{j*j} = pP + p\bar{P} = 0.72 + 0.18 = 0.9 = p$$

Example ( $(p_i)$  unknown)

Let  $P = 0.7$ . Assume  $(p_i)$  is uniform. Then

$$(R_{ij}) = \begin{pmatrix} p & 0 \\ 0 & \bar{p} \end{pmatrix} (P_{ij}) = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix} \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix} = \begin{pmatrix} 0.35 & 0.15 \\ 0.15 & 0.35 \end{pmatrix}.$$

*Ideal observer rule:*  $\Delta(0) = 0$ , and,  $\Delta(1) = 1$ .

$$\Pr_C = \sum_j R_{j*j} = pP + \bar{p}P = 0.35 + 0.35 = 0.7 = P$$

**Note:** In the unlikely event that  $(p_i)$  is far from uniformity and we assume  $(p_i)$  uniform,  $\Delta$  will *decide incorrectly* often.

Decision Rules

Definitions

Example

Improving  
ReliabilityMajority Decoding  
(Repetition code)

Generalisation

Transmission Rate

Peek Into Linear  
Block Codes

Example

Epilogue

# Sending through $\Gamma$

With the *decision rule*, our model becomes

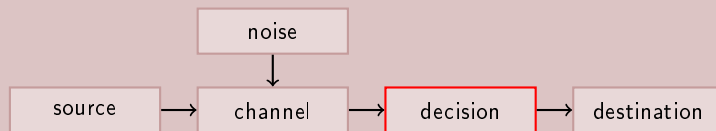


Figure: Communication System, with  $\Delta$

$\mathcal{A}$		$\mathcal{B}$		$\mathcal{A}$
0	$\rightarrow$	$\Gamma$	$\rightarrow$	0
1				1
$a_i$		$b_j$		$\Delta(b_j)$
				$= a_{j^*}$

## Lecture 7: Using $\Gamma$ Efficiently

Decision Rules

Definitions

Example

Improving  
ReliabilityMajority Decoding  
(Repetition code)

Generalisation

Transmission Rate

Peek Into Linear  
Block Codes

Example

Epilogue

# Improving Reliability: Repetition Code

Applying *repetition codes*, the model becomes

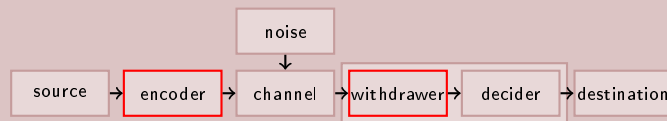
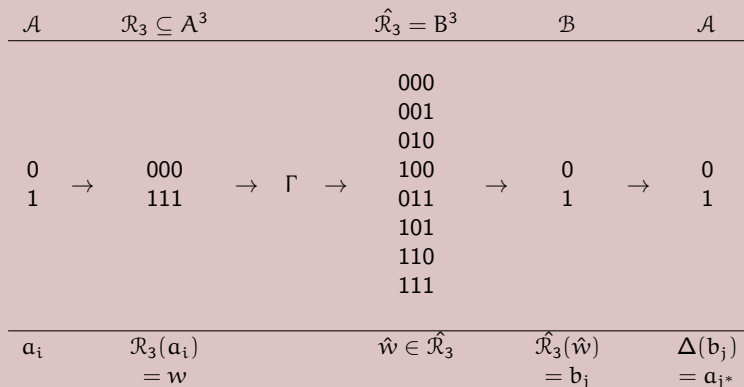


Figure: Communication System, using  $\mathcal{R}_n$



Example ( $\Gamma$  reliable, high input uncertainty)

**Recall:** For  $p = 0.6$  and  $P = 0.7$ . Then

$$(R_{ij}) = \begin{pmatrix} p & 0 \\ 0 & \bar{p} \end{pmatrix} (P_{ij}) = \begin{pmatrix} 0.6 & 0 \\ 0 & 0.4 \end{pmatrix} \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix} = \begin{pmatrix} 0.42 & 0.18 \\ 0.12 & 0.28 \end{pmatrix}.$$

*Ideal observer rule:*  $\Delta(0) = 0$ , and,  $\Delta(1) = 1$ , and  $\Pr_C = pP + \bar{p}P = P$

