Híradástechnika szigorlat
Kidolgozott tételek
INFORMATION AND CODING THEORY

Csutak Balázs

2017/18/2. félév

Contents

1 Describe the discrete memoryless source model (sampling, quantization, optimal Lloyd Max quantization)

The discrete memoryless source model can be seen as the result of digitization of a continuous analog signal. This digitization process involves three steps: sampling, quantization and coding.

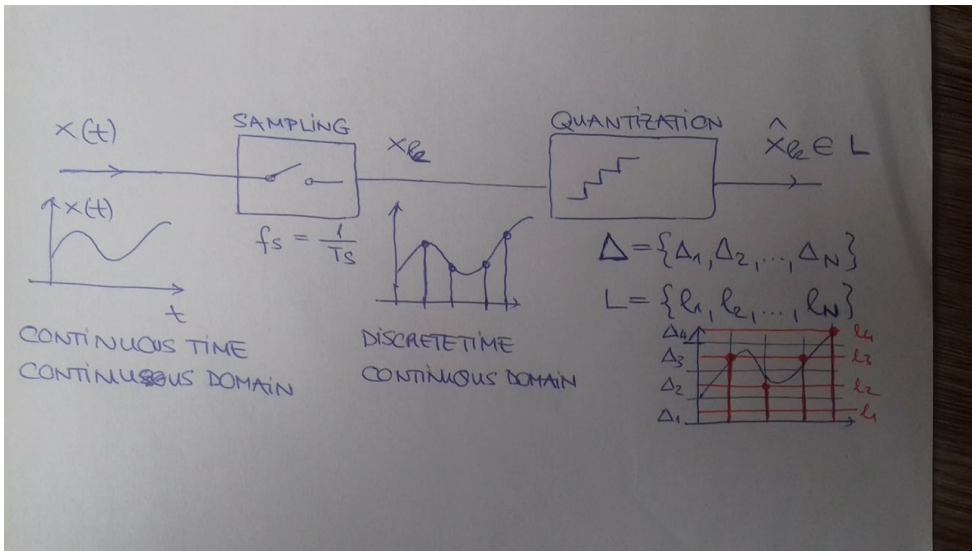

Figure 1: Discrete memoryless source model

Sampling

Definition. Let $x(t)$ be a continuous analog signal. The *sampled signal* is defined as: $x_k = x(t_0 + k \cdot T)$, where T is the sampling time and t_0 is the start time of the process.

Definition. The sampling is called *lossless*, if the original signal $x(t)$ can be fully reconstructed from the sampled x_k signal.

Definition. The *bandwidth* of a signal $x(t)$ is B, if the signal can be obtained from it's Fourier-transform as: $x(t) = \int_{-B}^{B} X(f) \cdot e^{j2\pi ft} df$.

Theorem. The sampling of a band-limitid $x(t)$ signal with bandwidth B is lossless, iff $2B \leq \frac{1}{T}$.

Proof. Let $x(t)$ be a continuos analog signal, with bandwidth B. This means:

$$x(t) = \int_{-B}^{B} X(f) e^{j2\pi ft} df$$

and

$$X(f) = 0 \forall f \notin [-B, B]$$

The spectrum of the sampled signal can be written as:

$$X_m(f) = \sum_{k=-\infty}^{+\infty} X(f + \frac{k}{T})$$

In order to losslessly reconstruct the original signal from the sampled one, we need $X(f) = X_m(f)$.
This means $\frac{1}{T} \geq 2B$. □

Note. Reconstruction of the signal happens by applying a lowpass filter to a series of Dirac-delta pulses:

$$x(t) = \sum_{k=-\infty}^{+\infty} x_k \cdot h(t - kT), \text{ where } h(t) = \frac{\sin(2\pi Bt)}{2\pi Bt}.$$

Quantization

Definition. The *quantization* of a sampled signal x_k is the process of mapping the continuous domain of the signal values to a discrete set of predefined quantization levels.

Definition. The quality of the quantization can be measured by its *signal-to-noise ration*, defined as: $SQNR = \frac{P_{signal}}{P_{noise}}$, where P is the power of the signal and the noise respectively.

In the following paragraphs we assume quantization happens between signal levels $[-C, C]$ with a number of N quantization levels noted by L=$\{l_1, l_2, ..., l_N\}$ and intervals $\Delta = \{\Delta_1, \Delta_2, ..., \Delta_N\}$. Endpoints of the intervals are noted by $\{x_0, x_1, \ldots, x_N\}$.

Equidistant quantization

Definition. The quantization is called *equidistant*, iff $\Delta = x_{k+1} - x_k = \frac{2C}{N} \forall k = 1...N$.

Theorem. The SQNR of the equidistant quantization: $SQNR = \frac{3}{2}2^{2n}$

Note. This is only true in case of completely kivezérelt signal with uniform probability distribution.

Proof. We note the quantization error: $\epsilon = x - \hat{x} \in [-\frac{\Delta}{2}, \frac{\Delta}{2}]$.

$$P_{noise} = \mathbb{E}(\epsilon^2) = \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} x^2 \cdot P(x)dx = \frac{\Delta^2}{12}$$

$$SQNR = \frac{P_{signal}}{P_{noise}} = \frac{C^2/2}{\Delta^2/12} = \frac{3}{2}N^2$$

□

Logarithmic quantization

In practical applications the signal's distribution is neither uniform, nor known. Logarithmic quantization ensures a suboptimal SQNR independent of the signal's properties.

Definition. The quantization is *logarithmic*, if quantization levels and intervals are distorted by the logarithmic function.

Theorem. The SQNR of the logarithmic quantization is:

$$SQNR = K \cdot \frac{\int_{-C}^{C} x^2 p(x)dx}{\int_{-C}^{C} \frac{1}{l'^2(x)} p(x)dx}$$

Proof.

$$\frac{d}{dx} l(x) \approx \frac{\Delta y}{\Delta x} \Rightarrow \Delta x = \frac{\Delta y}{l'(x)} = \frac{2C}{N \cdot l'(x)}$$

$$P_{noise} = \sum_{\Delta_i} \mathbb{E}(\epsilon | x \in \Delta_i) \cdot P(x \in \Delta_i) = \frac{1}{12} \sum_{\Delta_i} \Delta_i^2 p(x) = K \cdot \int_{-C}^{C} \frac{1}{l'^2(x)} p(x)dx$$

□

Optimal Lloyd-Max quantization

In this case, we try to find the optimal quantization levels, using an adaptive approach.

Definition. The optimal quantization levels are defined as: $l_{opt} : \max SQNR$. These levels are those minimizing the following criteria: $F(\Delta, L) = \sum_{i=1}^{n} \int_{\Delta_i} (x - l_i)^2 p(x)dx$.

The algorithm is iteratieve with the following two steps:

1. Calculate new intervals: $\Delta_{l,opt} = \{x : (x - q_l)^2 < (x - q_i)^2\}$.
2. Calculate new quantization levels: $q_{l,opt} = \mathbb{E}(x | x \in \Delta_l)$.

The algorithm can stick in local minimums, so it is *not* guaranteed that it converges to the optimal solution.

2 Derive the Nyquist criterion for ISI free communication over band limited channels

Theorem. The Nquist criterion states, that there is no intersymbol interference (ISI) over a band-limited chanel if and only if:

$$\frac{1}{T}\sum_{k=-\infty}^{+\infty} H(f-\frac{k}{T}) = 1,$$

where $H(f)$ is the frequency response function of the channel.

Proof. Let the impulse response function of the channel be $h(t)$.
This means, the recieved symbol y_l for a transmission of the x_k signal can be written as:

$$y_l = h_k * x_k = x_l \cdot h_0 + \underbrace{\sum_{k\neq l} x_k \cdot h_{l-k}}_{\text{ISI}}$$

The time-domain condition for ISI-free communication:

$$h_k = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases}$$

In order to transform the condition to the frequency domain, let's see it first in continuos time:

$$h(t) \cdot \sum_k \delta(t - kT) = \delta(t)$$

Now, Fourier-transforming both sides gives the Nquist criterion:

$$H(f) * \frac{1}{T}\sum_k \delta(f-\frac{k}{T}) = 1 \Leftrightarrow \frac{1}{T}\sum_k H(f-\frac{k}{T}) = 1$$

□

3 Describe the memoryless channel model (AWGN channel and BSC), derive the bit error probability as a function of the signal-to-noise ratio.

Channel model

Rajz!!

Definition. A channel is called *memoryless*, if the symbol received depends only on the symbol being transmitted in the very same time.

$$P(v_k = y_i | c_k = x_i, c_{k-1} = x_{i-1}, ...) = P(v_k = y_i | c_k = x_i)$$

BSC

Definition. A *binary simmetric channel* is a communication channel model, in which binary data is being transmitted, and probability of a bit being received incorrectly is independent of the bit itself.

