Appendix 3.A Rectangle and Sinc as Fourier Transform Pairs

The Fourier transform of a rectangular pulses is a sinc pulse. Often one has to go back and forth between such Fourier pairs. The purpose of this appendix is to make it easier to figure out the details.

First of all let us recall that a function g and its Fourier transform $g_{\mathcal{F}}$ are related by

$$\begin{aligned} g(u) &= \int g_{\mathcal{F}}(\alpha)\exp(j2\pi u\alpha)d\alpha \\ g_{\mathcal{F}}(v) &= \int g(\alpha)\exp(-j2\pi v\alpha)d\alpha. \end{aligned}$$

Notice that $g_{\mathcal{F}}(0)$ is the area under g and $g(0)$ is the area under $g_{\mathcal{F}}$.

Next let us recall that $\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$ is the function that equals 1 at $x = 0$ and equals 0 at all other integer values of x. Hence if $a, b \in \mathbb{R}$ are arbitrary constants, $a\text{sinc}(bx)$ equals a at $x = 0$ and and equals 0 at nonzero multiples of $1/b$.

If you could remember that the area under $a\text{sinc}(bx)$ is a/b then, from the two facts above, you could conclude that its Fourier transform, which you know is a rectangle, has hight equals a/b and area a. Hence the width of this rectangle must be b.

It is actually easy to remember that the area under $a\text{sinc}(bx)$ is a/b: it is the area of the triangle described by the main lobe of $a\text{sinc}(bx)$, namely the area of the triangle with coordinates $(-1/b, 0)$, $(0, a)$, $(1/b, 0)$.

Appendix 3.B White Gaussian Noise

We assume that you are familiar with the concept of White Gaussian Noise. The purpose of this appendix is just to write down what you absolutely need to remember, for the purpose of this Chapter, about White Gaussian Noise.

If $N(t)$ is White Gaussian Noise of double-sided spectral density $\frac{N_0}{2}$ then:

- Its covariance function is
$$K_N(\tau) \triangleq \frac{N_0}{2}\delta(\tau), \ \forall\tau.$$
- Its spectrum (the Fourier transform of the covariance function) is
$$S_N(f) = \frac{N_0}{2}.$$
- If
$$Z_i = \int_{-\infty}^{\infty} N(t)g_i(t)dt, \quad i = 1, \ldots, K,$$

then $(Z_1, \ldots, Z_N)$ is a zero-mean Gaussian random vector and for any $1 \le i,\ j \le K$,

$$\begin{aligned}
E[Z_i Z_j] &= E\left[\int_{-\infty}^{\infty} N(t) g_i(t) \int_{-\infty}^{\infty} N(\xi) g_i(\xi) d\xi\right] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E[N(t)N(\xi)] g_i(H g_j(\xi) dt\xi \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{N_0}{2}\delta(t-\xi) g_i(t) g_j(s) dt d\xi \\
&= \int_{-\infty}^{\infty} \frac{N_0}{2} g_i(t) g_j(t) dt
\end{aligned}$$

In particular, if $g_1(t), \ldots, g_k(t)$ are an orthonormal set then $Z_1, \ldots, Z_K$ are iid $\mathcal{N}(0, \frac{N_0}{2})$.

Chapter 4

Signal Design Trade-Offs

4.1 Introduction

It is time to shift our focus to the transmitter and take a look at some of the options we have in terms of choosing the signal constellation. The goal is to build up some intuition about the impact that those options have on fundamental performance measures such as transmission rate, bandwidth, power, and error probability. Throughout this section we assume that the channel is the AWGN channel and that the receiver implements a ML decision rule.

To put things into perspective, we mention from the outset that the problem of choosing a convenient signal constellation is not as clean-cut as the receiver design problem that has kept us busy until now. The reason is that the receiver design problem has a clear objective, namely to minimize the error probability, and an essentially unique solution, a MAP decision rule. In contrast, choosing a good signal constellation is making a tradeoff among conflicting objectives. For instance, if we could, we would choose a signal constellation that contains a very large number m of signals of very small duration T and very small bandwidth B. If we could choose these parameters at will, we could also achieve any desired rate $\frac{\log_2 m}{TB}$ (expressed in bits per second per Hz). In addition, we would choose our signals so that they use very little energy and result in a very small error probability. These are conflicting goals.

Besides the quantities already mentioned, other quantities that will come up in our discussion are the number $k = \log_2 m$ of bits associated to each signal, the average time $T_b = T/k$ it takes to transmit one bit, the dimensionality n of the signal space, the energy per bit $\mathcal{E}_b$, the block error probability P_e and the bit error probability P_b.

4.2 Transformations That Do Not Affect P_e

Two sets of waveforms can look very different yet lead to the same probability of error. In this section we take a look at some of the transformations that change the signal constellation without affecting the error probability. There are at least two obvious reasons why we may want to evoke such a transformation: (i) we may save ourselves some time if we recognize that the probability of error associated to the constellation we are using is the same as that of another constellation for which we have already determined the error probability or for which we know an easy way to determine it; (ii) given a signal constellation that has the desired probability of error, we may be able to transform it into one that has the same probability of error and uses less energy, or less bandwidth, or less time.

4.2.1 Isometric Transformations In $\mathbb{R}^n$

An isometry in $\mathbb{R}^n$ is a distance-preserving transformation $a : \mathbb{R}^n \to \mathbb{R}^n$. Hence for any two points $\boldsymbol{p}, \boldsymbol{q} \in \mathbb{R}^n$, the distance from $\boldsymbol{p}$ to $\boldsymbol{q}$ equals the distance from $a(\boldsymbol{p})$ to $a(\boldsymbol{q})$.

