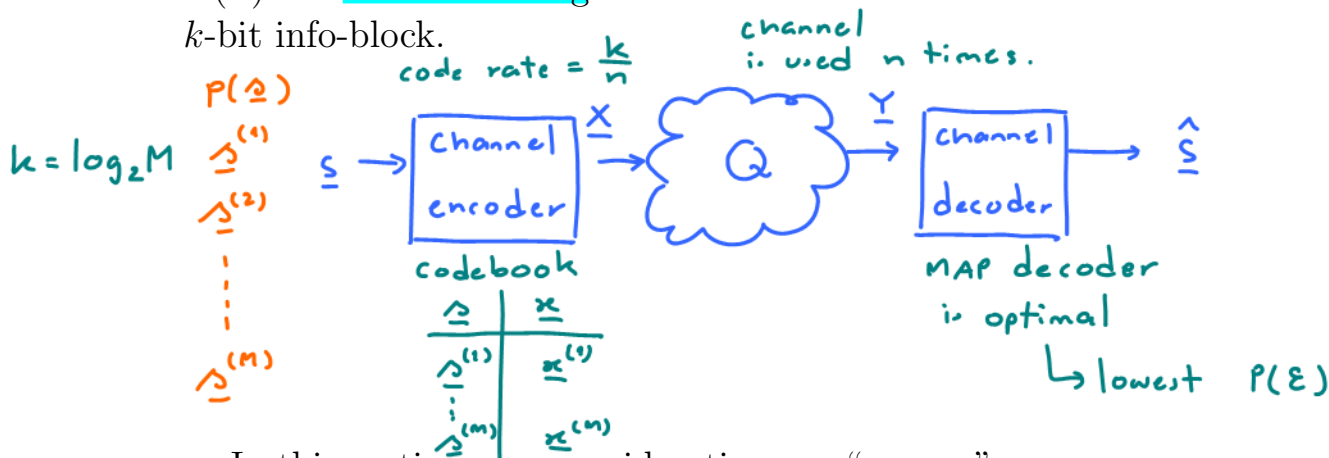


5.2 Operational Channel Capacity

5.10. In Section 3, we have studied how to compute the error probability $P(\mathcal{E})$ for digital communication systems over DMC. At the end of that section, we studied how to find the optimal decoder and the corresponding $P(\mathcal{E})$ for block encoding when the channel is used n times to transmit a k -bit info-block.



In this section, our consideration are “reverse”.

5.11. In this and the next subsections, we introduce a quantity called channel capacity which is crucial in benchmarking communication system. Recall that, in Section 2 where source coding was discussed, we were interested in the minimum rate (in bits per source symbol) to represent a source. Here, we are interested in the maximum rate (in bits per channel use) that can be sent through a given channel reliably.

We don't want a comm. system that can transmit @ 2 Gbps but has 75% error probability.

5.12. Here, reliable communication means arbitrarily small error probability can be achieved.

- This seems to be an impossible goal.
 - If the channel introduces errors, how can one correct them all?
 - * Any correction process is also subject to error, ad infinitum.

Definition 5.13. Given a DMC, its “operational” channel capacity is the maximum rate at which reliable communication over the channel is possible

it is possible to come up with appropriate n , encoder, and decoder.

- The channel capacity is the maximum rate in bits per channel use at which information can be sent with arbitrarily low error probability.

5.14. Claude Shannon showed, in his 1948 landmark paper, that this operational channel capacity is the same as the information channel capacity which we will discuss in the next subsection. From this, we can omit the words “operational” and “information” and simply refer to both quantities as the *channel capacity*.

Example 5.15. Soon, we will find that the capacity of a BSC with crossover probability $p = 0.1$ is approximately 0.531 bits per channel use. This means that for any rate $R < 0.531$ and any error probability $P(\mathcal{E})$ that we desire, as long as it is greater than 0, we can find a suitable n , a rate R encoder, and a corresponding decoder which will yield an error probability that is at least as low as our set value.

- Usually, for very low value of desired $P(\mathcal{E})$, we may need large value of n .

Example 5.16. Repetition code is not good enough.

