4 An Introduction to Channel Coding and Decoding over BSC

4.1. Recall that **channel coding** introduces, in a controlled manner, some *redundancy* in the (binary) information sequence that can be used at the receiver to overcome the effects of noise and interference encountered in the transmission of the signal through the channel.

Example 4.2. Repetition Code: Repeat each bit n times, where n is some positive integer.

- Use the channel n times to transmit 1 info-bit
- The (transmission) rate is $\frac{1}{n}$ [bpcu].
  - bpcu = bits per channel use

4.3. Two classes of channel codes

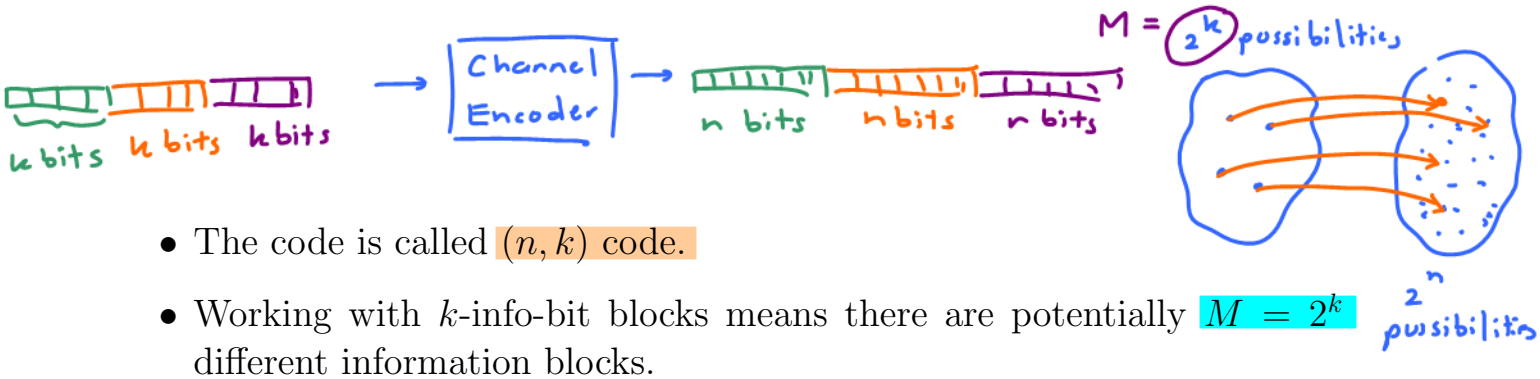
(a) Block codes

- To be discussed here.
- Realized by combinational/combinatorial circuit.

(b) Convolutional codes

- Encoder has memory.
- Realized by sequential circuit. (Recall state diagram, flip-flop, etc.)

Definition 4.4. Block Encoding: Take k (information) bits at a time and map each k-bit sequence into a (unique) n-bit sequence, called a **codeword**.



- The code is called (n, k) code.
- Working with k-info-bit blocks means there are potentially $M = 2^k$ different information blocks.
  - The table that lists all the 2^k mapping from the k-bit info-block $\underline{\mathbf{s}}$ to the n-bit codeword $\underline{\mathbf{x}}$ is called the **codebook**.
  - The M info-blocks are denoted by $\underline{\mathbf{s}}^{(1)}, \underline{\mathbf{s}}^{(2)}, \ldots, \underline{\mathbf{s}}^{(M)}$. The corresponding M codewords are denoted by $\underline{\mathbf{x}}^{(1)}, \underline{\mathbf{x}}^{(2)}, \ldots, \underline{\mathbf{x}}^{(M)}$ respectively.
- To have unique codeword for each information block, we need $n \geq k$. Of course, with some redundancy added to combat the error introduced by the channel, we need $n > k$.
  - The amount of redundancy is measured by the ratio $\frac{n}{k}$.
  - The number of redundant bits is $r = n - k$.
- Here, we use the channel n times to convey k (information) bits.
  - The ratio $\frac{k}{n}$ is called the rate of the code or, simply, the **code rate**.
  - The (transmission) rate is $R = \frac{k}{n} = \frac{\log_2 M}{n}$ [bpcu].