If we apply the same isometry to every point of a signal constellation and to every decoding region, the probability of error (for the AWGN channel) remains the same. This intuitive fact can be verified mathematically as follows. Let

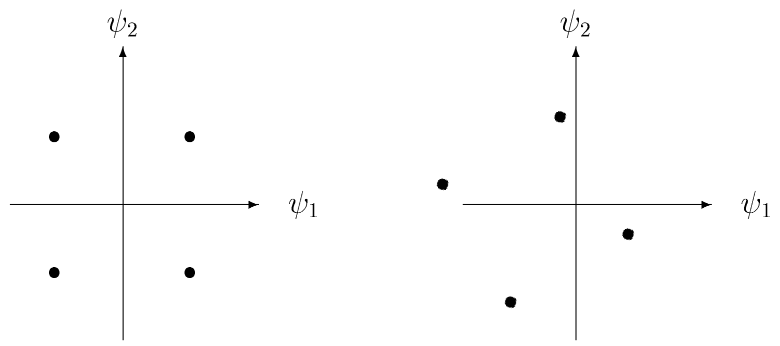
$$g(\gamma) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\gamma^2}{2\sigma^2}\right), \ \gamma \in \mathbb{R}$$

so that for $\boldsymbol{Z} \sim \mathcal{N}(0, \sigma^2 I_n)$ we can write $f_{\boldsymbol{Z}}(\boldsymbol{z}) = g(\|\boldsymbol{z}\|^2)$. Then for any isometry $a : \mathbb{R}^n \to \mathbb{R}^n$ we have

$$\begin{aligned}
P_c(i) &= Pr\{\boldsymbol{Y} \in \mathcal{R}_i | \boldsymbol{S} = \boldsymbol{s}_i\} \\
&= \int_{\boldsymbol{y} \in \mathcal{R}_i} g(\|\boldsymbol{y} - \boldsymbol{s}_i\|) d\boldsymbol{y} \\
&\overset{(a)}{=} \int_{\boldsymbol{y} \in \mathcal{R}_i} g(\|a(\boldsymbol{y}) - a(\boldsymbol{s}_i)\|) d\boldsymbol{y} \\
&\overset{(b)}{=} \int_{a(\boldsymbol{y}) \in a(\mathcal{R}_i)} g(\|a(\boldsymbol{y}) - a(\boldsymbol{s}_i)\|) d\boldsymbol{y} \\
&\overset{(c)}{=} \int_{\boldsymbol{\alpha} \in a(\mathcal{R}_i)} g(\|\boldsymbol{\alpha} - a(\boldsymbol{s}_i)\|) d\boldsymbol{\alpha} = Pr\{\boldsymbol{Y} \in a(\mathcal{R}_i) | \boldsymbol{S} = a(\boldsymbol{s}_i)\},
\end{aligned}$$

where in (a) we used the distance preserving property of an isometry, in (b) we used the fact that $y \in \mathcal{R}_i$ iff $a(\boldsymbol{y}) \in a(\mathcal{R}_i)$, and in (c) we made the change of variable $\boldsymbol{\alpha} = a(\boldsymbol{y})$ and used the fact that the Jacobian of an isometry is ± 1. The last line is the probability of decoding correctly when the transmitter sends $a(\boldsymbol{s}_i)$ and the corresponding decoding region is $a(\mathcal{R}_i)$.

Example 39. *The composition of a translation and a rotation is an isometry. The figure below shows an original signal set and a translated and rotated copy. Both have the same probability of error but not the same energy.*



□

Given $\{\boldsymbol{s}_0, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_{m-1}\}$, we are interested in finding the translating vector $\boldsymbol{a}$ so that the average energy of $\{\boldsymbol{s}'_0, \boldsymbol{s}'_1, \ldots, \boldsymbol{s}'_{m-1}\}$, where $\boldsymbol{s}'_i = \boldsymbol{s}_i - \boldsymbol{a}$, is minimized. The appropriate choice of $\boldsymbol{a}$ is (see Problem 1)

$$\boldsymbol{a} = \sum_i P_H(i)\boldsymbol{s}_i.$$

From now on we will use $\mathcal{E}$ to denote the average energy of the signal constellation at hand. Sometimes we will use $\mathcal{E}_b$ to denote the average energy per bit. Hence for a signal set $\{s_0, s_1, \ldots, s_{m-1}\}$, where signal s_i is used with probability $P_H(i)$, we have

$$\begin{aligned} \mathcal{E} &= \sum_i P_H(i)\|s_i\|^2 \\ \mathcal{E}_b &= \frac{\mathcal{E}}{\log m}, \end{aligned}$$

where $\log m$ is the number of bits of information that we convey when we communicate one of m possible choices.

When we make an isometric transformation as defined in this subsection, the signal space in which we are living does not change (the basis is the same). In the next section we consider isometric transformations that carry the signal space from one subspace of $\mathcal{L}_2$ to another.

4.2.2 Changing the Orthonormal Basis

The error probability depends solely on the position of the signals *in the signal space*. We may think of constructing various sets of time-domain signals in the following way.

We first choose a signal set in the signal space. From this set we construct a set of time-domain waveforms by selecting an orthonormal basis. We then construct a second set of time-domain waveforms by selecting a second orthonormal basis. The procedure may be repeated indefinitely. The resulting sets of waveforms may look very different. For instance one set may be of signals that have finite support (i.e. they vanish when t is outside a specified time interval), whereas another set may have infinite support. Nevertheless the associated average probability of error is identical for all signal sets obtained as described. Notice that changing the basis constitutes an isometric transformation in $\mathcal{L}_2$ (as opposed to an isometry in $\mathbb{R}^n$ as in the previous section).

EXAMPLE 40. *Let the signals in the signal space be* $s_0 = (\sqrt{\mathcal{E}}, 0)^T$ *and* $s_1 = (0, \sqrt{\mathcal{E}})^T$. *This choice completely determines the error probability and the average energy. It does not say anything, however, about the signals* $s_0(t)$ *and* $s_1(t)$, *except that they are orthogonal to one another. Example 38 shows four possible choices for* $s_0(t)$ *and* $s_1(t)$. *(There are infinitively many other possibilities).*

4.3 Time, Bandwidth, and Dimensionality

The bandwidth plays an important role in practice and should be included in our discussion on signal constellations. For now we focus on *baseband* signals, i.e. signals that have their spectral components centered around the origin.

One is tempted to define the bandwidth of a baseband signal $s(t)$ to be B if the support of $s_{\mathcal{F}}(t)$ is $[-\frac{B}{2}, \frac{B}{2}]$. This definition is not useful in practice since all man-made signals $s(t)$ have finite support and thus $s_{\mathcal{F}}(f)$ has infinite support.

