3.4 Optimal Block Decoding for Communications Over BSC

3.51. 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)

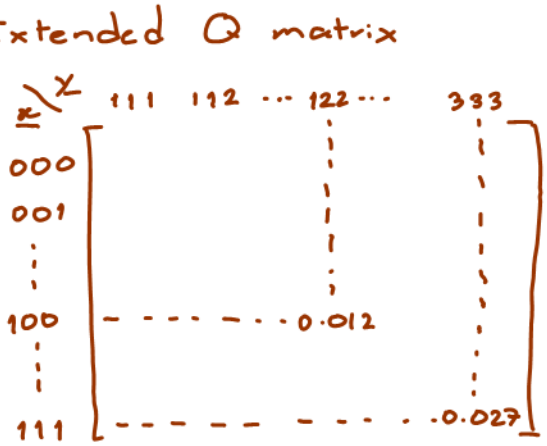
3.52. By the memoryless property of the channel,

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

Example 3.53. For a DMC in which $\mathcal{X} = \{0, 1\}$, $\mathcal{Y} = \{1, 2, 3\}$, $\mathbf{Q} = \begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.3 \end{bmatrix} \end{array}$, find

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

(b) $Q(333|111) = 0.3 \times 0.3 \times 0.3 = 0.027$



Example 3.54. For BSC, find

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

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

$d(\underline{x}, \underline{y})$ = # positions that $\underline{x}$ and $\underline{y}$ are different

$n - d(\underline{x}, \underline{y})$ = # positions that $\underline{x}$ and $\underline{y}$ are the same

3.55. For BSC,

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

Therefore,

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

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

Example 3.56. $d(101, 100) = 1$, $d(111, 111) = 0$, $d(00101, 01111) = 2$

3.57. 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:

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

- 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:

$$\hat{\underline{\mathbf{x}}}_{\mathrm{ML}}\left(\underline{\mathbf{y}}\right) = \arg\max_{\underline{\mathbf{x}}} Q\left(\underline{\mathbf{y}}\,|\underline{\mathbf{x}}\right). \tag{13}$$

Plugging-in

$$Q\left(\underline{\mathbf{y}}|\underline{\mathbf{x}}\right) = p^{d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right)}(1-p)^{n-d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right)} = \left(\frac{p}{1-p}\right)^{d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right)}(1-p)^{n}, \tag{14}$$

from (11), gives

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

does not depend on $\underline{\mathbf{x}}$

$$= \arg\max_{\underline{\mathbf{x}}} \left(\frac{p}{1-p}\right)^{d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right)} p\left(\underline{\mathbf{x}}\right). \tag{16}$$

and

$$\hat{\underline{\mathbf{x}}}_{\mathrm{ML}}\left(\underline{\mathbf{y}}\right) = \arg\max_{\underline{\mathbf{x}}} \left(\frac{p}{1-p}\right)^{d\left(\underline{\mathbf{x}},\underline{\mathbf{y}}\right)}. \tag{17}$$

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

From (17) (or directly from (11)), 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 $\underline{\mathbf{y}}$:

$$\hat{\underline{\mathbf{x}}}_{\mathrm{ML}}\left(\underline{\mathbf{y}}\right) = \arg\min_{\underline{\mathbf{x}}} d\left(\underline{\mathbf{x}}, \underline{\mathbf{y}}\right). \tag{18}$$

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

3.59. 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.

