

One Lesson of Information Theory

Prof. Dr.-Ing. Volker Kühn
Institute of Communications Engineering
Phone: 0381/498-7330, Room: W 8233
Email: volker.kuehn@uni-rostock.de

<http://www.int.uni-rostock.de/>

- **Lesson 1: One Lesson of Information Theory**
 - Principle structure of communication systems
 - Definitions of entropy, mutual information, ...
 - Channel coding theorem of Shannon

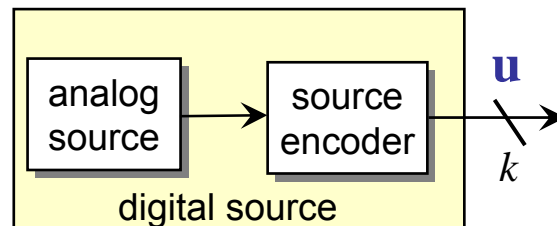
- **Lesson 2: Introduction to Error Correcting Codes**
 - Coding principles
 - Basics of error correcting codes
 - Examples: simple block codes and - if time permits - convolutional codes

- **Lesson 3: State-of-the-art channel coding**
 - Coding strategies to approach the capacity limits
 - Definition of soft-information and turbo decoding principle
 - Examples for state-of-the-art error correcting codes

- Lin/ Costello: „Error Control Coding: Fundamentals and Applications“
- Bossert: „Channel Coding“
- Johannesson/Zigangirov: „Fundamentals of Convolutional Codes“
- Richardson, Urbanke: „Modern Coding Theory“
- Neubauer, Freudenberger, Kühn: “Coding Theory – Algorithms, Architectures, and Applications”
- Johannesson: „Information Theory“
- Cover, Thomas: “Elements of Information Theory”

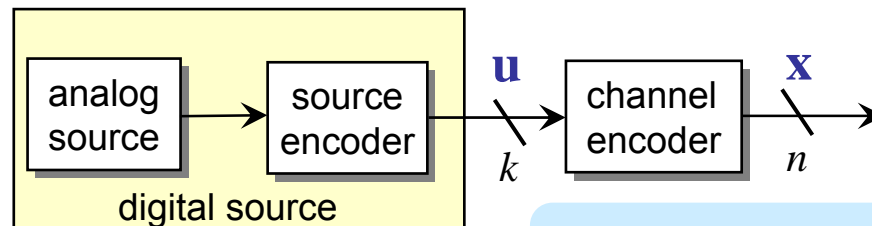
Principle structure of communication systems
Definitions of entropy, mutual information, ...
Channel coding theorem of Shannon

- Principle structure of digital communication system



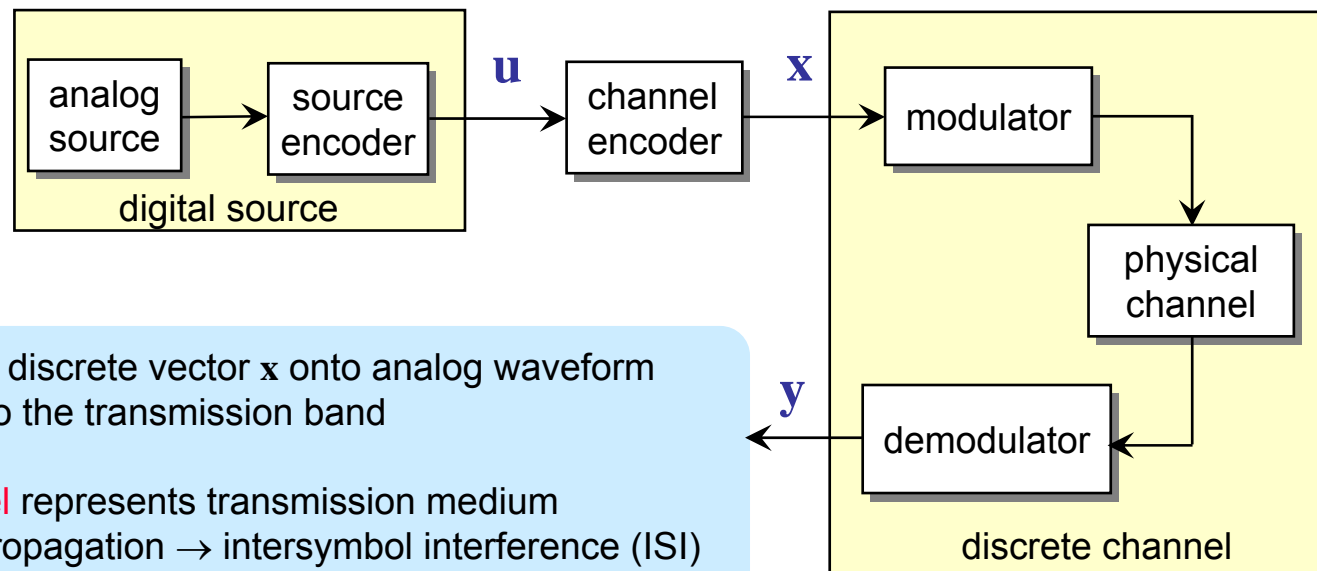
- Source** generates analog signal (e.g. voice, video)
- Source coding** samples, quantizes and compresses analog signal
- Digital Source**: comprises analog source and source coding, delivers digital data vector \mathbf{u} of length k

- Principle structure of digital communication system



- **Channel encoder** adds redundancy to \mathbf{u} resulting in code word \mathbf{x} of length n
- Channel encoder may consist of several constituent codes
- Code rate: $R_c = k / n$

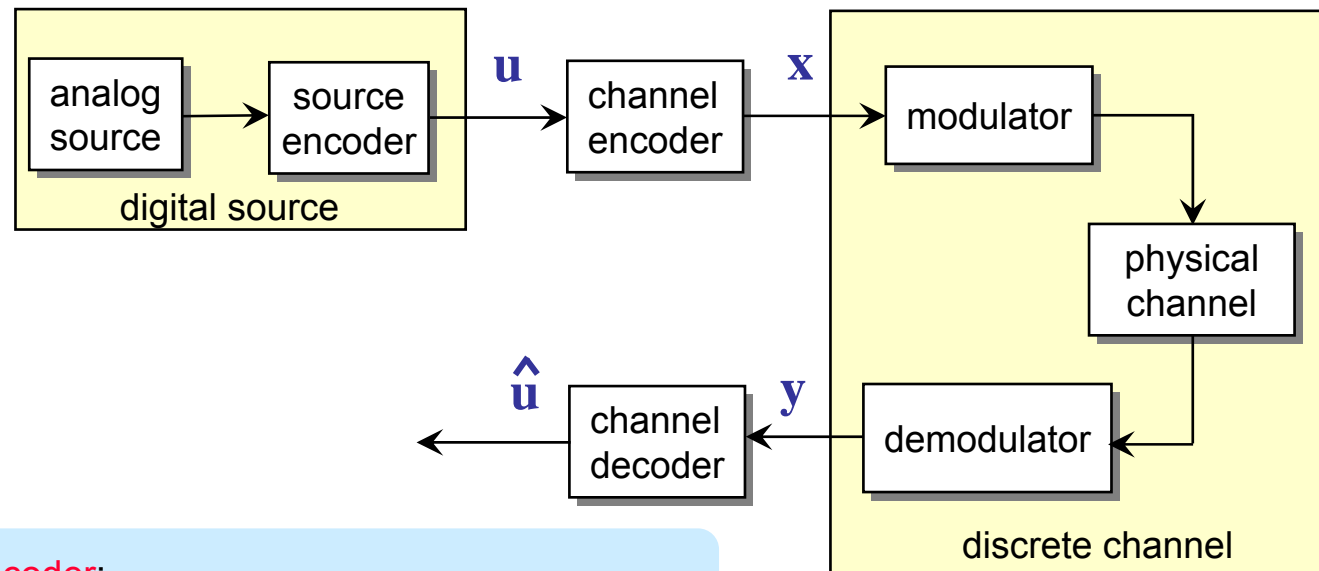
- Principle structure of digital communication system



- Modulator** maps discrete vector \mathbf{x} onto analog waveform and moves it into the transmission band
- Physical channel** represents transmission medium
 - Multipath propagation → intersymbol interference (ISI)
 - Time varying fading, i.e. deep fades
 - Additive noise
- Demodulator**: Moves signal back into baseband and performs lowpass filtering, sampling, quantization

Discrete channel:
comprises analog part of modulator, physical channel and analog part of demodulator

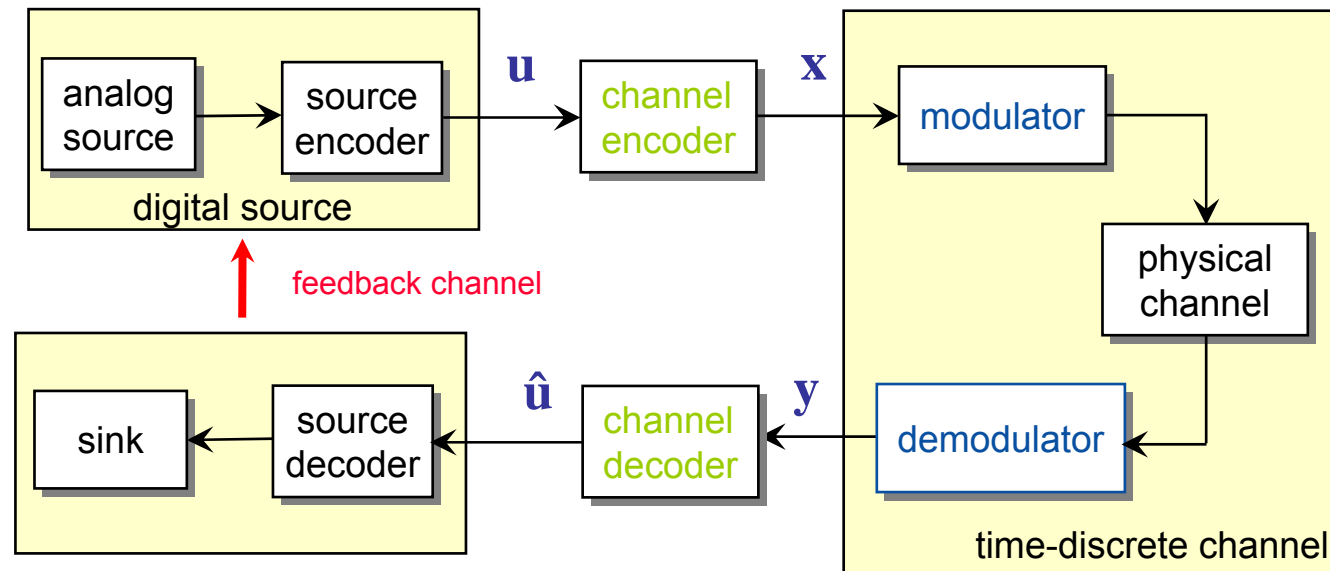
- Principle structure of digital communication system