A better definition (but not the only one that makes sense) is to fix a number $\eta \in (0,1)$ and say that the baseband signal $s(t)$ has bandwidth B if

$$\int_{-B}^{B} |s_{\mathcal{F}}(f)|^2 df = \|s\|^2(1-\eta).$$

In words, the signal has bandwidth B if B is the smallest number such that the interval $(-B, B)$ contains $100(1-\eta)\%$ of the signal power. The bandwidth changes if we change η . Reasonable values for η are $\eta = 0.1$ and $\eta = 0.01$.

This definition allows us to relate time, bandwidth, and dimensionality. If we let $\eta = \frac{1}{12}$ and define

$$\mathcal{L}_2(T,B) = \left\{ s(t) \in \mathcal{L}_2 : s(t) = 0, t \notin [-\frac{T}{2}, \frac{T}{2}] \text{ and } \int_{-B}^{B} |s_{\mathcal{F}}(f)|^2 df \geq \|s\|^2(1-\eta) \right\}$$

then one can show that the dimensionality of $\mathcal{L}_2(T,B)$ is $n = \lfloor 2TB+1 \rfloor$ (see Wozencraft & Jacobs for more on this). As $T \to \infty$, $\frac{n}{T} \to 2B$. Moreover, if one changes the value of η, then the constant in front of B changes but the essentially linear relationship between $\frac{n}{T}$ and B remains.

4.4 Examples

The aim of this section is to sharpen our intuition by looking at some examples.

Example 41. (Pulse Amplitude Modulation (PAM): Single Shot) *Let*

$$\boldsymbol{s}_i \in \left\{\pm\sqrt{E_w},\, \pm 3\sqrt{E_w},\, \pm 5\sqrt{E_w},\, \ldots, \pm(m-1)\sqrt{E_w}\right\}\; i \in \left\{0, 1, \ldots, (m-1)\right\}, \quad (4.1)$$

and let $s_i = \boldsymbol{s}_i \psi(t)$ *, where* ψ *is an arbitrary unit-energy waveform. Figure 4.1 shows the constellation in the signal space for* $m = 6$ *.*

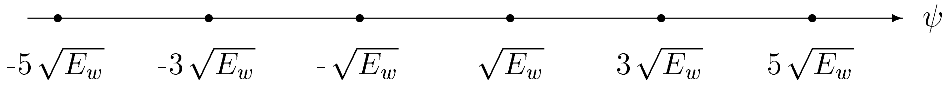


Figure 4.1: Signal Space Constellation for 6-ary PAM.

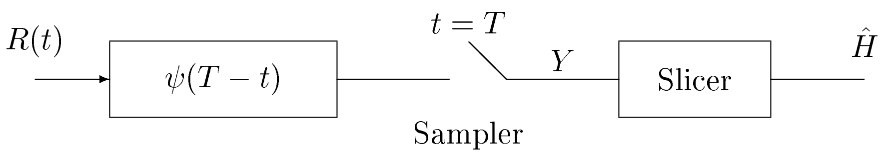


Figure 4.2: PAM Receiver

Notice that the receiver (shown in Figure 4.2) only needs a single matched filter since all the signals are in an inner product space of dimension 1 *. As we have seen, the filter projects the received signal onto* ψ *. Recall that the slicer is the device that finds (one of) the* i *for which* y *is as close to* s_i *as it is to any* s_j *,* $j \neq i$ *. The slicer needs the knowledge of the signal space constellation.*

Determining the probability of error is straightforward. If i *corresponds to one of the two end points,* $P_e(i) = Q(\frac{d}{2\sigma})$ *, where* $\frac{d}{2} = \sqrt{E_w}$ *and* $\sigma = \sqrt{\frac{N_0}{2}}$ *. For the remaining* $m-2$ *signal points, the probability of error is twice that of the end pints. Taking the average, we obtain*

$$P_e = \left(2 - \frac{2}{m}\right) Q\left(\sqrt{\frac{2E_w}{N_0}}\right). \quad (4.2)$$

Another quantity of interest is the average energy $\mathcal{E}$ *. A simple approximation to determine the average energy is obtained by computing the second moment of a random variable* X *that, instead of being uniformly distributed over the discrete set shown in*

(4.1), is uniformly distributed over the interval $[-m\sqrt{E_w}, m\sqrt{E_w}]$. *The second moment of this random variable is*

$$E[X^2] = \frac{1}{m\sqrt{E_w}} \int_0^{m\sqrt{E_w}} s^2 ds = \frac{m^2 E_w}{3}.$$

This approximation becomes better as the signal points move closer to one another, which is the case if we let the number k *of bits grow and let* E_w *scale so as to keep the average energy per bit equal to some constant* $\mathcal{E}_b$. *To see this, let* $\mathcal{E} = k\mathcal{E}_b$ *and use the continuous approximation to equate* $\mathcal{E} = EX^2 = \frac{m^2 E_w}{3}$. *Solving for* E_w *yields* $E_w = \frac{3\mathcal{E}}{m^2} = \frac{3k\mathcal{E}_b}{2^{2k}}$, *which indeed goes to zero exponentially fast as* k *goes to infinity. Since the signal points move closer to one another, the probability of error goes to* 1 *as* k *goes to infinity.* □

The next example uses a two-dimensionsional constellation.

Example 42. *(Phase-Shift-Keying (PSK): Single Shot) Let* $\tau = [0, T]$ *and define*

$$s_i(t) = \sqrt{\frac{2\mathcal{E}}{T}} \cos(2\pi f_0 t + \frac{2\pi}{m} i) 1_\tau(t), \quad i = 0, 1, \ldots, m-1. \tag{4.3}$$

We assume $f_0 T = \frac{k}{2}$ *for some integer* k, *so that* $\|s_i\|^2 = \mathcal{E}$ *for all* i. *The signal space representation may be obtained by using the trigonometric equivalence* $\cos(\alpha + \beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta)$ *to rewrite (4.3) as*

$$s_i(t) = s_{i,1}\psi_1(t) + s_{i,2}\psi_2(t),$$

where

$$s_{i1} = \sqrt{\mathcal{E}} \cos\left(\tfrac{2\pi i}{m}\right), \quad \psi_1(t) = \sqrt{\tfrac{2}{T}} \cos(2\pi f_0 t) 1_\tau(t),$$
$$s_{i2} = \sqrt{\mathcal{E}} \sin\left(\tfrac{2\pi i}{m}\right), \quad \psi_2(t) = -\sqrt{\tfrac{2}{T}} \sin(2\pi f_0 t) 1_\tau(t).$$

