

ECE 461: Digital Communication

Lecture 3: Histogram to Optimum Receiver

Introduction

In this lecture we focus our study on how to use the detailed statistical knowledge available in the histogram of the noise in doing reliable communication at a desired level of reliability. Though our specific interest will be on the Gaussian statistics, it helps (for later lectures) to study the more general situation. For a fixed transmission strategy, we will derive the *optimum* receiver in terms of minimizing the unreliability of communication. Towards doing this, we formally define what unreliability means by carefully looking at the different sources of randomness and what statistical assumptions we make about them. We conclude with a fundamental relation between the variance of the noise σ^2 , the transmit energy constraint E , and the reliability of communication.

Sources of Randomness

There are two sources of randomness from the perspective of the receiver: one intrinsic (the information – bits – itself is unknown) and the other extrinsic (the additive noise introduced by the channel). The receiver typically knows some statistical information about these sources of knowledge.

- *Statistics of the bit*: this is the fraction of bits that are 0. If there is some prior information on how likely the transmitted information bit is say, 1, then that could factor in the decision rule. In the extreme instance, if we somehow knew before the communication process that the information bit is 1 for sure, then we don't need to worry about the received voltage. We just decide at the receiver that the information bit is 1. Many a time, no such prior knowledge is available. In this case, we suppose that the information bit is *equally likely* to be 1 or 0.
- *Noise Statistics*: knowing whether the noise is more likely to be small or large will help the receiver make the decision. For instance, if the noise is more likely to be near zero than large, the receiver would likely pick the nearer of the two possible transmit voltages as compared to the received voltage (the so-called *nearest-neighbor* rule). One of the main conclusions at the end of Lecture 2 is that additive noise in the physical world is (far) more likely to be near its mean than away from it.

Figure 1 illustrates the action taken at the receiver.

Formal Definition of Reliable Communication

Consider a single bit to be communicated reliably. Figure 2 diagrammatically illustrates the familiar bits-to-voltages mapping at the transmitter.

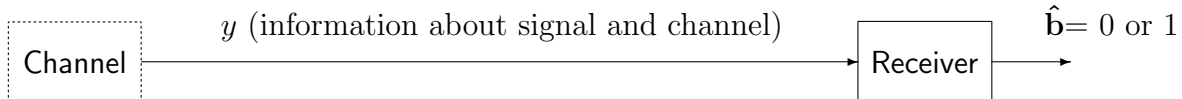


Figure 1: The basic block diagram of a receiver.



Figure 2: Mapping for sending 1 bit across an AGN channel.

The main job at the receiver is to decide on the information bit transmitted, denoted by say, \hat{b} , based on the received voltage y . The correct choice of the decision rule at the receiver is the one that maximizes the reliability of communication. Alternatively, we want to minimize the unreliability of communication. We will say an *error* occurs whenever communication is unreliable. In this case, the transmitted information is just one bit and there is only one way an error can occur. More generally, when we get around to sending multiple bits of information we will follow the convention that an error occurs *even if a single bit* is communicated erroneously. This convention is a natural byproduct of the nature of the digital world of information around us.

Actual sources of information (such as voice, images and video) have features that range the whole gamut from being very important to hardly any. For instance, if we consider digitizing voice with a 16-bit A/D converter the most significant bits (MSBs) are (almost by definition!) more important than the least significant bits (LSBs). Indeed, when communicating the 16-bit digital representation of the analog voice sample, we need to pay more attention to the reliability of the MSBs as compared to the LSBs.

On the other hand, the digital world around is organized very differently. Information collection is typically at a very different engineering level than information transmission: Information collection is done typically by microphones, cameras and camcorders. Information transmission is done typically over the ethernet or wireless. There are so many layers of separation between the engineering devices that do these two operations. Specifically, there are, starting from information collection and moving down to information transmission:

- the *application* layer, that decides whether the digital format for the voice is **.wav** or

.mp3;

- the *transport* layer, that decides whether the TCP/IP protocol is being used or a proprietary one used by cell phones and the corresponding impact on the digital representation of the analog voice sample;
- the *networking* and *physical* layers, that decide what format to finally package the digital voice data in.

So by the time the transmission of communication is initiated, the “analog” nature of the digital information (MSBs and LSBs) is entirely lost (or at least hidden underneath a whole lot of protocol layers). So, the communication problem is usually stated as trying to equally reliably send all the bits (whether they are MSBs, or LSBs, or formatting information corresponding to the different protocols involved). We will follow this tradition in this course by considering all the bits to be equally important.