**Repetition:** Forward probability for  $\hat{\mathcal{R}}_3$  determined by

$$\begin{pmatrix} P^3 & P^2Q & P^2Q & P^2Q & PQ^2 & PQ^2 & PQ^2 & Q^3 \\ Q^3 & PQ^2 & PQ^2 & PQ^2 & P^2Q & P^2Q & P^2Q & P^3 \end{pmatrix}$$

thus  $(P_{ij_{\mathcal{R}_3}})$  given by (sum quadrants)

$$(P_{ij_{\mathcal{R}_3}}) = \begin{pmatrix} P^3 + 3P^2Q & Q^3 + 3PQ^2 \\ Q^3 + 3Q^2P & P^3 + 3QP^2 \end{pmatrix} = \begin{pmatrix} 0.784 & 0.216 \\ 0.216 & 0.784 \end{pmatrix},$$

$$(R_{ij_{\mathcal{R}_3}}) = \begin{pmatrix} p & 0 \\ 0 & \bar{p} \end{pmatrix} (P_{ij_{\mathcal{R}_3}}) = \begin{pmatrix} 0.4704 & 0.1296 \\ 0.0864 & 0.3136 \end{pmatrix}.$$

$$\Pr_C = \sum_j R_{j^*j} = 0.4704 + 0.3136 = 0.784,$$

an improvement.



# Generalising The Idea

We do not *need* to *repeat*; in fact, we might be able to map  $a_i$  to  $w_i$  more efficiently. More *cleverly*. In general, the model is

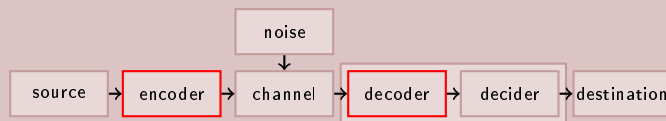


Figure: Communication System, using  $\mathcal{C}$

$\mathcal{A}$	$\mathcal{C} \subseteq A^n$	$\hat{\mathcal{C}} = B^n$	$\mathcal{B}$	$\mathcal{A}$
$a_1$	$w_1$	$00 \dots 0$	$b_1$	$a_1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$a_r$	$w_r$	$11 \dots 1$	$b_r$	$a_r$
<hr/>				
$a_i$	$\mathcal{C}(a_i)$ $= w$	$\hat{w} \in \hat{\mathcal{C}}$	$\hat{\mathcal{C}}(\hat{w})$ $= b_j$	$\Delta(b_j)$ $= a_{j^*}$

# How Fast is our Code?

$\mathcal{C}$  is an  $r$ -ary code of *length*  $n$ . If  $|\mathcal{C}| = r^k$ , then  $\mathcal{C}$  can encode  $k$ -th extension,  $\mathcal{A}^k$ , of  $\mathcal{A}$  (since  $|\mathcal{A}| = r^k$ ).

$$R = \frac{k}{n} = \frac{\log_r |\mathcal{C}|}{n}$$

We have  $0 \leq R \leq 1$ . Recall that

$$C = \max_{\mathcal{A}} I(\mathcal{A}, \mathcal{B})$$

and that

$$I(\mathcal{A}, \mathcal{B}) \geq 0.$$

We wish to find *clever* encoding which get  $R$  close to  $C$ .

## Lecture 7: Using $\Gamma$ Efficiently

Decision Rules

Definitions

Example

Improving  
Reliability

Majority Decoding  
(Repetition code)

Generalisation

**Transmission Rate**

Peek Into Linear  
Block Codes

Example

Epilogue

## Simple Example

Consider a source,  $\mathcal{A}$ , which produces a bit stream. We cut the emerging input stream into segments of 5 bits, and realise that the only patterns that emerge are

$$\begin{Bmatrix} 00000 \\ 01000 \\ 01010 \end{Bmatrix}$$

If we send these words across a binary symmetric channel  $\Gamma$ , some bits may be flipped. If we expect  $P = 0.8$ , and on the receiving end, 00100 appears, we *know* it must be 00000 with 1 bit flipped. But if we see 01000, we are in doubt; is it 01000, or one of 00000 and 01010 with 1 error? The “distance” between some words is 1.

If we map the above words to

$$\begin{Bmatrix} 10010 \\ 01001 \\ 00100 \end{Bmatrix},$$

We can *always* correct up to 1 error, since the “distance” between the words is 3.

### Lecture 7: Using $\Gamma$ Efficiently

Decision Rules

Definitions

Example

Improving

Reliability

Majority Decoding  
(Repetition code)

Generalisation

Transmission Rate

Peek Into Linear  
Block Codes

**Example**

Epilogue

## Summary

We now have a clear picture of noisy channels, and what can be done to suppress their malicious nature.