$2^k = M \Rightarrow k = \log_2 M$

4.5. When the mapping from the information block $\underline{\mathbf{s}}$ to the codeword $\underline{\mathbf{x}}$ is invertible, the task of the decoder can be separated into two steps:

- First, find $\hat{\underline{\mathbf{x}}}$ which is its guess of the $\underline{\mathbf{x}}$ value based on the observed value of $\underline{\mathbf{y}}$.
- Second, map $\hat{\underline{\mathbf{x}}}$ back to the corresponding $\hat{\underline{\mathbf{s}}}$ based on the codebook.

You may notice that it is more important to recover the index of the codeword than the codeword itself. Only its index is enough to indicate which info-block produced it.

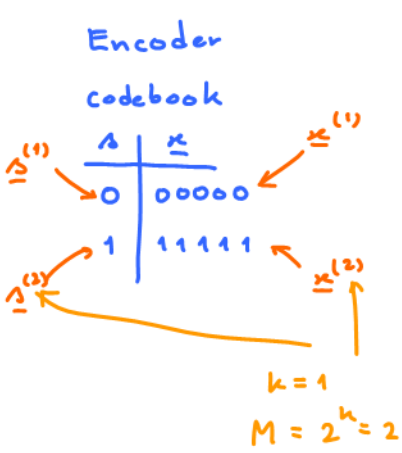
4.6. General idea for ML decoding of block codes over BSC: **minimum-distance decoder**

- To recover the value of $\underline{\mathbf{x}}$ from the observed value of $\underline{\mathbf{y}}$, we can apply what we studied about optimal detector in the previous section.
  - The optimal detector is again given by the MAP detector:
$$\hat{\underline{\mathbf{x}}}_{\mathrm{MAP}}\left(\underline{\mathbf{y}}\right) = \arg\max_{\underline{\mathbf{x}}} Q\left(\underline{\mathbf{y}}\,|\underline{\mathbf{x}}\right) p\left(\underline{\mathbf{x}}\right). \tag{8}$$
  - When the prior probabilities $p\left(\underline{\mathbf{x}}\right)$ is unknown or when we want simpler decoder, we may consider using the ML decoder:
$$\hat{\underline{\mathbf{x}}}_{\mathrm{ML}}\left(\underline{\mathbf{y}}\right) = \arg\max_{\underline{\mathbf{x}}} Q\left(\underline{\mathbf{y}}\,|\underline{\mathbf{x}}\right). \tag{9}$$
    In this section, we will mainly focus on the ML decoder.
- By the memoryless property of the channel,
$$Q\left(\underline{\mathbf{y}}|\underline{\mathbf{x}}\right) = Q(y_1|x_1) \times Q(y_2|x_2) \times \cdots \times Q(y_n|x_n).$$
  Furthermore, for BSC,
$$Q(y_i|x_i) = \begin{cases} p, & y_i \neq x_i, \\ 1-p, & y_i = x_i. \end{cases}$$
  Therefore,
$$Q\left(\underline{\mathbf{y}}|\underline{\mathbf{x}}\right) = p^{d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right)}(1-p)^{n-d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right)} = \left(\frac{p}{1-p}\right)^{d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right)}(1-p)^n, \tag{10}$$
  where $d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right)$ is the number of coordinates in which the two blocks $\underline{\mathbf{x}}$ and $\underline{\mathbf{y}}$ differ.
  Note that when $p < 0.5$, which is usually the case for practical systems, we have $p < 1-p$ and hence $0 < \frac{p}{1-p} < 1$. In which case, to maximize $Q\left(\underline{\mathbf{y}}|\underline{\mathbf{x}}\right)$, we need to minimize $d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right)$. In other words, $\hat{\underline{\mathbf{x}}}_{\mathrm{ML}}\left(\underline{\mathbf{y}}\right)$ should be the codeword $\underline{\mathbf{x}}$ which has the minimum distance from the observed $\underline{\mathbf{y}}$:
$$\hat{\underline{\mathbf{x}}}_{\mathrm{ML}}\left(\underline{\mathbf{y}}\right) = \arg\min_{\underline{\mathbf{x}}} d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right). \tag{11}$$
  In conclusion, for block coding over BSC with $p < 0.5$, the ML decoder is the same as the minimum distance decoder.

d(00101, 01111) = 2

distance (difference) → Hamming distance

Example 4.7. Repetition Code and Majority Voting: Back to Example 4.2.



$$Q(00001|00000) = (1-p)(1-p)(1-p)(1-p)p$$
$$= (1-p)^4 p$$

$$Q(01010|00000) = (1-p)^3 p^2$$
$$Q(y|00000) = (1-p)^{n_0} p^{n_1}$$
$$Q(y|11111) = (1-p)^{n_1} p^{n_0}$$