Channel decoder:

- Estimation of \mathbf{u} on the basis of received vector \mathbf{y}
- \mathbf{y} need not consist of hard quantized values (0,1)
- Since encoder may consist of several parts, decoder may also consist of several modules

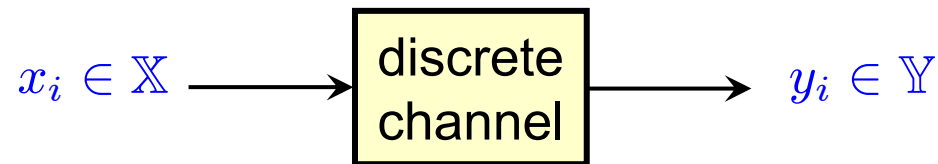
- Principle structure of digital communication system



Citation of Massey:

“The purpose of the **modulation system** is to create a good discrete channel from the modulator input to the demodulator output, and the purpose of the **coding system** is to transmit the information bits reliably through this discrete channel at the highest practicable rate.”

- Time-discrete channel comprises analog parts of modulator and demodulator as well as physical transmission medium



discrete input alphabet

$$\mathbb{X} = \{X_0, \dots, X_{|\mathbb{X}|-1}\}$$

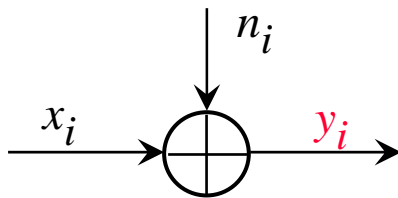
discrete or continuous output alphabets

$$\mathbb{Y} = \{Y_0, \dots, Y_{|\mathbb{Y}|-1}\}$$

$$\mathbb{Y} = \mathbb{R}$$

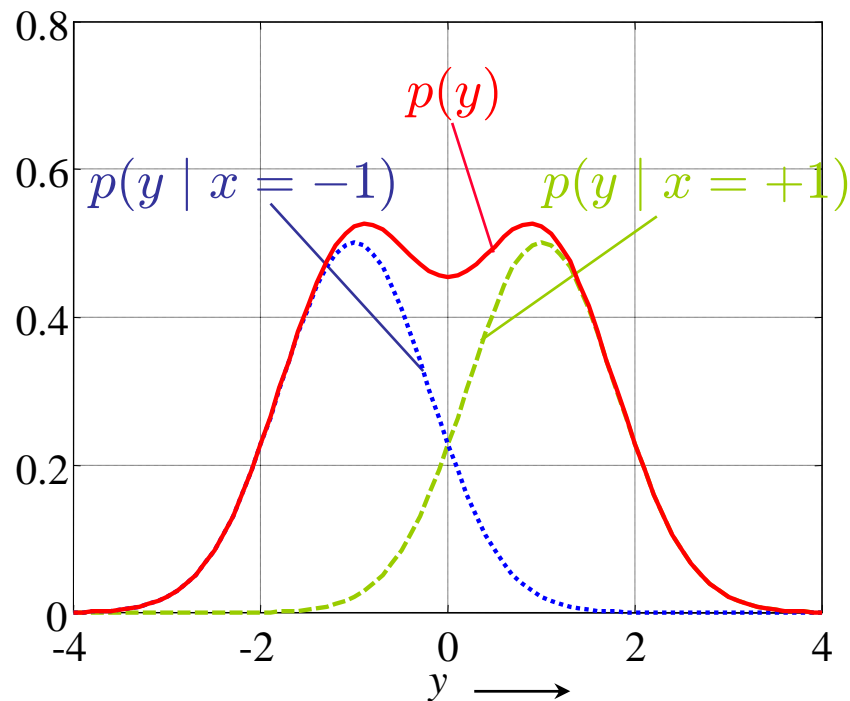
- Probabilities, probability densities: $\Pr\{X_\nu\}, \Pr\{Y_\mu\}$ $p(y)$
- Joint probability of event: $\Pr\{X_\nu, Y_\mu\}$ $p(x = X_\nu, y)$
- Transition probabilities: $\Pr\{Y_\mu | X_\nu\}$ $p(y | x = X_\nu)$
- A posteriori probabilities: $\Pr\{X_\nu | Y_\mu\}$ $\Pr\{X_\nu | y\}$

- Conditional probability density function of AWGN channel

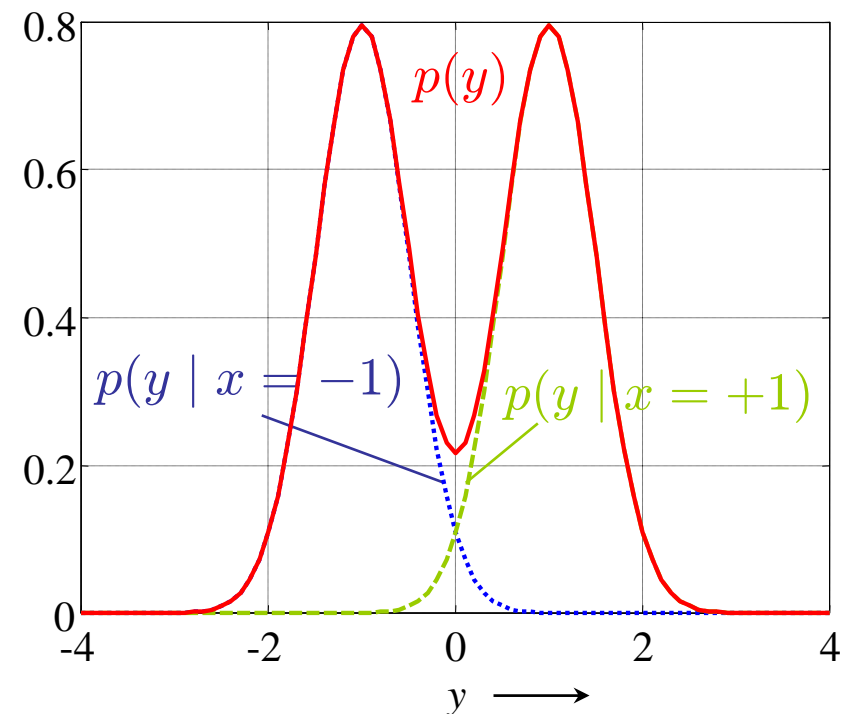


$$p(y | x = X_\nu) = \frac{1}{\sqrt{2\pi\sigma_N^2}} \cdot e^{-\frac{(y - X_\nu)^2}{2\sigma_N^2}}$$

signal-to-noise-ratio E_s/N_0 : 2 dB



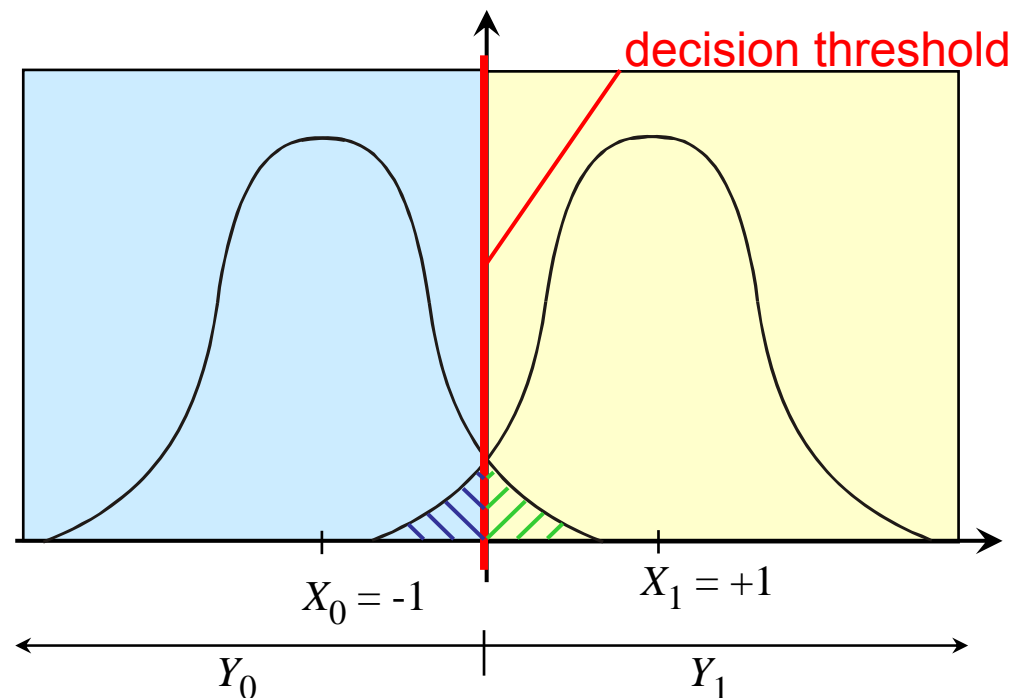
signal-to-noise-ratio E_s/N_0 : 6 dB



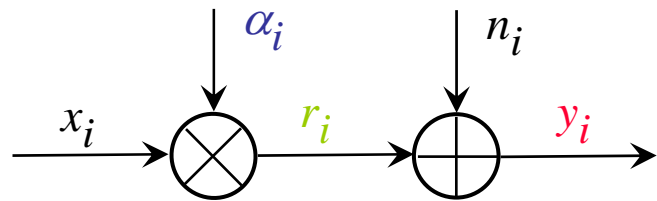
- Error probability for BPSK

$$P_s = \frac{1}{\pi} \cdot \int_{\sqrt{E_s/N_0}}^{\infty} e^{-\xi^2} d\xi = \frac{1}{2} \cdot \text{erfc} \left(\sqrt{\frac{E_s}{N_0}} \right)$$