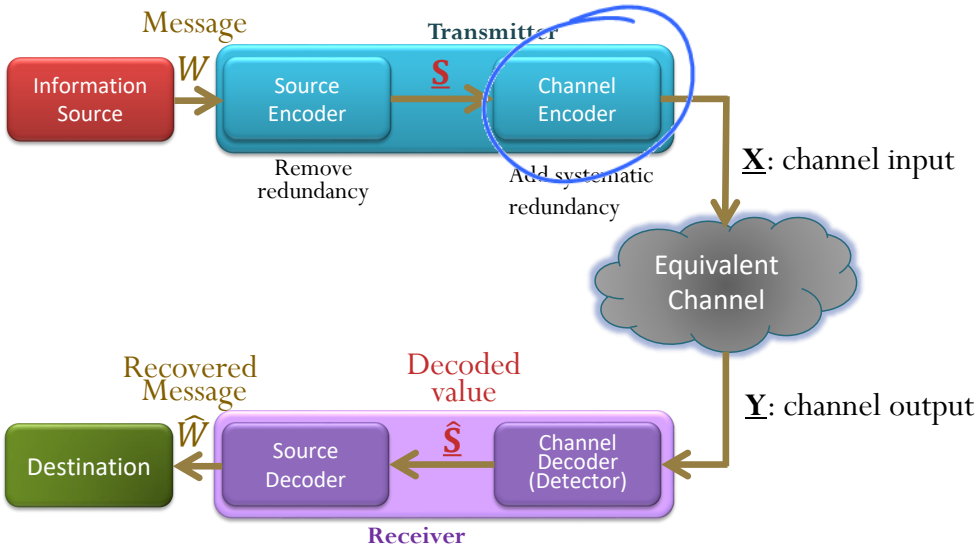


Figure 12: System model for Section 3.5. A channel encoder is added to improve the performance of the system considered in Figure 11 in Section 3.2.

- Note that variables X and Y are still used for the channel input and channel output, respectively. However, as in Section 3.4, we consider blocks (vectors) of them. Therefore, the variables used are $\underline{\mathbf{X}}$ and $\underline{\mathbf{Y}}$.
- Because we introduce another box between the source encoder and the (equivalent) channel, the output of the source encoder is not the same as the channel input anymore. Therefore, we rename the output of the source encoder as S. Again, when we consider a block of output from the source encoder, we denote it by $\underline{\mathbf{S}}$.
- The job of the decoder is now to (correctly) guess the value of $\underline{\mathbf{S}}$. Its output is now denoted by $\hat{\underline{\mathbf{S}}}$.
  - Usually, the mapping (by the channel encoder) from $\underline{\mathbf{S}}$ to $\underline{\mathbf{X}}$ is bijective[16]; so is the mapping from W to $\underline{\mathbf{S}}$ by the source encoder.

[16] A bijection, bijective function, or one-to-one correspondence is a function between the elements of two sets, where each element of one set is paired with exactly one element of the other set, and each element of the other set is paired with exactly one element of the first set.

Therefore, one can also say that, as before, the job of the decoder is still to (correctly) guess the value of $\underline{\mathbf{X}}$. Once we have the value of $\underline{\mathbf{X}}$, we can directly map it back to $\underline{\mathbf{S}}$ and then the original message W.

an example of channel encoder

3.60. Repetition Code: A simple example of channel encoding is to repeat each bit n times, where n is some positive integer.

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

3.61. Two classes of channel codes

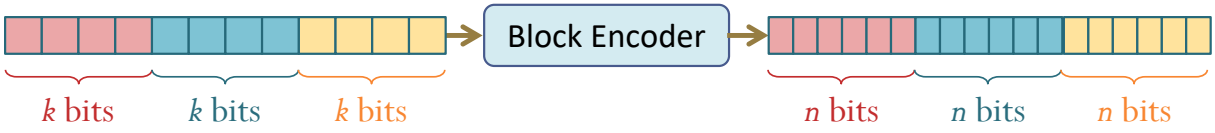
(a) Block codes

- 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 3.62. Block Encoding: Take k (information) bits at a time and map each k-bit sequence into a (unique) n-bit sequence, called a **codeword**[17].



- The code is called (n, k) code. code rate = $\frac{k}{n}$
- Working with k-info-bit blocks means there are potentially $M = 2^k$ different information blocks.

[17] Yes, we used this term already in Chapter 2. Both uses of the term "codeword" denote the outputs of the encoding processes.

- 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	$\underline{\mathbf{s}}^{(M)} = 111\ldots1$	$\underline{\mathbf{x}}^{(M)} =$

