Fundamentals of Wireless Communication

Information Theory From First Principles

Edward Kwao

Hanbat National University

26th January, 2022

Contents

❑Introduction to Information Theory

- Discrete Memoryless Channels
- Entropy, Conditional entropy and Mutual Information

❑Noisy Channel Coding Theorem

- Upper bound on reliable communication
- The capacity of Binary symmetric and Binary erasure Channels
- Analog Memoryless Channels and AWGN Capacity
- Capacity of Fast-Fading Channel

Introduction to Information Theory

- Information theory is the probabilistic treatment of the concepts, parameters and rules governing the transmission of messages through communication systems.

Key idea is to: increase *speed*, *capacity* and *reliability*

Minimum number of bps to represent information,

Entropy rate, $\boldsymbol{H}$

↑

Speed

Maximum number of bps that can be reliably communicated in the face of noise,

System capacity, **C**

↑

Capacity

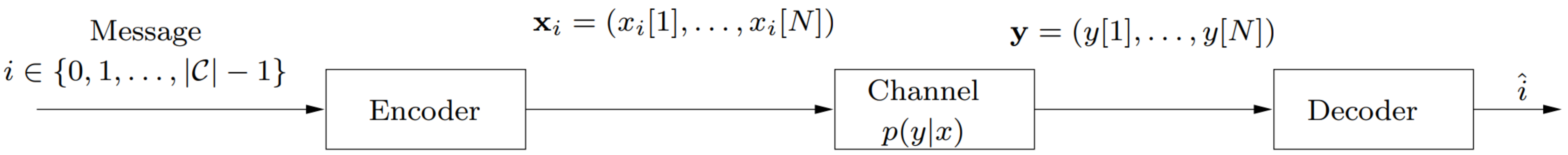
Reliable communication of information from source in the face of noise is possible if and only if:

$$\boldsymbol{H} < \boldsymbol{C}$$

↑

Reliability

Discrete Memoryless Channels



- $P(y|x)$, *the* channel transition probability, indicates each possible input-output paths.

Sender:
- One out of several equally likely messages.
- Use a codebook, $\mathcal{C}$ of block length N and size $|\mathcal{C}|$

Receiver:
- Receives vector y.
- Generates an estimate, $\hat{\imath}$ of the correct message.

Key performance measures

$$R = \frac{1}{N} \log |\mathcal{C}|$$
$$p_e = \mathbb{P}\left\{\hat{i} \neq i\right\}$$

Where $\mathcal{C} = \left\{\boldsymbol{x_1}, \ldots, \boldsymbol{x}_{|\mathcal{C}|}\right\}$, $\boldsymbol{x_i} =$ *codewords*, $p_e =$ *ML error probability*

Entropy, Conditional Entropy and Mutual Information

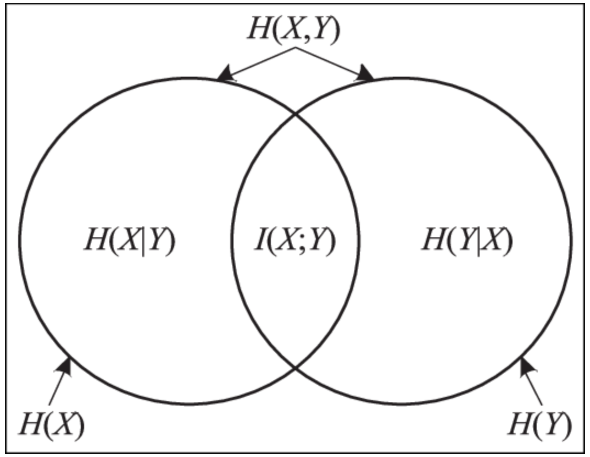
- Let x be a discrete random variable taking on values in X and with a probability mass function P_x

Entropy of x, $H(x) := \sum_{i \in X} P_x(i) \log\left(\frac{1}{P_x(i)}\right)$……………………(1)