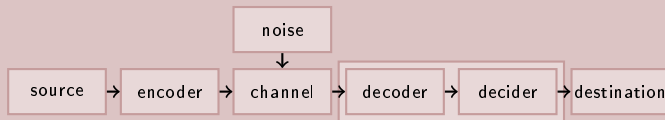


Figure: Communication System

Some approaches are not so clever ( $\mathcal{R}_n$ );  $R \rightarrow 0$  as  $n \rightarrow \infty$ . There are *linear codes* which do *better*. There are other codes that do *even better*.

**Shannon's Fundamental Theorem** says *how good* a code can ever hope to get:  $R \rightarrow C$  as  $n \rightarrow \infty$ .

We now have a framework within which we can *prove theoretical limit of error correction*. The coast is clear. Next time, we will see the proof.

# Shannon's Fundamental Theorem

The theorem referred to as

- Shannon's Fundamental Theorem,
- Shannon's Theorem,
- The Fundamental Theorem of Information Theory,
- The Channel Coding Theorem,

or a variation of these, *for BSC*, is stated as follows.

## Theorem

*Let  $\Gamma$  be a BSC with  $P > \frac{1}{2}$ , in which case  $\Gamma$  has capacity  $C = 1 - H(P) > 0$ , and let  $\delta, \epsilon > 0$  (small).*

# Shannon's Fundamental Theorem

The theorem referred to as

- Shannon's Fundamental Theorem,
- Shannon's Theorem,
- The Fundamental Theorem of Information Theory,
- The Channel Coding Theorem,

or a variation of these, *for BSC*, is stated as follows.

## Theorem

*Let  $\Gamma$  be a BSC with  $P > \frac{1}{2}$ , in which case  $\Gamma$  has capacity  $C = 1 - H(P) > 0$ , and let  $\delta, \epsilon > 0$  (small).*

*Then for all  $n$  sufficiently big, there is a code  $\mathcal{C} \subseteq \mathbb{Z}_2^n$  of rate  $R$ ;  $C - \epsilon \leq R < C$ , such that nearest neighbour decoding gives  $\Pr_E < \delta$ .*

**Meaning:** Let the BSC  $\Gamma$  be given. Pick  $\epsilon, \delta$  as small as you like. Then, for some big  $n$ , there is a code  $\mathcal{C}$  with rate  $R$  no more than  $\epsilon$  below  $C$ , which corrects errors as well as desired:  $\Pr_E < \delta$ .

## Informal

Consider  $\mathbb{Z}_2^n$  a “space” of words. Some words  $\mathbf{u}, \mathbf{v} \in \mathbb{Z}_2^n$  are close to each other, that is,  $d_H(\mathbf{u}, \mathbf{v})$  small), others are far away.

Recall that

$$R = \frac{\log_2 |\mathcal{C}|}{n}.$$

To make  $R$  “stay the same” as  $n$  increases and  $\mathcal{C}$  changes, we let  $|\mathcal{C}| = 2^{nR}$ . Then

$$R = \frac{\log_2 2^{nR}}{n} = \frac{nR}{n} = R.$$

As  $n$  increases,  $\mathcal{C}$ 's portion of  $\mathbb{Z}_2^n$  decreases (that is,  $\frac{|\mathcal{C}|}{|\mathbb{Z}_2^n|} = \frac{2^{nR}}{2^n}$ ).

Statement
Theorem
Explanation
Proof
Overview
Omissions
Epilogue

## Informal

Consider  $\mathbb{Z}_2^n$  a “space” of words. Some words  $\mathbf{u}, \mathbf{v} \in \mathbb{Z}_2^n$  are close to each other, that is,  $d_H(\mathbf{u}, \mathbf{v})$  small), others are far away.

Recall that

$$R = \frac{\log_2 |\mathcal{C}|}{n}.$$

To make  $R$  “stay the same” as  $n$  increases and  $\mathcal{C}$  changes, we let  $|\mathcal{C}| = 2^{nR}$ . Then

$$R = \frac{\log_2 2^{nR}}{n} = \frac{nR}{n} = R.$$

As  $n$  increases,  $\mathcal{C}$ 's portion of  $\mathbb{Z}_2^n$  decreases (that is,  $\frac{|\mathcal{C}|}{|\mathbb{Z}_2^n|} = \frac{2^{nR}}{2^n}$ ).