- From Figure 8b, it is clear that repetition coding can not achieve the capacity of 0.531 bits per channel use. In fact, when we require error probability to be less than 0.1, the required repetition code needs $n \geq 3$. (For simplicity, let’s assume only odd value of n can be used here.) However, this means the rate is $\leq \frac{1}{3} \approx 0.33$ which is a lot less than 0.531.
- In fact, for any rate > 0 , we can see from Figure 8b that communication system based on repetition coding is not “reliable” according to Definition 5.12. For example, for rate = 0.02 bits per channel use, repetition code can’t satisfy the requirement that the error probability must be less than 10^{-15} . In fact, Figure 8b shows that as we reduce the

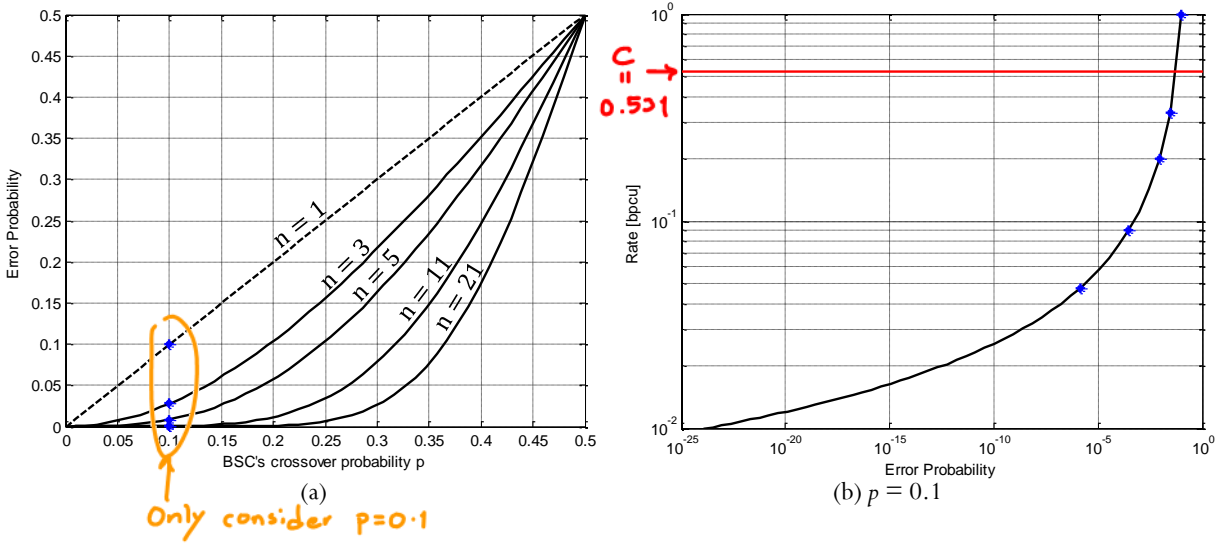


Figure 8: Performance of repetition coding with majority voting at the decoder

error probability to 0, the rate also goes to 0 as well. Therefore, there is no positive rate that works for all error probability.

However, because the channel capacity is 0.531, there must exist other encoding techniques which give better error probability than repetition code. Although Shannon's result gives us the channel capacity, it does not give us any explicit instruction on how to construct codes which can achieve that value.

5.3 Information Channel Capacity

5.17. In Section 5.1, we have studied how to compute the value of mutual information $I(X;Y)$ between two random variables X and Y . Here, X and Y are the channel input and output, respectively. We have also seen, in Example 5.8, how to compute $I(X;Y)$ when the joint pmf matrix \mathbf{P} is given. Furthermore, we have also worked on Example 5.9 in which the value of mutual information is computed from the prior probability vector $\underline{\mathbf{p}}$ and the channel transition probability matrix \mathbf{Q} . This second type of calculation is crucial in the computation of channel capacity. This kind of calculation is so important that we change the notation that we use for mutual information from $I(X;Y)$ to $I(\underline{\mathbf{p}}, \mathbf{Q})$.

↑↑
 This notation emphasizes the fact that the value of mutual information here depends on two quantities: the given matrix \mathbf{Q} and the row vector $\underline{\mathbf{p}}$.