error function complement

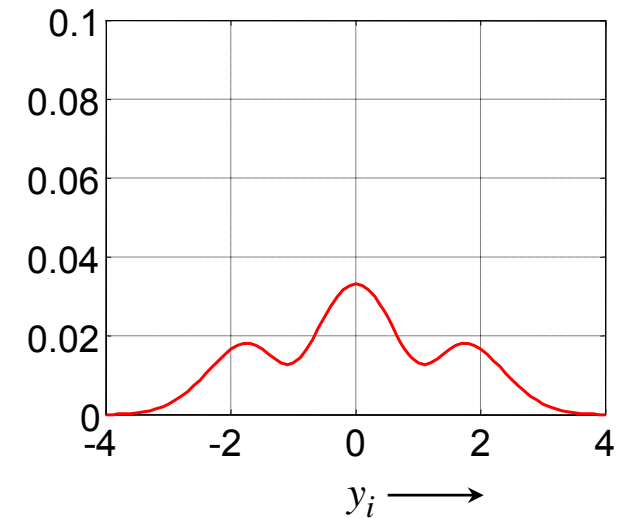
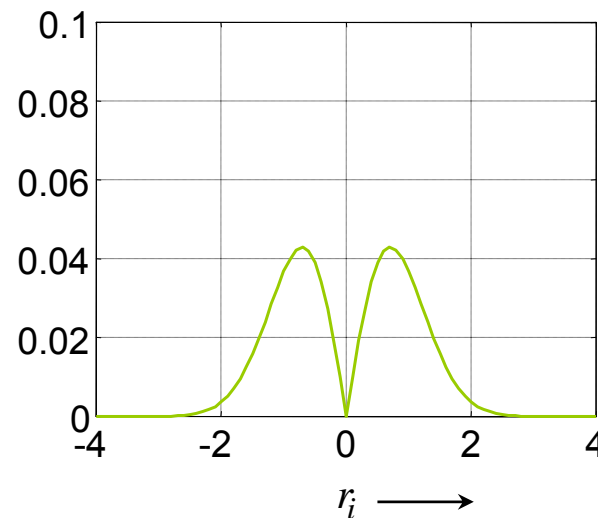
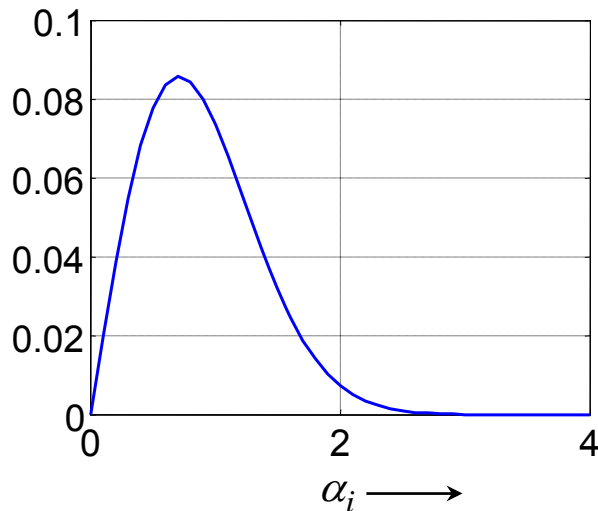


- Conditional probability density function of flat Rayleigh fading channel

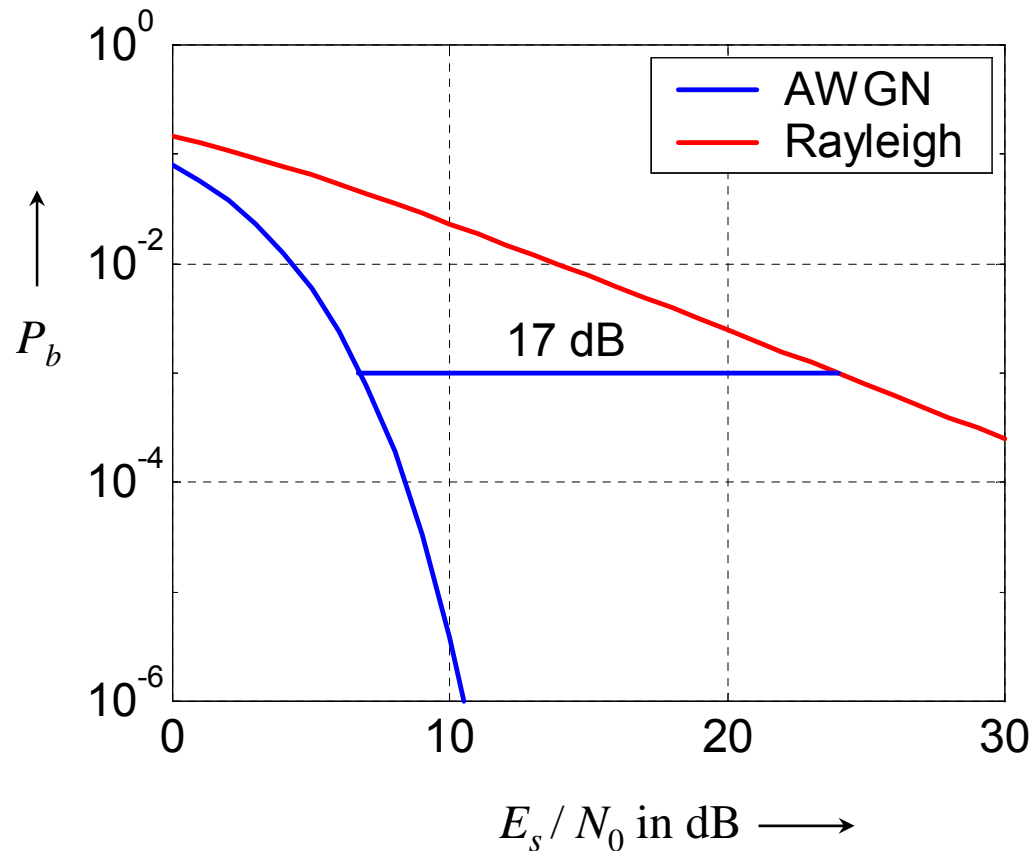


$$p_{|\alpha|}(\xi) = \begin{cases} 2\xi/\sigma_\alpha^2 \cdot \exp(-\xi^2/\sigma_\alpha^2) & \text{for } \xi \geq 0 \\ 0 & \text{else} \end{cases}$$

$$p_{|\alpha|^2}(\xi) = \begin{cases} 1/\sigma_\alpha^2 \cdot \exp(-\xi/\sigma_\alpha^2) & \text{for } \xi \geq 0 \\ 0 & \text{else} \end{cases}$$



- BPSK modulation



- AWGN channel:

$$P_b = \frac{1}{2} \cdot \operatorname{erfc} \left(\sqrt{\frac{E_s}{N_0}} \right)$$

- Flat Rayleigh fading channel:

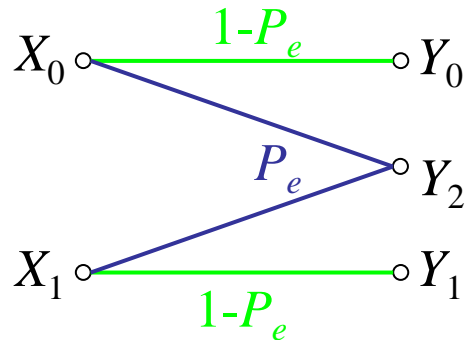
$$P_b = \frac{1}{2} \cdot \left[1 - \sqrt{\frac{E_s/N_0}{1 + E_s/N_0}} \right]$$

- Channel coding for fading channels essential (time diversity)

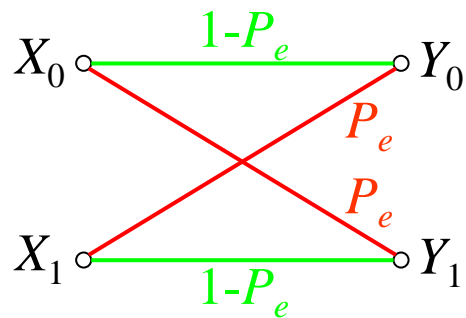
- Discrete channels arise from quantization of continuous channel output
- We consider binary antipodal transmission: $\mathbb{X} = \{X_0, X_1\} = \{+1, -1\}$
- Generally continuously distributed channel output: $\mathbb{Y} = \mathbb{R}$
- L -bit quantization due to finite precision of digital circuits delivers alphabet $\mathbb{Y} = \{Y_0, \dots, Y_{2^L-1}\}$
 - $L = 1$: Hard-Decision: $\mathbb{Y} = \{Y_0, Y_1\} = \{+1, -1\} = \mathbb{X}$
 - $L = 2$: four output symbols: $\mathbb{Y} = \{Y_0, Y_1, Y_2, Y_3\}$
 - $L = 3$: four output symbols: $\mathbb{Y} = \{Y_0, Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Y_7\}$

Discrete Channels (1)

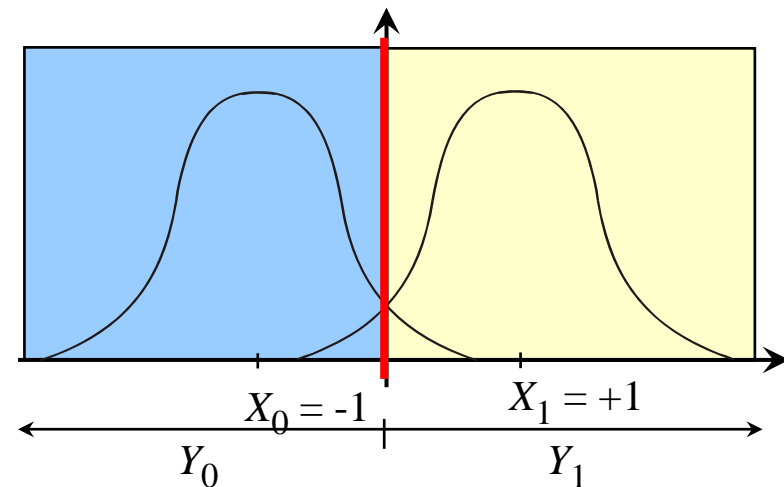
- Binary Erasure Channel (BEC)