## Example

Let  $R = 0.5$ . Then

$$\frac{2^{R \cdot 2}}{2^2} = 0.5 \quad \frac{2^{R \cdot 4}}{2^4} = 0.25 \quad \frac{2^{R \cdot 6}}{2^6} = 0.125 \quad \frac{2^{R \cdot 8}}{2^8} = 0.0625$$

If you pick  $\mathcal{C}$  such that words in  $\mathcal{C}$  are scattered evenly around in  $\mathbb{Z}_2^n$ , then distance between words in  $\mathcal{C}$  increases as  $n$  increases. Thus  $\Pr_E$  decreases as  $n$  increases.



# Proof Outline

**Spoiler:** The proof is a (long) series of inequalities, where we show that  $\Pr_E$ , in the end, is less than a term  $T$  dependent of  $n$  s.t.  $T \rightarrow 0$  as  $n \rightarrow \infty$ .

## Lecture 8: Shannon's Fundamental Theorem

Statement  
Theorem  
Explanation  
Proof  
**Overview**  
Omissions  
Epilogue

# Proof Outline

**Spoiler:** The proof is a (long) series of inequalities, where we show that  $\Pr_E$ , in the end, is less than a term  $T$  dependent of  $n$  s.t.  $T \rightarrow 0$  as  $n \rightarrow \infty$ .

## Proof

Pick  $R < C$  as desired. This gives  $\epsilon$ , since  $C - \epsilon \leq R < C$ . You can obtain a block code  $\mathcal{C}$  with word length  $n$  and rate  $R$  since  $R = \frac{\log_2 2^{nR}}{n} = \frac{\log_2 |\mathcal{C}|}{n}$ . Just pick  $2^{nR}$  words from  $\mathbb{Z}_2^n$  **randomly**.

# Proof Outline

**Spoiler:** The proof is a (long) series of inequalities, where we show that  $\Pr_E$ , in the end, is less than a term  $T$  dependent of  $n$  s.t.  $T \rightarrow 0$  as  $n \rightarrow \infty$ .

## Proof

Pick  $R < C$  as desired. This gives  $\epsilon$ , since  $C - \epsilon \leq R < C$ . You can obtain a block code  $\mathcal{C}$  with word length  $n$  and rate  $R$  since  $R = \frac{\log_2 2^{nR}}{n} = \frac{\log_2 |\mathcal{C}|}{n}$ . Just pick  $2^{nR}$  words from  $\mathbb{Z}_2^n$  **randomly**.

Pick a small  $\delta > 0$ . It remains to show that regardless of which  $\delta$  you choose,  $\Pr_E < \delta$  for  $n$  big enough.

# Proof Outline

**Spoiler:** The proof is a (long) series of inequalities, where we show that  $\Pr_E$ , in the end, is less than a term  $T$  dependent of  $n$  s.t.  $T \rightarrow 0$  as  $n \rightarrow \infty$ .

## Proof

Pick  $R < C$  as desired. This gives  $\epsilon$ , since  $C - \epsilon \leq R < C$ . You can obtain a block code  $\mathcal{C}$  with word length  $n$  and rate  $R$  since  $R = \frac{\log_2 2^{nR}}{n} = \frac{\log_2 |\mathcal{C}|}{n}$ . Just pick  $2^{nR}$  words from  $\mathbb{Z}_2^n$  **randomly**.

Pick a small  $\delta > 0$ . It remains to show that regardless of which  $\delta$  you choose,  $\Pr_E < \delta$  for  $n$  big enough.

Let  $Q = \bar{P} = 1 - P$ . Each symbol in  $\mathbf{u} \in \mathcal{C}$  has probability  $Q$  of being “flipped” when  $\mathbf{u}$  is sent over  $\Gamma$ . There are  $n$  symbols in  $\mathbf{u}$ . We thus expect  $Qn$  symbols to be flipped in  $\mathbf{u}$ . In fact, by *the law of large numbers*, for  $n \rightarrow \infty$ ,  $Qn$  is the nr. of symbols flipped in  $\mathbf{u}$  upon transmission over  $\Gamma$ . Let  $\mathbf{v} \in \mathcal{C}$  be the received word. Thus  $d_H(\mathbf{u}, \mathbf{v}) \approx Qn$  (“=” for  $n \rightarrow \infty$ ).