Important Fact : $H(X) \leq \log_2 |\mathcal{X}|$ with equality iff X is uniform on \mathcal{X} .

Definition 5.18. Given a DMC channel, we define its “information” channel capacity as

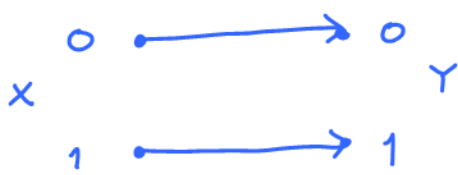
$$C = \max_{p(x)} I(X; Y) = \max_{\underline{p}} I(\underline{p}, Q),$$

where the maximum is taken over all possible input pmfs \underline{p} .

- Again, as mentioned in 5.14, Shannon showed that the “information” channel capacity defined here is equal to the “operational” channel capacity defined in Definition 5.13.

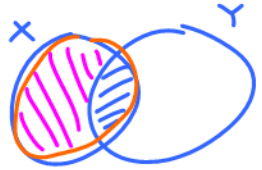
- Thus, we may drop the word “information” in most discussions of channel capacity.

Example 5.19. Find the channel capacity of a noiseless binary channel (a BSC whose crossover probability is $p = 0$).



$C = \max_{p(x)} I(X; Y) = \max_{\underline{p}} I(\underline{p}, Q)$

$I(X; Y) = I(\underline{p}, Q) = H(X) + H(Y) - H(X, Y)$
 $= H(X) - H(X|Y) = H(Y) - H(Y|X)$



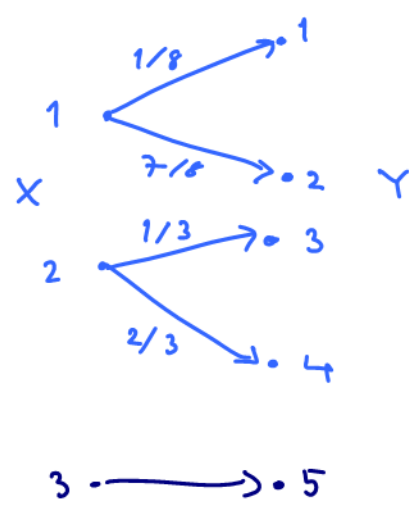
$H(X|Y) = \sum_Y q(y) H(X|y) = q(0) H(X|y=0) + q(1) H(X|y=1) = 0$

$I(X; Y) = H(X) - 0 = H(X) \leq \log_2 2 = 1$

become "=" when X is uniform. $C = 1$ is achieved by uniform X.

Example 5.20. Noisy Channel with Nonoverlapping Outputs: Find the channel capacity of a DMC whose

Example 5.20B



$Q = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{bmatrix} 1/8 & 7/8 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 2/3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix}$

$H(X|Y) = 0$

$I(X; Y) = H(X) - H(X|Y) = H(X)$

To maximize $H(X)$ (which in this case will also maximize $I(X; Y)$)

we use uniform $p(x)$

$C = \log_2 2^{56} = 1 \text{ bpcu}$
 $= \log_2 3 \text{ bpcu}$

Review: Some notation involving entropy

$$H(X) = \sum_x -p_x \log_2 p_x$$

$$H(p) = \sum_i -p_i \log_2 p_i$$

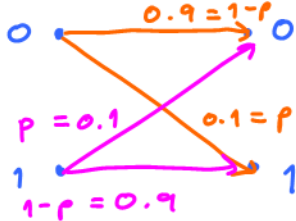
Binary entropy function

$$H(p) = -p \log_2 p - (1-p) \log_2 (1-p)$$

$$H([0.2 \ 0.8]) = -0.2 \log_2 0.2 - 0.8 \log_2 0.8$$

In this example, the channel appears to be noisy, but really is not. Even though the output of the channel is a random consequence of the input, the input can be determined from the output, and hence every transmitted bit can be recovered without error.

Example 5.21. Find the channel capacity of a BSC whose crossover probability is $p = 0.1$.



$$I(X;Y) = H(Y) - H(Y|X)$$

$$H(Y|X) = p(0)H(Y|X=0) + p(1)H(Y|X=1) = \underbrace{H(p)}_{\text{Binary Entropy Function}}$$

$$= p(0) \times 0.469 + p(1) \times 0.469 = 0.469$$

So, $H(Y|X)$ does not depend on p .