Hence, the n*-tuple representation of the signals is*

$$\boldsymbol{s}_i = \sqrt{\mathcal{E}} \begin{pmatrix} \cos 2\pi i/m \\ \sin 2\pi i/m \end{pmatrix}.$$

In Example 7 we have already studied this constellation and derived the following upper bound to the error probability

$$P_e \leq 2Q\left(\sqrt{\frac{\mathcal{E}}{\sigma^2}} \sin \frac{\pi}{m}\right),$$

where $\sigma^2 = \frac{N_0}{2}$ *is the variance of the noise in each coordinate.*

From the decoding regions plotted in Example 7 we also immediately see that for each $0 \leq i, j \leq m-1$, *there is an isometry* $a : \mathbb{R}^2 \to \mathbb{R}^2$ *such that* $a(\boldsymbol{s}_i) = \boldsymbol{s}_j$ *and* $a(\mathcal{R}_i) = \mathcal{R}_j$.

Thus, as an application of what we have learned in the previous section, we can tell the rather obvious fact that $P_e(i)$ *is the same for all* $i \in \mathcal{H}$.

As in the previous example we are interested in understanding what happens as k *goes to infinity while* $\mathcal{E}_b$ *remains constant. Since* $\mathcal{E} = k\mathcal{E}_b$ *grows linearly with* k, *the circle that contains the signal points has radius* $\sqrt{\mathcal{E}} = \sqrt{k\mathcal{E}_b}$ *that grows with* $\sqrt{k}$, *while the number* $m = 2^k$ *of points on this circle grows exponentially with* k. *Hence the minimum distance between points goes to zero (indeed exponentially fast). As a consequence, the argument of the* Q *function that upperbounds the probability of error for PSK goes to* 0 *and the probability of error goes to* 1. *Recall from Example 7 that the upperbound becomes tight as* m *grows.* □

As they are, the signal constellations used in the above two examples are not suitable to transmit a large amount of data. Indeed, to do so, we would have to let m be large enough so that $\log_2 m$ is the number of bits we want to transmit. As m grows, the probability of error goes to 1. The problem with these two examples is that, as m grows, we are trying to pack more and more signal points into a space that also grows in size but does not grow fast enough. The space becomes "crowded" as m grows, meaning that the minimum distance becomes smaller, and the probability of error increases.

In the next example we try to do better. So far we have not made use of the fact that we should expect to use more time to transmit more bits. In both of the above examples, the length T of the time interval used to communicate was constant. In the next example we let T grow linearly with the number of bits. This will free up a number of dimensions that grows linearly with k. (Recall that $n = 2BT$ is possible.) Each dimension may be used with the signal constellation of Example 41. Alternatively, every two dimensions may be used with the constellation of Example 42.

Example 43. *(Bit by Bit on a Pulse Train) The idea is to transmit a signal of the form*

$$s_i(t) = \sum_{j=1}^{k} s_{ij}\psi_j(t). \tag{4.4}$$

by letting $\psi_j(t)$ *be a time-translated version of a basic pulse* $\psi(t)$. *Hence,*

$$s_i(t) = \sum_j s_{ij}\psi(t - jT_s) \tag{4.5}$$

where

$$s_{ij} \in \{\pm\sqrt{\mathcal{E}_b}\}$$
$$\langle \psi(t - jT), \psi(t - lT)\rangle = \delta_{jl}.$$

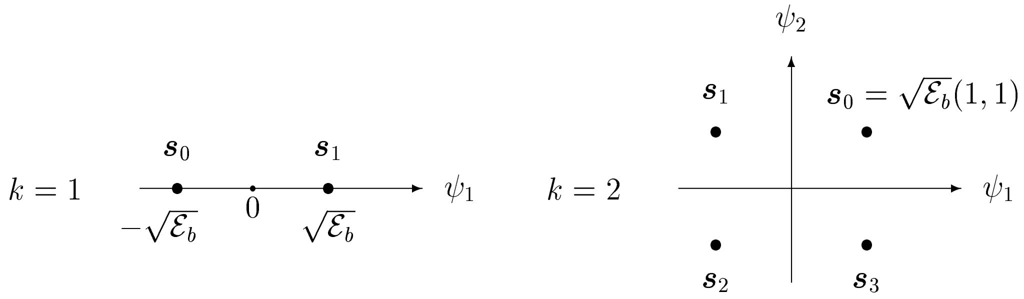
To be specific, we let

$$s_{ij} = (2d_j - 1)\sqrt{\mathcal{E}_b}$$

where $d_j \in \{0,1\}$ *is* j *th source bit. The subscript* b *indicates that* $\mathcal{E}_b$ *is the energy per (source) bit. For obvious reasons, the above signaling method will be called* bit-by-bit on a pulse train.

There are various possible choices for $\psi(t)$. *Common choices for* $\psi(t)$ *are sinc pulses, rectangular pulses, and raised-cosine pulses (to be defined later).*

The signal space representation of s_i *is* $\boldsymbol{s}_i = (s_{i1}, s_{i2}, \ldots, s_{ik})^T \in \{\pm\sqrt{\mathcal{E}_b}\}^k$. *This is a vertex of a* k *-dimensional hypercube as shown in the figures below for* $k = 1, 2$.