Let $\underline{\mathbf{0}}$ and $\underline{\mathbf{1}}$ denote the n-dimensional row vectors $00\ldots0$ and $11\ldots1$, respectively. Observe that

$$d\left(\underline{\mathbf{x}}, \underline{\mathbf{y}}\right) = \begin{cases} \#1 \text{ in } \underline{\mathbf{y}}, & \text{when } \underline{\mathbf{x}} = \underline{\mathbf{0}}, \\ \#0 \text{ in } \underline{\mathbf{y}}, & \text{when } \underline{\mathbf{x}} = \underline{\mathbf{1}}. \end{cases}$$

Therefore, the minimum distance detector is

$$\hat{\underline{\mathbf{x}}}_{\mathrm{ML}}\left(\underline{\mathbf{y}}\right) = \begin{cases} \underline{\mathbf{0}}, & \text{when } \#1 \text{ in } \underline{\mathbf{y}} < \#0 \text{ in } \underline{\mathbf{y}}, \\ \underline{\mathbf{1}}, & \text{when } \#1 \text{ in } \underline{\mathbf{y}} > \#0 \text{ in } \underline{\mathbf{y}}. \end{cases}$$

Equivalently,

$$\hat{s}_{\mathrm{ML}}\left(\underline{\mathbf{y}}\right) = \begin{cases} 0, & \text{when } \#1 \text{ in } \underline{\mathbf{y}} < \#0 \text{ in } \underline{\mathbf{y}}, \\ 1, & \text{when } \#1 \text{ in } \underline{\mathbf{y}} > \#0 \text{ in } \underline{\mathbf{y}}. \end{cases}$$

This is the same as taking a majority vote among the received bit in the $\underline{\mathbf{y}}$ vector.

The corresponding error probability is

$$P\left(\mathcal{E}\right) = \sum_{c=\left\lceil \frac{n}{2} \right\rceil}^{n} \binom{n}{c} p^c (1-p)^{n-c}.$$

For example, when $p = 0.01$, we have $P(\mathcal{E}) \approx 10^{-5}$. Figure 6 compares the error probability when different values of n are used.

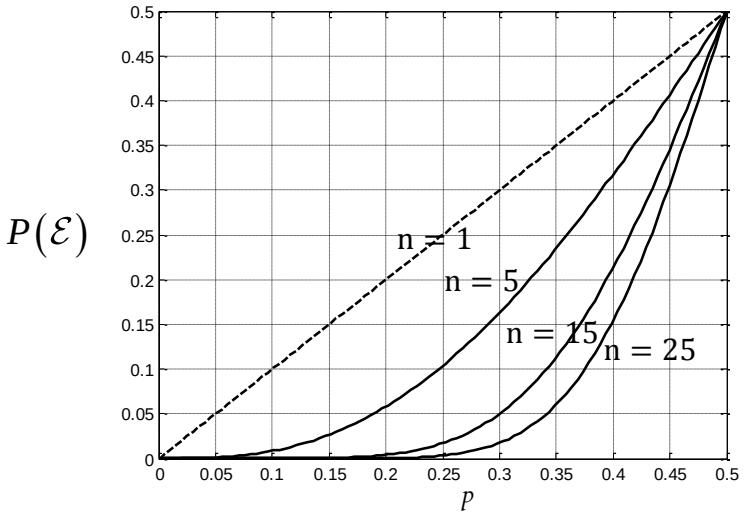


Figure 6: Error probability for a system that uses repetition code at the transmitter (repeat each info-bit n times) and majority voting at the receiver. The channel is assumed to be binary symmetric with crossover probability p.

- Notice that the error probability decreases to 0 when n is increased. It is then possible to transmit with arbitrarily low probability of error using this scheme.
- However, the (transmission) rate $R = \frac{k}{n} = \frac{1}{n}$ is also reduced as n is increased.

So, in the limit, although we can have very small error probability, we suffer tiny (transmission) rate.

We may then ask "what is the maximum (transmission) rate of information that can be *reliably* transmitted over a communications channel?" Here, reliable communication means that the error probability can be made arbitrarily small. Shannon provided the solution to this question in his seminal work. We will revisit this question in the next section.

