3.4 Optimal Block Decoding for Communications Over BSC

3.45. The decoding techniques (MAP and ML) discussed in the previous section can be extended to the case in which we simultaneously consider n consecutive channel output symbols resulted from having n input symbols.

Notation-wise, this simply means we consider an input-output vector pair $(\underline{\mathbf{X}}, \underline{\mathbf{Y}})$ instead of an input-output symbol pair (X, Y)

$(x_1, x_2, \cdots, x_n) = \underline{x} \rightarrow$ DMC $\rightarrow \underline{Y} = (Y_1, Y_2, \cdots, Y_n)$

3.46. By the memoryless property of the channel,

$$Q\left(\underline{\mathbf{y}}|\underline{\mathbf{x}}\right) = Q(y_1|x_1) \times Q(y_2|x_2) \times \cdots \times Q(y_n|x_n).$$

Example 3.47. For a DMC in which $\mathcal{X} = \{0, 1\}$, $\mathcal{Y} = \{1, 2, 3\}$, $\mathbf{Q} = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}$, find

(a) $Q(122|100) = 0.2 \times 0.2 \times 0.3 = 0.012$

(b) $Q(333|111) = (0.3)^3 = 0.027$

Extended Q matrix

$\underline{x} \backslash \underline{y}$	111	112	121	⋯	122	⋯	333
000							
001							
010							
⋮							
100					0.012		
⋮							
111							0.027

$n-d$ positions that $\underline{x}$ and $\underline{y}$ are the same

different

$d(\underline{x}, \underline{y})$

Example 3.48. For BSC, find

(a) $Q(101|100) = p \times (1-p) \times (1-p) = (1-p)^2 p^1$

(b) $Q(111|111) = (1-p)^3 p^0$

3.49. For BSC,

$$Q(y_i|x_i) = \begin{cases} p, & y_i \neq x_i, \\ 1-p, & y_i = x_i. \end{cases}$$

Therefore,

$$Q\left(\underline{\mathbf{y}}|\underline{\mathbf{x}}\right) = p^{d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right)}(1-p)^{n-d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right)} = \left(\frac{p}{1-p}\right)^{d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right)}(1-p)^n, \tag{8}$$

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.

distance (difference)

Hamming distance

$d(101, 100) = 1$

$d(111, 111) = 0$

$d(00101, 01111) = 2$

3.50. To recover the value of $\underline{\mathbf{x}}$ from the observed value of $\underline{\mathbf{y}}$, we can apply the vector version of what we studied about optimal decoder in the previous section.

- The optimal decoder is again given by the MAP decoder:

$$\hat{\underline{\mathbf{x}}}_{\mathrm{MAP}}\left(\underline{\mathbf{y}}\right) = \arg\max_{\underline{\mathbf{x}}} Q\left(\underline{\mathbf{y}}\,|\underline{\mathbf{x}}\right) p\left(\underline{\mathbf{x}}\right). \tag{9}$$

- 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:

$$\hat{\underline{\mathbf{x}}}_{\mathrm{ML}}\left(\underline{\mathbf{y}}\right) = \arg\max_{\underline{\mathbf{x}}} Q\left(\underline{\mathbf{y}}\,|\underline{\mathbf{x}}\right). \tag{10}$$

no $p(x)$ term in MLD

Plugging-in

$$Q\left(\underline{\mathbf{y}}|\underline{\mathbf{x}}\right) = p^{d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right)}(1-p)^{n-d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right)} = \left(\frac{p}{1-p}\right)^{d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right)}(1-p)^{n}, \tag{11}$$

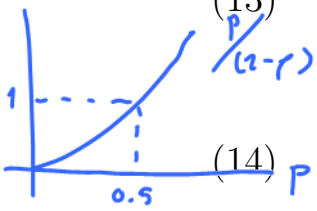
from (8), gives

$$\hat{\underline{\mathbf{x}}}_{\mathrm{MAP}}\left(\underline{\mathbf{y}}\right) = \arg\max_{\underline{\mathbf{x}}} \left(\frac{p}{1-p}\right)^{d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right)}(1-p)^{n} p\left(\underline{\mathbf{x}}\right) \tag{12}$$

$$= \arg\max_{\underline{\mathbf{x}}} \left(\frac{p}{1-p}\right)^{d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right)} p\left(\underline{\mathbf{x}}\right). \tag{13}$$

and

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



3.51. **Minimum-distance decoder** as a ML decoding of block codes over BSC:

From (14) (or directly from (8)), 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 $\underline{\mathbf{y}}$:

$$\hat{\underline{\mathbf{x}}}_{\mathrm{ML}}\left(\underline{\mathbf{y}}\right) = \arg\min_{\underline{\mathbf{x}}} d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right). \tag{15}$$

In conclusion, for block coding over BSC with $p < 0.5$, the ML decoder is the same as the minimum distance decoder.

3.5 Repetition Code for Channel Coding in Communications Over BSC

3.52. 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.

3.53. **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

3.54. Two classes of channel codes

(a) Block codes

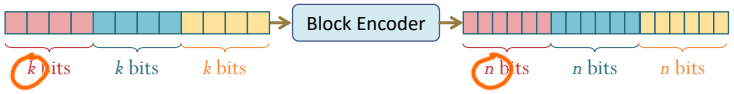
- To be discussed here.
- Realized by combinational/combinatorial circuit.

