4.2 Operational Channel Capacity

4.16. In Chapter 3, we have studied how to compute the error probability $P(\mathcal{E})$ for digital communication systems over DMC. At the end of that chapter, we studied block encoding where the channel is used n times to transmit a k-bit info-block.

In this section, our consideration is "reverse".

4.17. In this and the next sections, we introduce a quantity called channel capacity which is crucial in benchmarking communication system. Recall that, in Chapter 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*.

4.18. 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 4.19. Given a DMC, its **"operational" channel capacity** is the maximum rate at which *reliable* communication over the channel *is possible.*

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

4.20. 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 section. From this, we can omit the words

"operational" and "information" and simply refer to both quantities as the *channel capacity.*

Example 4.21. In Example 4.35, 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 desired value of $P(\mathcal{E})$, we may need large value of n.

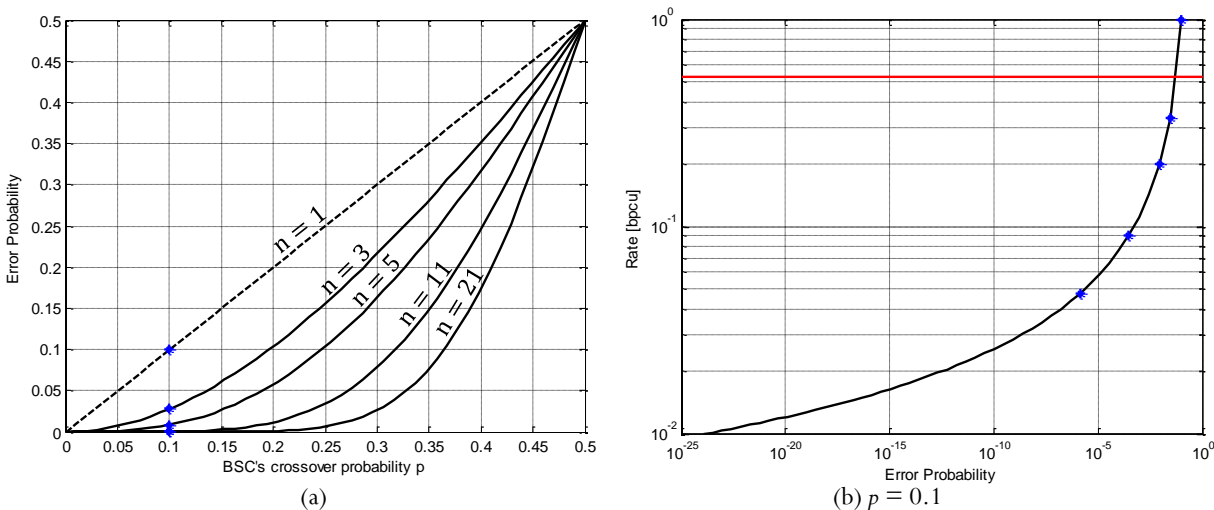**Example 4.22.** Repetition code is not good enough.

Figure 17: Performance of repetition coding with majority voting at the decoder

- Continue from Example 4.21.

- In Figure 17b, with repetition code, trying to reduce the error probability to be less than the original p even a little bit already causes the rate to drop far below the capacity level indicated by the red horizontal line.

- In fact, for any rate > 0, we can see from Figure 17b that communication system based on repetition coding is not "reliable" according

to Definition 4.18. 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 17b shows that as we reduce the 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 [bpcu], 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.

4.3 Information Channel Capacity

4.23. In Section 4.1, we have studied how to compute the value of mutual information $I(X;Y)$ between two random variables X and Y. Recall that, here, X and Y are the channel input and output, respectively. We have also seen, in Example 4.14, how to compute $I(X;Y)$ when the joint pmf matrix $\mathbf{P}$ is given. Furthermore, we have also worked on Example 4.15 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 may write the mutual information $I(X;Y)$ as $I(\underline{\mathbf{p}}, \mathbf{Q})$.

Definition 4.24. Given a DMC channel, we define its "information" channel capacity as

$$C = \max_{\underline{\mathbf{p}}} I(X;Y) = \max_{\underline{\mathbf{p}}} I(\underline{\mathbf{p}}, \mathbf{Q}), \tag{34}$$

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

- Again, as mentioned in 4.20, Shannon showed that the "information" channel capacity defined here is equal to the "operational" channel capacity defined in Definition 4.19.

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

4.5 Shannon's Coding theorem

4.42. 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

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

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.

4.43. 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).

4.44. 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

4.45. 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.