The amount of uncertainty associated with x is $H(x)$

For two discrete-random variables x and y, the entropy of x conditioned on $y =$ j is:

$$H(x|y = j) := \sum_{i \in X} P_{x|y}(i|j) \log\left(\frac{1}{P_{x|y}(i|j)}\right) \dots\dots\dots\dots\dots\dots\dots (2)$$



Total uncertainty in x and y:

$$H(x,y) = H(x) + H(y|x) = H(y) + H(x|y) \ldots\ldots.(3)$$

If x and y are independent, $\boldsymbol{H(x) = H(x|y)}$, $\quad hence\ H(x,y) = H(x) + H(y) \ldots\ldots\ldots\ldots.\ldots(4)$

Conditioning reduces uncertainty: $\boldsymbol{H(x) \geq H(x|y)} \Rightarrow (3) \leq (4)$

With equality and independence of x and y: $\boldsymbol{H(x) - H(x|y)}$

Hence mutual information, $I(x;y) \coloneqq H(y) - H(y|x) = H(x) - H(x|y) \ldots\ldots\ldots\ldots\ldots\ldots\ldots..(5)$

Noisy Channel Coding Theorem

- From (5), to decode the transmitted message correctly with high probability, the conditional entropy, $H(x|y)$ must be close to zero.

- $H(x)$ is then = the number of bits conveyed ie. $\log|\mathcal{C}|$ = NR, *where* $R = data\ rate$

For reliable communication, $H(x|y) \approx 0$, which implies

$$R \approx \frac{1}{N} I(x;y) \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (6)$$

- $I(x;y)$ depends on the distribution of the random input $\boldsymbol{x}$ and this distribution is in turn a function of the code $\mathcal{C}$**.** By optimizing over all input distribution instead of all codes,

- **Upper bound on reliable rate of communication**: $\bar{C} \coloneqq \max_{P_x} \frac{1}{N} I(x;y)$

- If the inputs are made independent over time,

$$\bar{C} = \frac{1}{N}\sum_{m=1}^{N} \max_{P_{x[m]}} I(x[m]; y[m]) = \max_{P_{x[1]}} I(x[1]; y[1]) \ldots\ldots\ldots\ldots\ldots\ldots.. (7)$$

- To achieve $\bar{C}$, one must find a code whose $\frac{I(x;y)}{N}$ is close to $\bar{C}$ to satisfy (6)

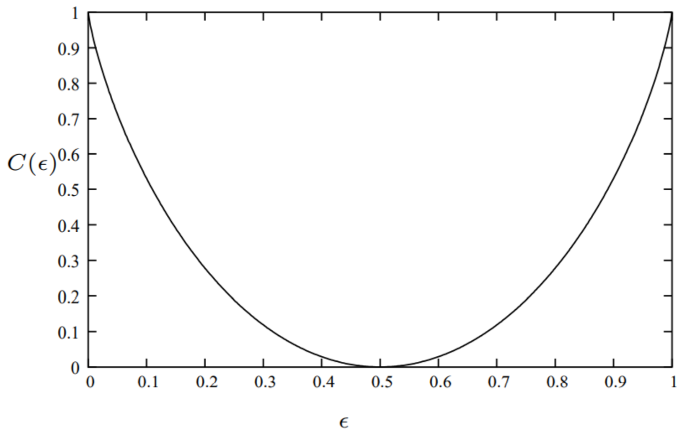
- These codes exist $\boldsymbol{if\ the\ block\ length\ N\ is\ chosen\ sufficiently\ large}$.

The capacity of Binary symmetric and Binary erasure Channels

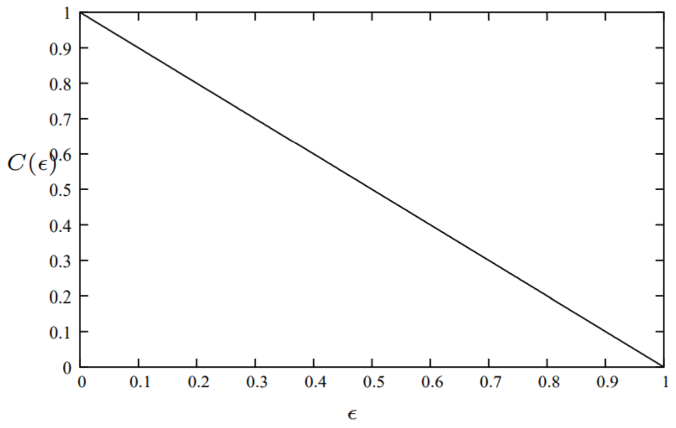
- The optimal input distribution for the BSC is uniform because of the symmetry.
- Similar symmetry exists in the BEC.