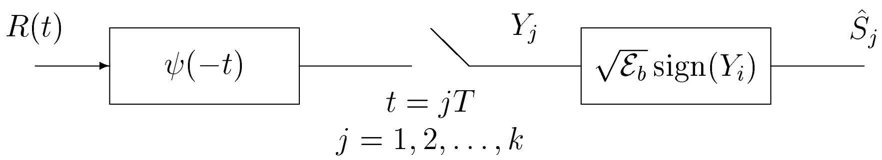
The ML receiver decides for

$$\begin{aligned}\hat{H}_{ML}(\boldsymbol{y}) &= \arg\max_i \langle \boldsymbol{y}, \boldsymbol{s}_i \rangle - \frac{\|\boldsymbol{s}_i\|^2}{2} \\ &= \arg\max_i \langle \boldsymbol{y}, \boldsymbol{s}_i \rangle\end{aligned}$$

where we used the fact that $\|\boldsymbol{s}_i\|^2 = k\mathcal{E}_b$, *independently of* i. *The maximum is achieved with the* $\boldsymbol{s}_i$ *for which its* j *th coordinate has the same sign as* y_j. *(If* $y_j = 0$ *then it does not matter whether we consider* y_j *as positive or negative. Either way the probability of error will be the same.) Thus,*

$$\boldsymbol{s}_{\hat{H}(\boldsymbol{y})} = \sqrt{\mathcal{E}_b}(\mathrm{sign}(Y_1), \mathrm{sign}(Y_2), \ldots, \mathrm{sign}(Y_k))^T.$$

This can be implemented as shown in the next figure. Notice that we need only one matched filter to do the $n = k$ *projections. This is one of the reasons why we choose* $\psi_i(t) = \psi(t - iT_s)$.



In the figure above, $\hat{S}_j$ *stands for the* j *th component of* $\boldsymbol{S}_{\hat{H}(\boldsymbol{y})}$.

We now compute the error probability. As usual, we first do so for a fixed transmitted signal $s_i = (s_{i1}, \ldots, s_{ik})^T$. *If* $s_{ij} = -\sqrt{\mathcal{E}_b}$, *the* j *th component of* s_i *will be decoded correctly if the* j *th noise component fulfills* $Z_j < \sqrt{\mathcal{E}_b}$. *This happens with probability* $1 - Q(\frac{\sqrt{\mathcal{E}_b}}{\sigma})$. *Similarly, If* $s_{ij} = \sqrt{\mathcal{E}_b}$, *the* j *th component of* s_i *will be decoded correctly if* $Z_j > -\sqrt{\mathcal{E}_b}$. *This also happens with probability* $1 - Q(\frac{\sqrt{\mathcal{E}_b}}{\sigma})$. *The probability that all* k *symbols are decoded correctly is*

$$P_c(i) = \left[1 - Q\left(\frac{\sqrt{\mathcal{E}_b}}{\sigma}\right)\right]^k.$$

Since this probability does not depend on s_i *we have that*

$$P_c = \left[1 - Q\left(\frac{\sqrt{\mathcal{E}_b}}{\sigma}\right)\right]^k.$$

Notice that $P_c \to 0$ *as* $k \to \infty$. *However, the probability that a specific symbol (bit) be decoded incorrectly is* $Q(\frac{\sqrt{\mathcal{E}_b}}{\sigma})$. *This is constant with respect to* k.

The following properties (due to our choice $\psi_j(t) = \psi(t - jT)$*) are worth noticing:*

- k *may be arbitrary and may vary from one message to the other without changing the structure of the transmitter and the receiver. (This would not be true with a general choice of* $\psi_1, \ldots, \psi_k$.*)*
- *The transmitter does not have to wait until it has received all* k *information bits to start transmission. This is important in real time applications, e.g. speech, video, etc.*
- *A ML receiver decides for each bit independently. Moreover, it can decide bit* i *as soon as the signal transmitted in the* i *th time interval has been received.*

All of the above properties are desirable for practical systems. □

The drawback of bit-by-bit signaling is that $P_c \to 0$ as $k \to \infty$. Hence, as it is, it is not appropriate to communicate long files either. We are, however, in a better situation than with the first two examples of this section. In those examples the probability of error was going to one since signal points were getting closer as k increased. To the contrary, in bit-by-bit on a pulse train the probability that we make an error in decoding one or more of the k bits goes to one because the number of neighbors increases. Coding will fix this problem by ensuring that the distance between neighboring signal points grows enough to compensate for the growing number of neighbors.

While in the last example we have chosen to transmit a single bit per dimension, we could have transmitted multiple bits per dimension as done in the previous two examples. In that case we call the signaling scheme *symbol by symbol on a pulse train.* Symbol by symbol on a pulse train will come up often in the remainder of this course. In fact it is the basis for most digital communication systems.

The following question seems natural at this point: Is it possible to map k bits into a signal s_i and avoid that $P_c \to 0$ as $k \to \infty$? The next example shows that it is indeed possible.

EXAMPLE 44. *(Frequency Shift Keying (FSK): An Example Of Orthogonal Signaling) Let $n = m = 2^k$. We do this by using m equal-norm orthogonal functions $s_1(t), \ldots, s_m(t)$:*

$$s_i = \sqrt{\mathcal{E}}\psi_i, \qquad \langle \psi_i, \psi_j \rangle = \delta_{ij}.$$

This is called block orthogonal signaling. The name stems from the fact that one collects a block of k bits and maps them into one of 2^k orthogonal waveforms.

There are many ways to choose the 2^k waveforms ψ_i. One way is to choose $\psi_i(t) = \psi(t - iT)$ for some basic pulse $\psi(t)$ such that $\langle \psi(t-iT), \psi(t-jT) \rangle = \delta_{ij}$ as in bit-by-bit signaling. Notice, however, that now we need 2^k such shifts of ψ as opposed to only k such shifts. Another way is what is called m-FSK (m-ary frequency shift keying). Specifically,

$$s_i(t) = \sqrt{\frac{2\mathcal{E}}{T}} \cos(2\pi f_i t) 1_\tau(t) \tag{4.6}$$

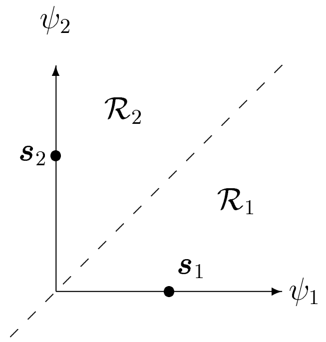
for some $\tau = [0, T]$ and $i = 1, 2, \ldots, m$. (For FSK it is convenient to index this way rather than letting $i = 0, 1, \ldots, m-1$ as usual.) For convenience we assume $f_i T = k_i$ for some integer k_i such that $k_i \neq k_j$ if $i \neq j$. Then