We now have a formal definition of how reliable communication is. It is the *average* probability (averaged over the a priori probabilities with which the information bits take different values) with which all the bits are correctly received. We will next see the decision rule at the receiver that is optimal in the sense of allowing the most reliable communication.

The Optimal Decision Rule: MAP

To begin with, let us list all the information that the receiver has.

1. The a priori probabilities of the two values the information bit can take. We will normally consider these to be equal (to 0.5 each).
2. The received voltage y . While this is an analog value, i.e., any real number, in engineering practice we *quantize* it at the same time the waveform is converted into a discrete sequence of voltages. For instance if we are using a 16-bit ADC for the discretization, then the received voltage y can take one of 2^{16} possible values. We will start with this discretization model first.
3. The *encoding rule*. In other words we need to know how the information bit is mapped into voltages at the transmitter. For instance, this means that the mapping in illustrated in Figure 2 should be known to the receiver. This could be considered part of the protocol that both the transmitter and receiver subscribe to. In engineering practice, all widespread communication devices subscribe to a universally known *standard*. For example, Verizon cell phones subscribe to a standard known as **CDMA**.

Assuming L possible discrete received voltage levels, Figure 3 illustrates the two possible transmit voltages and the chance that they lead to the discrete shows a plot of possible transmitted and the chance with which they could lead to the L possible received voltages (here $L = 3$). The additive Gaussian noise channel model combined with the discretization of the received voltage level naturally leads to a statistical characterization of how likely a

certain received voltage level is given a certain transmit voltage level. In Figure 3, we have written these probabilities in the most general form; in a homework exercise you are asked to calculate these values for a specific way of discretization of the received voltage.

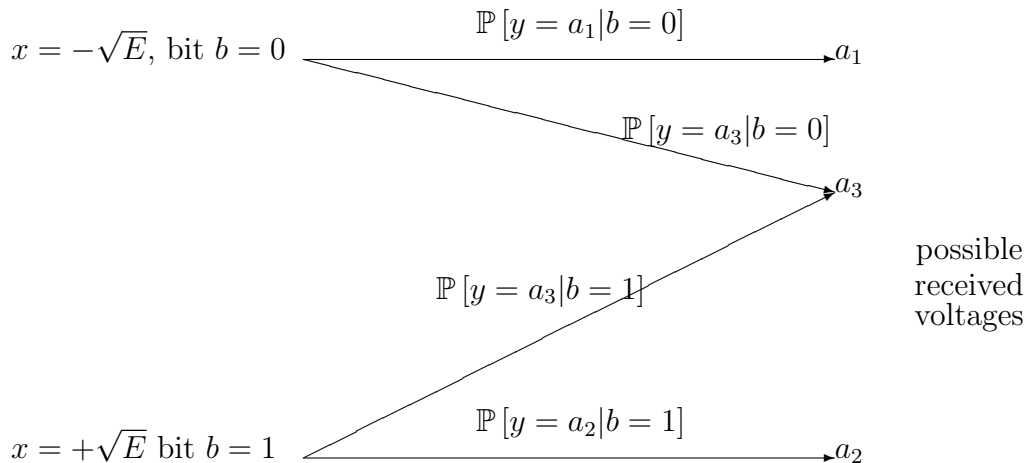


Figure 3: Sent and received voltage pairs along with their conditional probabilities

The probability that the information bit is i (either 1 or 0) *and* the received voltage is a (one of L possible values, denoted by a_1, \dots, a_j) is simply

$$\mathbb{P}[b = i, y = a] = \mathbb{P}[b = i | y = a] \mathbb{P}[y = a], \quad (1)$$

where the unconditional probability that the received voltage is a ,

$$\mathbb{P}[y = a] = \mathbb{P}[b = 0, y = a] + \mathbb{P}[b = 1, y = a], \quad (2)$$

does *not depend* on the actual value of the information bit b . The quantity $\mathbb{P}[b = i | y = a]$ in Equation 1 is known as the *a posteriori* probability of the information bit being equal to i . This captures the role of the communication process: the received voltage level alters our perception of what the information bit could possibly be.

The decision rule at the receiver then is to map every possible received voltage level to a particular estimate $\hat{b}(a)$ of what was sent. The reliability of communication conditioned on a specific received voltage level (say, a) is simply the *a posteriori* probability of the information bit b being equal to the estimate \hat{b} :

$$\mathbb{P}[\mathcal{C} | y = a] = \mathbb{P}[b = \hat{b}(a) | y = a]. \quad (3)$$

We want to maximize $\mathbb{P}[\mathcal{C} | y = a]$, so we should just choose $\hat{b}(a)$ to be that value (1 or 0) which has the larger *a posteriori* probability.