$$C = \max_{P_x} H(y) - H(y|x)$$
$$= \max_{P_x} H(y) - H(\epsilon)$$
$$1 - H(\epsilon)\ bits\ per\ channel\ use$$

$$C = 1 - \epsilon\ bits\ per\ channel\ use$$



(a)



(a)

Analog Memoryless Channels and AWGN Capacity

- For continuous random variable x with $pdf\ f_x$, differential entropy of x:

$$h(x) \coloneqq \int_{-\infty}^{\infty} f_x(u) log\left(\frac{1}{f_x(u)}\right) du \ldots\ldots\ldots..(8)$$

- The conditional differential entropy of x given y is:

$$h(x|y) \coloneqq \int_{-\infty}^{\infty} f_{x,y}(u,v) log\left(\frac{1}{f_{x|y}(u|v)}\right) dudv \ldots\ldots\ldots..(9)$$

The mutual information is again defined as:

$$I(x;y) \coloneqq h(x) - h(x|y)$$

Capacity: $\bar{C} \coloneqq \max_{f_x} I(x;y)$

- Every codeword, x_n in the codebook must satisfy:

$$\frac{1}{N}\sum_{m=1}^{N} c(x_n[m]) \leq A,$$

where $c: X \rightarrow \Re$ is the cost constraint function on the transmitted codewords for the channel

- The minimum rate of reliable communication subject to this constraint on the codewords is:

$$C = \max_{f_x: E[c(x)] \leq A} I(x; y) \ldots\ldots\ldots\ldots\ldots\ldots\ldots. (10)$$

- For real power-constrained AWGN channel: $y = x + w$, with cost function $c(x) = x^2$,

The differential entropy of w, $h(w) = \frac{1}{2} log(2\pi e \sigma^2)$

- $I(x;y) = h(y) - h(y|x) = h(y) - \frac{1}{2} log(2\pi e \sigma^2)$

Capacity: $C = \max_{f_x : E[x^2] \leq P} I(x;y)$

$h(y)$ is maximized when y is $\mathcal{N}(0, P + \sigma^2)$ which is achieved by choosing x to be $\mathcal{N}(0, P)$

$$\therefore C = \frac{1}{2} log(2\pi e(P + \sigma^2)) - \frac{1}{2} log(2\pi e \sigma^2) = \frac{1}{2} log(1 + \frac{P}{\sigma^2}) \ldots \ldots \ldots \ldots \ldots \ldots \ldots (11)$$

Capacity of the Fast-Fading Channel

- For fast fading channel: $y[m] = h[m]x[m] + w[m]$

The capacity of the power-constrained fast fading channel with receiver CSI:

$$C = \max_{P_x:\mathbb{E}[x^2]\leq P} I(x; y, h)$$

- Since h is independent of the input, $I(x; y|h) = 0$,

$$I(x; y, h) = I(x, h) + I(x; y|h) = I(x; y|h) \dots\dots\dots\dots\dots\dots..(12)$$

Conditioned on fading coefficient, $h\ SNR = \frac{P|h|^2}{N_0}$

$$C = \mathbb{E}_h\left[log\left(1 + \frac{P|h|^2}{N_0}\right)\right] \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (13)$$