Therefore, to maximize $I(X;Y)$, we need to maximize $H(Y)$.

$$Q = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix} = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$$

To maximize $H(Y)$, we first try to make uniform Y .

Therefore,

$$q = pQ \Rightarrow \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

$$C = 1 - H(p)$$

$$= 1 - 0.469$$

$$= 1 - H(0.1) = 0.531$$

Try uniform X
Get uniform Y !

The BSC is a member of a class of channels called symmetric channel.

Definition 5.22. A DMC is called **symmetric** if (1) all the rows of its probability transition matrix Q are permutations of each other and (2) so are the columns.

Example 5.23. For each of the following Q , is the corresponding DMC symmetric?

$$\begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}$$

symmetric



weakly symmetric

$$\begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.2 & 0.3 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}$$

not symmetric



not weakly symmetric

$$\begin{bmatrix} 1/3 & 1/6 & 1/2 \\ 1/3 & 1/2 & 1/6 \end{bmatrix}$$

$\frac{2}{3} \quad \frac{2}{3} \quad \frac{2}{3}$
not symmetric

but weakly symmetric

Example 5.24. Find the channel capacity of a DMC whose

The capacity C of a given DMC can be found by

$$C = \max_{\mathbf{p}} I(X;Y)$$

$$Q = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}$$

Here,

$$I(X;Y) = H(Y) - H(Y|X)$$

$$H([0.2 \ 0.3 \ 0.5]) = H(\mathbf{r}) = 1.4855$$

$H(Y|X=x)$ is this value for all x

$$C = \log_2 3 - 1.4855 = 1.5850 - 1.4855 = 0.0995 \text{ bpcu}$$

So, we see that to maximize $I(X;Y)$, we need to maximize $H(Y)$.

Of course, we know that the maximum value of $H(Y)$ is $\log_2 |\mathcal{Y}|$ which happens when Y is uniform. If we can find \mathbf{p} which makes Y uniform, then it is the \mathbf{p} that gives capacity.

$$\mathbf{q}_y = \mathbf{p} Q$$

$$\begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}$$

Try uniform X

Get uniform Y ! ← possible because the column sums of Q are all the same.

Remark: If we can't find \mathbf{p} that makes Y uniform then $C < \log_2 |\mathcal{Y}| + H(\mathbf{r})$ and we have to find a different technique to calculate C .

Definition 5.25. A DMC is called **weakly symmetric** if (1) all the rows of its probability transition matrix Q are permutations of each other and (2) all the column sums are equal.

It should be clear from the definition that a symmetric channel is automatically weakly symmetric.

5.26. For a weakly symmetric channel,

$$C = \log_2 |\mathcal{Y}| - H(\mathbf{r}),$$

where \mathbf{r} is any row from the Q matrix. The capacity is achieved by a uniform pmf on the channel input.

5.27. Properties of channel capacity

(a) $C \geq 0$

(b) $C \leq \max_{in} \{\log |\mathcal{X}|, \log |\mathcal{Y}|\}$

Ex. Suppose $Q = \begin{bmatrix} 0.9 & 0.05 & 0.05 \\ 0.05 & 0.9 & 0.05 \\ 0.025 & 0.025 & 0.95 \end{bmatrix}$

Find C

- (a) 1.0944
- (b) 1.5944
- (c) 2.0944
- (d) 2.5944

Summary

① Channel capacity

Shannon showed that the two quantities are actually the same.

"operational": max rate at which reliable communication is possible

↳ arbitrarily small $P(E)$ can be achieved.

"information": $\max_P I(X;Y)$ [bpcu]

(relatively)

② Special cases in which capacity values can be found easily

(a) Noisy channel with nonoverlapping outputs [NO^2]

↳ only one non-zero element in each column of the Q matrix

$C = \log_2 |X|$ is achieved by uniform X

$$P = \left[\frac{1}{|x|} \quad \frac{1}{|x|} \quad \dots \quad \frac{1}{|x|} \right]$$

Important special case: Noiseless channel

(b) Weakly symmetric channel

↳ (1) all the rows of Q are permutations of each other

and

(2) all the column sums are equal

$C = \log_2 |Y| - H(\underline{r})$ is achieved by uniform X .

