Digital Communications Engineering 1

(COMM2108)

Signal Space Analysis & Error Performance

Signal Space Analysis

- Signal space analysis provides a mathematically elegant and highly insightful tool for the study of digital signal transmission.
- Signal space analysis permits a general geometric framework for the interpretation of digital signaling that includes both baseband and bandpass signaling schemes.
- Furthermore, it provides an intuitive insight into the error performance and spectral efficiency characteristics of the various digital signaling schemes.

Signal Space Analysis

- Before introducing the Signal Space analysis technique, a brief review of digital transmission is necessary.
- The transmitter takes the message source output m_i and codes it into a distinct signal $s_i(t)$ suitable for transmission over the communications channel.
- The signal $s_i(t)$ has a duration of T seconds and more importantly is a real-valued energy signal where

$$E_i = \int_0^T s_i^2(t)\,dt$$

Signal Space Analysis

- The transmission channel is perturbed by zero-mean additive white Gaussian noise (AWGN).
- The AWGN channel is one of the simplest mathematical models for various physical communications channels.
- The received signal $r(t)$ is given by

$$r(t) = s_i(t) + n(t) \qquad \text{for } 0 \leq t \leq T$$

- The receiver has the task of observing the received signal $r(t)$ for a duration of T seconds and making the best estimate of the transmitted signal $s_i(t)$.

Signal Space Analysis

- However, owing to the presence of channel noise, the decision making process is statistical in nature with the result that the receiver will make occasional errors.
- The requirement is to design the receiver so as to minimize the *average* probability of symbol error which is defined as

$$P_e = \sum_{i=1}^{M} p_i P(\hat{m} \neq m_i | m_i)$$

- where m_i is the transmitted symbol and $\hat{m}$ is the estimate produced by the receiver.

Signal Space Analysis

- The basic idea behind the geometric representation of signals is to represent any set of M energy signals $\{S_i(t)\}$ as linear combinations of N orthonormal basis functions, where $N \le M$.
- Given a set of real-valued energy signals $\{s_1(t),\ s_2(t), \ldots, s_M(t)\}$, each of duration T seconds, it is possible to write

$$s_i(t) = \sum_{i=1}^{N} a_{ij}\phi_j(t) \quad \begin{cases} 0 \le t \le T \\ i = 1, \ldots, M \end{cases}$$

Signal Space Analysis

- The coefficients a_{ij} are defined as

$$a_{ij} = \int_0^T s_i(t)\phi_j(t)\,dt \qquad \begin{cases} i = 1,\dots, M \\ j = 1,\dots, N \end{cases}$$

- The real-valued basis functions $\phi_1(t),\ \phi_2(t),\dots,\ \phi_N(t)$ are *orthonormal* which means that

$$\int_0^T \phi_i(t)\phi_j(t)\,dt = \delta_{ij} = \begin{cases} 1 & if \quad i = j \\ 0 & if \quad i \neq j \end{cases}$$

- where δ_{ij} is the Kronecker delta function.

Signal Space Analysis

- The basis functions $\phi_1(t),\ \phi_2(t),\ldots,\ \phi_N(t)$ are orthogonal and have unit energy.
- From a geometric point of view each basis functions $\phi_j(t)$ is mutually perpendicular to each of the other basis functions.
- The set of basis functions $\{\phi_1(t),\ \phi_2(t),\ldots,\ \phi_N(t)\}$ characterises an N-*dimensional* signal space in which there are *N* mutually perpendicular axes labelled $\phi_1(t),\ \phi_2(t),\ldots,\ \phi_N(t)$.

Signal Space Analysis

- The set of signal waveforms $\{s_1(t), s_2(t), \ldots, s_M(t)\}$ can be plotted as a set of N-*dimensional* vectors with coordinates $(a_{i1}, a_{i2}, \ldots, a_{iN})$ within the signal space spanned by $\{\phi_1(t), \phi_2(t), \ldots, \phi_N(t)\}$.