But how does one calculate this quantity at the receiver, using the three quantities that the receiver has access to (enumerated at the beginning of this lecture)? For any received

voltage level a in the set $\{a_1, \dots, a_L\}$, the a posteriori probability for the information bit b being equal to, say 1, can be written using the *Bayes* rule as:

$$\mathbb{P}[b = 1|y = a] = \frac{\mathbb{P}[y = a|b = 1] \mathbb{P}[b = 1]}{\mathbb{P}[y = a]}. \quad (4)$$

Similarly the a posteriori probability for the information bit b being equal to 0, given that the received voltage is the same a , is

$$\mathbb{P}[b = 0|y = a] = \frac{\mathbb{P}[y = a|b = 0] \mathbb{P}[b = 0]}{\mathbb{P}[y = a]}. \quad (5)$$

Since the denominator is common to both the two a posteriori probabilities and the decision rule is only based on the relative comparison, we only need the numerators to form the decision rule. The a priori probabilities $\mathbb{P}[b = 1]$ and $\mathbb{P}[b = 0]$ sum to unity and is part of the information the receiver has ahead of time. The *likelihoods*

$$\mathbb{P}[y = a|b = 1] \quad \text{and} \quad \mathbb{P}[y = a|b = 0] \quad (6)$$

is to be calculated based on the statistical knowledge of the channel noise. We will do this shortly for the Gaussian noise, but a couple of quick digressions are in order before we do that.

ML Decision Rule

As we discussed earlier, a common situation in communication is that the a priori probabilities of the information bit are equal to each other. In this (typical) situation, the MAP rule simplifies even more. It now suffices to just compare the two likelihoods (the two quantities in Equation 6). The decision rule is then to decide that \hat{b} is 1 if

$$\mathbb{P}[y = a|b = 1] > \mathbb{P}[y = a|b = 0], \quad (7)$$

and 0 otherwise. This rule is called the *maximum likelihood* rule. Due to its typicality, this will be the decision rule we will use throughout this course at the receiver.

MAP and ML Rules for the AGN Channel

Given the universality of the Gaussian statistics for additive noise models, it is of immediate interest to calculate these rules for such a statistical channel model. The only potential hurdle is that the statistics are described for *analog* valued noise (and hence received voltage) levels. In our setup so far, we only considered a discrete set of voltage levels. We now have one of two options: either generalize the previous description to analog values (a whole continuous range of voltage levels than a finite number) or deduce the statistics of the discrete noise levels as induced by the Gaussian statistics on the analog noise level and the ADC. We take the former approach below.

The generalization required is only a matter of calculating the a posteriori probabilities conditioned on a whole continuous range of received voltage levels, than just a finite number. Following the earlier calculation in Equation 4, we see the main technical problem:

$$\mathbb{P}[y = a|b = 1] \quad \text{and} \quad \mathbb{P}[y = a] \quad (8)$$

are both zero: the chance that an analog noise level is exactly a value we want is simply zero. So we cannot use Bayes rule naively. Since we only need the ratio of these two quantities (cf. Equation 8) in the MAP rule, we can use the L'Hopital's rule:

$$\frac{\mathbb{P}[y = a|b = 1]}{\mathbb{P}[y = a]} = \lim_{\epsilon \rightarrow 0} \frac{\mathbb{P}[y \in (a - \epsilon, a + \epsilon)|b = 1]}{\mathbb{P}[y \in (a - \epsilon, a + \epsilon)]} = \frac{f_y(a|b = 1)}{f_y(a)}. \quad (9)$$

Here $f_y(\cdot)$ is the PDF of the *analog* received voltage y and $f_y(\cdot|b = 1)$ is the PDF of the received voltage conditioned on the event that the information bit b is 1. So, the MAP rule when the received voltage is equal to a is:

$$\begin{aligned} &\text{decide } \hat{b} = 1 \text{ if} \\ &\quad \mathbb{P}[b = 1] f_y(a|b = 1) \geq \mathbb{P}[b = 0] f_y(a|b = 0) \\ &\text{and 0 otherwise.} \end{aligned} \quad (10)$$