4.8. General idea for MAP decoding of block codes over BSC: Of course, when the prior probabilities are known, the optimal decoder is given by the MAP decoder:

$$\underline{\hat{\mathbf{x}}}_{\text{MAP}}\left(\underline{\mathbf{y}}\right) = \arg\max_{\underline{\mathbf{x}}} Q\left(\underline{\mathbf{y}}\,|\underline{\mathbf{x}}\right) p\left(\underline{\mathbf{x}}\right). \tag{12}$$

By definition, $p\left(\underline{\mathbf{x}}\right) \equiv P\left[\underline{\mathbf{X}} = \underline{\mathbf{x}}\right]$. Therefore,

$$p\left(\underline{\mathbf{x}}^{(i)}\right) \equiv P\left[\underline{\mathbf{X}} = \underline{\mathbf{x}}^{(i)}\right] = P\left[\underline{\mathbf{S}} = \underline{\mathbf{s}}^{(i)}\right].$$

Again, because the BSC is memoryless, from (10), we have

$$Q\left(\underline{\mathbf{y}}|\underline{\mathbf{x}}\right) = p^{d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right)}(1-p)^{n-d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right)} = \left(\frac{p}{1-p}\right)^{d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right)}(1-p)^{n}, \tag{13}$$

Therefore,

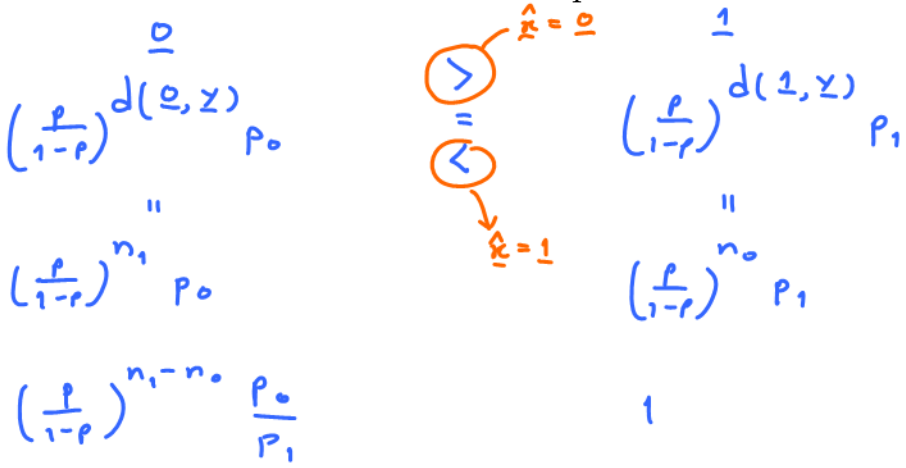
$$\hat{\underline{\mathbf{x}}}_{\mathrm{MAP}}\left(\underline{\mathbf{y}}\right) = \arg\max_{\underline{\mathbf{x}}} \left(\frac{p}{1-p}\right)^{d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right)} (1-p)^{n} p\left(\underline{\mathbf{x}}\right) \tag{14}$$

$$= \arg\max_{\underline{\mathbf{x}}} \left(\frac{p}{1-p}\right)^{d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right)} p\left(\underline{\mathbf{x}}\right). \tag{15}$$

Example 4.9. Consider a communication over a BSC with $p = \frac{1}{3}$. Suppose repetition code is used with $n = 3$. Assume that the info-bit has $P\left[S=0\right] = \frac{3}{4}$. Find the ML and MAP decoders and their error probabilities.

$P[\underline{x} = 000] = \frac{3}{4} = p_0$

$P[\underline{x} = 111] = \frac{1}{4} = p_1$



$n - n_1$ = 3

$\frac{p}{1-p} = \frac{1/3}{2/3} = \frac{1}{2}$

$\frac{p_0}{p_1} = \frac{3/4}{1/4} = 3$

n_1	n_0	$n_1 - n_0$	$\left(\frac{p}{1-p}\right)^{n_1-n_0}$	$\times \frac{p_0}{p_1}$	$\hat{s}_{\mathrm{MAP}}\left(\underline{\mathbf{y}}\right)$	$\hat{s}_{\mathrm{ML}}\left(\underline{\mathbf{y}}\right)$
0	3	-3	8 > 1	24 > 1	0	0
1	2	-1	2 > 1	6 > 1	0	0
2	1	1	1/2 < 1	3/2 > 1	0	1
3	0	3	1/8 < 1	3/8 < 1	1	1