- The geometrical representation of a set of energy signals provides the mathematical basis for the noise analysis of digital communications systems in a visually intuitive manner.

Signal Space Analysis

- The length of the vector is a measure of the signal energy transmitted during a symbol duration.
- By definition, the energy of a signal $s_i(t)$ of duration T seconds is

$$E_i = \int_0^T s_i^{\,2}(t)\,dt$$

- Substituting for $s_i(t)$ we get

$$E_i = \int_0^T \left(\sum_{j=1}^{N} a_{ij}\phi_j(t) \right)^2 dt$$

Signal Space Analysis

$$
\begin{aligned}
E_i &= \int_0^T \sum_{j=1}^{N} a_{ij}\phi_j(t) \sum_{k=1}^{N} a_{ik}\phi_k(t)\,dt \\
&= \sum_{j=1}^{N} \sum_{k=1}^{N} a_{ij} a_{ik} \int_0^T \phi_k \phi_j\,dt \\
&= \sum_{j=1}^{N} \sum_{k=1}^{N} a_{ij} a_{ik} \delta_{jk} \\
&= \sum_{j=1}^{N} a_{ij}^2 \\
&= \|s_i\|^2
\end{aligned}
$$

- In other words, the energy of the signal is given by the square of the Euclidean length of the vector representing it.

Signal Space Analysis

- Another useful relation involving the vector representation of the signals $s_i(t)$ and $s_k(t)$ is the Euclidean distance d_{ik} and is defined as:

$$\left\| s_i - s_k \right\|^2 = \sum_{j=1}^{N} \left(s_{ij} - s_{kj} \right)^2$$

$$= \int_0^T \left(s_i(t) - s_k(t) \right)^2 dt$$

Signal Space Analysis

- It is also possible to represent noise within this signal space analysis.
- Additive white Gaussian noise can be expressed as a linear combination of orthonormal basis functions.
- However, it is convenient to first partition the noise into two components

$$n(t) = \hat{n}(t) + \tilde{n}(t)$$

Signal Space Analysis

- Where $\hat{n}(t)$ is the noise that falls within the signal space and is given by

$$\hat{n}(t) = \sum_{j=1}^{N} n_j \phi_j(t)$$

- And $\tilde{n}(t)$ is the noise that falls outside of the signal space and may be thought of as the noise that is effectively tuned out by the receiver.
- As a result only the noise $\hat{n}(t)$ within the signal space will interfere with the detection process.

Signal Space Analysis

- The noise waveform

$$n(t) = \sum_{j=1}^{N} n_j \phi_j(t) + \tilde{n}(t)$$

- where

$$n_j = \int_0^T n(t)\phi_j(t)\,dt$$

- $n(t)$ can be expressed as a noise vector within the signal space with coordinates $(n_1, n_2, \ldots, n_N)$. The noise vector n will have zero mean and a Gaussian distribution.

Signal Space Analysis

- Next, the statistical characteristics of the correlator outputs is considered.
- The received signal $r(t)$ is given by

$$r(t) = s_i(t) + n(t) \quad for\ 0 \leq t \leq T$$

- where $s_i(t)$ is the transmitted waveform and $n(t)$ is AWGN with zero mean and a power spectral density $N_0/2$.
- Since $r(t)$ has been corrupted by noise it is considered to be a random variable and consequently the correlator output is also considered to be a random variable.
- Therefore it only meaningful to consider the correlator outputs in terms of their statistics properties, i.e. in terms of the mean and variance.

Signal Space Analysis

- The output from correlator j will be given by

$$z_j(T) = \int_0^T r(t)\phi_j(t)\,dt$$

$$= a_{ij} + n_j$$

- where a_{ij} is a deterministic component determined by the transmitted signal $s_i(t)$ and is defined by

$$a_{ij} = \int_0^T s_i(t)\phi_j(t)\,dt$$

- And n_j is a random variable (due to the AWGN) and is defined by

$$n_j = \int_0^T n(t)\phi_j(t)\,dt$$

Signal Space Analysis

- The *mean value* of the correlator output is given by

$$\overline{z_j(T)} = a_{ij}$$

- The mean value of the correlator output depends on the transmitted signal $s_i(t)$ only (since noise AGWN has zero mean).