Cont.

## Proof (continued)

**Established:**  $d_H(\mathbf{u}, \mathbf{v}) \approx Qn$ ; “=” for  $n \rightarrow \infty$ .  $\mathcal{C}$  is random.

$\Delta(\mathbf{v})$  is the  $\hat{\mathbf{u}}$  in  $\mathcal{C}$  closest to  $\mathbf{v}$ . If incorrect decision, then  $\hat{\mathbf{u}} \neq \mathbf{u}$ . In fact,  $d_H(\hat{\mathbf{u}}, \mathbf{v}) \leq d_H(\mathbf{u}, \mathbf{v})$ .

Lecture 8:  
Shannon's  
Fundamental  
Theorem

Statement  
Theorem  
Explanation  
Proof  
Overview  
Omissions  
Epilogue

## Proof (continued)

**Established:**  $d_H(\mathbf{u}, \mathbf{v}) \approx Qn$ ; “=” for  $n \rightarrow \infty$ .  $\mathcal{C}$  is random.

$\Delta(\mathbf{v})$  is the  $\hat{\mathbf{u}}$  in  $\mathcal{C}$  closest to  $\mathbf{v}$ . If incorrect decision, then  $\hat{\mathbf{u}} \neq \mathbf{u}$ . In fact,  $d_H(\hat{\mathbf{u}}, \mathbf{v}) \leq d_H(\mathbf{u}, \mathbf{v})$ . Probability of incorrectly decoding  $\mathbf{u}$  to  $\hat{\mathbf{u}}$  no greater than probability of  $\hat{\mathbf{u}}$  existing ( $\mathcal{C}$  random). Thus

$$\begin{aligned} \Pr_E &\leq \sum_{\hat{\mathbf{u}} \neq \mathbf{u}} \Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq d_H(\mathbf{u}, \mathbf{v})) \\ &\approx \sum_{\hat{\mathbf{u}} \neq \mathbf{u}} \Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn) \\ &= (|\mathcal{C}| - 1) \Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn) \\ &< 2^{nR} \Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn) \end{aligned}$$

(terms in the sum equal, since  $\mathcal{C}$  is random).

## Proof (continued)

**Established:**  $d_H(\mathbf{u}, \mathbf{v}) \approx Qn$ ; “=” for  $n \rightarrow \infty$ .  $\mathcal{C}$  is random.

$\Delta(\mathbf{v})$  is the  $\hat{\mathbf{u}}$  in  $\mathcal{C}$  closest to  $\mathbf{v}$ . If incorrect decision, then  $\hat{\mathbf{u}} \neq \mathbf{u}$ . In fact,  $d_H(\hat{\mathbf{u}}, \mathbf{v}) \leq d_H(\mathbf{u}, \mathbf{v})$ . Probability of incorrectly decoding  $\mathbf{u}$  to  $\hat{\mathbf{u}}$  no greater than probability of  $\hat{\mathbf{u}}$  existing ( $\mathcal{C}$  random). Thus

$$\begin{aligned} \Pr_E &\leq \sum_{\hat{\mathbf{u}} \neq \mathbf{u}} \Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq d_H(\mathbf{u}, \mathbf{v})) \\ &\approx \sum_{\hat{\mathbf{u}} \neq \mathbf{u}} \Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn) \\ &= (|\mathcal{C}| - 1) \Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn) \\ &< 2^{nR} \Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn) \end{aligned}$$

(terms in the sum equal, since  $\mathcal{C}$  is random).  $\Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn)$  equals portion of words in  $\mathbb{Z}_2^n$  satisfying  $d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn$ .

## Proof (continued)

**Established:**  $d_H(\mathbf{u}, \mathbf{v}) \approx Qn$ ; “=” for  $n \rightarrow \infty$ .  $\mathcal{C}$  is random.