The ML rule is simpler, as usual:

$$\begin{aligned} &\text{decide } \hat{b} = 1 \text{ if} \\ &\quad f_y(a|b = 1) \geq f_y(a|b = 0) \\ &\text{and 0 otherwise.} \end{aligned} \quad (11)$$

For the additive noise channel, it is a straightforward matter to calculate the conditional PDFs of the received voltage. Indeed

$$f_y(a|b = 1) = f_y(a|x = +\sqrt{E}) \quad (12)$$

$$= f_w(a - \sqrt{E}|x = +\sqrt{E}) \quad (13)$$

$$= f_w(a - \sqrt{E}). \quad (14)$$

In the first step we used the knowledge of the mapping between the information bit to transmit voltage levels (cf. Figure 2). The second step is simply using the fact that $w = y - x$. The third step used the statistical independent of the additive noise and the voltage transmitted. So, the MAP and ML rules for the additive noise channel are:

$$\begin{aligned} &\text{MAP: decide } \hat{b} = 1 \text{ if} \\ &\quad \mathbb{P}[b = 1] f_w(a + \sqrt{E}) \geq \mathbb{P}[b = 0] f_w(a - \sqrt{E}) \\ &\text{and 0 otherwise;} \end{aligned} \quad (15)$$

and

ML: decide $\hat{b} = 1$ if

$$f_w(a + \sqrt{E}) \geq f_w(a - \sqrt{E}) \quad (16)$$

and 0 otherwise.

We can simplify the rules even further given some more knowledge of the statistics of the noise. For example, suppose we know that the noise w is more likely to be small in magnitude than large (since the the mean was supposed to be zero, this means that the noise is more likely to be near the average value than farther away):

$$f_w(a) \geq f_w(b), \quad |a| \geq |b|. \quad (17)$$

This property is definitely true for the Gaussian statistics. Then the ML rule simplifies significantly: decide $\hat{b} = 0$ if

$$\begin{aligned} f_w(a + \sqrt{E}) &\geq f_w(a - \sqrt{E}) \\ (a + \sqrt{E})^2 - 0^2 &\leq (a - \sqrt{E})^2 - 0^2 \\ 4\sqrt{E}a &\leq 0 \\ a &\leq 0. \end{aligned}$$

In other words, the ML decision rule take the received voltage $y = a$ and estimates:

$$\begin{aligned} a \leq 0 &\Rightarrow 0 \text{ was sent} \\ \text{Else, } &1 \text{ was sent.} \end{aligned}$$

Figure 4 illustrates the ML decision rule when superposed on the “bits to voltage” mapping (cf. Figure 2). The decision rule picks that transmitted voltage level that is closer to the received voltage (closer in the usual sense of Euclidian distance). Hence, the maximum likelihood (ML) rule is also known as the minimum distance rule or the *nearest neighbor* rule.

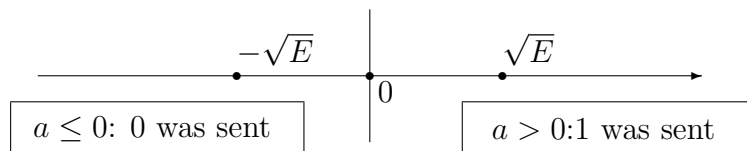


Figure 4: The ML rule superposed on Figure 2.

In the rest of this lecture we look at two natural extensions of the material developed painstakingly so far:

1. an evaluation of the performance of the ML rule and the reliability to communication it affords. Our focus will be on understanding the relation between the energy constraint at the transmitter and the noise variance in deciding the reliability level.
2. move forward towards sending multiple bits at the same time instant. There is a natural generalization of the nearest-neighbor rule and the corresponding level of reliability to communication.

Reliability of Communication

The receiver makes an error if it decides that a 1 was sent when a 0 was sent, or vice versa. The *average* error is a weighted sum of the probabilities of these two types of error events, with the weights being equal to the a priori probabilities of the information bit:

$$\mathbb{P}[\mathcal{E}] = \mathbb{P}[\mathcal{E}|b=0] \mathbb{P}[b=0] + \mathbb{P}[\mathcal{E}|b=1] \mathbb{P}[b=1]. \quad (18)$$

We suppose the a priori probabilities are equal (to 0.5 each). Let us focus on one of the error events by supposing that the information bit was actually 0. Then with the nearest neighbor rule,

$$\begin{aligned} P[\mathcal{E}|b=0] &= P[\hat{b}=1|b=0] \\ &= P[y > 0|b=0] \\ &= P[x+w > 0|b=0] \\ &= P[w > \sqrt{E}] \\ &= Q\left(\frac{\sqrt{E}}{\sigma}\right). \end{aligned}$$