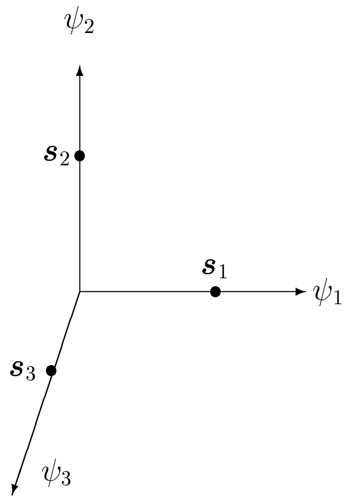
$$\begin{aligned}
\langle s_i, s_j \rangle &= \frac{2\mathcal{E}}{T} \int_0^T \cos(2\pi f_i t) \cos(2\pi f_j t) dt \\
&= \frac{2\mathcal{E}}{T} \int_0^T \left[\frac{1}{2} \cos[2\pi(f_i + f_j)t] + \frac{1}{2} \cos[2\pi(f_i - f_j)t] \right] dt \\
&= \frac{\mathcal{E}}{T} \int_0^T \cos[2\pi(f_i - f_j)t] dt \\
&= \mathcal{E}\delta_{ij}.
\end{aligned}$$

Letting $\psi_i(t) = \sqrt{\frac{2}{T}} \cos 2\pi f_i t 1_\tau(t)$ we obtain

$$s_i = \sqrt{\mathcal{E}}\psi_i, \qquad i = 1, \ldots, m. \tag{4.7}$$

Hence we have an orthogonal signal set as desired.





When $m \geq 3$*, it is not easy to visualize the decoding regions. However we can proceed analytically:*

$$\begin{aligned}\hat{H}_{MAP}(\boldsymbol{y}) &= \arg\max_i \langle \boldsymbol{y}, \boldsymbol{s}_i \rangle - \frac{\mathcal{E}}{2} \\ &= \arg\max_i \langle \boldsymbol{y}, \boldsymbol{s}_i \rangle \\ &= \arg\max_i y_i.\end{aligned}$$

When $H = i$*,*

$$Y_j = \begin{cases} Z_j & \text{if } j \neq i, \\ \sqrt{\mathcal{E}} + Z_j & \text{if } j = i. \end{cases}$$

Then

$$P_c(i) = Pr\{Y_1 > Z_2, Y_1 > Z_3, \ldots, Y_1, > Z_m | H = 1\}.$$

To evaluate the right side we first determine the probability of being correct conditioned on $H = 1$ *and* $Y_1 = \alpha$*, where* $\alpha \in \mathbb{R}$ *is an arbitrary number*

$$\begin{aligned}Pr\{c|H = 1, Y_1 = \alpha\} &= Pr\{\alpha > Z_2, \ldots, \alpha > Z_m\} \\ &= \left[1 - Q\left(\frac{\alpha}{\sqrt{N_0/2}}\right)\right]^{m-1}\end{aligned}$$

and then remove the conditioning on Y_1 *obtaining*

$$\begin{aligned}P_c(1) &= \int_{-\infty}^{\infty} f_{Y_1|H}(\alpha|1) \left[1 - Q\left(\frac{\alpha}{\sqrt{N_0/2}}\right)\right]^{m-1} d\alpha \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi N_0}} e^{-\frac{(\alpha - \sqrt{\mathcal{E}})^2}{N_0}} \left[1 - Q\left(\frac{\alpha}{\sqrt{N_0/2}}\right)\right]^{m-1} d\alpha,\end{aligned}$$

where we used the fact that when $H = 1$, $Y_1 \sim \mathcal{N}(\sqrt{\mathcal{E}}, \frac{N_0}{2})$. *The above expression for* $P_c(1)$ *cannot be simplified further. By symmetry,*

$$P_c = P_c(1) = P_c(i)$$

for all i.

The union of events bound is especially useful when the signal set $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m\}$ *is completely symmetric, like for orthogonal signals. In this case:*

$$\begin{aligned} P_e = P_e(i) &\leq (m-1)Q\left(\frac{d}{2\sigma}\right) \\ &= (m-1)Q\left(\sqrt{\frac{\mathcal{E}}{N_0}}\right) \\ &< 2^k \exp\left[-\frac{\mathcal{E}}{2N_0}\right] \\ &= \exp\left[-k\left(\frac{\mathcal{E}/k}{2N_0} - \ln 2\right)\right], \end{aligned}$$

where we used $\sigma^2 = \frac{N_0}{2}$ *and* $d = \sqrt{2\mathcal{E}}$. *The latter follows from* $d = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|$ *and*

$$\|\boldsymbol{s}_i - \boldsymbol{s}_j\|^2 = \|\boldsymbol{s}_i\|^2 + \|\boldsymbol{s}_j\|^2 - 2\langle \boldsymbol{s}_i, \boldsymbol{s}_j\rangle = \|\boldsymbol{s}_i\|^2 + \|\boldsymbol{s}_j\|^2 = 2\mathcal{E}.$$

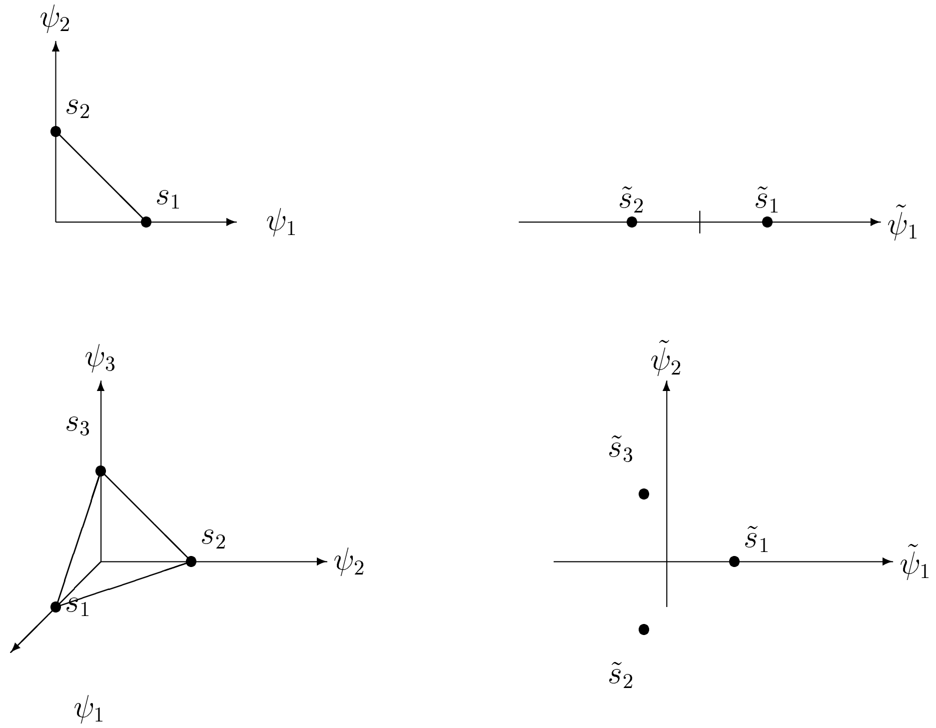
Here $\mathcal{E}$ *is the signal's energy. If we let* $\mathcal{E} = \mathcal{E}_b k$, *meaning that we let the signal's energy grow linearly with the number of bits as in bit-by-bit signaling, then we obtain*