Definition. These channels can be described by their *bit error probability*, defined as:

$$P_b = P(v = 1 | c = 0) = P(v = 0 | c = 1).$$

AWGN

Definition. An *Additive White Gaussian Noise* channel is a communication model, in which the received signal is modeled as the sum of a random variable with normal distribution and the transmitted symbol. The channel can be described by the σ^2 deviance of the distribuiton.

Definition. The *signal-to-noise ratio* of the AWGN channel: $SNR_{[\text{dB}]} = 10 \cdot \log\left(\frac{P_{signal}}{\sigma^2}\right)$.

Theorem. The bit error probability of a BSC can be calculated as a function of the SNR as follows:

$$P_b = \Phi\left(-0.5 \cdot \sqrt{\frac{10^{\frac{1}{10} SNR_{[\text{dB}]}}}{P_{signal}}}\right)$$

Proof. Let's assume we have a treshhold 0.5 for the symbols (which means a received value below 0.5 is rounded to 0, a value above is rounded to 1).
This means:

$$P(v = 0 | c = 1) = P(e < -0.5) = \Phi\left(\frac{-0.5}{\sigma}\right) \text{ and}$$

$$P(v = 1 | c = 0) = P(e > 0.5) = 1 - \Phi\left(\frac{0.5}{\sigma}\right) = \Phi\left(\frac{-0.5}{\sigma}\right).$$

$$P_b = P(c=0) \cdot P(v=1|c=0) + P(c=1) \cdot P(v=0|c=1) = \Phi\left(\frac{-0.5}{\sigma}\right).$$

Now, we can express the deviance from the SNR:

$$SNR_{[\text{dB}]} = 10 \cdot \log\left(\frac{P_{signal}}{\sigma^2}\right) \Rightarrow \sigma = \sqrt{\frac{P_{signal}}{10^{\frac{1}{10}SNR_{[\text{dB}]}}}}$$

By substitution we arrive to the desired expression.

□

4 Define and describe the properties of entropy, joint entropy, conditional entropy and mutual information

Definition. Information of a symbol is defined as: $\mathrm{I}(x) = \mathrm{ld}\,\frac{1}{p(x)}$

Definition. The entropy of a source is the average information of the symbols emitted:

$$\mathrm{H}(X) = \mathbb{E}(\mathrm{I}(x)) = \sum_{x\in X} p(x)\cdot \mathrm{I}(x) = \sum_{x\in X} p(x)\,\mathrm{ld}\,\frac{1}{p(x)}$$

Theorem. For any source: $0 \le \mathrm{H}(X) \le \mathrm{ld}\,N$

Proof. Insert proof here □

Definition. Joint entropy of two sources: $\mathrm{H}(X,Y) = \sum_x \sum_y p(x,y)\,\mathrm{ld}\,\frac{1}{p(x,y)}$

Definition. Conditional entropy of two sources: $\mathrm{H}(X|Y) = \sum_x \sum_y p(x,y)\,\mathrm{ld}\,\frac{1}{p(x|y)}$

Theorem. The joint entropy: $\mathrm{H}(X,Y) = \mathrm{H}(X) + \mathrm{H}(Y|X) = \mathrm{H}(Y) + \mathrm{H}(X|Y)$

Proof. The proof of the first part, based on: $p(x,y) = p(x)\cdot p(y|x)$

$$\begin{aligned}
\mathrm{H}(X,Y) &= \sum_x \sum_y p(x,y)\,\mathrm{ld}\,\frac{1}{p(x,y)} = \sum_x \sum_y p(x,y)\left(\mathrm{ld}\,\frac{1}{p(x)} + \mathrm{ld}\,\frac{1}{p(y|x)}\right)\\
&= \sum_x \sum_y p(x,y)\,\mathrm{ld}\,\frac{1}{p(x)} + \sum_x \sum_y p(x,y)\,\mathrm{ld}\,\frac{1}{p(y|x)}\\
&= \sum_x p(x)\,\mathrm{ld}\,\frac{1}{p(x)} + \sum_x \sum_y p(x,y)\,\mathrm{ld}\,\frac{1}{p(y|x)} = \mathrm{H}(X) + \mathrm{H}(Y|X)
\end{aligned}$$

Second part can be proven using: $p(x,y) = p(y)\cdot p(x|y)$ □

Note. For independent sources, this means: $\mathrm{H}(X,Y) = \mathrm{H}(X) + \mathrm{H}(Y)$

Definition. The Kullback-Leibler distance of two probability distributions:

$$\mathbb{D}(p(x)||q(x)) = \sum_x p(x)\,\mathrm{ld}\,\frac{p(x)}{q(x)}$$

Definition. Mutual information of two sources: $\mathrm{I}(X,Y) = \mathbb{D}(p(x,y)||p(x)\cdot(y))$

Theorem. The mutual information: $\mathrm{I}(X,Y) = \mathrm{H}(X) - \mathrm{H}(X|Y) = \mathrm{H}(Y) - \mathrm{H}(Y|X)$

Proof. Proof of the first part.

$$\mathrm{I}(X,Y) = \sum_x \sum_y p(x,y) \operatorname{ld} \frac{p(x,y)}{p(x)p(y)} = \sum_x \sum_y p(x,y) \operatorname{ld} \frac{p(x|y)p(y)}{p(x)p(y)}$$
$$= \sum_x p(x) \operatorname{ld} \frac{1}{p(x)} - \sum_x \sum_y p(x,y) \operatorname{ld} \frac{1}{p(x|y)}$$

Second part can be proven using $p(x,y) = p(x) \cdot p(y|x)$ □

Note. Theorems can be easily memorized based on Figure 2

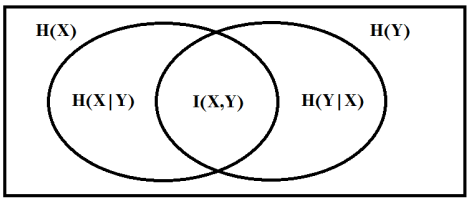

Figure 2: Entropy, conditional entropy and mutual information

5 Define the typical set of an IT source (AEP) and derive its properties.

Definition. Asymptotic equipartition property (AEP): $\lim_{n\to\infty} \frac{1}{n} \operatorname{ld} p(x_1, ..., x_n) = -\operatorname{H}(X)$

Definition. The typical set belonging to the AEP IT source X is defined as:

$$A = \{\mathbf{x} = (x_1, x_2, ..., x_n) \mid 2^{-n\operatorname{H}(X)-\varepsilon} \leq p(\mathbf{x}) \leq 2^{-n\operatorname{H}(X)+\varepsilon}\}$$

Theorem. Probability of a random vector belonging to the set: $(1-\varepsilon) \leq P(\mathbf{x} \in A) \leq 1$

Proof. Stochastic convergence of the AEP definition implies:

$$P(|\operatorname{H}(x) - \frac{1}{n} ld \frac{1}{p(x_1, ..., x_n)}| < \varepsilon) > 1 - \varepsilon$$

□

Theorem. The size of the typical set: $(1-\varepsilon) \cdot 2^{n\operatorname{H}(X)-\varepsilon} \leq |A| \leq 2^{n\operatorname{H}(X)+\varepsilon}$

Proof. Proof of the left side:

$$|A| \cdot 2^{-n\operatorname{H}(X)+\varepsilon} \geq \sum_{x\in A} p(\mathbf{x}) \geq (1-\varepsilon) \Rightarrow (1-\varepsilon) \cdot 2^{n\operatorname{H}(X)-\varepsilon} \leq |A|$$

Proof of the right side:

$$|A| \cdot 2^{-n\operatorname{H}(X)-\varepsilon} \leq \sum_{x\in A} p(\mathbf{x}) \leq \sum_{x} p(\mathbf{x}) = 1 \Rightarrow |A| \leq 2^{n\operatorname{H}(X)+\varepsilon}$$

□

Theorem. The probability density function defined above the typical set is approixi-mately uniform.

6 Define the properties of uniquely decodable codes.

Definition. A code is unequely decodable if it's prefix free, meaning that $\forall i \neq j : c_i \not\prec c_j$

Definition. The average code length of a prefix-free code: $L = \sum_x p(x) \cdot l(x)$

Theorem. (*Kraft inequality*) For any prefix-free code: $\sum_x 2^{-l(x)} \leq 1$

Proof. Let the codeword lengths be $L = (l_1 \leq l_2 \leq ... \leq l_N)$. We construct a prefix-free code step-by-step, choosing the codeword for each symbol. At step k there exist 2^{l_k} possible codewords. To avoid one of the previously chosen codewords to be prefix of this one, $\sum_{i=1}^{k-1} 2^{l_k - l_i}$ codewords are forbidden.
To be able to construct the code, we must have a codeword to choose at each step, which means:

$$\sum_{i=1}^{k-1} 2^{l_k - l_i} + 1 \leq 2^{l_k} \Leftrightarrow \sum_{i=1}^{k} 2^{l_k - l_i} \leq 2^{l_k} \Leftrightarrow \sum_{i=1}^{k} 2^{-l_i} \leq 1$$

□

Note. This is a special case of the Kraft-McMillian inequality, which states the following: Given a list of positive integers $(n_1, n_2, ..., n_r)$ there exists a uniquely decodable code with these codeword lengths, if and only if:

$$\sum_{i=1}^{r} s^{-n_i} \leq 1$$

where s is the alphabet size (2 for binary codes).