↑ any row of Q

Important special case: BSC

$$C = 1 - H(p)$$

↑ cross-over probability for BSC

↑ binary entropy function

$$Q = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

③ The key property that was used to find capacity values of all the special cases above is that:

for any RV X , $H(X) \leq \log_2 |X|$ with equality iff X is uniform

Another way to see this :

$$Q(y|x) \text{ does not depend on } x \Rightarrow X \perp\!\!\!\perp Y \Rightarrow I(X;Y) = 0$$

Example 5.28. Another case where capacity can be easily calculated: Find the channel capacity of a DMC of which **all the rows of its Q matrix are the same.**

$$Q = \begin{bmatrix} 1-a & a \\ 1-a & a \end{bmatrix} \quad \text{or} \quad Q = \begin{bmatrix} \underline{r} \\ \vdots \\ \underline{r} \end{bmatrix}$$

$$I(X;Y) = H(Y) - H(Y|X) = 0$$

for any p .

$$H(Y|x) = H(\underline{r}) \Rightarrow H(Y|X) = \sum_x p(x) \underbrace{H(Y|x)}_{H(\underline{r})} = H(\underline{r})$$

$$C = \max_P \underbrace{I(X;Y)}_0 = 0 \text{ bpcu}$$

$$I(X;Y) = H(Y) - H(Y|X)$$

$$\underline{y} = \underline{p} Q = [p_1 \ p_2 \ \dots] Q = p_1 \underline{r} + p_2 \underline{r} + \dots = \underline{r} (\sum_x p(x)) = \underline{r}$$

$$H(Y) = H(\underline{r})$$

5.29. So far, we worked with toy examples in which finding capacity is relatively easy. In general, there is **no closed-form solution for computing capacity.** When we have to deal with cases that do not fit in any special family of **Q** described in the examples above, the maximum can be found by standard **nonlinear optimization** techniques. **MATLAB: fmincon**

Example 5.30. The capacity of a BAC whose $Q(1|0) = 0.9$ and $Q(0|1) = 0.4$ can be found by first realizing that $I(X;Y)$ here is a function of a single variable: p_0 . The plot of $I(X;Y)$ as a function of p_0 gives some rough estimates of the answers. One can also solve for the optimal p_0 by simply taking derivative of $I(X;Y)$ and set it equal to 0. This gives the capacity value of 0.0918 bpcu which is achieved by $\underline{p} = [0.5376, 0.4624]$.

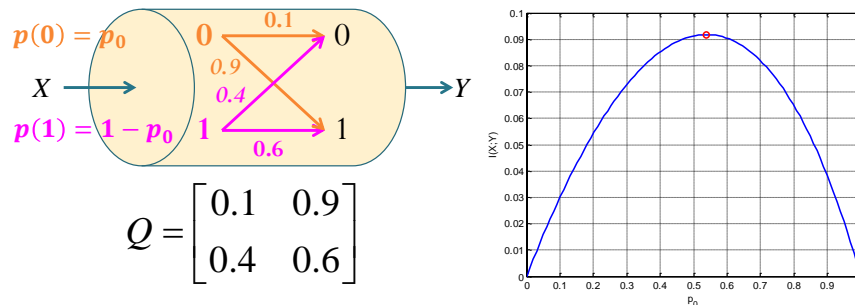


Figure 9: Maximization of mutual information to find capacity of a BAC channel. Capacity of 0.0918 bits is achieved by $\underline{p} = [0.5376, 0.4624]$