$$P_e < e^{-k(\frac{\mathcal{E}_b}{2N_0} - \ln 2)}.$$

Here $P_e \to 0$ *as* $k \to \infty$, *provided that* $\frac{\mathcal{E}_b}{N_0} > 2\ln 2$. *(*$2\ln 2$ *is approximately* 1.39.*)*

$\square$

A useful application of the energy minimization idea (See Problem 1) applied to an orthogonal signal constellation leads to the *simplex* signal set.



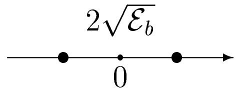
4.5 Bit By Bit Versus Block Orthogonal

In the last two examples of the previous section we have considered a case in which the number of dimensions n increased linearly with the number k of bits and one in which n increased exponentially with k. Let us compare the two.

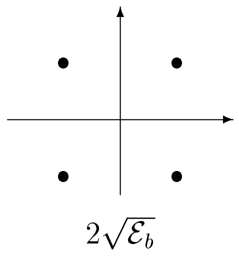
In bit-by-bit on a pulse train the bandwidth is constant (we have not proved this yet, but this is consistent with the asymptotic limit $2B = n/T$ seen in Section 4.3 applied with $T = nT_s$), and the time and the energy increased linearly with k. These are all desirable properties. (We have also seen that the delay at the sender and at the receiver are small and that we need only one matched filter to do the projections but we will not take complexity and delay into this discussion).

The drawback of bit-by-bit on a pulse train was the fact that the probability of error goes to 1 as k goes to infinity. The union of events bound is a useful tool to understand what is going on. Let us use it to bound the probability of error when $H = i$. The union of events bound has one term for each alternative j. The dominating terms in the bound are those that correspond to signals $\boldsymbol{s}_j$ that are the closest neighbors to $\boldsymbol{s}_i$. There are k closest neighbors, obtained by changing $\boldsymbol{s}_i$ in exactly one component, and each of them is at distance $2\sqrt{\mathcal{E}_b}$ from $\boldsymbol{s}_i$ (see the figure below). As k increases, the number of dominant terms goes up and so does the probability of error.

$k = 1$



$k = 2$



Let us now consider block orthogonal signaling. Since the dimensionality of the space it occupies grows exponentially with k, the expression $n = 2BT$ tells us that either the time or the bandwidth has to grow exponentially also. This is a significant drawback. Now let us consider the error probability. Using the bound

$$Q\left(\frac{\sqrt{2k\mathcal{E}_b}}{2\sqrt{\frac{N_0}{2}}}\right) = Q\left(\sqrt{\frac{k\mathcal{E}_b}{N_0}}\right) < \frac{1}{2}\exp\left[-\frac{k\mathcal{E}_b}{2N_0}\right]$$

we see that the probability that the noise carries a signal closer to a specific neighbor goes down as $\exp\left(-\frac{k\mathcal{E}_b}{2N_0}\right)$. There are $2^k - 1 = e^{k\ln 2} - 1$ nearest neighbors (all alternative signals are nearest neighbors). For $\frac{\mathcal{E}_b}{2N_0} > k\ln 2$, the growth in distance dominates the probability of error behavior. For $\frac{\mathcal{E}_b}{2N_0} < k\ln 2$ the number of neighbors dominates. Finally notice that the bit error probability P_b can not be larger than the block error probability P_e. Indeed they are the same iff every time that the decoder selects a wrong message the bit sequence that corresponds to this message has all bits flipped with respect to the bit sequence that corresponds to the correct message.

4.6 Conclusion

We have discussed some of the trade-offs between the number of transmitted bits, the duration, the bandwidth, and the energy of the signal we use to transmit those bits, and the resulting error probability. We have seen that, rather surprisingly, it is possible to transmit an increasing number k of bits at a fixed energy per bit $\mathcal{E}_b$ and make the probability that even a single bit is decoded incorrectly go to zero as k increases. However, the scheme we used to prove this has the undesirable property of requiring an exponential growth of the bandwidth. Ideally we would like to make the probability of error go to zero with a scheme similar to bit by bit on a pulse train. Is it possible? The answer is yes and the technique to do so is coding. We will give an example of coding in Chapter 9.

The study of the fundamental relationships between the rate at which we want to communicate (e.g. in bits per second per Hz), the power of the signal (measured at the receiver), and the probability of error that can be achieved is a typical subject of information theory. For instance, the *capacity* of the additive white Gaussian noise channel of power spectral density N_0 is

$$C = B \log_2 \left(1 + \frac{P}{N_o B}\right) \quad [bits/sec],$$

where B is the bandwidth [Hz] and P is the power (energy per second) [Watts] that we are allowed to use. One can show that at rates smaller than C one can make the communication arbitrarily reliable. This is not possible at rates above C.

4.7 Problems

Problem 1. (Minimum-energy Signals.)

Consider a given signal constellation consisting of vectors $\{s_1, s_2, \ldots, s_m\}$. *Let signal* s_i *occur with probability* p_i. *In this problem, we study the influence of moving the origin of the coordinate system of the signal constellation. That is, we study the properties of the signal constellation* $\{s_1 - a, s_2 - a, \ldots, s_m - a\}$ *as a function of* a.

(i) *Draw a sample signal constellation, and draw its shift by a sample vector* a.