- The *variance* of the correlator output is given by

$$\mathrm{var}\big(z_j(T)\big) = \frac{N_0}{2} \quad for \quad all \quad j$$

- This is an important result since it shows that all the correlator outputs have a variance equal to the power spectral density of the noise process $n(t)$.

Signal Space Analysis

- To summarise the results of signal space analysis.
- In each time slot of duration T seconds, one of M possible signals $s_1(t), s_2(t),\ldots,s_M(t)$ is transmitted.
- Each of the signals $s_i(t)$ may be represented by a point in Euclidean space of dimension $N \le M$ spanned by a set of N orthonormal basis functions.
- These points are referred to as *transmitted signal points* or *message points*.
- The set of message points corresponding to the set of transmitted signals $\{s_1(t), s_2(t),\ldots,s_M(t)\}$ is called a *signal constellation*.

Signal Space Analysis

- The representation of the received signal $r(t)$ is complicated by the presence of AWGN.
- AWGN may also be represented by a vector plotted within the signal space for the transmitted signals, however it is only the part of the noise that will actually interfere with the detection process that can be represented in this way.
- In a digital receiver, the received signal $r(t)$ is applied to a bank of N correlators whose outputs $z_j(T)$ define an *observation vector* $Z(T) = \{z_1(T), z_2(T), \ldots, z_N(T)\}$.
- The observation vector $Z(T)$ represents the received signal $r(t)$ in the same N-dimensional space used to represent the transmitted signals, this is known as the *received signal point*.

Signal Space Analysis

- The observation vector may therefore be viewed as the sum of the signal vector and the noise vector.
- Since the noise vector is a random variable, the received signal point will “wander” about the message point in a completely random fashion.
- The received signal point may lie anywhere inside a Gaussian-distributed “cloud” centred on the message point.
- The “cloud” is dense in the centre and becomes sparse with increasing distance from the message point.
- The task of the digital receiver is to determine <u>*which*</u> of the signal vectors does the observation vector most closely resemble.

Signal Space Analysis

- The process of deciding which of the signal vectors does the observation vector most closely resemble is equivalent to determining which of the message points is the received signal vector closest to.
- The detection process can be thought of as a *distance* measurement.
- The analysis of all digital demodulation or detection schemes involves the concept of the *distance* between a received waveform and a set of possible transmitted waveforms.
- A possible decision rule for deciding which signal was transmitted in any given signaling interval is to choose the message point *closest* to the received signal point.

Signal Space Analysis

- In other words, to choose the transmitted signal s_i such that the distance

$$d(r, s_i) = \|r - s_i\| \quad \text{is} \quad \text{minimised}$$

- This decision rule can be represented in terms of decision regions constructed within the signal vector space.
- By constructing a perpendicular bisector to the line joining the message points s_i and s_j, the signal space may be partitioned into decision regions.
- The decision rule for the receiver may be stated in terms of decision regions, if the received signal r fall into region i choose signal i and so on.

Signal Space Analysis

- In practical terms, this means that in a correlation receiver it is possible to replace the bank of M correlators with a bank of N correlators where the set of orthonormal basis functions $\{\phi_j(t)\}$ form the reference signals.
- Moreover, since $N \leq M$ this can represent a cost effective implementation since a smaller number of correlator circuits are required.
- Typically, $N << M$ which results in very significant cost savings. In most cases $N = 2$, while M can be as large as 64, 128 or 256.