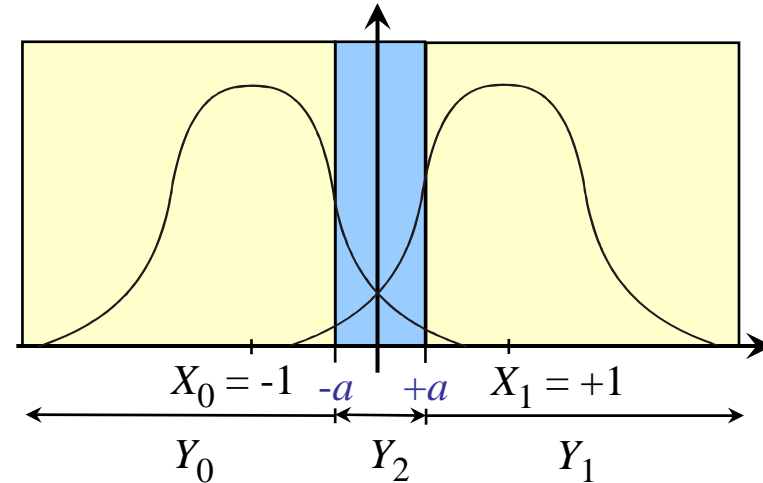
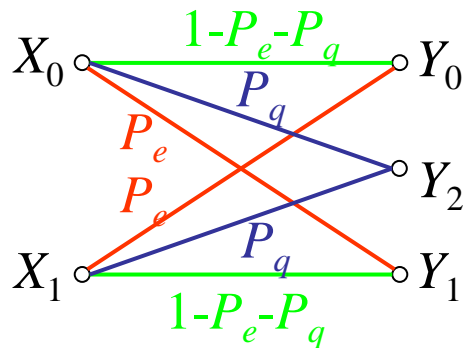
- Binary Symmetric Channel (BSC)



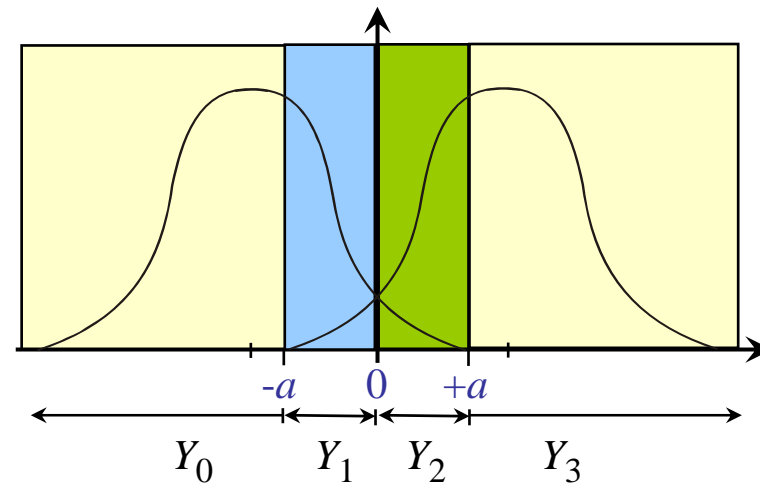
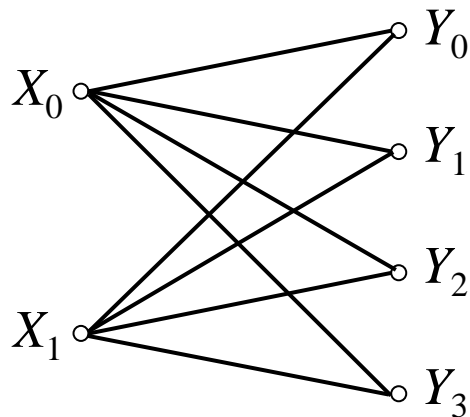
$$P_e = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{E_s}{N_0}} \right)$$



- Binary Symmetric Erasure Channel (BSEC)



- 2-Bit-Quantization



Principle structure of communication systems
Definitions of entropy, mutual information, ...
Channel coding theorem of Shannon

- Amount of information should depend on probability:

$$I(X_\nu) = f(\Pr\{X_\nu\})$$

- For independent events:

$$\Pr\{X_\nu, Y_\mu\} = \Pr\{X_\nu\} \cdot \Pr\{Y_\mu\} \quad \Rightarrow \quad I(X_\nu, Y_\mu) = I(X_\nu) + I(Y_\mu)$$

- Logarithm is sole function that maps product onto a sum

$$I(X_\nu) = -\log_2(\Pr\{X_\nu\}) = \log_2\left(\frac{1}{\Pr(X_\nu)}\right) \geq 0$$

- Entropy:

$$H(\mathbb{X}) = -\sum_\nu \Pr\{X_\nu\} \cdot \log_2(\Pr\{X_\nu\}) = -\mathbb{E}\{\log_2(\Pr\{X\})\}$$

- Entropy is a measure of uncertainty

Examples for Entropy

- Set of events: $\mathbb{X} = \{X_1, X_2, X_3, X_4, X_5\}$
- Each event occurs with certain probability

$$\Pr\{X_1\} = 0.30 \quad \Rightarrow \quad I(X_1) = 1.7370$$

$$\Pr\{X_2\} = 0.20 \quad \Rightarrow \quad I(X_1) = 2.3219$$

$$\Pr\{X_3\} = 0.20 \quad \Rightarrow \quad I(X_1) = 2.3219$$

$$\Pr\{X_4\} = 0.15 \quad \Rightarrow \quad I(X_1) = 2.7370$$

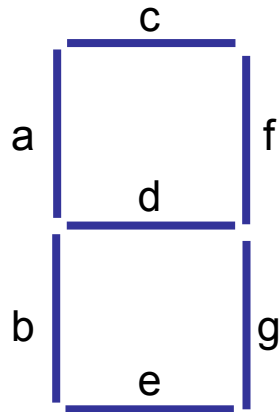
$$\Pr\{X_5\} = 0.15 \quad \Rightarrow \quad I(X_1) = 2.7370$$

$$H(\mathbb{X}) = - \sum_{\nu} \Pr\{X_{\nu}\} \cdot \log_2(\Pr\{X_{\nu}\}) = 2.271 \text{ bit}$$

- Entropy of a set is maximized, when all M elements are equally likely

$$\max_{\Pr\{\mathbb{X}\}} H(\mathbb{X}) = H_{\text{equal}}(\mathbb{X}) = \sum_{\nu=0}^{M-1} \frac{1}{M} \cdot \log_2(M) = \log_2(M) \text{ bit}$$

Example: LCD for 10 Digits



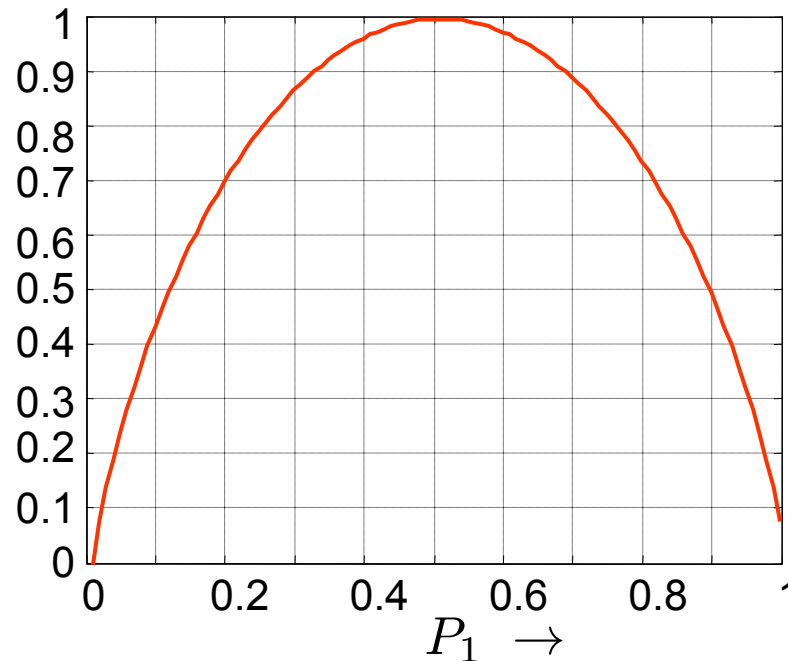
digit	0	1	2	3	4	5	6	7	8	9
a	1	0	0	0	1	1	1	0	1	1
b	1	0	1	0	0	0	1	0	1	0
c	1	0	1	1	0	1	1	1	1	1
d	0	0	1	1	1	1	1	0	1	1
e	1	0	1	1	0	1	1	0	1	1
f	1	1	1	1	1	0	0	1	1	1
g	1	1	0	1	1	1	1	1	1	1

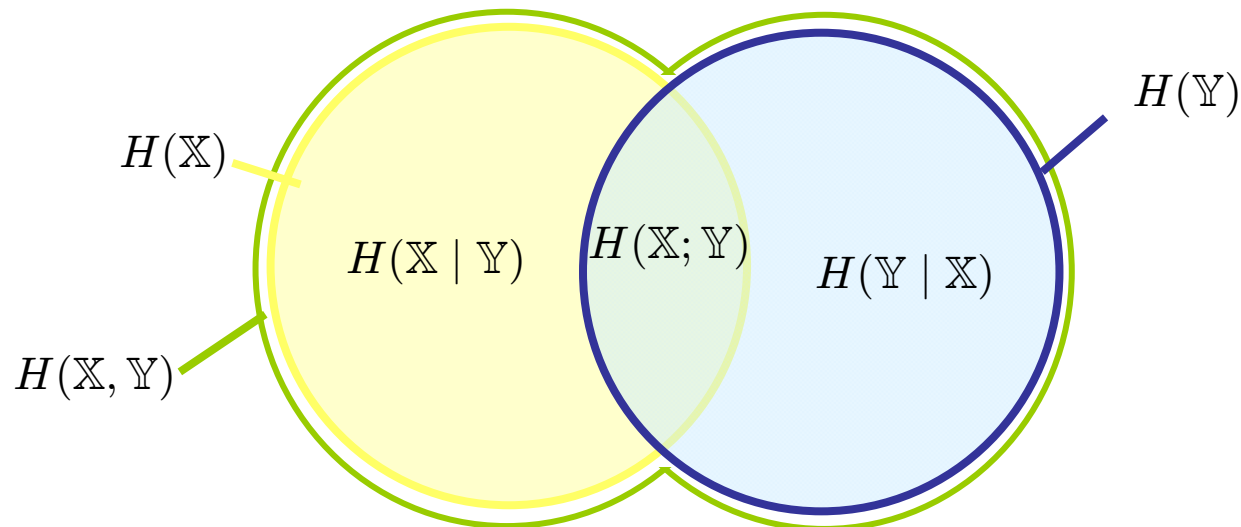
- All digits with same probability: $\Pr\{X_\nu\} = 0.1$
- Amount of information per digit: $I(X_\nu) = -\log_2(\Pr\{X_\nu\}) = \log_2(10) = 3.32$ bit
- Entropy of alphabet: $H(\mathbb{X}) = \sum_\nu \Pr\{X_\nu\} \cdot I(X_\nu) = 3.32$ bit
- Absolute redundancy: $R = m - H(\mathbb{X}) = 7 \text{ bit} - 3.32 \text{ bit} = 3.68 \text{ bit}$
- Relative redundancy: $r = R/m = 3.68 \text{ bit}/7 \text{ bit} = 52.54\%$