7 Discuss the source coding theorem

Theorem. Source coding theorem: $\mathrm{H}(X) \leq L = \sum_x p(x) \cdot l(x)$

Proof. Let $q(x) = \frac{2^{-l(x)}}{\sum_y 2^{-l(y)}}$ an artificial probability distribution. As $\sum_x q(x) = 1$, this is a distribution indeed.
The Kullback-Leibler distance of the distributions:

$$\begin{aligned}\mathbb{D}(p(x)||q(x)) &= \sum_x p(x) \operatorname{ld} \frac{p(x) \sum_y 2^{-l(y)}}{2^{-l(x)}} \leq^* \sum_x p(x) \operatorname{ld} \frac{p(x) \cdot 1}{2^{-l(x)}} = \\ &= \sum_x p(x) \{\operatorname{ld} p(x) - \operatorname{ld} 2^{-l(x)}\} = -\sum_x p(x) \frac{1}{\operatorname{ld} p(x)} + \sum_x p(x) l(x) = \\ &= -\mathrm{H}(x) + L\end{aligned}$$

Now, $0 \leq \mathbb{D}(p(x)||q(x)) \Rightarrow 0 \leq -\mathrm{H}(x) + L \Rightarrow \mathrm{H}(x) \leq L$.
*using the Kraft inequality for uniquely decodable codes (page 12). □

8 Describe the Shannon-Fano, Huffman, and arithmetic coding and discuss their performance

Shanon-Fano code

Definition. Code length of an SF code: $l(x) = \left\lceil \mathrm{ld}\, \frac{1}{p(x)} \right\rceil$

Theorem. Bounds for the code lenght: $\mathrm{H}(X) \leq L_{SF} \leq \mathrm{H}(X) + 1$

Shanon-Fano block code

Definition. Symbols of source Y are defined as a vector of K consecutive symbols of source X: $y = (x_1, x_2, ..., x_K)$. As X is memoryless source, $x_1, ..., x_K$ are independent variables, meaning that $p(y) = \prod_{i=1}^{K} p(x_i)$

Theorem. Bounds for the SF block code length: $\mathrm{H}(X) \leq L_{SF}^{NEW}(X) \leq \mathrm{H}(X) + \frac{1}{K}$

Proof.

$$K \cdot \mathrm{H}(X) = \mathrm{H}(Y) \leq L_{SF}^{BLOCK}(Y) \leq \mathrm{H}(Y) + 1 = K \cdot \mathrm{H}(X) + 1$$
$$\mathrm{H}(X) = \frac{1}{K}\mathrm{H}(Y) \leq L_{SF}^{NEW}(X) \leq \frac{1}{K}\mathrm{H}(Y) + \frac{1}{K} = \mathrm{H}(X) + \frac{1}{K}$$

□

Note. By using block codes, SF is asymptotically optimal.

Note. Limits of block coding: LUT complexity becomes $\mathcal{O}(N^K)$

Huffmann code

The technique works by creating a binary tree of nodes. A node can be either a leaf node or an internal node. Initially, all nodes are leaf nodes, which contain the symbol itself, the weight (frequency of appearance) of the symbol. Internal nodes contain a weight, and the links to two child nodes. The process begins with the leaf nodes containing the probabilities of the symbol they represent. Then, the process takes the two nodes with smallest probability, and creates a new internal node having these two nodes as children. The weight of the new node is set to the sum of the weight of the children. We then apply the process again, on the new internal node and on the remaining nodes (i.e., we exclude the two leaf nodes), we repeat this process until only one node remains, which is the root of the Huffman tree.

Theorem. Huffman codes are asymptotically optimal.

Note. Construction of a Huffmann code has $\mathcal{O}(N^2)$ complexity.

Comparison between codes

$$\mathrm{H}(X) \leq L_H \leq L_{SF}$$

Arithmetic coding ???

9 Describe the LZ based compression algorithms

LZ77 and LZ78 are the two lossless data compression algorithms, which does not require the knowledge of the IT source's probability distribution. Both of them are dictionary-based coders, with slightly different implementation. The LZ78 constructs an explicit dictionary for the whole data, LZ77 operates with local dictionaries obtained by the usage of a sliding window. The compression methods are equivalent if the entire data is to be compressed and decomressed at a time. Both algorithms are asymptotically optimal.

10 Define the channel capacity and elaborate on its calculation for symmetric channels

Definition. Channel capacity (denoted by C) is the tight upper bound on the rate at which information can be reliably transmitted over a communications channel.

Theorem. Channel coding theorem: $C = \max_{p(x)} I(x,y) = \max_{p(x)}\{\mathrm{H}(Y) - \mathrm{H}(Y|X)\}$

Proof. See detailed description in section 11 (page 18). □

Capacity of symmetric channels

Definition. The transition matrix of a channel: $\mathbf{P} : p_{ij} = P(Y = y_j | X = x_i)$.

Note. This means the sum of elements in a column is always 1.

Note. The probability of a received symbol: $P(Y = y_j) = \sum_{i=1}^{N} \mathbf{P}_{ij} \cdot P(X = x_i)$

Note. Transition matrix of BSC: $\mathbf{P} = \begin{pmatrix} 1-P_b & P_b \\ P_b & 1-P_b \end{pmatrix}$

Definition. A channel is symmetric, if columns of $\mathbf{P}$ are permutations of each other.

Definition. Conditional entropy of the symmetric channel: $\mathrm{H}(\mathbf{r}) = \sum_{i=1}^{N} \mathbf{P}_{ij} \operatorname{ld} \frac{1}{\mathbf{P}_{ij}}$

Note. As columns are permutations of each other, this is independent of j.

Theorem. Capacity of symmetric channel: $C = \operatorname{ld} N - \mathrm{H}(\mathbf{r})$

Proof. We simply substitute $\operatorname{ld}(N)$ and $\mathrm{H}(\mathbf{r})$ into $\mathrm{H}(Y)$ and $\mathrm{H}(Y|X)$ respectively. □

Capacity of BSC - naive approach

Definition. Entropy of the error source: $\mathrm{H}(P_b) = P_b \cdot \operatorname{ld} \frac{1}{P_b} + (1-P_b) \cdot \operatorname{ld} \frac{1}{1-P_b}$

In this approach, we make the assumption, that the error vector $\mathbf{e}$ generated by the channel is known at transmitter side. In order to ensure reliable transmission, this error data is also sent over the channel, compressed as good as possible. Moreover, error data corrupting this additional transmission is also compressed and sent (recursively).
This means, in order to transmit k bits of useful information, we need:

$$n = k + k \cdot \mathrm{H}(P_b) + k \cdot \mathrm{H}(P_b)\,\mathrm{H}(P_b) + ... = k \cdot \sum_{i=0}^{\infty} \mathrm{H}^i(P_b) = k \cdot \frac{1}{1-\mathrm{H}(P_b)}$$

Rearranging the equation gives maximum channel capacity: $\frac{k}{n} \leq 1 - H(P_b)$.

11 Describe the channel coding theorem

Theorem. Channel coding theorem: $C = \max_{p(x)} I(x, y)$

Proof. We have already discussed in section 5 (page 11), that an AEP IT source emits typical sequences "almost all time". Thus, we give a proof for typical sets only.

Given symbols with length k at the transmitter side, we have 2^k symbols.

At the receiver side, each symbol detected falls into a typical set belonging to it, having a size of $2^{n\,\mathrm{H}(Y|X)}$.

Now, as seen in figure 3, the size of all typical sequences must be greater or equal than the sum of these tipical sets belonging to the symbols.

This means, $2^k \cdot 2^{nH(Y|X)} \leq 2^{nH(Y)}$.

By rearranging the equation: $\frac{k}{n} \leq \max_{p(x)} \mathrm{H}(Y) - \mathrm{H}(Y|X) = \max_{p(x)} \mathrm{I}(x, y)$.

