Channel Coding

Consider transmission of blocks of length N.
Denote:

$$\boldsymbol{X}_N = (X_1, \dots, X_N) \text{ input random vector of length } N$$
$$\boldsymbol{Y}_N = (Y_1, \dots, Y_N) \text{ output random vector of length } N$$

where $X_1, \dots, X_N \in \mathcal{X}$, $Y_1, \dots, Y_N \in \mathcal{Y}$.

Only a subset of all possible blocks of length N is used as input, the channel code.

Definition 4.9.
A set of M codewords of length N, denoted by

$$\mathcal{C}_N = \{\boldsymbol{c}_1, \dots, \boldsymbol{c}_M\} \subseteq \mathcal{X}^N$$

is called (N, M)-code.

$$R = \frac{\log M}{N}$$

is called the *code rate*. It represents the number of bits per channel use.

Channel Coding

Transmission is characterized by

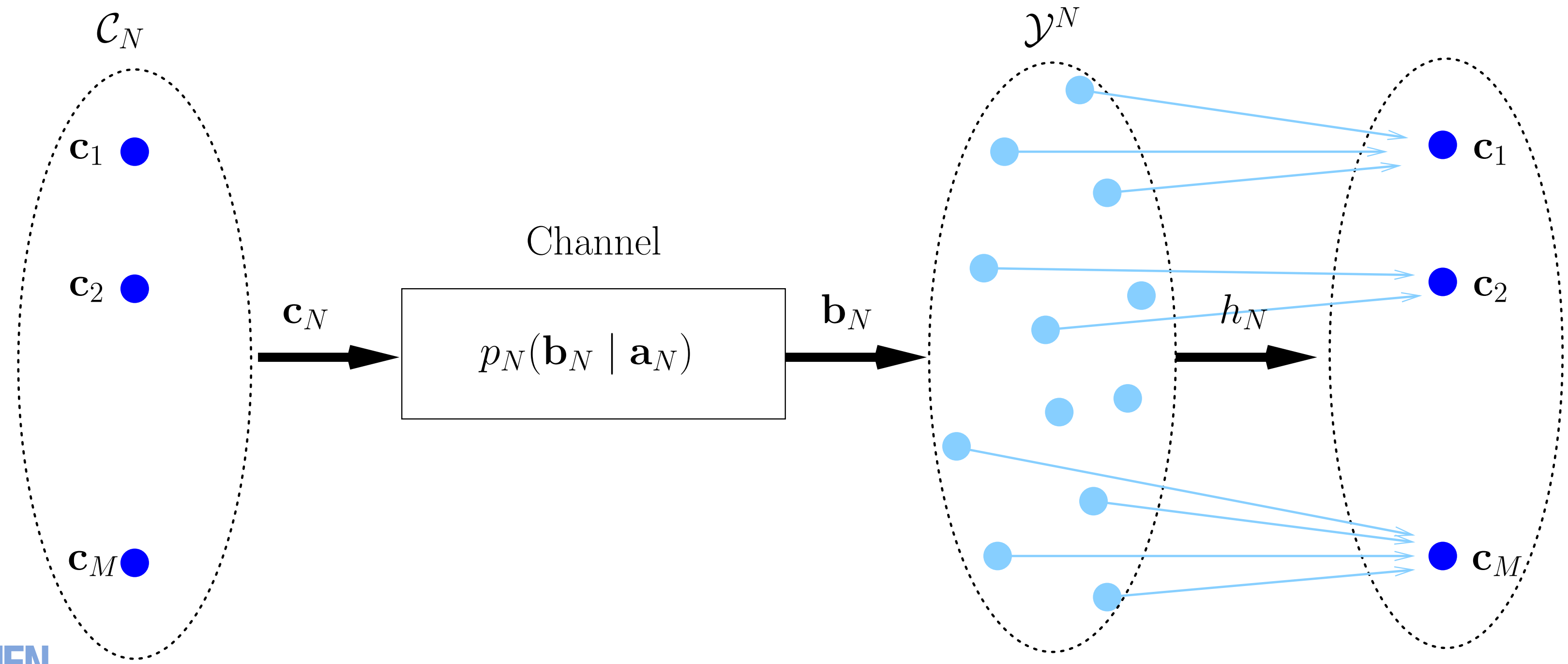
- the channel code $\mathcal{C}_N = \{\boldsymbol{c}_1, \dots, \boldsymbol{c}_M\}$
- transmission probabilities

$$p_N(\boldsymbol{b}_N \mid \boldsymbol{a}_N) = P\big(\boldsymbol{Y}_N = \boldsymbol{b}_N \mid \boldsymbol{X}_n = \boldsymbol{a}_N\big)$$

- the decoding rule

$$h_N : \mathcal{Y}^N \to \mathcal{C}_N : \boldsymbol{b}_N \mapsto h_N(\boldsymbol{b}_N)$$

Graphically:



Decoding Rules

Definition 4.10.

A decoding rule $h_N : \mathcal{Y}^N \to \mathcal{C}_n$ is called *minumum error rule (ME)* or *ideal observer* if

$$\boldsymbol{c}_j = h_N(\boldsymbol{b}) \Rightarrow P(\boldsymbol{X}_N = \boldsymbol{c}_j \mid \boldsymbol{Y}_N = \boldsymbol{b}) \geq P(\boldsymbol{X}_N = \boldsymbol{c}_i \mid \boldsymbol{Y}_N = \boldsymbol{b})$$

for all $i = 1, \ldots, M$. Equivalently,

$$\begin{aligned}\boldsymbol{c}_j = h_N(\boldsymbol{b}) \Rightarrow P(\boldsymbol{Y}_N = \boldsymbol{b} \mid \boldsymbol{X}_N = \boldsymbol{c}_j)P(\boldsymbol{X}_N = c_j) \\ \geq P(\boldsymbol{Y}_N = \boldsymbol{b} \mid \boldsymbol{X}_N = \boldsymbol{c}_i)P(\boldsymbol{X}_N = c_i)\end{aligned}$$

for all $i = 1, \ldots, M$.