Binary Entropy Function

- Set of events: $\mathbb{X} = \{X_1, X_2, \}$
- Each event occurs with certain probability: $\Pr\{X_1\} = P_1$
 $\Pr\{X_2\} = 1 - P_1$
- Binary entropy function

$$H(\mathbb{X}) = H_2(P_1) = -P_1 \cdot \log_2(P_1) - (1 - P_1) \cdot \log_2(1 - P_1)$$





- $H(\mathbb{X})$: entropy of source alphabet
- $H(\mathbb{Y})$: entropy of sink alphabet
- $H(\mathbb{X}, \mathbb{Y})$: joint entropy of source and sink
- $H(\mathbb{X} | \mathbb{Y})$: equivocation: information lost during transmission
- $H(\mathbb{Y} | \mathbb{X})$: irrelevance, information originating not from source
- $H(\mathbb{X}; \mathbb{Y})$: mutual information: information correctly received at sink

- Joint Information $I(X_\nu, Y_\mu) = -\log_2 (\Pr\{X_\nu, Y_\mu\})$

- Joint Entropy of source and sink:

$$\begin{aligned} H(\mathbb{X}, \mathbb{Y}) &= -\sum_{\nu} \sum_{\mu} \Pr\{X_\nu, Y_\mu\} \cdot \log_2 (\Pr\{X_\nu, Y_\mu\}) \\ &= -\mathbf{E}\{\log_2 (\Pr\{X_\nu, Y_\mu\})\} \end{aligned}$$

- Equivocation: Information lost during transmission

$$\begin{aligned} H(\mathbb{X} | \mathbb{Y}) &= H(\mathbb{X}, \mathbb{Y}) - H(\mathbb{Y}) = -\sum_{\nu} \sum_{\mu} \Pr\{X_\nu, Y_\mu\} \cdot \log_2 (\Pr\{X_\nu | Y_\mu\}) \\ &= -\mathbf{E}\{\log_2 (\Pr\{X_\nu | Y_\mu\})\} \end{aligned}$$

- Irrelevance

$$\begin{aligned} H(\mathbb{Y} | \mathbb{X}) &= H(\mathbb{X}, \mathbb{Y}) - H(\mathbb{X}) = -\sum_{\nu} \sum_{\mu} \Pr\{X_\nu, Y_\mu\} \cdot \log_2 (\Pr\{Y_\mu | X_\nu\}) \\ &= -\mathbf{E}\{\log_2 (\Pr\{Y_\mu | X_\nu\})\} \end{aligned}$$

- Definition of Mutual Information

$$\begin{aligned} H(\mathbb{X}; \mathbb{Y}) &= H(\mathbb{X}) - H(\mathbb{X} | \mathbb{Y}) = H(\mathbb{X}) - [H(\mathbb{X}, \mathbb{Y}) - H(\mathbb{Y})] \\ &= H(\mathbb{Y}) - H(\mathbb{Y} | \mathbb{X}) = H(\mathbb{Y}) - [H(\mathbb{X}, \mathbb{Y}) - H(\mathbb{X})] \\ &= H(\mathbb{X}) + H(\mathbb{Y}) - H(\mathbb{X}, \mathbb{Y}) \\ &= \sum_{\nu} \sum_{\mu} \Pr\{Y_{\mu} | X_{\nu}\} \cdot \Pr\{X_{\nu}\} \cdot \log_2 \left(\frac{\Pr\{Y_{\mu} | X_{\nu}\}}{\Pr\{Y_{\mu}\}} \right) \\ &= \mathbb{E} \left\{ \log_2 \left(\frac{\Pr\{Y_{\mu} | X_{\nu}\}}{\Pr\{Y_{\mu}\}} \right) \right\} \end{aligned}$$

- Mutual information is the amount of information common to X and Y
- Mutual information is the reduction of uncertainty in X due to the knowledge of Y

- Chain rule of entropies

$$H(\mathbb{X}_1, \mathbb{X}_2) = H(\mathbb{X}_1) + H(\mathbb{X}_2 | \mathbb{X}_1)$$

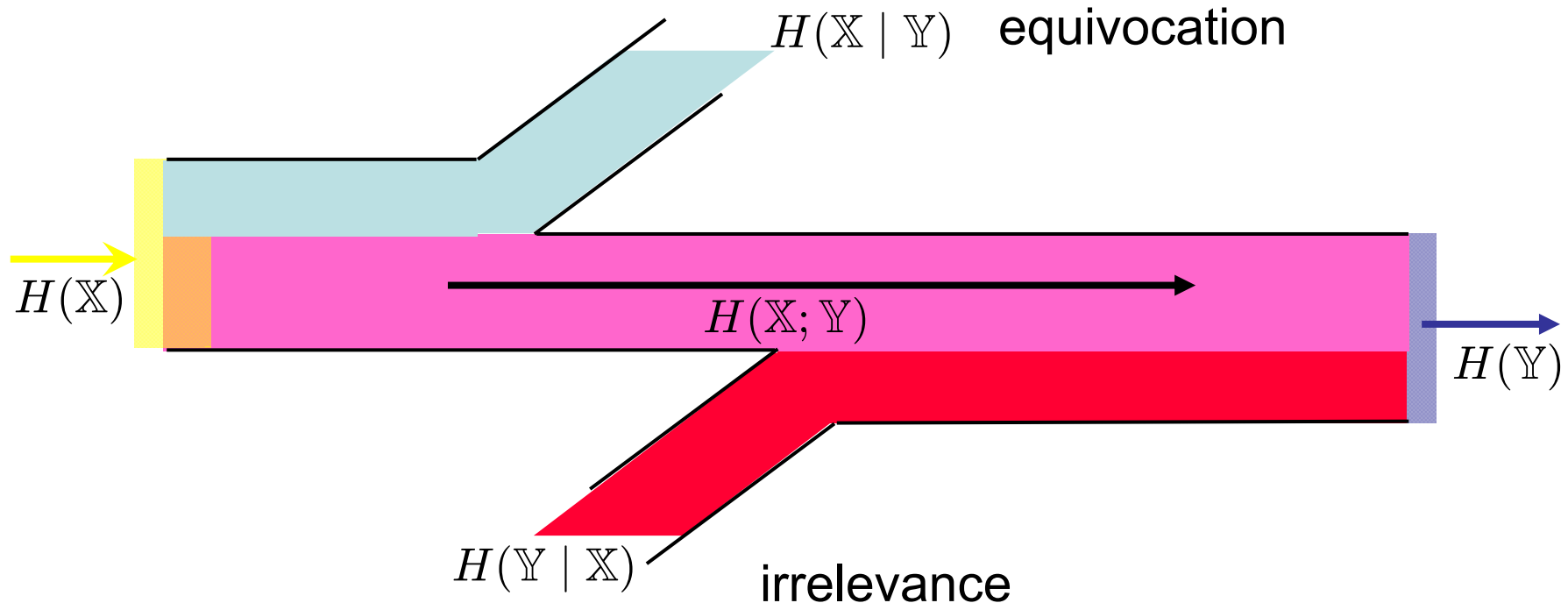
$$H(\mathbb{X}_1, \mathbb{X}_2, \mathbb{X}_3) = H(\mathbb{X}_1) + H(\mathbb{X}_2, \mathbb{X}_3 | \mathbb{X}_1) = H(\mathbb{X}_1) + H(\mathbb{X}_2 | \mathbb{X}_1) + H(\mathbb{X}_3 | \mathbb{X}_1, \mathbb{X}_2)$$

•
•
•

$$H(\mathbb{X}_1, \mathbb{X}_2, \dots, \mathbb{X}_n) = \sum_{i=1}^n H(\mathbb{X}_i | \mathbb{X}_{i-1}, \dots, \mathbb{X}_1)$$

- Chain rule of information

$$H(\mathbb{X}_1, \dots, \mathbb{X}_n; \mathbb{Y}) = \sum_{i=1}^n H(\mathbb{X}_i; \mathbb{Y} | \mathbb{X}_{i-1}, \dots, \mathbb{X}_1)$$



- Maximization of mutual information with respect to source statistics delivers channel capacity:

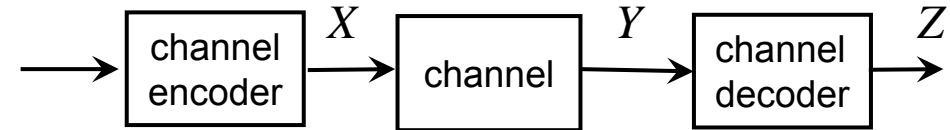
$$C = \sup_{\Pr\{\mathbb{X}\}} \sum_{\nu} \sum_{\mu} \Pr\{Y_{\mu} | X_{\nu}\} \cdot \Pr\{X_{\nu}\} \cdot \log_2 \frac{\Pr\{Y_{\mu} | X_{\nu}\}}{\Pr\{Y_{\mu}\}}$$



Principle structure of communication systems
Definitions of entropy, mutual information, ...
Channel coding theorem of Shannon

- Shannon, 1948: “A Mathematical Theory of Communication”
- If a channel has the capacity C , there exist a code with rate $R_c \leq C$ for which the probability of a decoding error can be made arbitrary small.
- Converse Theorem:
If a channel has the capacity C , a reliable (error-free) communication cannot be achieved for codes with rates $R_c > C$.
- Theorems are not constructive, i.e. they do not provide a construction guideline for powerful codes