□

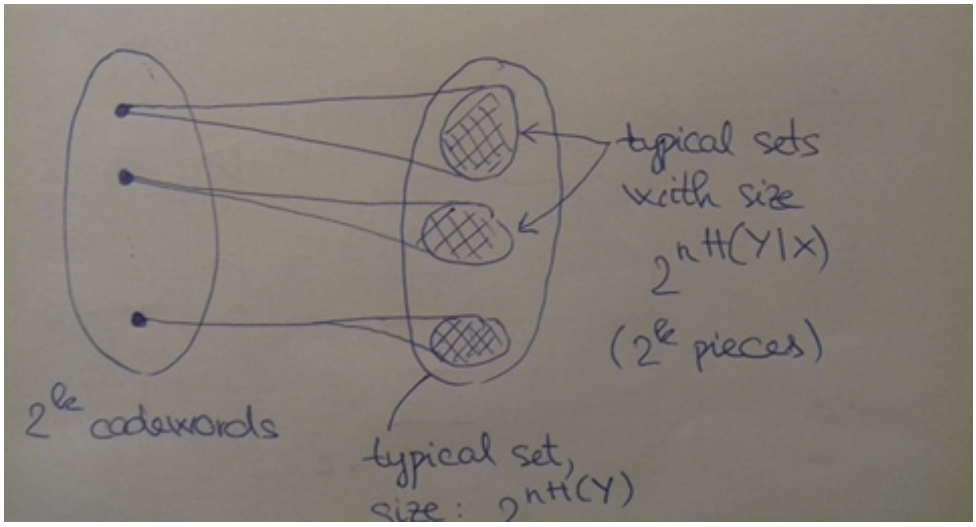

Figure 3: Channel coding theorem: size of typical sets

12 Define and explain the relationship between the following properties and parameters of error correcting codes: minimum code distance; code-length and message-length versus performance (Singleton and Hamming bounds); general algorithmic complexity of coding with tables

Code parameters

Definition. Minimum code distance: $d_{min} : \min_{\mathbf{c} \neq \mathbf{c'}} d(c, c')$

Definition. Number of errors a code can detect: $l = d_{min} - 1$

Definition. Number of errors a code can correct: $t = \left\lfloor \frac{d_{min} - 1}{2} \right\rfloor$

Theorem. Singleton bound (for linear codes): $d_{min} \leq n - k + 1$

Proof. Let $\mathbf{C}$ be a set of binary codewords with minimum distance d. If we delete the first $d-1$ bits of every codeword in $\mathbf{C}$, they should remain different. As the number of distinct codewords with length $n - (d-1)$ is 2^{n-d+1}, the number of elements in $\mathbf{C}$ should be less then this. Rearranging the inequality: $|\mathbf{C}| \leq 2^{n-d+1} \Leftrightarrow 2^k \leq 2^{n-d+1} \Leftrightarrow d \leq n - k + 1$. □

Definition. Maximum distance separable (MDS) code: $d_{min} = n - k + 1$

Theorem. Hamming bound: $\sum_{i=0}^{t} \binom{n}{i} \leq 2^{n-k}$

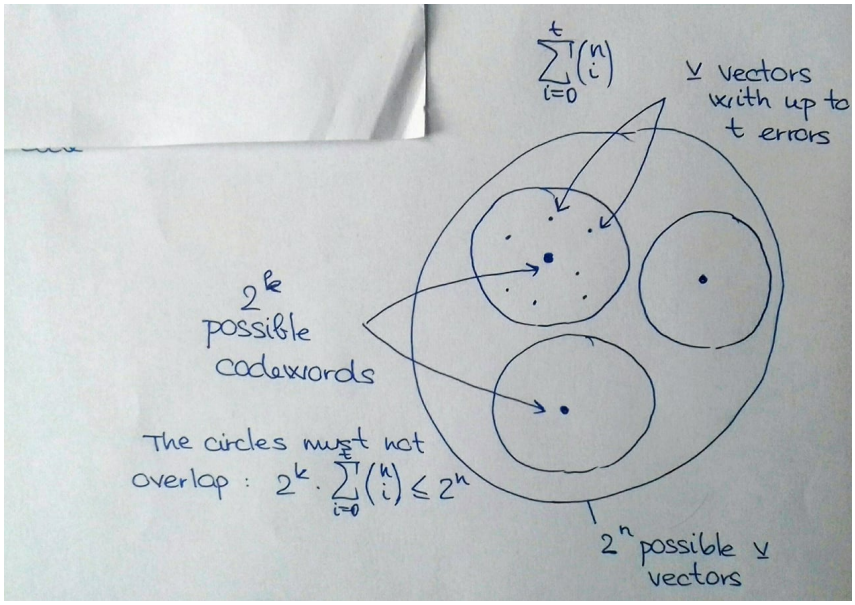

Figure 4: Graphical representation of Hamming bound

Proof. See figure 4. □

Note. Codes which attain to the Hamming-bound are called *perfect codes*.

General coding scheme

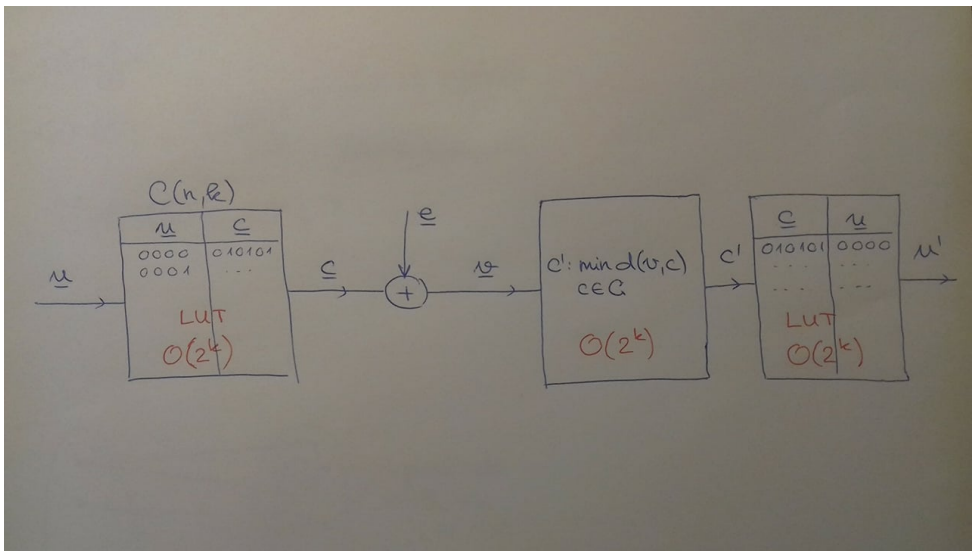

Figure 5: General coding scheme with lookup tables

Complexity of lookup tables and search operation: $3 \cdot \mathcal{O}(2^k)$

Complexity of optimal code construction (offline): $\mathcal{O}\left(\binom{2^n}{2^k}\binom{2^k}{2}\right)$

13 Introduce the concept of linear block coding and explain the meaning of systematic codes; generator matrix, parity check matrix and their relationship; algorithmic complexity of coding with tables

Definition. For a linear $C(n,k)$ code the generator space is defined as:

$$G = \{g^{(1)}, g^{(2)}, ..., g^{(k)}\}, \quad \dim\left(g^{(i)}\right) = n.$$

Definition. The codewords are the linear combination of the elements of G: $C = \mathcal{L}c\{G\}$

Definition. The encoding operation: $\boldsymbol{c} = \sum_{i=1}^{k} u_i \cdot g^{(i)}$.

Theorem. In case of linear codes, $d_{min} = \min_{\boldsymbol{c}\neq\boldsymbol{0}} w(\boldsymbol{c})$

Proof. As codewords form a vector space, the difference between any pair of codewords is a codeword itself. □

Definition. A code is called *systematic*, if codewords can be written as the message itself and some complementary bits:

$$\mathbf{c} = (u_1, u_2, ..., u_k, p_1, p_2, ..., p_{n-k}).$$

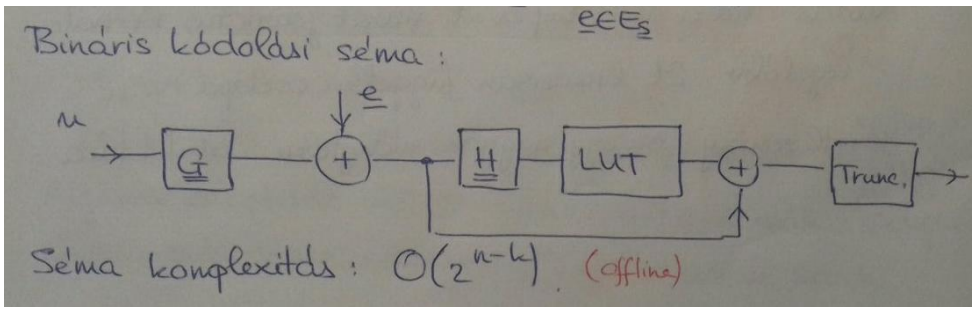

Figure 6: Linear binary coding scheme

Definition. The *generator matrix* of a linear binary code:

$$\boldsymbol{G}_{k\times n} = \begin{bmatrix} g^{(1)} \\ g^{(2)} \\ \dots \\ g^{(k)} \end{bmatrix}$$

Note. The generator matrix of a systematic code:

$$\boldsymbol{G}_{k\times n} = \left[\boldsymbol{I}_{k\times k} | \boldsymbol{B}_{k\times(n-k)}\right]$$

Definition. The *parity check matrix* of the linear binary code is defined as:

$$\boldsymbol{H}_{(n-k)\times n} : \boldsymbol{H} \cdot \boldsymbol{c}^T = \boldsymbol{0} \quad \forall \boldsymbol{c} \in C$$