Signal Space Analysis

- As an example of how the signal space representation may be applied to a set of signals, consider the set of four signal waveforms $\{s_1(t), s_2(t), s_3(t), s_4(t)\}$ shown in the handout.
- The set of $M = 4$ signal waveforms may be expressed in terms of a linear combination of $N = 3$ basis functions.
- The signal waveforms may be expressed as

$$s_1(t) = \sqrt{2}\,\phi_1(t)$$

$$s_2(t) = \sqrt{2}\,\phi_2(t)$$

$$s_3(t) = -\sqrt{2}\,\phi_2(t) + \phi_3(t)$$

$$s_4(t) = \sqrt{2}\,\phi_1(t) + \phi_3(t)$$

Signal Space Analysis

- The set of signal waveforms may be also be expressed as a set of vectors $\{s_1, s_2, s_3, s_4\}$ whose coordinates in the signal space are

$$s_1 = (\sqrt{2}, 0, 0)$$
$$s_2 = (0, \sqrt{2}, 0)$$
$$s_3 = (0, -\sqrt{2}, 1)$$
$$s_4 = (\sqrt{2}, 0, 1)$$

- However, the set of basis functions is not unique. An alternative set may also be derived.
- It is worth noting that although the set of basis functions may vary, the dimensionality of the signal space will not change (i.e. $N = 3$).

Signal Space Analysis

- Next we consider some specific examples of baseband digital signaling formats.
- For example, in binary polar baseband signaling a binary 1 is represented by a pulse of amplitude $+A$ volts while a binary 0 is represented by a pulse of amplitude $-A$ volts. The signal waveforms are

$$s_1(t) = +A \quad for \quad 0 \le t \le T$$

$$s_2(t) = -A \quad for \quad 0 \le t \le T$$

- The signal set $\{s_1(t),\ s_2(t)\}$ may be expressed in terms of the basis function

$$\phi_1(t) = \frac{1}{\sqrt{T}} \quad for \quad 0 \le t \le T$$

Signal Space Analysis

- The two signal waveforms can be expressed in terms of $\phi_1(t)$ as follows

$$s_1(t) = A\sqrt{T}\,\phi_1(t)$$

$$s_2(t) = -A\sqrt{T}\,\phi_1(t)$$

- However, before considering the signal space representation of the two polar signals, it is instructive to consider the energy in the two signals.

$$E_{s_1} = \int_0^T s_1(t)^2\,dt = A^2 T$$

$$= E_{s_2}$$

$$= E_s$$

Signal Space Analysis

- The two signal waveforms can be expressed in terms of the signal energy E_s and $\phi_1(t)$ as follows

$$s_1(t) = \sqrt{E_s}\phi_1(t)$$
$$s_2(t) = -\sqrt{E_s}\phi_1(t)$$

- The two signal waveforms may be represented as two vectors s_1 and s_2 in a one-dimensional (i.e. $N = 1$) signal space, another words a <u>*line*</u>.
- The two signal vectors $s_1 = -s_2$ are a special case of signaling known as *binary anti-podal* signaling and as such they achieve the maximum separation between the signal points where the distance d_{12} between the points is

$$d_{12} = 2\sqrt{E_s}$$

Signal Space Analysis

- In binary unipolar baseband signaling, a binary 1 is represented by a pulse of amplitude $+A$ volts and a binary 0 is represented by a pulse of 0 volts. The signal waveforms are

$$s_1(t) = A \quad for \quad 0 \le t \le T$$

$$s_2(t) = 0 \quad for \quad 0 \le t \le T$$

- The signal set $\{s_1(t), s_2(t)\}$ may be expressed in terms of the basis function

$$\phi_1(t) = \frac{1}{\sqrt{T}} \quad for \quad 0 \le t \le T$$

Signal Space Analysis

- The two signal waveforms can be expressed in terms of the signal energy E_s and $\phi_1(t)$ as follows

$$s_1(t) = \sqrt{E_s}\,\phi_1(t)$$

$$s_2(t) = 0$$

- The two signal waveforms may be represented as two vectors s_1 and s_2 in a one-dimensional (i.e. N = 1) signal space, another words a *line*.
- The distance d_{12} between the signal points is given by

$$d_{12} = \sqrt{E_s}$$

- It is worth noting that the distance d_{12} is smaller for unipolar signaling than for polar signaling.