- Markov Chain: $X \rightarrow Y \rightarrow Z$
- Joint probability



$$p(X, Y, Z) = p(X) \cdot p(Y | X) \cdot p(Z | Y)$$

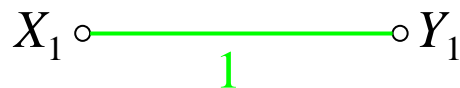
- Random variable Z is independent of X if Y is known

$$p(X, Z | Y) = \frac{p(X, Y) \cdot p(Z | Y)}{p(Y)} = p(X | Y) \cdot p(Z | Y)$$

- Data processing inequality: $H(X; Y) \geq H(X; Z)$

No processing of Y (random or deterministic) can increase the mutual information that Y contains on X

- Statistics of channel



$$\Pr \{X_\nu\} = \begin{cases} P_0 & \text{for } \mu = 0 \\ 1 - P_0 & \text{for } \mu = 1 \end{cases}$$

$$\Pr \{Y_\mu | X_\nu\} = \begin{cases} 1 & \text{for } \mu = \nu \\ 0 & \text{for } \mu \neq \nu \end{cases}$$

$$\Pr \{Y_\mu\} = \begin{cases} P_0 & \text{for } \mu = 0 \\ 1 - P_0 & \text{for } \mu = 1 \end{cases}$$

- Mutual information

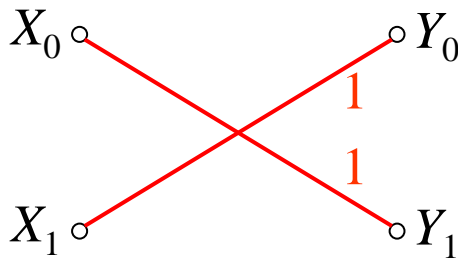
$$H(\mathbb{X}; \mathbb{Y}) = P_0 \log_2 \frac{1}{P_0} + (1 - P_0) \log_2 \frac{1}{1 - P_0} = H_2(P_0) = H(\mathbb{X})$$

– Hint: $0 \cdot \log_2(0) = 0$

- Perfect transmission without any errors!**



- Statistics of channel



$$\Pr \{X_\nu\} = \begin{cases} P_0 & \text{for } \mu = 0 \\ 1 - P_0 & \text{for } \mu = 1 \end{cases}$$

$$\Pr \{Y_\mu | X_\nu\} = \begin{cases} 0 & \text{for } \mu = \nu \\ 1 & \text{for } \mu \neq \nu \end{cases}$$

$$\Pr \{Y_\mu\} = \begin{cases} 1 - P_0 & \text{for } \mu = 0 \\ P_0 & \text{for } \mu = 1 \end{cases}$$

- Mutual information

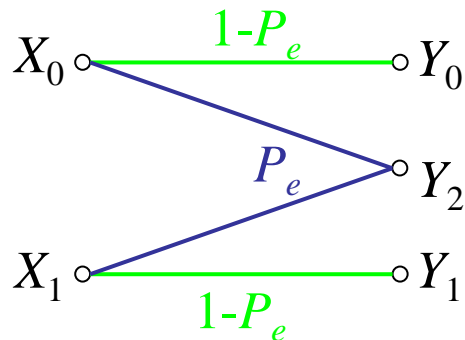
$$H(\mathbb{X}; \mathbb{Y}) = P_0 \log_2 \frac{1}{P_0} + (1 - P_0) \log_2 \frac{1}{1 - P_0} = H_2(P_0) = H(\mathbb{X})$$

– Hint: $0 \cdot \log_2(0) = 0$

- Perfect transmission without any errors!**



- Statistics of BEC channel



$$\Pr \{X_\nu\} = \begin{cases} P_0 & \text{for } \mu = 0 \\ 1 - P_0 & \text{for } \mu = 1 \end{cases}$$

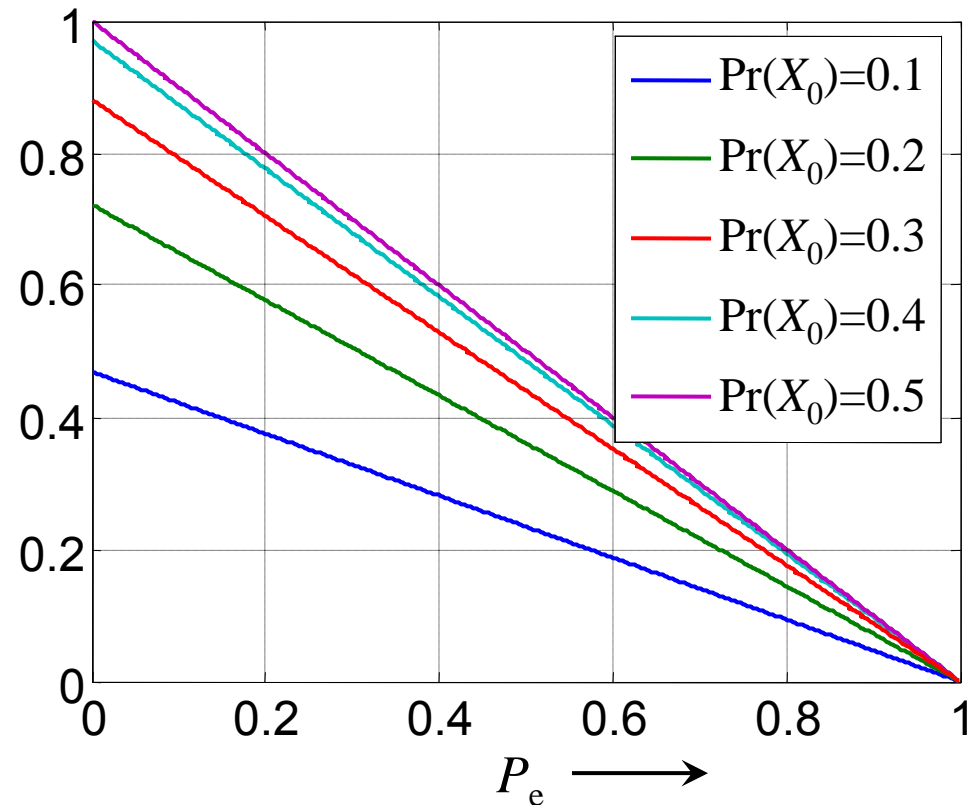
$$\Pr \{Y_\mu\} = \begin{cases} P_0 \cdot (1 - P_e) & \text{for } \mu = 0 \\ P_e \cdot (P_0 + 1 - P_0) = P_e & \text{for } \mu = 2 \\ (1 - P_0) \cdot (1 - P_e) & \text{for } \mu = 1 \end{cases}$$

- Mutual information of BEC

$$\begin{aligned} I(\mathbb{X}; \mathbb{Y}) &= (1 - P_e)P_0 \log_2 \frac{1 - P_e}{P_0(1 - P_e)} + P_e P_0 \log_2 \frac{P_e}{P_e(P_0 + 1 - P_0)} \\ &+ (1 - P_e)(1 - P_0) \log_2 \frac{1 - P_e}{(1 - P_0)(1 - P_e)} + P_e(1 - P_0) \log_2 \frac{P_e}{P_e} \\ &= (1 - P_e) \cdot H_2(P_0) \end{aligned}$$

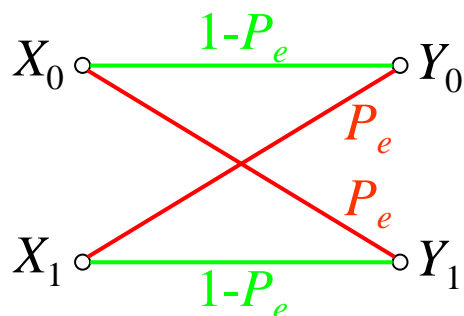


- Mutual information of BEC and different statistics of input signal



- Capacity of BEC for uniform input distribution: $C_{\text{BEC}} = (1 - P_e)$

- Statistics of BSC channel for uniform input distribution



$$\Pr \{X_0\} = \Pr \{X_1\} = \frac{1}{2}$$

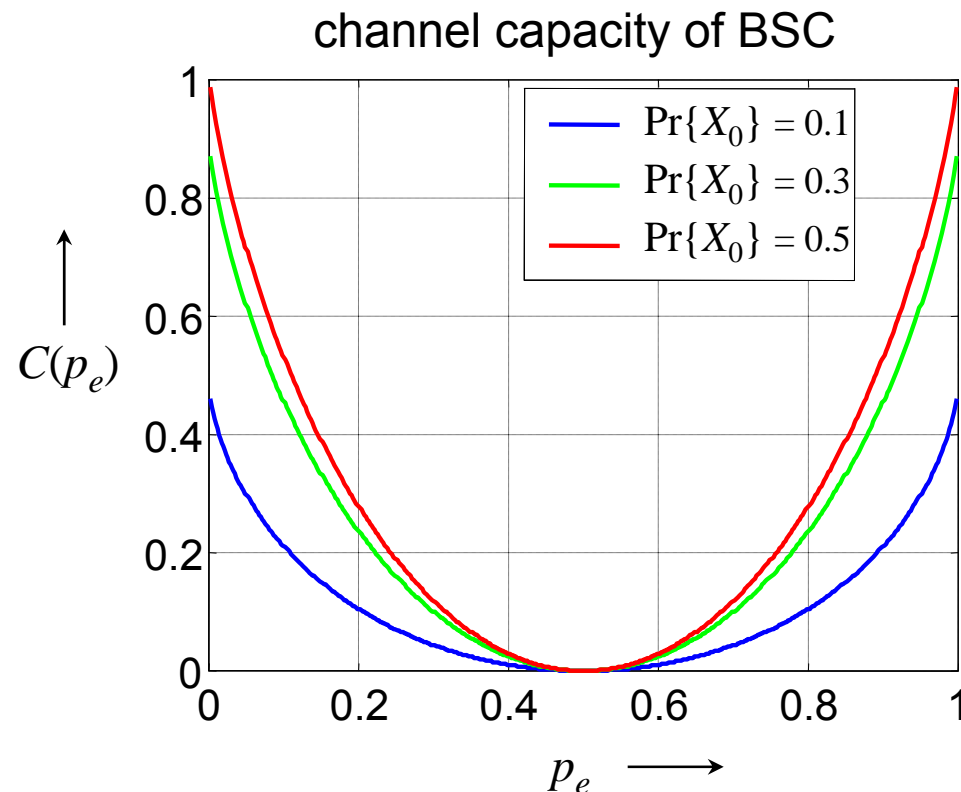
$$\Pr \{Y_\mu | X_\nu\} = \begin{cases} 1 - P_e & \text{for } \mu = \nu \\ P_e & \text{for } \mu \neq \nu \end{cases}$$

$$\Pr \{Y_0\} = \Pr \{Y_1\} = \frac{1}{2}$$

- Mutual information of BSC

$$\begin{aligned} C_{\text{BSC}} &= 2 \cdot (1 - P_e) \cdot \frac{1}{2} \cdot \log_2 [2(1 - P_e)] + 2 \cdot P_e \cdot \frac{1}{2} \cdot \log_2 (2P_e) \\ &= (1 - P_e) \cdot [1 + \log_2(1 - P_e)] + P_e \cdot [1 + \log_2(P_e)] \\ &= 1 + (1 - P_e) \cdot \log_2(1 - P_e) + P_e \cdot \log_2(P_e) \\ &= 1 - H_2(P_e) \end{aligned}$$