Due to the complete symmetry of the mapping from the bit values to the voltage levels and the decision rule, the probability of the other error event is also the same:

$$\begin{aligned} P[\mathcal{E}|b=1] &= P[\hat{b}=0|b=1] \\ &= P[y < 0|b=1] \\ &= P[x+w < 0|b=1] \\ &= P[w < -\sqrt{E}] \\ &= Q\left(\frac{\sqrt{E}}{\sigma}\right). \end{aligned}$$

The average probability of error is also equal to the same $Q\left(\frac{\sqrt{E}}{\sigma}\right)$.

SNR and Reliability of Communication

The first observation we make from the expression for the unreliability of communication is that it depends only on the *ratio*, of the transmit energy E and the noise variance σ^2 : the error probability is

$$Q\left(\sqrt{\text{SNR}}\right). \quad (19)$$

We have already seen this phenomenon before in Lecture 1, albeit in a deterministic setting. This ratio is called the *signal to noise ratio*, or simply SNR. Basically, the communication engineer can design for a certain reliability level by choosing an appropriate SNR setting. While the $Q(\cdot)$ function can be found in standard statistical tables, it is useful for the communication engineer to have a rule of thumb for how sensitive this SNR “knob” is in terms of the reliability each setting offers. For instance, it would be useful to know by how much the reliability increases if we double the SNR setting. To do this, it helps to use the following approximation (cf. Question 3(e) in Homework 1):

$$Q(a) \approx \frac{1}{2} \exp\left(-\frac{a^2}{2}\right). \quad (20)$$

This approximation implies that the unreliability level

$$Q\left(\sqrt{\text{SNR}}\right) \approx \frac{1}{2} e^{\frac{-\text{SNR}}{2}} \quad (21)$$

Equation (21) is saying something very interesting: it says that the SNR has an *exponential* effect on the probability of error. For instance, supposing we double the SNR setting the error probability

$$Q\left(2\sqrt{\text{SNR}}\right) \approx \left(Q\left(\sqrt{\text{SNR}}\right)\right)^2, \quad (22)$$

is a *square* of what it used to be before.

Transmitting Multiple Bits

Let us consider the same transmit energy constraint as before and see by how much the reliability is reduced when we transmit multiple bits in the same single time sample. As in Lecture 1, let us start with mapping the bits to voltage levels that are as far apart from each other: this is illustrated in Figure 5 for 2 bits (and hence 4 voltage levels).

The ML rule is the same nearest neighbor one: pick that transmit voltage level that is closest to the received voltage level. Figure 6 provides a short justification.

Reliability of Communication

A look at the bits-to-voltage mapping in Figure 5 suggests that the inner two voltage levels (\clubsuit and \spadesuit) are less reliable than the outer ones (\heartsuit and \diamondsuit): the inner levels have neighbors on *both* sides while the outer ones have only one neighbor. We can calculate the probability

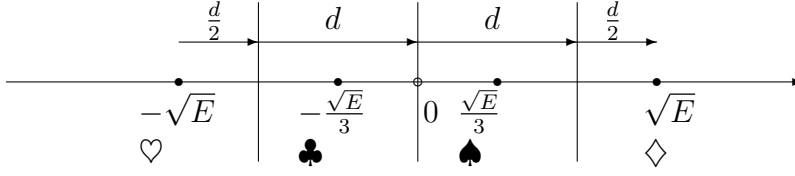


Figure 5: Sending 2 information bits across an AGN channel.