Signal Space Analysis

- The signal waveforms that have been considered so far have all been one dimensional (i.e. $N = 1$).
- Next, two-dimensional (i.e. $N = 2$) signals will be considered.
- An example of two-dimensional signals are the *orthogonal* signals.
- Two signals waveforms $s_i(t)$ and $s_k(t)$ are considered to be orthogonal if

$$\int_0^T s_i(t)s_k(t)dt = 0 \quad for \quad i \neq k$$

Signal Space Analysis

- The orthogonal signal set $\{s_1(t),\ s_2(t)\}$ in the hand-out can be expressed in terms of two basis functions given by

$$\phi_1(t) = \begin{cases} \sqrt{2/T} & 0 \le t \le T/2 \\ 0 & otherwise \end{cases}$$

$$\phi_2(t) = \begin{cases} \sqrt{2/T} & T/2 \le t \le T \\ 0 & otherwise \end{cases}$$

- The signal waveforms may be represented in terms of vectors in the signal space as follows

$$s_1 = (A\sqrt{T/2}, A\sqrt{T/2})$$

$$s_2 = (A\sqrt{T/2}, -A\sqrt{T/2})$$

Signal Space Analysis

- Expressing the signal vectors in terms of the signal energy $E_s = A^2T$ gives

$$s_1 = (\sqrt{E_s/2}, \sqrt{E_s/2})$$

$$s_2 = (\sqrt{E_s/2}, -\sqrt{E_s/2})$$

- The separation between the signal vectors d_{12} is given by

$$d_{12} = 2\sqrt{E_s/2}$$

$$= \sqrt{2E_s}$$

- It is worth noting that the distance d_{12} for orthogonal signaling is smaller than that polar signaling, but greater than that for unipolar signaling.

Error Performance

- Next, it is necessary to derive an expression for the average probability of symbol error P_e.
- It is assumed that the signal space is partitioned into M decision regions.
- An error is considered to occur if a signal vector s_i is transmitted and the received signal vector does not fall inside the decision region i associated with signal s_i.

$$P_e = \sum_{i=1}^{M} p_i P(Z(T) \text{ does not lie in region i} \mid s_i \text{ sent})$$

Error Performance

- The average probability of symbol error P_e is obtained by averaging over all the M symbols which can be shown to give the following expression

$$P_e = \sum_{i=1}^{M} \sum_{\substack{k=1 \\ k \neq i}}^{M} p_i Q\left(\frac{d_{ik}}{\sqrt{2N_0}} \right)$$

- where $Q(x)$ is the *complementary error function*.
- Furthermore, by defining a minimum distance d_{min} as the smallest Euclidean distance between any two transmitted signal points in the constellation where

$$d_{min} = \min_{i \neq k} \ d_{ik}$$

Error Performance

- This expression may be simplified by assuming equiprobable symbols, $p_i = 1/M$, which gives

$$P_e = \frac{1}{M} \sum_{i=1}^{M} \sum_{\substack{k=1 \\ k \neq i}}^{M} Q\left(\frac{d_{min}}{\sqrt{2N_0}} \right)$$

- For binary signaling schemes where $M = 2$ and hence $p_i = 0.5$, we obtain

$$P_e = Q\left(\frac{d_{min}}{\sqrt{2N_0}} \right)$$

Error Performance

- By definition, *the complementary error function* $Q(x)$ is

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-\frac{u^2}{2}} du$$

- This expression <u>*cannot*</u> be evaluated analytically. However, it can be solved numerically and the results are widely available in tabular form.