- Mutual information of BSC and different statistics of input signal

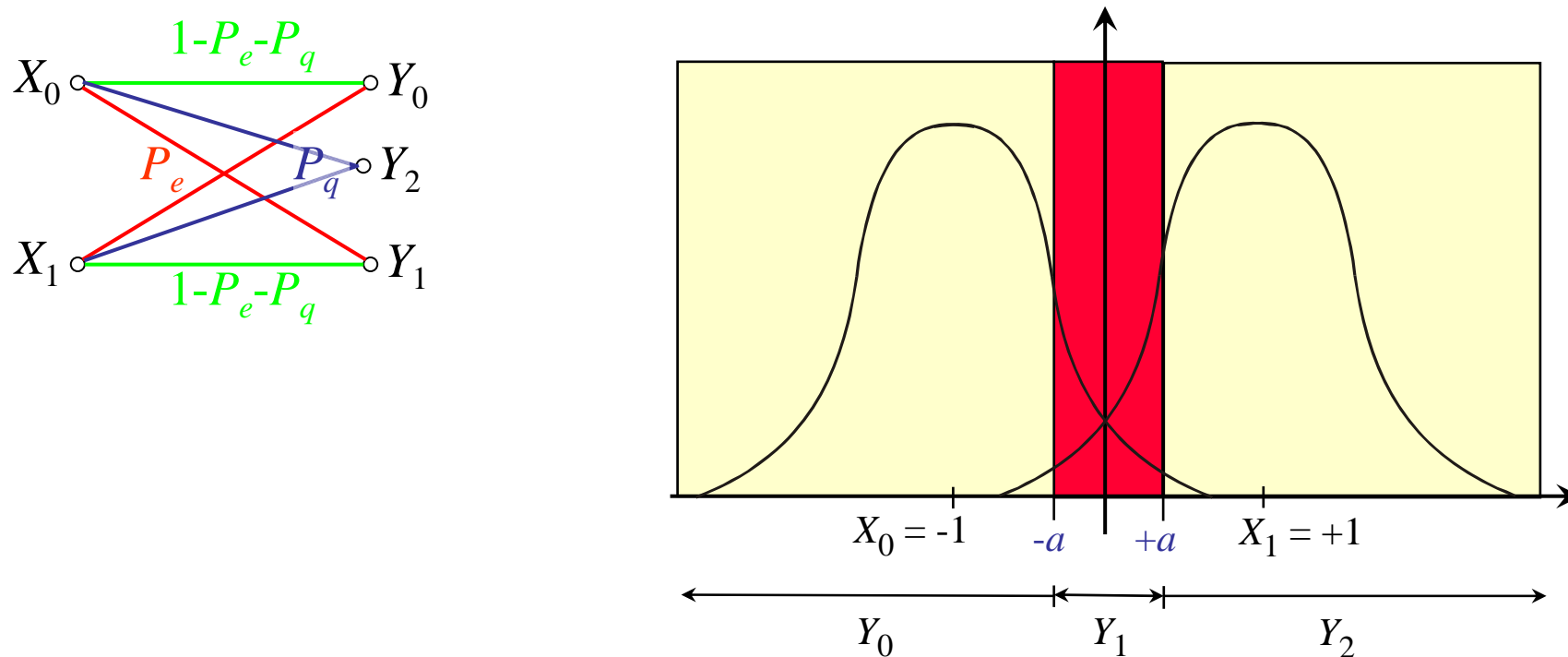


- Capacity of BSC for uniform input distribution

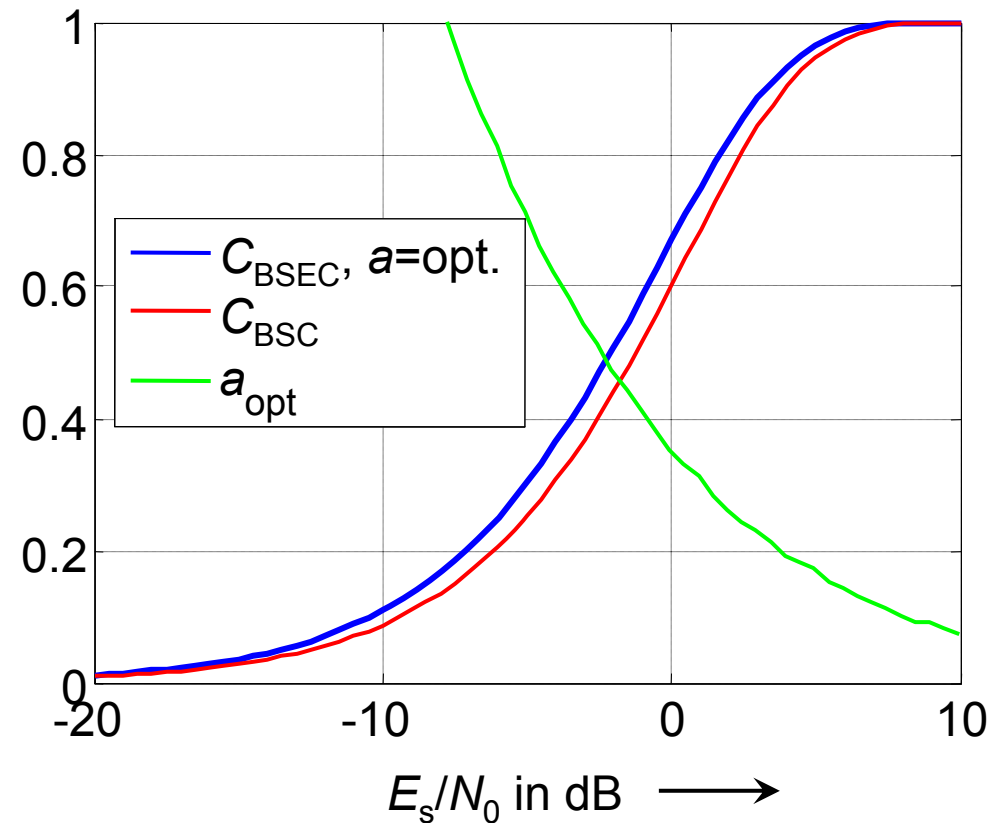
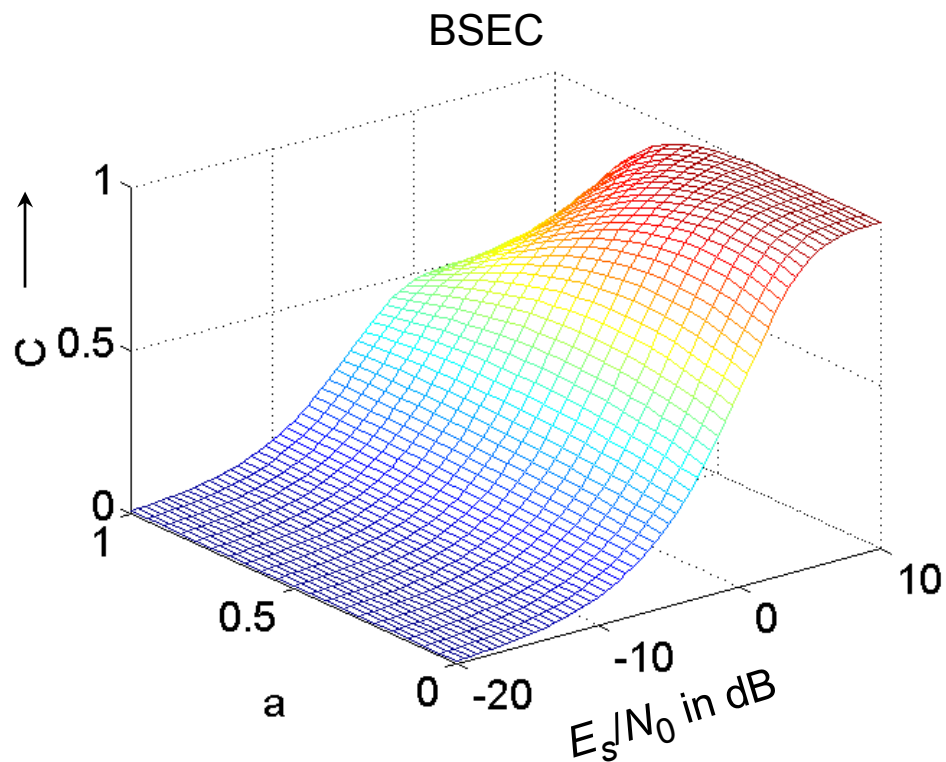
$$C_{\text{BSC}} = 1 + P_e \cdot \log_2(P_e) + (1 - P_e) \cdot \log_2(1 - P_e) = 1 - H_2(P_e)$$

Binary Symmetric Erasure Channel (BSEC)

- Quantization parameter a has to be optimized with respect to channel capacity C
- Optimal choice depends on signal-to-noise-ratio E_s/N_0

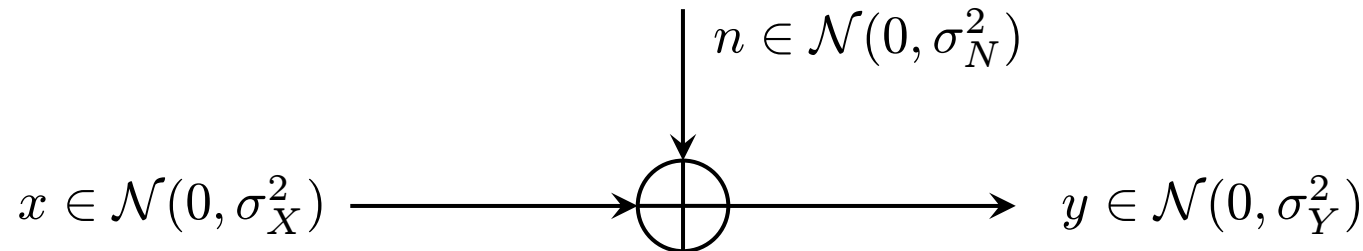


$$C_{\text{BSEC}} = 1P_q + P_e \log_2(P_e) + (1 - P_e - P_q) \cdot \log_2(1 - P_e - P_q) - (1 - P_q) \cdot \log_2(1 - P_q)$$