Theorem. The connection between the two matrices: $\boldsymbol{H} \cdot \boldsymbol{G}^T = \boldsymbol{0}$

Proof.

$$\boldsymbol{H} \cdot \boldsymbol{c}^T = \boldsymbol{H} \cdot (\boldsymbol{u} \cdot \boldsymbol{G})^T = \boldsymbol{H} \cdot \boldsymbol{G}^T \cdot \boldsymbol{u} = 0 \quad \forall \boldsymbol{u} \Rightarrow \boldsymbol{H} \cdot \boldsymbol{G}^T = \boldsymbol{0}.$$

□

Theorem. The parity check matrix of systematic codes:

$$\boldsymbol{H} = \left[\boldsymbol{A}_{(n-k)\times k} | \boldsymbol{I}_{(n-k)\times(n-k)}\right], \text{ where } \boldsymbol{A} = \boldsymbol{B}^T$$

Proof.

$$\boldsymbol{H} \cdot \boldsymbol{G}^T = \boldsymbol{A} \cdot \boldsymbol{I} + \boldsymbol{I} \cdot \boldsymbol{B}^T = \boldsymbol{A} + \boldsymbol{B}^T = \boldsymbol{0} \Rightarrow \boldsymbol{A} = -\boldsymbol{B}^T = \boldsymbol{B}^T.$$

□

Theorem. The key equation: $\boldsymbol{H} \cdot \boldsymbol{v}^T = \boldsymbol{H} \cdot \boldsymbol{e}^T = \boldsymbol{s}^T$.

Proof.

$$\boldsymbol{H} \cdot \boldsymbol{v}^T = \boldsymbol{H} \cdot (\boldsymbol{c} + \boldsymbol{e})^T = \boldsymbol{H} \cdot \boldsymbol{c}^T + \boldsymbol{H} \cdot \boldsymbol{e}^T = \boldsymbol{H} \cdot \boldsymbol{e}^T = \boldsymbol{s}^T.$$

□

Definition. The error group belonging to a syndrome vector $\boldsymbol{s}$:

$$E_{\boldsymbol{s}} = \{\boldsymbol{e} : \boldsymbol{H} \cdot \boldsymbol{e}^T = \boldsymbol{s}^T\}$$

Note. An error group has 2^k elements.

Definition. The error group leader: $\boldsymbol{e_s} = \min_{\boldsymbol{e} \in E_{\boldsymbol{s}}} \mathrm{w}(\boldsymbol{e}) \longrightarrow$ this is uploaded to LUT

Algorithmic complexity of detection (offline): $\mathcal{O}(2^{n-k})$.
Algorithmic complexity of encoding and decoding (online): $\mathcal{O}(n \cdot k)$.
Algorithmic complexity of the whole scheme: 1 LUT (detection) + 2 matrix-vector multiplication + truncation.

14 Give the construction of binary Hamming codes (define the corresponding matrices and the error correcting capability).

Hamming codes are linear codes with error correction capability $t = 1$ (see linear codes in section 13, page 21).
Detection: in case of a single error, the syndrom vector has a total mach with one of the column vectors in $\boldsymbol{H}$.

Definition. The parity check matrix of a Hamming-code:

$$\boldsymbol{H} = [a^{(1)}, a^{(2)}, ..., a^{(k)} | \boldsymbol{I}_{(n-k)\times(n-k)}], \text{ where } \boldsymbol{a}^{(i)} \neq \boldsymbol{0} \text{ and } \boldsymbol{a}^{(i)} \neq \boldsymbol{a}^{(j)} \ \forall i \neq j.$$

Theorem. Hamming codes are perfect, as $\sum_{i=0}^{t} \binom{n}{i} = n + 1 = 2^{n-k}$.

Proof. There exist $2^{n-k} - 1$ syndrome vectors indicating an error. Each single error to be corrected matches with one of the n columns of $\boldsymbol{H}$. □

Example: for the code C(3,1), $3 \leq d_{min} \leq n - k + 1 = 3 \Rightarrow$ this is an MDS code.

Theorem. The bit error probability of a channel encoded with a Hamming-code:

Correct block probability: $(1 - P_b')^k = n \cdot P_b \cdot (1 - P_b)^{n-1} + (1 - P_b)^n$

Based on this: $P_b' = \Psi(P_b, n, k) \leq 10^{-\gamma}$ QoS requirement can be fulfilled.

Multiple error correction

Theorem. If a Hamming-code can correct every t error, $\boldsymbol{A}$ must have at least $2t$ independent column vectors.

Detection rule for $t = 2$: the syndrome vector matches with one of the columns (single error) or with one of the two-element-sums of the column vectors. This is highly inefficient to check.

15 Describe the Reed-Solomon codes (generator matrix, parity check matrix, performance)

Mathematical background: Galois fields

Definition. Let q be a prime number. The *finite field* or *Galois field* over q:

$$GF(q) = \{0, 1, 2, ..., q-1\}$$

Operations $(+, \cdot)$ are defined using modulo q arithmetic.

Definition. The *order* of an element:

$$ord(\alpha) = \min_{1 \leq k} \{\alpha^k = 1\}$$

Theorem. $\forall \alpha \in GF(q) \backslash \{0\} : \alpha^{q-1} = 1$.

Definition. $\alpha \in GF(q)$ is a *primitive element* if $\alpha, \alpha^2, ..., \alpha^{q-1}$ expand the field.
Equivalent definition: α is a primitive element, if $ord(\alpha) = q - 1$.

Theorem. Every Galois field has at least one primitive element.

Definition. A *polynom* over GF(q):

$$a(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2 + \cdots + a_n \cdot x^n$$

This polynom can also be represented as a vector:

$$a(x) \leftrightarrow \boldsymbol{a} = (a_0, a_1, \ldots, a_n).$$

Reed-Solomon codes

The Reed-Solomon codes are defined over $GF(q)$, where q is prime, $n = q - 1$.
As discussed above: $u = (u_0, u_1, \ldots, u_{k-1}) \rightarrow u(x) = u_0 + u_1 \cdot x + \cdots + u_{k-1} \cdot x^{k-1}$.
Encoding operation: $c_i = u(\alpha_i)$.

Definition. Generator matrix of Reed-Solomon code:

$$G_{k \times n} = \begin{bmatrix} 1 & 1 & ... & 1 \\ \alpha_1 & \alpha_2 & ... & \alpha_n \\ \alpha_1^2 & \alpha_2^2 & ... & \alpha_n^2 \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_1^{k-1} & \alpha_2^{k-1} & ... & \alpha_n^{k-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & ... & 1 \\ 1 & \alpha & ... & \alpha^{n-1} \\ 1 & \alpha^2 & ... & \alpha^{2(n-1)} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & \alpha^{k-1} & ... & \alpha_n^{(k-1)(n-1)} \end{bmatrix}$$

Let α be the primitive element of the field, and $\alpha_i = \alpha^i$
Parity check operation: $c(\alpha_i) = 0 \ \ \forall i = 1, 2, ..., n-1$.

Definition. The parity check matrix of the Reed-Solomon code:

$$H_{(n-k)\times n} = \begin{bmatrix} 1 & \alpha & \alpha^2 & ... & \alpha^{(n-1)} \\ 1 & \alpha^2 & \alpha^4 & ... & \alpha^{2(n-1)} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & \alpha^{n-k} & \alpha^{2(n-k)} & ... & \alpha^{(n-1)(n-k)} \end{bmatrix}$$

Theorem. The Reed-Solomon codes are always MDS.

Proof. As $\deg(u(x)) = k-1 \Rightarrow u(x)$ has maximum $k-1$ roots. This means, that $\forall \boldsymbol{c}$ codeword has maximum $k-1$ nonzero components.
As Reed-Solomon codes are linear, $d_{min} = \min_{\boldsymbol{c}} \mathrm{w}(\boldsymbol{c}) \geq n-(k-1) = n-k+1$.
However, due to the singleton bound: $d_{min} \leq n-k+1$.
Consequently, $d_{min} = n-k+1$, which meand this is MDS code indeed. □

The only problem: complexity of the lookup table used: $\mathcal{O}(q^{n-k})$.