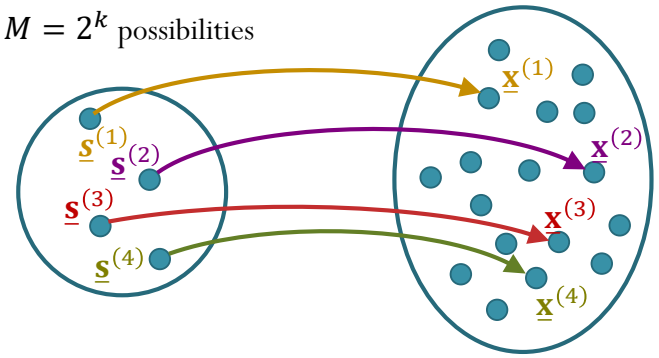


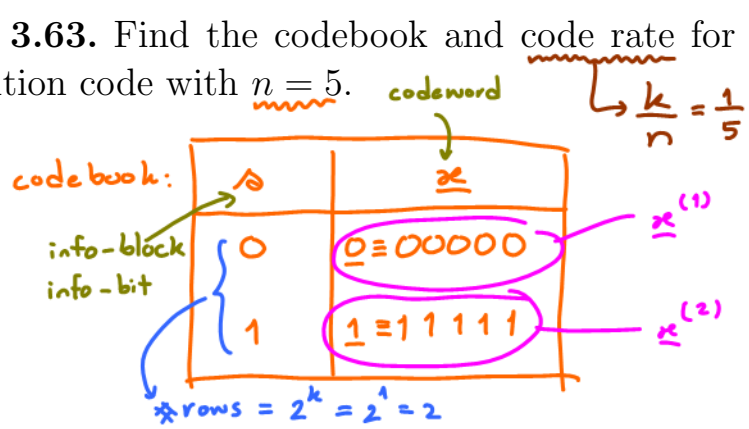
Figure 13: The mapping for block encoding.

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

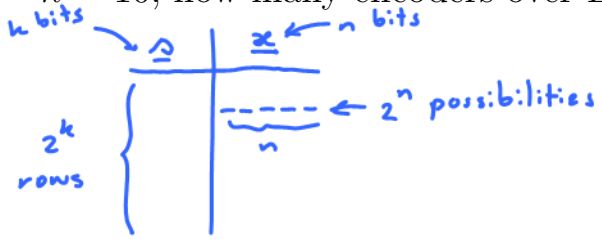
$$p_i \equiv p\left(\underline{\mathbf{x}}^{(i)}\right) \equiv P\left[\underline{\mathbf{X}} = \underline{\mathbf{x}}^{(i)}\right] = P\left[\underline{\mathbf{S}} = \underline{\mathbf{s}}^{(i)}\right].$$

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



Example 3.64. 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 over BSC are possible?



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

3.65. 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. Knowing its index is enough to indicate which info-block produced it.

Example 3.66. Repetition Code and Majority Voting: Back to Example 3.60. on the decoding side

First recall that

(1) MAP decoder is optimal. (It minimizes $P(\mathcal{E})$).

(2) ML decoder is suboptimal. However, it can be optimal (same $P(\mathcal{E})$ as the MAP decoder) when the codewords are equally-likely.

(3) ML decoder is the same as the minimum distance decoder when the crossover probability of the BSC p is < 0.5 (which is usually the case).

$$\hat{x}_{\text{min\_d}}(\underline{y}) = \begin{cases} \underline{0} & \text{when } d(\underline{0},\underline{y}) < d(\underline{1},\underline{y}) \\ \underline{1} & \text{when } d(\underline{0},\underline{y}) > d(\underline{1},\underline{y}) \end{cases}$$

Therefore, minimum distance decoder can be optimal in many situations.

In this example, assume $p < 0.5$. Let $\underline{\mathbf{0}}$ and $\underline{\mathbf{1}}$ denote the n-dimensional row vectors $00\ldots0$ and $11\ldots1$, respectively. Observe that

Ex. $\underline{y} = 01010$

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

$d(\underline{0},\underline{y}) = 2 = \#1$'s in $\underline{y}$

$d(\underline{1},\underline{y}) = 3 = \#0$'s in $\underline{y}$

Therefore, the minimum distance decoder is

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

Equivalently,

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

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

The corresponding error probability is

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

For example, when $p = 0.01$, we have $P(\mathcal{E}) \approx 10^{-5}$. Figure 14 shows how we can view this as having the original BSC channel replaced by a new one with better crossover probability.

Figure 15 compares the error probability when different values of n are used.

- 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.