$\Delta(\mathbf{v})$  is the  $\hat{\mathbf{u}}$  in  $\mathcal{C}$  closest to  $\mathbf{v}$ . If incorrect decision, then  $\hat{\mathbf{u}} \neq \mathbf{u}$ . In fact,  $d_H(\hat{\mathbf{u}}, \mathbf{v}) \leq d_H(\mathbf{u}, \mathbf{v})$ . Probability of incorrectly decoding  $\mathbf{u}$  to  $\hat{\mathbf{u}}$  no greater than probability of  $\hat{\mathbf{u}}$  existing ( $\mathcal{C}$  random). Thus

$$\begin{aligned} \Pr_E &\leq \sum_{\hat{\mathbf{u}} \neq \mathbf{u}} \Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq d_H(\mathbf{u}, \mathbf{v})) \\ &\approx \sum_{\hat{\mathbf{u}} \neq \mathbf{u}} \Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn) \\ &= (|\mathcal{C}| - 1) \Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn) \\ &< 2^{nR} \Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn) \end{aligned}$$

(terms in the sum equal, since  $\mathcal{C}$  is random).  $\Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn)$  equals portion of words in  $\mathbb{Z}_2^n$  satisfying  $d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn$ . For any  $\mathbf{v}$  and  $i \in \mathbb{Z}$ , the nr. of  $\hat{\mathbf{u}}$  satisfying  $d_H(\hat{\mathbf{u}}, \mathbf{v}) = i$  is  $\binom{n}{i}$  since  $\mathbf{v}, \hat{\mathbf{u}}$  are binary strings.



## Proof (continued)

**Established:**  $d_H(\mathbf{u}, \mathbf{v}) \approx Qn$ ; “=” for  $n \rightarrow \infty$ .  $\mathcal{C}$  is random.

$\Delta(\mathbf{v})$  is the  $\hat{\mathbf{u}}$  in  $\mathcal{C}$  closest to  $\mathbf{v}$ . If incorrect decision, then  $\hat{\mathbf{u}} \neq \mathbf{u}$ . In fact,  $d_H(\hat{\mathbf{u}}, \mathbf{v}) \leq d_H(\mathbf{u}, \mathbf{v})$ . Probability of incorrectly decoding  $\mathbf{u}$  to  $\hat{\mathbf{u}}$  no greater than probability of  $\hat{\mathbf{u}}$  existing ( $\mathcal{C}$  random). Thus

$$\begin{aligned} \Pr_E &\leq \sum_{\hat{\mathbf{u}} \neq \mathbf{u}} \Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq d_H(\mathbf{u}, \mathbf{v})) \\ &\approx \sum_{\hat{\mathbf{u}} \neq \mathbf{u}} \Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn) \\ &= (|\mathcal{C}| - 1) \Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn) \\ &< 2^{nR} \Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn) \end{aligned}$$

(terms in the sum equal, since  $\mathcal{C}$  is random).  $\Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn)$  equals *portion* of words in  $\mathbb{Z}_2^n$  satisfying  $d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn$ . For any  $\mathbf{v}$  and  $i \in \mathbb{Z}$ , the nr. of  $\hat{\mathbf{u}}$  satisfying  $d_H(\hat{\mathbf{u}}, \mathbf{v}) = i$  is  $\binom{n}{i}$  since  $\mathbf{v}, \hat{\mathbf{u}}$  are binary strings. As such, the nr. of words in  $\mathbb{Z}_2^n$  satisfying  $d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn$  is  $\sum_{i=0}^{nQ} \binom{n}{i}$ . So

$$\Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn) = \frac{1}{2^n} \sum_{i=0}^{nQ} \binom{n}{i}$$

Cont.

## Detour

We need to simplify  $\frac{1}{2^n} \sum_{i=0}^{nQ} \binom{n}{i}$ . Fortunately, we can!

## Exercise

When  $\lambda + \mu = 1$  and  $0 \leq \lambda \leq \frac{1}{2}$ , then

$$1 \geq \sum_{i=0}^{\lfloor \lambda n \rfloor} \binom{n}{i} \lambda^i \mu^{n-i} \geq \sum_{i=0}^{\lfloor \lambda n \rfloor} \binom{n}{i} \lambda^{\lambda n} \mu^{\mu n}$$

and thus,

$$\sum_{i=0}^{\lfloor \lambda n \rfloor} \binom{n}{i} \leq 2^{nH(\lambda)}.$$

Statement

Theorem

Explanation

Proof

Overview

Omissions

Epilogue

## Solution

Recall the binomial theorem.

$$(x + y)^n = \sum_{i=0}^n \binom{n}{i} x^i y^{n-i} \quad \text{ex: } x = y = 1; 2^n = \dots$$

### Lecture 8: Shannon's Fundamental Theorem

Statement  
Theorem  
Explanation  
Proof  
**Overview**  
Omissions  
Epilogue

## Solution

Recall the binomial theorem.