16 Describe the steps of the Error Trapping Algorithm for detection in case of cyclic codes

Polynom operations on shift registers

LFFSR: Linear feedforward shift register

This circuit implements multiplication of polynoms over $GF(q)$.

$$v_i = \sum_{j=0}^{n} a_j \cdot u_{i-j} = (a * u)_i \Rightarrow v(x) = a(x) \cdot u(x)$$

LFBSR: Linear feedback shift register

This circuit implements division without remainder of polynoms over $GF(q)$

$$v_i = u_i + (1 - a_0)v_i - \sum_{j=1}^{n} a_j \cdot v_{i-j} \Rightarrow u_i = (a * v)_i \Rightarrow u(x) = a(x) \cdot v(x) \Rightarrow v(x) = \frac{u(x)}{a(x)}$$

With another coefficients, the LFBSR implements division with remainder:

$$v_i = u_i + (1 - a_n)v_i - \sum_{j=1}^{n} a_{n-j} v_{i-j}$$
$$v(x) = r(x), \text{ where } u(x) = a(x) \cdot q(x) + r(x), \deg(r) < \deg(q)$$

Cyclic codes

Definition. The *cyclic shift operation* of a vector:

$$S\boldsymbol{c} = S(c_0, c_1, \ldots, c_{n-1}) = (c_{n-1}, c_0, \ldots, c_{n-2}).$$

Definition. The *cyclic shift operation* as polynom multiplication:

$$S\boldsymbol{c} \to Sc(x) = c'(x) = x \cdot c(x) \mod (x^n - 1) \to \boldsymbol{c}'$$

Theorem. If $\boldsymbol{c} \in C \Rightarrow \boldsymbol{c}' = S\boldsymbol{c} \in C$. Moreover, $\exists g(x)$ generator polynom, with these properties:

1. $\deg(g(x)) = n - k$
2. $g_{n-k} = 1$
3. $\forall c(x) = u(x) \cdot g(x)$
4. $g(x) \mid (x^n - 1)$

Proof. Let $a(x) \in C$ be the minimum degree code polynom, $\deg(a(x)) = m$.
Let $g(x) = a_m^{-1}a(x)$. It can be clearly seen, that $g_m = 1$.
Now, we proove that $g(x)$ is the generator polynom.
First, we proove that $u(x)g(x) \in C \forall u(x)$ with maximum degree $n-1-m$:.

$$u(x)g(x) = u_0g(x)+u_1xg(x)+\dots u_{n-1-m}x^{n-1-m}g(x) = \mathcal{L}c\{g(x), xg(x), \dots, x^{n-1-m}g(x)\} \in C$$

Then, we proove that all code polinoms are multiples of g(x). Assuming $c(x) = g(x)q(x)+r(x) \in C$ exists:

$$r(x) = c(x) - u(x)g(x) = \mathcal{L}c\{c(x), g(x)\} \in C$$

But $\deg(r(x)) < \deg(g(x))$, which leads to contradiction.
Finally, as message polynomials have a degree of $k-1$, $k-1 = n-1-m \Rightarrow m = n-k$.
□

The Error Trapping Algorithm

The ETA is a method for identifying cyclic code error polynomials without use of lookup tables. However it does not provide an equivalent replacement: only error vectors with *generic error configuration* (also called *burst errors*) can be detected this way.
The concept behind the algorithm is the following similarity:

$$v(x) = a(x)g(x) + r(x) \text{ (division with remainder on receiver side)}$$
$$v(x) = u(x)g(x) + e(x) \text{ (directly from the definition)}$$

The condition, which makes $r(x) = e(x)$: $\deg(e(x)) < \deg(g(x)) = n-k$. This is a too strict condition. However, if the error is burst error, so that the first and last error bit in $e(x)$ has a maximum distance of $n-k$, $e(x) = x^{i_0}r(x)$, where i_0 is the first error position.
To find this first error, we can use the *key equation*:

$$v(x) \mod g(x) = e(x) \mod g(x)$$

$$x^{-i}v(x) = a(x)g(x) + r(x)$$
$$x^{-i}e(x) = b(x)g(x) + r(x)$$

As $b(x)g(x)$ is a codeword, it has a minimum weight of $2t+1$.
As we correct up to t errors, weight of $e(x)$ is maximum t.
From the above two statements, $\mathrm{w}(r(x)) \geq t+1$.
This means, $x^{-i_0}e(x) = r(x)$ if and only if $\mathrm{w}(r(x)) \leq t$.

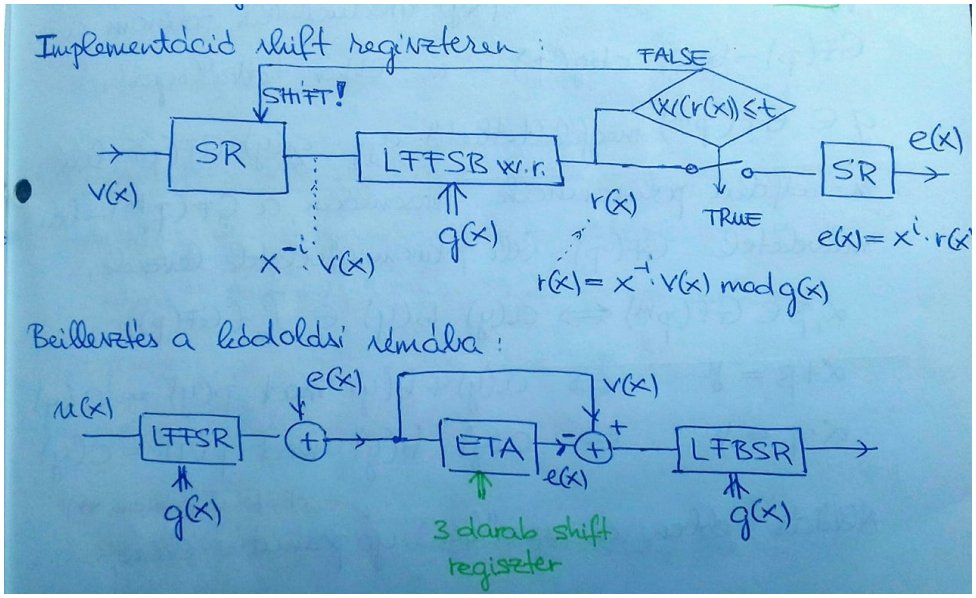

Figure 7: Error trapping scheme

17 Describe the cyclic RS codes (generator polynom, parity check polynom, implementation)

Theorem. The Reed-Solomon codes are cyclic.

Definition. The generator polynom: $g(x) = \prod_{i=1}^{n-k}(x - \alpha^i)$.

Definition. The parity check polynom: $h(x) = \prod_{i=n-k+1}^{n}(x - \alpha^i)$.

Implementation in binary domain

Definition. A polynom over $GF(q)$ is irreducible, if it cannot be written as a product of two lower-degree polynoms.

Theorem. Let $p(y)$ be an irreducible polynom with degree m in $GF(p)$, p prime number. Then, any element q of the field $GF(p^m)$ can be uniquely mapped to a polynom $q(y)$ over $GF(p)$, with the following operation:

$$\begin{aligned} \alpha, \beta \in GF(p^m) &\longleftrightarrow a(y), b(y) \in \mathcal{P}(GF(p)) \\ \gamma = \alpha + \beta &\longleftrightarrow c(y) = a(y) + b(y) \mod p(y) \\ \gamma = \alpha \cdot \beta &\longleftrightarrow c(y) = a(y) \cdot b(y) \mod p(y) \end{aligned}$$

Theorem. In this field, the first-order polynom y is be a primitive element.

Definition. Standard form of a polynom in $GF(p^m)$:

$$\begin{aligned} \alpha(x) &= \alpha_0 + \alpha_1 \cdot x + \cdots + \alpha_n \cdot x^n \\ \alpha(x) &= a_0(y) + a_1(y) \cdot x + \cdots + a_n(y) \cdot x^n \\ \alpha(x) &= y^{i_0} \cdot x + y^{i_1} \cdot x + \cdots + y^{i_n} \cdot x^n \end{aligned}$$

Construction of optimal binary Reed-Solomon codes: $t \to (n, k); p = 2, q = 2^m, n = 2^{m-1}$

18 Describe the CDMA/FH system

The abbreviation CDMA stands for *Code Division Multiple Access*, and it is a channel access solution used in almost all modern telecommunication systems. CDMA allows multiple users to access the same channel simultaneously, without disturbing each others transmissions.

The CDMA-FH (CDMA *Frequency Hopping*) works by transmitting radio signals by rapidly switching a carrier among many frequency channels, using a pseudorandom sequence known to both transmitter and receiver. This approach ensures a spread spectrum, much more resistant to deliberate jamming and disturbances like fading/multipath propagation as well. The frequency hopping also provides enhanced privacy: without the hopping scheme being known, the transmission can hardly be distinguished from random noise.

The CDMA-FH system works by applying so called *code tables* to the message being transmitted.