- $a > 1$ leads only to minor improvement of channel capacity

- Additive White Gaussian Noise Channel



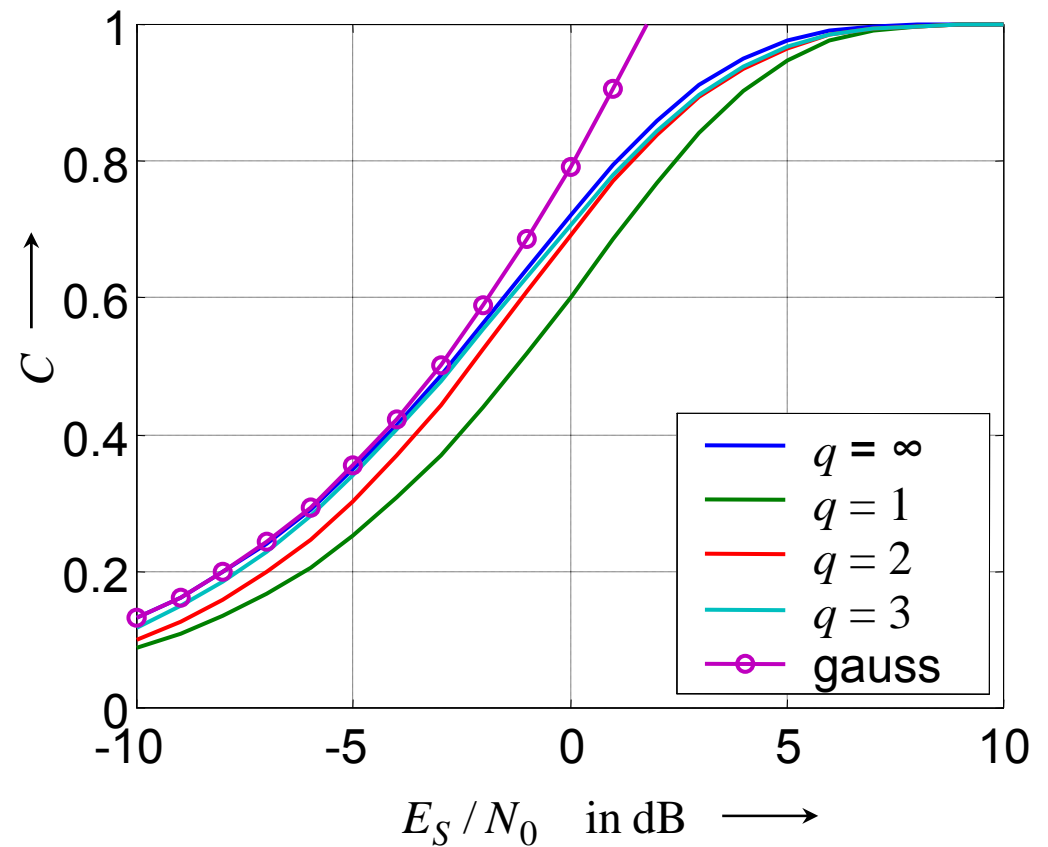
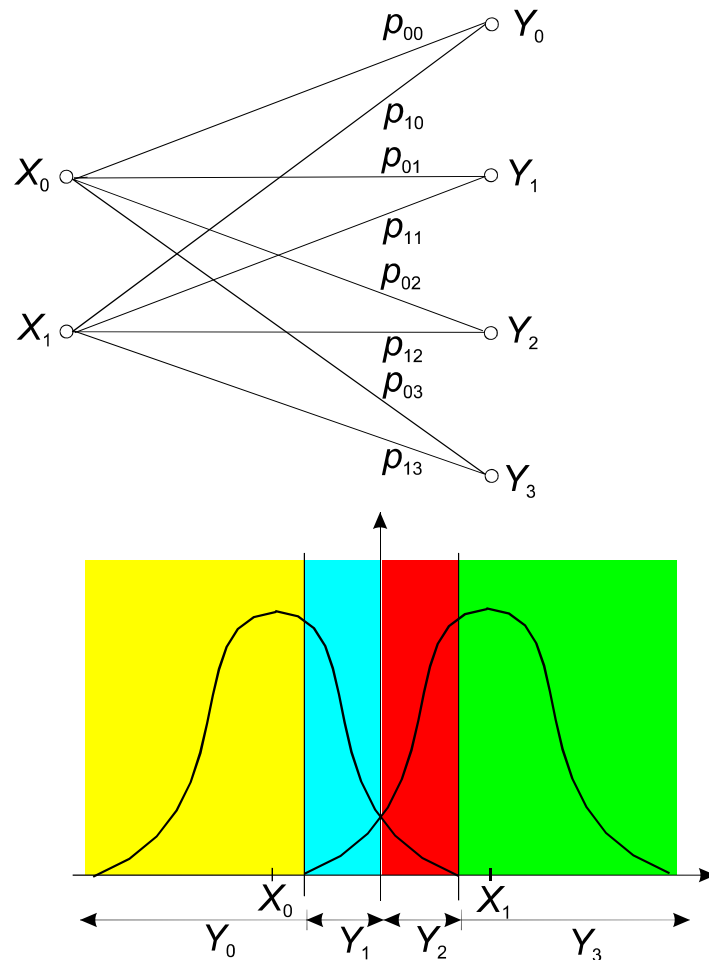
- Differential entropy of Gaussian random process

$$h(\mathbf{X}) = - \int_{-\infty}^{\infty} p_N(\xi) \cdot \log_2 p_N(\xi) d\xi = \frac{1}{2} \cdot \log_2(2\pi e \sigma_X^2)$$

- Capacity of AWGN channel

$$\begin{aligned} C &= h(\mathbb{Y}) - h(\mathbb{Y} | \mathbb{X}) = h(\mathbb{Y}) - h(\mathbb{N}) \\ &= \frac{1}{2} \cdot \log_2(2\pi e \sigma_Y^2) - \frac{1}{2} \cdot \log_2(2\pi e \sigma_N^2) \\ &= \frac{1}{2} \cdot \log_2 [2\pi e (\sigma_X^2 + \sigma_N^2)] - \frac{1}{2} \cdot \log_2(2\pi e \sigma_N^2) = \frac{1}{2} \cdot \log_2 (1 + \sigma_X^2 / \sigma_N^2) \end{aligned}$$

- Influence of quantization



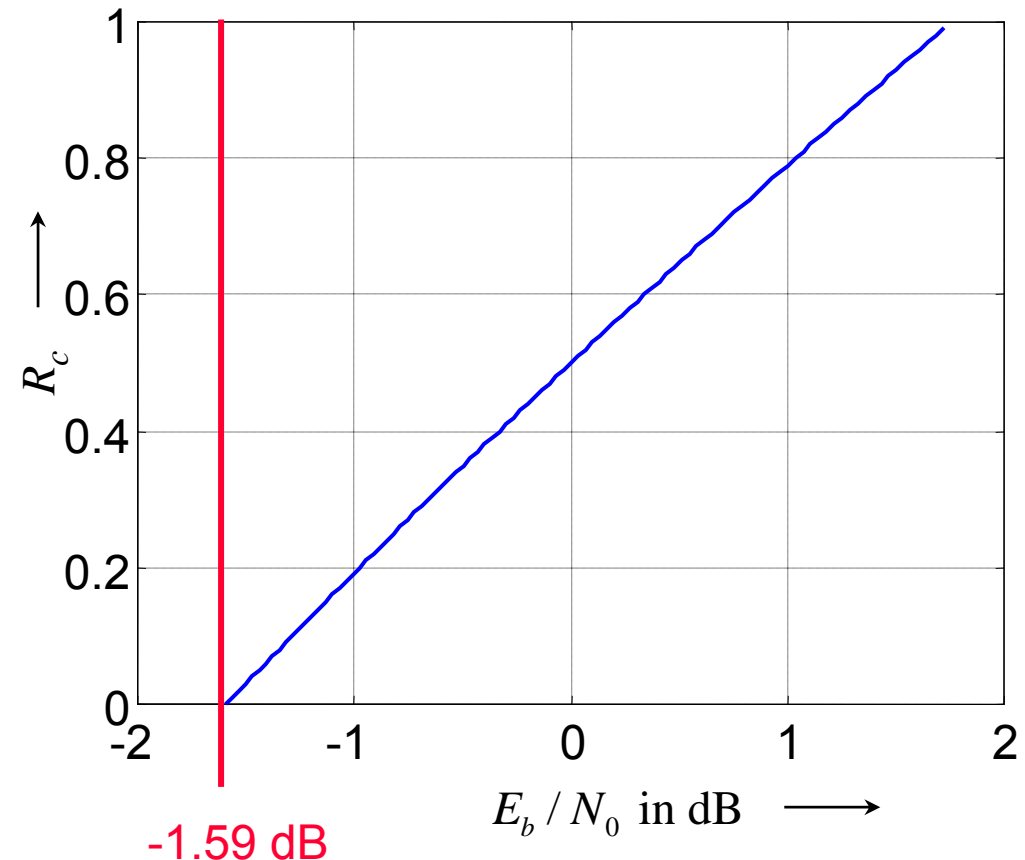
Ultimate Communication Limit

- Energy per information bit: $E_b = E_s / C \Rightarrow E_s = C \cdot E_b$
- Capacity of 1-D AWGN channel

$$\begin{aligned} C &= \frac{1}{2} \cdot \log_2 \left(1 + 2 \cdot \frac{E_s}{N_0} \right) \\ &= \frac{1}{2} \cdot \log_2 \left(1 + 2C \cdot \frac{E_b}{N_0} \right) \end{aligned}$$

- Minimum signal to noise ratio for information bits

$$\begin{aligned} \frac{E_b}{N_0} &= \frac{2^{2C} - 1}{2C} \xrightarrow{C \rightarrow 0} \ln(2) \\ &\approx -1.59 \text{ dB} \end{aligned}$$



Thanks for your attention!