Error Performance

- Another form of the complementary error function that is often used is

$$erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-u^2} du$$

- The two forms of the complementary error function are related as follows

$$erfc(x) = 2Q\left(x\sqrt{2}\right)$$

$$Q(x) = \frac{1}{2} erfc\left(\frac{x}{\sqrt{2}}\right)$$

Error Performance

- By obtaining the minimum distance d_{min} between any signaling schemes we can quickly obtain an expression for the probability of error P_e.
- For example in the case of binary polar signaling (or antipodal signaling) we found that

$$d_{12}=d_{min}=2\sqrt{E_s}$$

- This results in an expression for the average probability of error P_e given by

$$P_e = Q\left(\sqrt{\frac{2E_s}{N_0}}\right)$$

Error Performance

- In the case of binary unipolar signaling we found that

$$d_{12}=d_{min}=\sqrt{E_s}$$

- This produces an expression for the average probability of error P_e given by

$$P_e = Q\left(\sqrt{\frac{E_s}{2N_0}}\right)$$

- Compared with binary polar signaling, it can be seen that binary unipolar signaling requires four times (i.e. 6 dB) the value of (E_s/N_0) to realise the same P_e performance.
- In other words, binary polar signaling out-performs binary unipolar signaling by 6 dB in terms of the required (E_s/N_0).

Error Performance

- In the case of binary orthogonal signaling we found that

$$d_{12} = d_{min} = \sqrt{2E_s}$$

- This produces an expression for the average probability of error P_e given by

$$P_e = Q\left(\sqrt{\frac{E_s}{N_0}}\right)$$

- Compared with binary polar signaling, it can be seen that binary unipolar signaling requires twice (i.e. 3 dB) the value of (E_s/N_0) to realise the same P_e performance.
- In other words, binary polar signaling out-performs binary orthogonal signaling by 3 dB in terms of the required (E_s/N_0).

Error Performance

- It is worth noting that split-phase Manchester coding is also an example of anti-podal signaling with a symbol energy of

$$E_s = A^2 T$$

- The separation between signal points $d_{\min}$ is

$$d_{\min} = 2\sqrt{E_s}$$

- The average probability of symbol error P_e for Manchester signaling is

$$P_e = Q\left(\sqrt{\frac{2E_s}{N_0}}\right)$$

Error Performance

- To summarize – We have obtained an expression for the probability of error P_e that shows the critical role played by the minimum distance d_{min} in determining the error performance of a signaling scheme.

$$P_e = Q\left(\frac{d_{min}}{\sqrt{2N_0}}\right)$$

- Since the complementary error function $Q(x)$ is a monotonically decreasing function, the greater the value of d_{min} the smaller the probability of error.
- Intuitively this makes sense, since the minimum distance d_{min} is related to the energy difference between the signals and the greater this energy difference the more difficult it is for the noise to distort a signal such that it will be mistaken for another signal.

Error Performance

- The greater the value of d_{min}, the greater the *immunity* to noise, i.e. the more rugged is the signaling scheme.
- The anti-podal signaling schemes achieve the greatest possible separation between signal points and so have the best noise performance.
- The E_s/N_0 ratio can be thought of as a signal-to-noise ratio (SNR) metric for digital systems.
- The error performance of signaling schemes may be compared in terms of the E_s/N_0 ratio required to realise a given probability of error P_e.

Error Performance

- For example, comparing the different baseband binary signaling schemes considered so far, we find that the E_s/N_0 ratios required to produce the same value of P_e are related as follows:

$$4\left(\frac{E_s}{N_0}\right)_{unipolar} = 2\left(\frac{E_s}{N_0}\right)_{orthogonal} = \left(\frac{E_s}{N_0}\right)_{polar}$$

- In other words, polar (or anti-podal) schemes *outperform* orthogonal schemes by a factor of 3 dB (factor of 2) and *outperform* unipolar schemes by a factor of 6 dB (factor of 4) in terms of required E_s/N_0.

Error Performance

- In many cases, the superior error performance of a signaling scheme can be traded-in for a reduction in the transmitter power which may be of considerable practical benefit where portable battery powered equipment is concerned.
- A further reduction in the E_s/N_0 requirement may be achieved through the use of *error control* schemes – known as *channel coding*.
- Channel coding involves the addition of extra bits in a code word to allow for the detection and correction of transmission errors.