CDMA-FH encoding

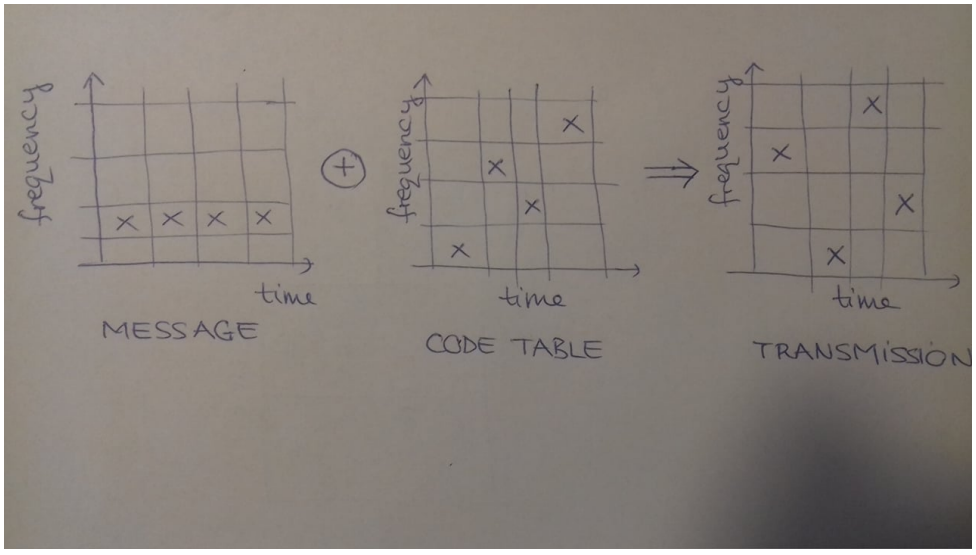

Figure 8: CDMA-FH transmission using code tables

CDMA-FH decoding

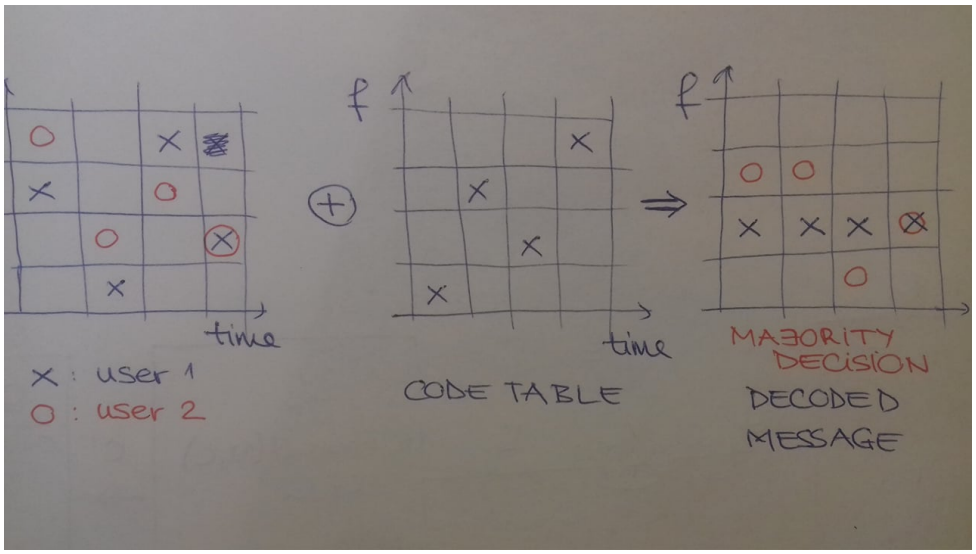

Figure 9: CDMA-FH decoding

As seen in figure 9, parts of the message belonging to the coding table used by the first user align into line, transmisson of the second user appears as noise, spread over the whole spectrum. Detection happens based on majority decision. Even though parts of the message overlap in timeslot 4, both users are able to decode their transmission. In practice, tables with much more time and frequency slots are used.

The code tables used are *binary superimposed* codes, which means their bitwise OR result must be different at a predescribed level. In a CDMA-FH system, $M = \mathcal{O}(N^2)$ different codes can be used.

19 Describe the CDMA/DS system and the Walsh-Hadamard codes

The abbreviation CDMA stands for *Code Division Multiple Access*, and it is a channel access solution used in almost all modern telecommunication systems. CDMA allows multiple users to access the same channel simultaneously, without disturbing each others transmissions.

In the CDMA-DS (CDMA *Direct sequence*) system a method of achieving the spreading of a given signal is provided by the modulation scheme.The message signal is used to modulate a bit sequence called *signature signal*; this pseudorandom code consists of radio pulses much shorter in durationthan the original message signal. This modulation of the message signal scrambles and spreads the pieces of data, and thereby resulting in a bandwidth size much larger than the message signal. The duration of the radio pulses are called *chiptime*. The smaller this duration, the larger the bandwidth of the resulting signal.

Formal description of the system

Users of the system: $i = 1, 2, ..., M$.

Definition. The *message symbol* being sent by user i: $y_i \in \{-1, 1\}$.

Definition. The *codeword* of user i: $c^{(i)} \in \{-1, 1\}^N$, where

$$N = \frac{T_{sample}}{T_{chip}} = \frac{f_{chip}}{f_{sample}}.$$

Note. Typical values for N are 1000 in (low-speed) vioce transmissions, and as low as 3-5 in high speed data applications.

Definition. The *signature signal* belonging to the codeword is defined as a square signal: $s_i(t) = c_k^{(i)}$ for $t \in [(k-1)T_{chip}, kT_{chip}]$.

Definition. The *received* or *multiplexed (MUX)* signal for a transmitted symbol is:

$$x(t) = \sum_{j=1}^{M} y_j \cdot s_j(t) + \nu(t)$$

In a more complex model, taking in consideration attenuation and delay:

$$x(t) = \sum_{j=1}^{M} y_j \cdot (\alpha_j \cdot s_j(t - \tau_j)) + \nu(t)$$

Using an even more more complex model (each channel having its own impulse response function $g_j(t)$):

$$x(t) = \sum_{j=1}^{M} y_j \cdot (s_j(t) * g_j(t)) + \nu(t)$$

Below, for the sake of simplicity, we use the first model.

At the receiver side, the signals are distinguished based on their correlation to the respective signature signals. To ensure unique detection, correlation of signature signals must be examined.

Definition. The *correlation matrix* of the codewords:

$$\mathbf{R} : R_{ij} = \frac{1}{T_s}\int_0^{T_s} s_i(t)\cdot s_j(t)dt = \frac{1}{N}\mathbf{c}^{(i)T}\cdot\mathbf{c}^{(j)}$$

Note. The main diagonal of the matrix: $R_{ii} = \frac{1}{N}\sum_{j=1}^{N}(c_j^{(i)})^2 = 1$.

Definition. The *correlation vector* of the random noise:

$$\boldsymbol{\nu} : \nu_j = \frac{1}{T_s}\int_0^{T_s} \nu(t)\cdot s_j(t)dt$$

.

Using the above definitions, the received symbol can be written as:

$$x_i = \frac{1}{T_s}\int_0^{T_s} s_i(t)x(t)dt = y_i + \sum_{j=1,j\neq i}^{M} y_j\cdot\frac{1}{T_s}\int_0^{T_s} s_i(t)s_j(t)dt + \int_0^{T_s}\nu(t)\cdot s_j(t)dt =$$

$$= y_i + \underbrace{\sum_{j=1,j\neq i}^{M} R_{ij}\cdot y_j}_{MUI} + \nu_i, \text{where MUI stand for MultiUser Interference.}$$

Written in vector-matrix form: $\mathbf{x} = \mathbf{Ry} + \boldsymbol{\nu}$.

Walsh-Hadamard codes

Walsh-Hadamard method allows construction of $M = 2^k$ size ortogonal codes.

Initialization: $\boldsymbol{C}(0) = 1$. Iterative step: $\boldsymbol{C}(k+1) = \left[\begin{array}{c|c} \boldsymbol{C}(k) & \boldsymbol{C}(k) \\ \hline \boldsymbol{C}(k) & -\boldsymbol{C}(k) \end{array}\right]$

Walsh-Hadamard codes are orthogonal, which means $R_{ij} = 0 \quad \forall i \neq j$.

For this reason, there is no multiuser interference, and $x_i = y_i + \nu_i$

Definition. The *deviation matrix* of the noise can be written as: $\boldsymbol{K} : K_{ij} = \mathbb{E}(\nu_i\nu_j)$

Recognition of the symbol happens using the concept of maximum likelyhood:

$$\hat{y}^{(i)} : \max_{\boldsymbol{y}\in\{-1,1\}} p(\boldsymbol{y}|\boldsymbol{x}) = \max\frac{p(\boldsymbol{x}|\boldsymbol{y})p(\boldsymbol{y})}{p(\boldsymbol{x})}$$

As $p(\boldsymbol{x})$ and $p(\boldsymbol{y})$ are constant, this depends only on $p(\boldsymbol{x}|\boldsymbol{y})$.

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{1}{\sqrt{(2\pi)^N det(\boldsymbol{K})}}\cdot e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{Ry})^T\cdot\boldsymbol{K}^{-1}(\boldsymbol{x}-\boldsymbol{Ry})}$$

To maximize the expression above, the following problem should be solved:

$$\hat{y}^{(i)} : \min_{\boldsymbol{y}} \boldsymbol{y}^T \boldsymbol{R}\boldsymbol{y} - \boldsymbol{x}^T \boldsymbol{y}$$

Complexity of solution (mathematically): $\mathcal{O}(2^M)$.
Complexity using approximation (Hopfield net): $\mathcal{O}(M^2)$.
State transition rule of the Hopfield net:

$$y_l(k+1) = -\text{sgn}\left\{\sum_j R_{lj} y_j(k) - x_l + \beta_l\right\}$$

20 Describe the CDMA/DS system with random codes

Formal description

Formal description of CDMA/DS systems is in section 19 on page 32.