- (13) holds if $h[m]$ is **stationary** and **ergodic.**

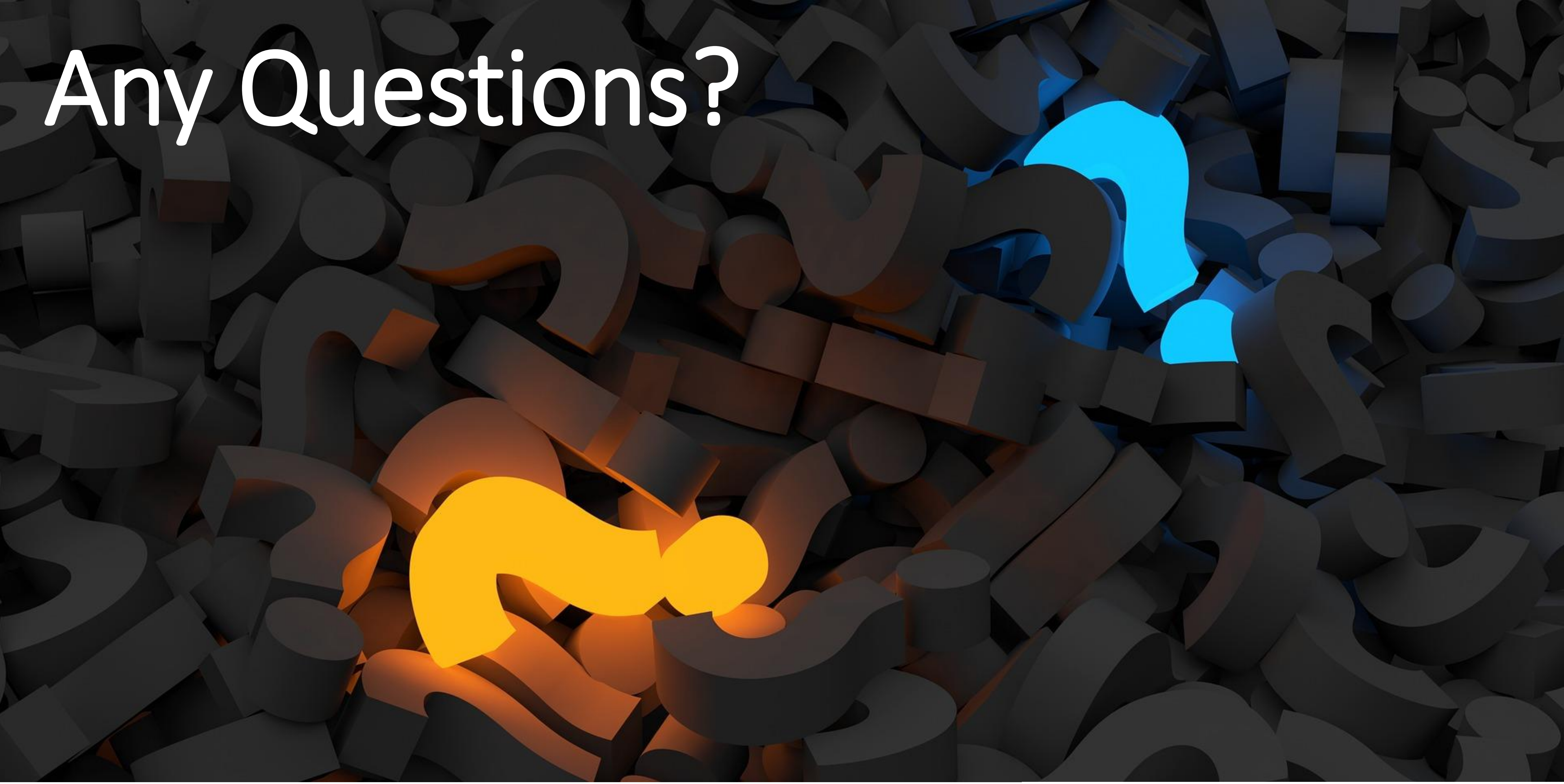
Rate of reliable communication is the average rate of flow of $I(x;y)$

$$\frac{1}{N} I(x;y) = \frac{1}{N} \sum_{m=1}^{N} log(1 + |h[m]|^2 SNR)$$

- For large N, due to the ergodicity of the fading process,

$$\frac{1}{N} \sum_{m=1}^{N} log(1 + |h[m]|^2 SNR) \rightarrow \mathbb{E}[log(1 + |h|^2 SNR)]$$

Any Questions?



E-mail: edkwao10@gmail.com