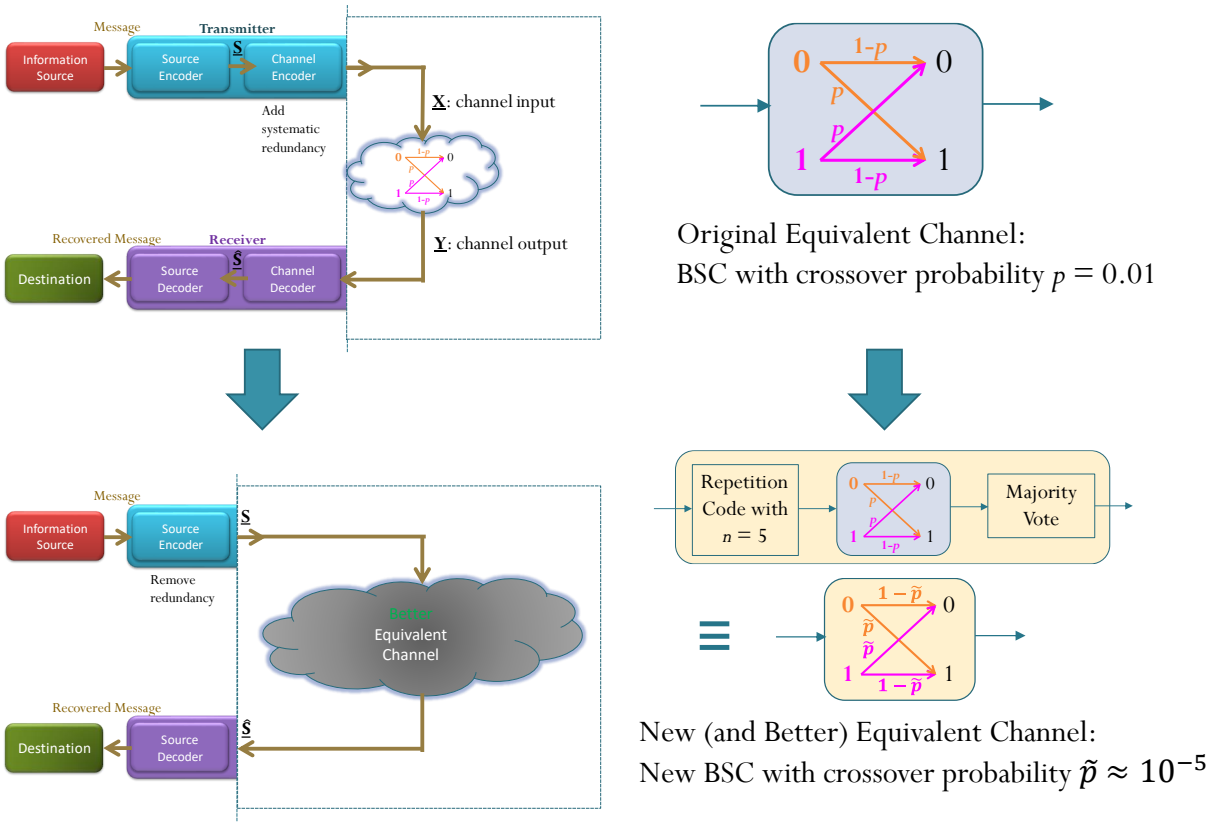


Figure 14: With the addition of channel encoder and channel decoder, the performance of the system is improved. The original BSC, when combined with the channel encoder and channel decoder, can be viewed as a new equivalent BSC with better crossover probability.

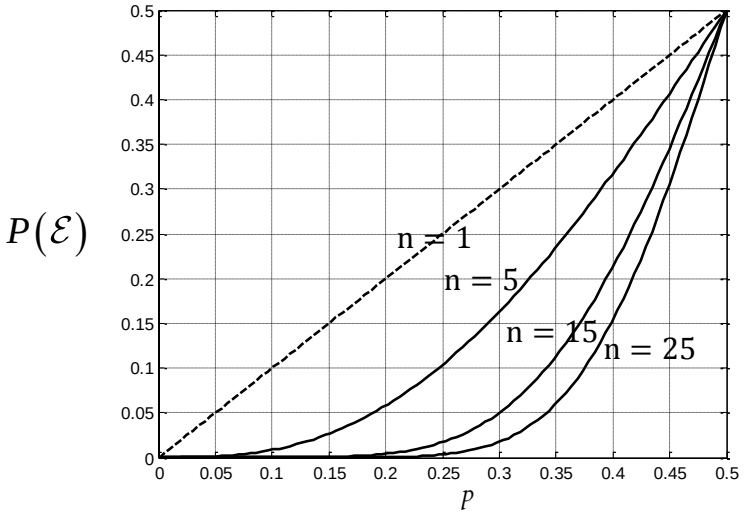


Figure 15: Error probability for a system that uses repetition code at the transmitter (repeat each info-bit n times) and majority voting at the receiver. The channel is assumed to be binary symmetric with crossover probability p.

3.67. 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 chapter.