Random codes

Using CDMA/DS with Walsh-Hadamard orthogonal codes minimizes the multiuser interference, but has a major drawback: number of users is limited by $N = \frac{f_{chip}}{f_{sample}} \approx 3 - 5$ in high-speed data transmissions.

To overcome this constraint, randomly generated, *quasi orthogonal* codes can be used. These codes are generated using an RNG following Bernoulli distribution, which means $P(c_k^{(i)} = 1) = P(c_k^{(i)} = 0) = 0.5$. This approach does not eliminate multiuser interference (MUI), so this will negatively affect bit error probability.

$$
\begin{aligned}
P_b &= P(\hat{y}_l \neq y_l) = \frac{1}{2}P(\hat{y}_l = 1|y_l = -1) + \frac{1}{2}P(\hat{y}_l = -1|y_l = 1) = \\
&= \frac{1}{2}P\left(\operatorname{sgn}\left\{-1 + \sum_{j\neq l} R_{lj}y_j + \nu_l\right\} = 1\right) + \frac{1}{2}P\left(\operatorname{sgn}\left\{+1 + \sum_{j\neq l} R_{lj}y_j + \nu_l\right\} = -1\right) = \\
&= \frac{1}{2}\sum_{z\in\{-1,1\}^{M-1}} \frac{1}{2^{M-1}}\left[P\left(\operatorname{sgn}\left\{-1 + \sum_{j\neq l} R_{lj}z_j + \nu_l\right\} > 0\right) + P\left(\ldots < 0\right)\right] = \\
&= \frac{1}{2^M}\sum_{z\in\{-1,1\}^{M-1}} \Big[P(\nu_l > 1 - \sum R_{lj}z_j) + P(\nu_l < 1 - \sum R_{lj}z_j)\Big] = \\
&= \frac{1}{2^M}\sum_{z\in\{-1,1\}^{M-1}} \left[\Phi\left(\frac{-1 - \sum R_{lj}z_j}{\sqrt{N_0}}\right) + \Phi\left(\frac{-1 + \sum R_{lj}z_j}{\sqrt{N_0}}\right)\right] = \\
&= \Psi(\boldsymbol{R}) = \Psi\left(\frac{1}{N}\boldsymbol{C}^T\boldsymbol{C}\right).
\end{aligned}
$$

To construct an optimal code, this $\Psi(\boldsymbol{R})$ expression should be minimised.

Offline complexity (mathematically): $\mathcal{O}(2^{M-1})$.

Approximation using Monte-Carlo simulation can produce better results.

21 Describe the OTP method for cryptography

In cryptography, the one-time pad (OTP) is an encryption technique that cannot be cracked, but requires the use of a one-time pre-shared key the same size as, or longer than, the message being sent. In this technique, a plaintext is paired with a random secret key (also referred to as a one-time pad). Then, each bit or character of the plaintext is encrypted by combining it with the corresponding bit or character from the pad using modular addition. If the key is truly random, and is never reused in whole or in part, then the resulting ciphertext will be impossible to decrypt or break. Digital versions of one-time pad ciphers are used by nations for virtually all top secret diplomatic and military communication, but the problems of secure key distribution have made them impractical for less critical applications.

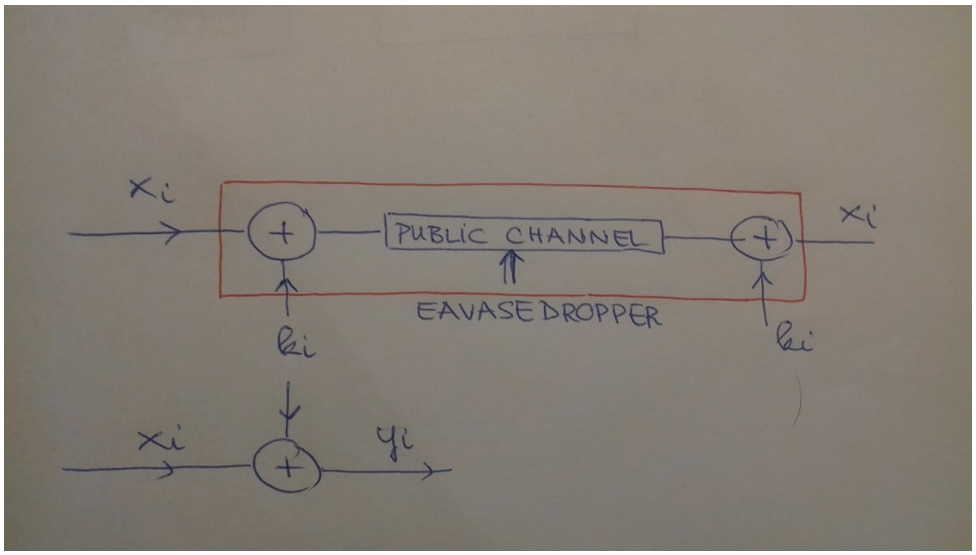

Figure 10: One-time-pad as binary simmetric channel

As seen in the figure, for an eavesdropper without access to the one-time-pad, this type of encryption can be seen as a binary simmetric channel with a bit error probability equal to the probability of a bit being 1 in the key. If the key is truly random, $P(k_i = 0) = P(k_i = 1) = 0.5$.

Now, it can be easily concluded, that for a bit error probability $P_b = 0.5$, the conditional entropy of the channel is equal to $\mathrm{H}(P_b) = P_b \cdot \mathrm{ld}\,\frac{1}{P_b} + (1 - P_b) \cdot \mathrm{ld}\,\frac{1}{1-P_b} = 1$, resulting in $C = 0$ channel capacity. In another words, the mutual informaiton of the transmitted and the eavasedropped symbols is $I(X, Y) = 0$.

In conclusion, if the key is never reused, the OTP is impossible to be broken.

22 Describe the RSA algorithm

The RSA is one of the first public-key cryptography algorithms, widely used for secure data transmission and digital signitures.

Mathematical background

Theorem (Euclidean division algorithm)**.** Given $b, p \in \mathbb{N}^*, b > p$, there exists $q, r \in \mathbb{N}, r < p$ such that $b = p \cdot q + r$.

Theorem (Fermat's little theorem)**.** If $p, c \in \mathbb{N}, p$ is prime, and $p \nmid c$, then $c^{p-1} \equiv 1 \pmod{p}$

Proof. First, we consider $c, 2c, ..., (p-1)c \bmod p$.

Assuming $0 < i, j \leq p-1$ exist, such that $ic = jc$, leads to $(i-j)c \equiv 0 \pmod{p} \Rightarrow p \mid c$ contradiction. Thus, these numbers are pairwise different.

Now, if they are different, these numbers are in fact $1, 2, 3, ..., (p-1)$ in different order. Consequently,

$$\begin{aligned} c \cdot 2c \cdot ... \cdot (p-1)c &= 1 \cdot 2 \cdot ... \cdot (p-1) \pmod{p} \\ c^{p-1} \cdot (p-1)! &= (p-1)! \pmod{p} \\ c^{p-1} &= 1 \pmod{p} \end{aligned}$$

□

Theorem (Extended Fermat theorem)**.** If $p_1, p_2, c \in \mathbb{N}, p_1, p_2$ is prime, and $p_1 \nmid c, p_2 \nmid c$, then $c^{(p_1-1)(p_2-1)} \equiv 1 \pmod{p_1 \cdot p_2}$

Proof. Using Fermat's little theorem twice:

$$\left.\begin{aligned} (c^{p_2-1})^{p_1-1} &\equiv 1 \pmod{p_1} \\ (c^{p_1-1})^{p_2-1} &\equiv 1 \pmod{p_2} \end{aligned}\right\} \Rightarrow c^{(p_1-1)(p_2-1)} \equiv 1 \pmod{p_1 \cdot p_2}$$

□

Definition. For $n = \prod_{i=1}^{N} p_i$, where p_i prime, the *Euler operator* is defined as: $\Phi(n) = \prod_{i=1}^{N} (p_i - 1)$

The RSA algorithm

Key generation

1. First, we pick two large, random prime numbers: p_1, p_2.
2. We calculate $m = p_1 \cdot p_2$ and $\Phi(m) = (p_1 - 1)(p_2 - 1)$.

3. We choose $0 < e \leq \Phi(m)$, so that g.c.d.$(e, m) = 1$.
4. We calculate $d = e^{-1} \mod \Phi(m)$.

Now, the *public key* is $k^p = (e, m)$, the *private* or *secret key* is $k^s = (d, p_1, p_2)$.

Encryption and decryption

Encryption: the *cyphertext* $c = x^e \mod m$.
Decryption: the *plaintext* $x = c^d \mod m$.

Proof.

$$c^d = (x^e)^d = x^{ed} = x^{q\Phi(m)+1} = x \cdot x^{(p_1-1)(p_2-1)} = x \pmod{p_1 \cdot p_2}$$

□