Biorthogonal Signaling

- It is possible to combine anti-podal and orthogonal signaling to produce *biorthogonal* signaling.
- The biorthogonal signal set contains four signal waveforms $\{s_1(t), s_2(t), -s_1(t), -s_2(t)\}$.
- Since the pair $s_1(t)$ and $s_2(t)$ are orthogonal and the pair $-s_1(t)$ and $-s_2(t)$ are orthogonal, the signal set is termed biorthogonal.
- By using biorthogonal signaling it is possible to transmit 2 bits of information in the signaling interval of T seconds.

M-ary Signaling

- By using a set of M signal waveforms where $M = 2^k$ it is possible to transmit k bits of information in the signaling interval T seconds or each transmitted signal waveform is carrying k bits of information.
- This leads to what is known as M-*ary* digital signaling, i.e. where each symbol is carrying more than 1 bit of data.
- For example, when $M = 2$ we have bin<u>ary</u> or when $M = 3$ we have tern<u>ary</u> etc.
- The procedure for constructing larger sets of signal waveforms is relatively straightforward by continuing to use $N = 2$ orthonormal basis functions $\{\phi_1(t), \phi_2(t)\}$.

M-ary Signaling

- For example, if $M = 8$, three bits of information can be transmitted in a signaling interval T seconds.
- Assuming all signal waveforms of energy E_s, the signal points can arranged on a circle centred on the origin and of radius $\sqrt{E_s}$.
- If the condition that all signals have equal energy is relaxed, it is possible to have 4 biorthogonal waveforms of energy E_1 and another 4 biorthogonal waveforms of energy E_2.
- The signal constellation will comprise two sets of signals arranged on two concentric circles with radii given by $\sqrt{E_1}$ and $\sqrt{E_2}$ respectively.

M-ary Signaling

- Previously, it was shown that $M = 2^k$ signal waveforms can be constructed in two-dimensions.
- However, it is also possible to construct a set of $M = 2^k$ signal waveforms having more that two-dimensions.
- If the M signals are all mutually orthogonal, the dimensionality of the signal space $N = M$.
- An example of orthogonal M*-ary* signaling is Pulse Position Modulation (PPM).
- The geometrical representation of M*-ary* orthogonal signals is relatively straightforward with each of the signal waveforms lying on each of the axes.

M-ary Signaling

- The coordinates of the signal vectors will have the following form

$$
\begin{aligned}
s_1 &= (\sqrt{E_s}, 0, 0, \ldots\ldots, \; 0) \\
s_2 &= (0, \sqrt{E_s}, 0, \ldots\ldots, \; 0) \\
s_3 &= (0, 0, \sqrt{E_s}, \ldots\ldots, \; 0) \\
&: \\
s_M &= (0, 0, 0, \ldots\ldots, \; \sqrt{E_s})
\end{aligned}
$$

- The separation between the signal vectors $d_{\min}$ is given by

$$d_{\min} = \sqrt{2E_s}$$

M-ary Signaling

- It has been shown how $M = 4$ biorthogonal signals can be constructed in $N = 2$ dimensions.
- In general, a set of M biorthogonal signals can be constructed from a set of $M/2$ orthogonal signals $s_i(t)$ for $i = 1, 2, 3, \ldots, M/2$ and their negatives $-s_i(t)$ for $i = 1, 2, 3, \ldots, M/2$.
- The channel bandwidth required to transmit an information sequence by using M biorthogonal signals is half that required for M orthogonal signals.
- For this reason biorthogonal signals are preferred to orthogonal signals in certain applications where bandwidth conservation is critical.