AND, OR, NOT, NAND, etc gates and wires

(b) Convolutional codes

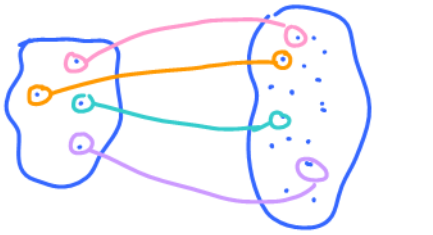
- Encoder has memory.
- Realized by sequential circuit. (Recall state diagram, flip-flop, etc.)

Definition 3.55. 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.

index i	info-block $\underline{\mathbf{s}}$	codeword $\underline{\mathbf{x}}$
1	$\underline{\mathbf{s}}^{(1)} = 000\ldots0$	$\underline{\mathbf{x}}^{(1)} =$
2	$\underline{\mathbf{s}}^{(2)} = 000\ldots1$	$\underline{\mathbf{x}}^{(2)} =$
$\vdots$	$\vdots$	$\vdots$
M	$\underline{\mathbf{s}}^{(M)} = 111\ldots1$	$\underline{\mathbf{x}}^{(M)} =$



  - By the bijective mapping from $\underline{\mathbf{s}}$ to $\underline{\mathbf{x}}$,

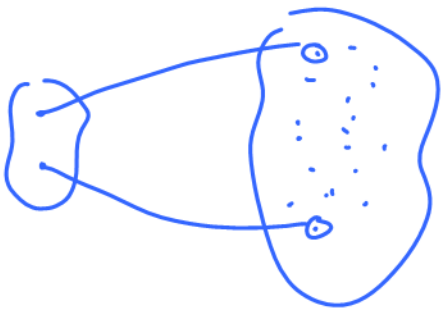
$$p_i \equiv p\left(\underline{\mathbf{x}}^{(i)}\right) \equiv P\left[\underline{\mathbf{X}} = \underline{\mathbf{x}}^{(i)}\right] = P\left[\underline{\mathbf{S}} = \underline{\mathbf{s}}^{(i)}\right].$$

- 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].

Example 3.56. Find the codebook and code rate for the encoder which uses repetition code with $n = 5$.

codebook

s	x
0	00000 ≡ 0
1	11111 ≡ 1



Example 3.57. To get some idea about the difficulty of finding an optimal encoder, we need to consider the size of our search space. For $k = 5$ and $n = 10$, how many encoders are possible?

$$\binom{2^n}{2^k} = \binom{2^{10}}{2^5} = \binom{1024}{32} \approx 5 \times 10^{60}$$

index	$\underline{s}$	$\underline{x}$
1	00000	----------
2	00001	n positions
⋮	00010	
	⋮	
$2^k = 2^5 = 32$	11111	

3.58. Decoding: When the mapping from the information block $\underline{\mathbf{s}}$ to the codeword $\underline{\mathbf{x}}$ is invertible, the task of any 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.

Example 3.59. Repetition Code and Majority Voting: Back to Example 3.53.

Recall: ① MAP decoder is optimal (min $P(\mathcal{E})$)
② ML decoder is suboptimal. However, it can be optimal (the same as the MAPD) when the codewords are equally-likely.
③ ML decoder is the same as the min-distance decoder when the crossover probability p of BSC is < 0.5.

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

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

Ex $\underline{Y} = 01010$

Therefore, the minimum distance decoder is

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

$d(\underline{0}, \underline{Y}) = 2 = \#1$ in $\underline{Y}$

$d(\underline{1}, \underline{Y}) = 3 = \#0$ in $\underline{Y}$

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

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

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.

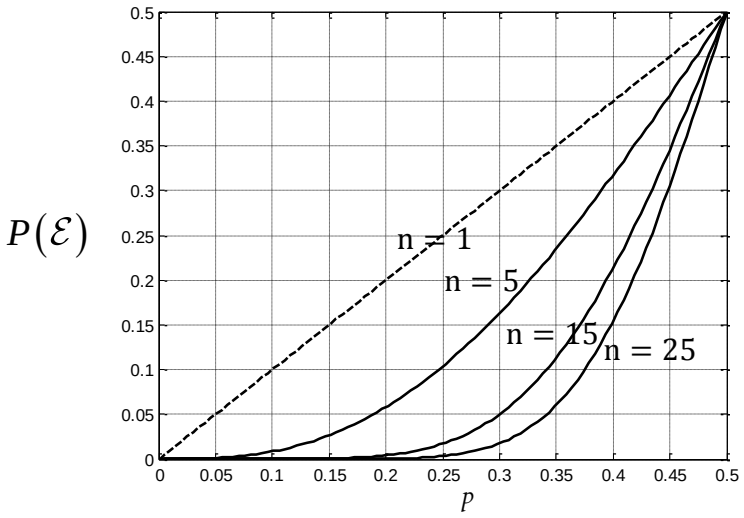


Figure 6: Error probability for a system that uses repetition code at the transmitter (repeat each info-bit n times) and majority voting at the receiver. The channel is assumed to be binary symmetric with crossover probability p.

- 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.

3.60. 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.