(ii) *Does the average error probability,* $Pr\{e\}$, *depend on the value of* a? *Explain.*

(iii) *The average energy per symbol depends on the value of* a. *For a given signal constellation* $\{s_1, s_2, \ldots, s_m\}$ *and given signal probabilities* p_i, *prove that the value of* a *that minimizes the average energy per symbol is the centroid (the center of gravity) of the signal constellation, i.e.,*

$$a \;=\; \sum_{i=1}^{m} p_i s_i. \tag{4.8}$$

Hint: First prove that if X *is a real-valued zero-mean random variable and* $b \in \mathbb{R}$, *then* $E[X^2] \leq E[(X-b)^2]$ *with equality iff* $b = 0$. *Then extend your proof to vectors and consider* $\boldsymbol{X} = \boldsymbol{S} - E[\boldsymbol{S}]$ *where* $\boldsymbol{S} = \boldsymbol{s}_i$ *with probability* p_i.

Problem 2. (Orthogonal Signal Sets.)

Consider the following situation: A signal set $\{s_j(t)\}_{j=0}^{m-1}$ *has the property that all signals have the same energy* $\mathcal{E}_s$ *and that they are mutually orthogonal:*

$$\langle s_i, s_j \rangle \;=\; \mathcal{E}_s \delta_{ij}. \tag{4.9}$$

Assume also that all signals are equally likely. The goal is to transform this signal set into a minimum-energy signal set $\{s_j^*(t)\}_{j=0}^{m-1}$. *It will prove useful to also introduce the unit-energy signals* $\phi_j(t)$ *such that* $s_j(t) = \sqrt{\mathcal{E}_s}\phi_j(t)$.

(i) *Find the minimum-energy signal set* $\{s_j^*(t)\}_{j=0}^{m-1}$.

(ii) *What is the dimension of* $span\{s_0^*(t), \ldots, s_{m-1}^*(t)\}$? *For* $m = 3$, *sketch* $\{s_j(t)\}_{j=0}^{m-1}$ *and the corresponding minimum-energy signal set.*

(iii) *What is the average energy per symbol if* $\{s_j^*(t)\}_{j=0}^{m-1}$ *is used? What are the savings in energy (compared to when* $\{s_j(t)\}_{j=0}^{m-1}$ *is used) as a function of* m?

Problem 3. (Antipodal Signaling with Rayleigh Fading.)

Suppose that we use antipodal signaling (i.e $s_0(t) = -s_1(t)$*). When the energy per symbol is* $\mathcal{E}_b$ *and the power spectral density of the additive white Gaussian noise in the channel is* $N_0/2$, *then we know that the average probability of error is*

$$Pr\{e\} = Q\left(\sqrt{\frac{\mathcal{E}_b}{N_0/2}}\right). \tag{4.10}$$

In mobile communications, one of the dominating effects is fading. *A simple model of fading is as follows: Let the channel attenuate the signal by a random variable* A. *Specifically, if* $\boldsymbol{s}_i$ *is transmitted, the received signal is* $Y = A\boldsymbol{s}_i + N$. *The probability density function of* A *depends on the particular channel that is to be modeled.*[1] *Suppose* A *assumes the value* a. *From the receiver point of view this is as if there is no fading and the transmitter uses the signals* $as_0(t)$ *and* $-as_0(t)$. *Hence,*

$$Pr\{e|A = a\} = Q\left(\sqrt{\frac{a^2\mathcal{E}_b}{N_0/2}}\right). \tag{4.11}$$

The average probability of error can thus be computed by taking the expectation over the random variable A, *i.e.*

$$Pr\{e\} = E_A[Pr\{e|A\}] \tag{4.12}$$

An interesting, yet simple model is to take A *to be a Rayleigh random variable, i.e.*

$$f_A(a) = \begin{cases} 2ae^{-a^2}, & \text{if } a \geq 0, \\ 0, & \text{otherwise..} \end{cases} \tag{4.13}$$

This type of fading, which can be justified especially for wireless communications is called Rayleigh fading.

(i) *Compute the average probability of error for antipodal signaling subject to Rayleigh fading.*

(ii) *Comment on the difference between Eqn. (4.10) (the average error probability without fading) and your result in* (i) *(the average error probability with Rayleigh fading). Is it significant? For an average error probability* $Pr\{e\} = 10^{-5}$, *find the necessary* $\mathcal{E}_b/N_0$ *for both cases.*

[1] In a more realistic model, not only the amplitude, but also the phase of the channel transfer function is a random variable.

PROBLEM 4. *(i) The root-mean square (rms) bandwidth of a low-pass signal $g(t)$ of finite energy is defined by*

$$W_{rms} = \left[\frac{\int_{-\infty}^{\infty} f^2|G(f)|^2 df}{\int_{-\infty}^{\infty} |G(f)|^2 df}\right]^{1/2}$$

where $|G(f)|^2|$ is the energy spectral density of the signal. Correspondingly, the root mean-square (rms) duration of the signal is defined by

$$T_{rms} = \left[\frac{\int_{-\infty}^{\infty} t^2|g(t)|^2 dt}{\int_{-\infty}^{\infty} |g(t)|^2 dt}\right]^{1/2}.$$

Using these definitions and assuming that $|g(t)| \to 0$ faster than $1/\sqrt{|t|}$ as $|t| \to \infty$, show that

$$T_{rms}W_{rms} \geq \frac{1}{4\pi}.$$

Hint: Use Schwarz's inequality

$$\left\{\int_{-\infty}^{\infty} [g_1^*(t)g_2(t) + g_1(t)g_2^*(t)]dt\right\}^2 \leq 4\int_{-\infty}^{\infty} |g_1(t)|^2 dt \int_{-\infty}^{\infty} |g_2(t)|^2 dt$$

in which we set

$$g_1(t) = tg(t)$$

and

$$g_2(t) = \frac{dg(t)}{dt}.$$

(ii) Consider a Gaussian pulse defined by

$$g(t) = \exp(-\pi t^2).$$

Show that for this signal, the equality

$$T_{rms}W_{rms} = \frac{1}{4\pi}$$

can be reached.

Hint:

$$\exp(-\pi t^2) \stackrel{\mathcal{F}}{\longleftrightarrow} \exp(-\pi f^2).$$