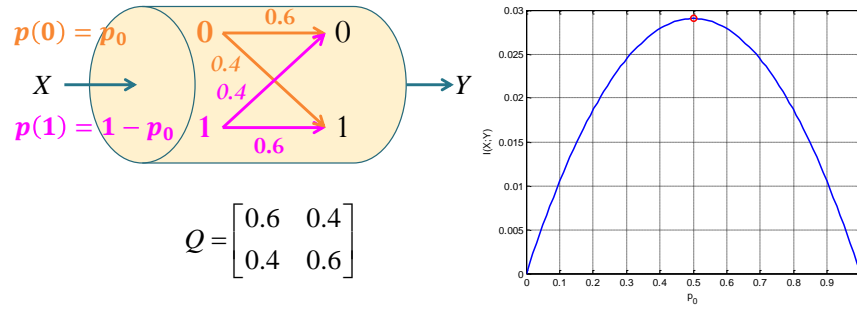


Figure 10: Maximization of mutual information to find capacity of a BSC channel. Capacity of 0.029 bits is achieved by $\underline{\mathbf{p}} = [0.5, 0.5]$

5.31. Blahut-Arimoto Algorithm [4, Section 10.8]: Alternatively, in 1972, Arimoto [1] and Blahut [2] independently developed an iterative algorithm to help us **approximate the pmf $\underline{\mathbf{p}}^*$** which **achieves capacity C** . To do this, start with any (guess) input pmf $p_0(x)$, define a sequence of pmfs $p_r(x)$, $r = 0, 1, \dots$ according to the following iterative prescription:

(a) $q_r(y) = \sum_x p_r(x) Q(y|x)$ for all $y \in \mathcal{Y}$.

(b) $c_r(x) = 2 \left(\sum_y Q(y|x) \log_2 \frac{Q(y|x)}{q_r(y)} \right)$ for all $x \in \mathcal{X}$.

(c) It can be shown that

$$\underbrace{\log_2 \left(\sum_x p_r(x) c_r(x) \right)}_A \leq C \leq \underbrace{\log_2 \left(\max_x c_r(x) \right)}_B$$

- If the lower-bound and upper-bound above are close enough. We take $p_r(x)$ as our answer and the corresponding capacity is simply the average of the two bounds.
- Otherwise, we compute the pmf

$$p_{r+1}(x) = \frac{p_r(x) c_r(x)}{\sum_x p_r(x) c_r(x)} \quad \text{for all } x \in \mathcal{X}$$

and repeat the steps above with index r replaced by $r + 1$.

5.32. Shannon's (Noisy Channel) Coding theorem [Shannon, 1948]

- (a) Reliable communication over a (discrete memoryless) channel is possible if the communication rate R satisfies $R < C$, where C is the channel capacity.

In particular, for any $R < C$, there exist codes (encoders and decoders) with sufficiently large n such that

$$P(\mathcal{E}) \leq 2^{-n \times E(R)},$$

where $E(R)$ is

- a positive function of R for $R < C$ and
- completely determined by the channel characteristics

- (b) At rates higher than capacity, reliable communication is impossible.

5.33. Significance of Shannon's (noisy channel) coding theorem:

- (a) Express the limit to reliable communication
- (b) Provides a yardstick to measure the performance of communication systems.
- A system performing near capacity is a near optimal system and does not have much room for improvement.
 - On the other hand a system operating far from this fundamental bound can be improved (mainly through coding techniques).

5.34. Shannon's nonconstructive proof for his coding theorem

- Shannon introduces a method of proof called **random coding**.
- Instead of looking for the best possible coding scheme and analyzing its performance, which is a difficult task,
 - all possible coding schemes are considered
 - * by generating the code randomly with appropriate distribution
 - and the performance of the system is averaged over them.
 - Then it is proved that if $R < C$, the average error probability tends to zero.

- Again, Shannon proved that
 - as long as $R < C$,
 - at any arbitrarily small (but still positive) probability of error,
 - one can find (there exist) at least one code (with sufficiently long block length n) that performs better than the specified probability of error.
- If we used the scheme suggested and generate a code at random, the code constructed is likely to be good for long block lengths.
- No structure in the code. Very difficult to decode

5.35. Practical codes:

- In addition to achieving low probabilities of error, useful codes should be “simple”, so that they can be encoded and decoded efficiently.
- Shannon’s theorem does not provide a practical coding scheme.
- Since Shannon’s paper, a variety of techniques have been used to construct good error correcting codes.
 - The entire field of coding theory has been developed during this search.
- Turbo codes have come close to achieving capacity for Gaussian channels.