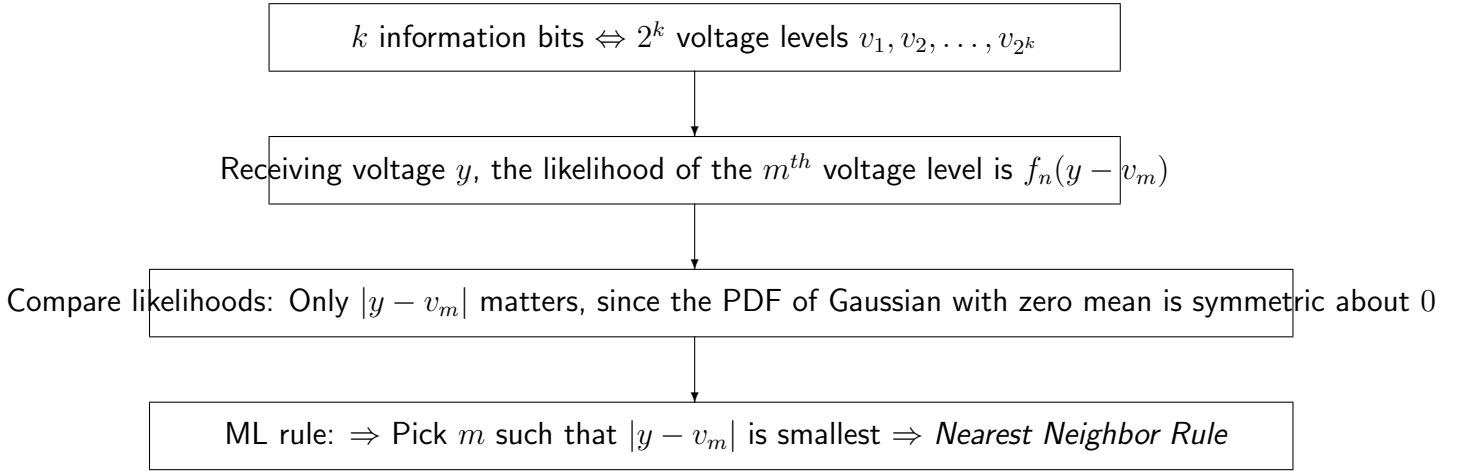


Figure 6: ML Rule for k information bits is the nearest neighbor rule.

of making an error with the ML rule given the transmission of an outer voltage level (say, the ♥) exactly as in the earlier part of this lecture:

$$\mathbb{P}[\mathcal{E}|\heartsuit] = \mathbb{P}\left[w > \frac{d}{2}\right] \quad (23)$$

$$= Q\left(\frac{\sqrt{E}}{3\sigma}\right). \quad (24)$$

On the other hand, the probability of making an error with the ML rule given the transmission of an inner voltage level (say, the ♠) is:

$$\mathbb{P}[\mathcal{E}|\spadesuit] = \mathbb{P}\left[\left\{w > \frac{d}{2}\right\} \cup \left\{w < \frac{-d}{2}\right\}\right] \quad (25)$$

$$= 2Q\left(\frac{\sqrt{E}}{3\sigma}\right). \quad (26)$$

Finally, the average probability of making an error (averaged over all the 4 different voltage levels) is

$$\mathbb{P}[\mathcal{E}] = 2 \times \frac{1}{4} \times Q\left(\frac{\sqrt{E}}{3\sigma}\right) + 2 \times \frac{1}{4} \times 2Q\left(\frac{\sqrt{E}}{3\sigma}\right) \quad (27)$$

$$= \frac{3}{2} Q\left(\frac{\sqrt{E}}{3\sigma}\right). \quad (28)$$

We have done this analysis for transmitting $k = 2$ information bits. The same analysis carries over to larger k . Indeed, the error probability is readily calculated to be, as an extension of Equation (28):

$$\mathbb{P}[\mathcal{E}] = \left(2 - \frac{1}{2^{k-1}}\right) Q\left(\frac{d}{2\sigma}\right), \quad (29)$$

where the minimum distance between two of the equally spaced voltage levels is

$$d = \frac{\sqrt{E}}{2^k - 1}. \quad (30)$$

As before the error probability is determined only by the SNR of the channel.

An Engineering Conclusion

At the end of Lecture 1, we noted a relationship between completely reliable communication of number of bits, transmit energy, bounds to the additive noise: the required energy to maintain the same reliability essentially quadrupled when we looked to transmit one extra bit. We have relaxed our definition of reliability, and replaced the noise bounds by a statistical quantity (an appropriate multiple of the standard deviation σ). But the essential relationship between transmit energy and the number of information bits you can reliably transmit (at any reliability level) is unchanged: a linear increase in the number of transmit bits requires an exponential increase in the required transmit energy while maintaining the same reliability level.

Looking Ahead

We notice from Equation (29) that the reliability goes down to zero as the number of information bits sent k increases. This picture suggests that increasing the rate of communication invariably leads to a degradation of the reliability. We will see in the next few lectures that this is only an artifact of the specific communication scheme we have chosen and not a fundamental relation. Specifically, we will see that buffering the information bits and jointly communicating them over multiple time instants will improve the reliability significantly, as compared to sending the bits on a time-sample-by-time-sample basis. Sure, there is no free lunch: the cost paid here is the delay involved in buffering the information and then communicating them jointly.