$$(x + y)^n = \sum_{i=0}^n \binom{n}{i} x^i y^{n-i} \quad \text{ex: } x = y = 1; 2^n = \dots$$

Also, since  $\frac{\lambda}{\mu} \leq 1$  and  $i \leq \lambda n$ , we have

$$\begin{aligned} \lambda^i \mu^{n-i} &= \lambda^i \mu^{-i} \mu^n = \lambda^i \frac{1}{\mu^i} \mu^n = \left(\frac{\lambda}{\mu}\right)^i \mu^n \geq \left(\frac{\lambda}{\mu}\right)^{\lambda n} \mu^n \\ &= \lambda^{\lambda n} \mu^{n-\lambda n} = \lambda^{\lambda n} \mu^{n(1-\lambda)} = \lambda^{\lambda n} \mu^{\mu n} \end{aligned}$$

# Lecture 8: Shannon's Fundamental Theorem

Statement  
Theorem  
Explanation  
Proof  
**Overview**  
Omissions  
Epilogue

## Solution

Recall the binomial theorem.

$$(x + y)^n = \sum_{i=0}^n \binom{n}{i} x^i y^{n-i} \quad \text{ex: } x = y = 1; 2^n = \dots$$

Also, since  $\frac{\lambda}{\mu} \leq 1$  and  $i \leq \lambda n$ , we have

$$\begin{aligned} \lambda^i \mu^{n-i} &= \lambda^i \mu^{-i} \mu^n = \lambda^i \frac{1}{\mu^i} \mu^n = \left(\frac{\lambda}{\mu}\right)^i \mu^n \geq \left(\frac{\lambda}{\mu}\right)^{\lambda n} \mu^n \\ &= \lambda^{\lambda n} \mu^{n-\lambda n} = \lambda^{\lambda n} \mu^{n(1-\lambda)} = \lambda^{\lambda n} \mu^{\mu n} \end{aligned}$$

By insertion into binomial theorem ( $x = \lambda$ ,  $y = \mu$ ),

$$(\lambda + \mu)^n = \sum_{i=0}^n \binom{n}{i} \lambda^i \mu^{n-i} \geq \sum_{i=0}^{\lfloor \lambda n \rfloor} \binom{n}{i} \lambda^i \mu^{n-i} \geq \sum_{i=0}^{\lfloor \lambda n \rfloor} \binom{n}{i} \lambda^{\lambda n} \mu^{\mu n}$$

Statement  
Theorem  
Explanation

Proof

Overview  
Omissions

Epilogue

## Solution

Recall the binomial theorem.

$$(x + y)^n = \sum_{i=0}^n \binom{n}{i} x^i y^{n-i} \quad \text{ex: } x = y = 1; 2^n = \dots$$

Also, since  $\frac{\lambda}{\mu} \leq 1$  and  $i \leq \lambda n$ , we have

$$\begin{aligned} \lambda^i \mu^{n-i} &= \lambda^i \mu^{-i} \mu^n = \lambda^i \frac{1}{\mu^i} \mu^n = \left(\frac{\lambda}{\mu}\right)^i \mu^n \geq \left(\frac{\lambda}{\mu}\right)^{\lambda n} \mu^n \\ &= \lambda^{\lambda n} \mu^{n-\lambda n} = \lambda^{\lambda n} \mu^{n(1-\lambda)} = \lambda^{\lambda n} \mu^{\mu n} \end{aligned}$$

By insertion into binomial theorem ( $x = \lambda$ ,  $y = \mu$ ),

$$(\lambda + \mu)^n = \sum_{i=0}^n \binom{n}{i} \lambda^i \mu^{n-i} \geq \sum_{i=0}^{\lfloor \lambda n \rfloor} \binom{n}{i} \lambda^i \mu^{n-i} \geq \sum_{i=0}^{\lfloor \lambda n \rfloor} \binom{n}{i} \lambda^{\lambda n} \mu^{\mu n}$$

Divide by the constant  $\lambda^{\lambda n} \mu^{\mu n}$  to get

$$\sum_{i=0}^{\lfloor \lambda n \rfloor} \binom{n}{i} \leq \frac{1}{\lambda^{\lambda n} \mu^{\mu n}} = \lambda^{-\lambda n} \mu^{-\mu n} = (\lambda^{-\lambda} \mu^{-\mu})^n.$$

Cont.