M-ary Signaling

- The geometrical representation of M-*ary* biorthogonal signals starts with $M/2$ orthogonal vectors in $N = M/2$ dimensions and then appending their negatives.
- The signal vectors will have the following form

$$
\begin{aligned}
s_1 &= (\sqrt{E_s}, 0, 0, \ldots, 0) \\
s_2 &= (0, \sqrt{E_s}, 0, \ldots, 0) \\
&\vdots \\
s_{\frac{M}{2}} &= (0, 0, 0, \ldots, \sqrt{E_s}) \\
s_{\frac{M}{2}+1} &= (-\sqrt{E_s}, 0, 0, \ldots, 0) \\
&\vdots \\
S_M &= (0, 0, 0, \ldots, -\sqrt{E_s})
\end{aligned}
$$

E_b/N_0 and BER

- Often it is desirable to express the error performance in terms of the average energy per bit E_b or more appropriately the average energy per bit to noise power spectral density ratio E_b/N_0.
- The probability of a bit error P_B or the bit error rate (BER) performance is one of the main parameters of any modulation/coding scheme.
- In the case of binary systems, the probability of bit error P_B is the same as that for symbol error P_e since each symbol is used to transmit a single data bit.
- For simplicity, we will restrict ourselves to the case of binary systems only, since for M-ary systems the relationship between P_B and P_e is somewhat more complicated.

E_b/N_0 Ratio

- For polar (or anti-podal) signaling, the average energy per bit E_b is given by

$$E_b = \tfrac{1}{2}\left(A^2 T\right)_{binary\ 0} + \tfrac{1}{2}\left(A^2 T\right)_{binary\ 1}$$

$$= A^2 T$$

$$= E_s$$

- Consequently, the average probability of bit error P_B for binary polar signaling is given by

$$P_B = Q\left(\sqrt{\frac{2E_b}{N_0}}\right)$$

E_b/N_0 Ratio

- For unipolar signaling, the average energy per bit E_b is given by

$$
\begin{aligned}
E_b &= \tfrac{1}{2}(0)_{binary\ 0} + \tfrac{1}{2}\left(A^2T\right)_{binary\ 1} \\
&= \tfrac{1}{2}A^2T \\
&= \tfrac{1}{2}E_s
\end{aligned}
$$

- Consequently, the average probability of bit error P_B for binary unipolar signaling is given by

$$P_B = Q\left(\sqrt{\frac{E_b}{N_0}}\right)$$

E_b/N_0 and E_s/N_0

- When comparing the BER performance in terms of the E_b/N_0 parameter, it can be seen that binary polar signaling outperforms binary unipolar signaling by a factor of 2 (i.e. 3 dB).
- This compares with a factor of 4 (i.e. 6 dB) when using the E_s/N_0 parameter.
- Consequently, one needs to be very careful when making comparisons between the error performances of different signaling systems, especially for baseband signaling systems.
- Generally, the E_b/N_0 would be the more widely used parameter for conducting a comparison between signaling systems since it results in a more system-independent measure.

E_b/N_0 and S/N

- The E_b/N_0 parameter can be expressed in terms of the average signal power to average noise power ratio S/N .
- By introducing the signal bandwidth W, one can write the following identities:

$$\frac{E_b}{N_0} = \frac{ST}{N_0} = \frac{S}{RN_0} = \frac{SW}{RN_0 W} = \frac{S}{N}\left(\frac{W}{R}\right)$$

- where $N = N_0 W$, S is the average signal power, T is the bit duration, and $R = 1/T$ is the bit rate.
- The *spectral efficiency* η is defined as the bit rate per unit bandwidth and has units of bps/Hz, i.e.

$$\eta = \frac{R}{W}$$

BER and E_b/N_0

- The BER is usually plotted as logarithmic plot against E_b/N_0 and leads to the characteristic “waterfall curve”.
- The shape of the curve depends on the channel and on the modulation scheme.
- For example, most curves would be plotted for an ideal AWGN channel.
- There is an inherent trade-off between spectral efficiency η and average energy per bit to average noise power spectral density ratio E_b/N_0 where the greater the spectral efficiency η, the greater the required E_b/N_0 is likely to be.
- This tends to lead to two types of modulation/coding scheme, optimized for either power or bandwidth efficiency.
- As a result communications systems tend to be characterised as being either *power-limited* or *bandwidth limited*.