With ME-decoding, $\boldsymbol{b}$ is decoded as the codeword $\boldsymbol{c}_j$ which has greatest conditional probability of having been sent given $\boldsymbol{b}$ is received. Hence,

$$h_N(\boldsymbol{b}) \in \arg \max_{i=1,\ldots,M} P\big(\boldsymbol{X}_N = \boldsymbol{c}_i \mid \boldsymbol{Y}_N = \boldsymbol{b}\big).$$

ME decoding rules depend on the input distribution. This is avoided by maximum likelihood decoding, see next slide.

Decoding Rules

Definition 4.11.
A decoding rule $h_N : \mathcal{Y}^N \to \mathcal{C}_n$ is called *maximum likelihood rule (ML)* if

$$\boldsymbol{c}_j = h_N(\boldsymbol{b}) \Rightarrow P(\boldsymbol{Y}_N = \boldsymbol{b} \mid \boldsymbol{X}_N = \boldsymbol{c}_j) \geq P(\boldsymbol{Y}_N = \boldsymbol{b} \mid \boldsymbol{X}_N = \boldsymbol{c}_i)$$

for all $i = 1, \dots, M$.

With ML-decoding, $\boldsymbol{b}$ is decoded as the codeword $\boldsymbol{c}_j$ which has greatest conditional probability of $\boldsymbol{b}$ being received given that $\boldsymbol{c}_j$ was sent. Hence,

$$h_N(\boldsymbol{b}) \in \arg \max_{i=1,\dots,M} P\big(\boldsymbol{Y}_N = \boldsymbol{b} \mid \boldsymbol{X}_N = \boldsymbol{c}_i\big).$$

Error Probabilities

For a given Code $\mathcal{C}_N = \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_M\}$,

- $$e_j(\mathcal{C}_N) = P\big(h_N(\boldsymbol{Y}_N) \neq \boldsymbol{c}_j \mid \boldsymbol{X}_N = \boldsymbol{c}_j\big)$$

  is the probability for a decoding error of code word $\boldsymbol{c}_j$.

- $$e(\mathcal{C}_N) = \sum_{j=1}^{M} e_j(\mathcal{C}_N)\, P(\boldsymbol{X}_N = \boldsymbol{c}_j)$$

  is the error probability of code $\mathcal{C}_N$.

- $$\hat{e}(\mathcal{C}_N) = \max_{j=1,\ldots,M} e_j(\mathcal{C}_N)$$

  is the maximum error probability.

Discrete Memoryless Channel

Definition 4.12.
A discrete channel is called *memoryless (DMC)* if

$$P\big(\boldsymbol{Y}_N = \boldsymbol{b}_N \mid \boldsymbol{X}_N = \boldsymbol{a}_N\big) = \prod_{i=1}^{N} P\big(Y_1 = b_i \mid X_1 = a_i\big)$$

for all $N \in \mathbb{N}$, $\boldsymbol{a}_N = (a_1, \ldots, a_N) \in \mathcal{X}^N$, $\boldsymbol{b}_N = (b_1, \ldots, b_N) \in \mathcal{Y}^N$.

Lemma 4.13.
From the above definition it follows that the channel

- is memoryless and nonanticipating
- transition probablities of symbols are the same at each position
- transition probabilities of blocks only depend on the channel matrix

The Noisy Coding Theorem

Theorem 4.14. (Shannon 1949)

Given some discrete memoryless channel of capacity C. Let $0 < R < C$ and $M_N \in \mathbb{N}$ be a sequence of integers such that the rate

$$\frac{\log M_N}{N} < R. \qquad \left(\Leftrightarrow M_n < m^{NR} \right)$$

There exists a sequence of (N, M_N)-codes with M_N codewords of length N and a constant $a > 0$ such that

$$\hat{e}(\mathcal{C}_N) \leq e^{-Na}.$$

Hence, the maximum error probability tends to zero exponentially fast as the block length N tends to infinity.

Example: BSC

Consider the BSC with $\varepsilon = 0.03$.

$$C = 1 + (1-\varepsilon)\log_2(1-\varepsilon) + \varepsilon \log_2 \varepsilon = 0.8056$$

Choose $R = 0.8$

$$\frac{\log_2 M_N}{N} < R \Leftrightarrow M_N < 2^{NR}$$

hence choose

$$M_N = \lfloor 2^{0.8N} \rfloor$$

N	10	20	30
$\lvert\mathcal{X}^N\rvert = 2^N$	1 024	1 048 576	$1.0737 \cdot 10^9$
$M_N = \lfloor 2^{0.8N} \rfloor$	256	65 536	$16.777 \cdot 10^6$
Percentage of used codewords	25%	6.25%	1.56%

Strong Converse of the Noisy Coding Theorem

Theorem 4.15. (Wolfowitz 1957)

Given some discrete memoryless channel of capacity C. Let $R > C$ and $M_N \in \mathbb{N}$ be a sequence of integers such that

$$\frac{\log M_N}{N} > R.$$

For any sequence of (N, M_N)-codes with M_N codewords of length N it holds that that

$$\lim_{N \to \infty} e(\mathcal{C}_N) = 1.$$

Hence, such codes tend to be fully unreliable.