Statement
Theorem
Explanation
Proof
Overview
Omissions
Epilogue

## Solution (continued)

**Established:**  $\sum_{i=0}^{\lfloor \lambda n \rfloor} \binom{n}{i} \leq (\lambda^{-\lambda} \mu^{-\mu})^n.$

**Aim:**  $\sum_{i=0}^{\lfloor \lambda n \rfloor} \binom{n}{i} \leq 2^{nH(\lambda)}$

# Lecture 8: Shannon's Fundamental Theorem

Statement
Theorem
Explanation
Proof
<b>Overview</b>
Omissions
Epilogue

## Solution (continued)

**Established:**  $\sum_{i=0}^{\lfloor \lambda n \rfloor} \binom{n}{i} \leq (\lambda^{-\lambda} \mu^{-\mu})^n.$

**Aim:**  $\sum_{i=0}^{\lfloor \lambda n \rfloor} \binom{n}{i} \leq 2^{nH(\lambda)}$

Since  $\log_b c^p = p \log_b c$ , taking  $\log_2$  yields

$$\begin{aligned} \log_2 \sum_{i=0}^{\lfloor \lambda n \rfloor} \binom{n}{i} &\leq n \log_2 (\lambda^{-\lambda} \mu^{-\mu}) = n (\log_2 \lambda^{-\lambda} + \log_2 \mu^{-\mu}) \\ &= n (-\lambda \log_2 \lambda - \mu \log_2 \mu) = n (-\lambda \log_2 \lambda - (1 - \lambda) \log_2 (1 - \lambda)) \\ &= n H_2(\lambda). \end{aligned}$$

Thus,

$$\sum_{i=0}^{\lfloor \lambda n \rfloor} \binom{n}{i} \leq 2^{nH(\lambda)},$$

as desired.

Statement  
Theorem  
Explanation  
Proof  
Overview  
Omissions  
Epilogue



## Back On Track...

## Proof (continued).

**Established:**

- $\Pr_E < 2^{nR} \Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn)$
- $\Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn) = \frac{1}{2^n} \sum_{i=0}^{nQ} \binom{n}{i}$
- $\sum_{i=0}^{\lfloor n\lambda \rfloor} \binom{n}{i} \leq 2^{nH(\lambda)}$

**Lecture 8:  
Shannon's  
Fundamental  
Theorem**

Statement

Theorem

Explanation

Proof

**Overview**

Omissions

Epilogue

## Back On Track...

Lecture 8:  
Shannon's  
Fundamental  
Theorem

## Proof (continued).

**Established:**

- $\Pr_E < 2^{nR} \Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn)$
- $\Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn) = \frac{1}{2^n} \sum_{i=0}^{nQ} \binom{n}{i}$
- $\sum_{i=0}^{\lfloor n\lambda \rfloor} \binom{n}{i} \leq 2^{nH(\lambda)}$

Let  $\lambda = Q$ . Then by insertion, we get

$$\begin{aligned}
 \Pr_E &< 2^{nR} \Pr(d_H(\hat{\mathbf{u}}, \mathbf{u}) \leq Qn) = (2^n)^R \frac{1}{2^n} \sum_{i=0}^{nQ} \binom{n}{i} \\
 &\leq \frac{(2^n)^R}{2^n} (2^{nH(Q)}) = 2^{n(R-1)+n(H(Q))} \\
 &= 2^{n(R-1)+H(Q)} = 2^{n(R-(1-H(Q)))} = 2^{n(R-(1-H(P)))} \\
 &= 2^{n(R-C)}.
 \end{aligned}$$

Statement

Theorem

Explanation

Proof

Overview

Omissions

Epilogue

## Main Results

We have now seen all the main results in this field!

**Source Coding:** We can get  $L(\mathcal{C})$  for  $\mathcal{C}$  arbitrarily close to  $H(\mathcal{S})$  by applying source extension.

- Input distribution known: *Source Extension* and Huffman.
- Input distribution unknown: Lempel-Ziv.

**Channel Coding:** We *can* get  $R$  for  $\mathcal{C}$  to approach  $C$  with  $\Pr_E \rightarrow \infty$ . *And* there is no  $C' > C$  that satisfies this condition.

- **How?**

**Combined Source-Channel Coding:** (Theoretically) just as effective as separating them.

- In practice, combining the two, with the channel coding schemes discovered to date, gets us closer to theoretical bounds.

Statement  
Theorem  
Explanation  
Proof  
Overview  
Omissions  
Epilogue