A Note on Comparison of Signaling Schemes.

Bit error Probability vs Symbol error Probability

Motivation:

Comparison of signaling schemes utilizing diff # of possible signals should be done on equivalent basis. One way to do this is to define an "equivalent bit error prob", $P_B$. For example, given a word (symbol) in any scheme,

\[ \begin{array}{cccccccc}
\text{x} & \text{a} & \text{a} & \text{a} & \text{a} & \text{x} & \text{---} & \text{x} \\
\end{array} \]

$P_B$ denotes the prob that a bit is in error. 

the object is to find $P_b(\text{e})$ vs $P_s(\text{e})$.

1. Orthogonal signals (MFSK)

Consider $M = 2^n \implies$ Each symbol in the $M$-ary is equivalent to an $n$-bit word for a binary system.
For example for \( n = 3 \) we set

<table>
<thead>
<tr>
<th>M-ary symbol</th>
<th>Binary</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>000</td>
</tr>
<tr>
<td>1</td>
<td>001</td>
</tr>
<tr>
<td>2</td>
<td>010</td>
</tr>
<tr>
<td>3</td>
<td>011</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>101</td>
</tr>
<tr>
<td>6</td>
<td>110</td>
</tr>
<tr>
<td>7</td>
<td>111</td>
</tr>
</tbody>
</table>

Transmitted symbols

Note: selecting any one of the \((n-1)\) erroneous symbols is equally likely. (Read that this is orthogonal.)

Note that in any column of this table, there are 4 0's and 4 1's. In general, there are \(2^{n-1}\) zeros and \(2^{n-1}\) ones. From Bayes' rule we have

\[
P_B(e) = \frac{P(B|W)P(e)}{P(W|B)}
\]

This is clearly one!
\[ P(B|W) = ? \quad \iff \quad \text{Prob of (hit error | word error)} \]

Looking at the table, it is seen that a word detected in error is \( \pm 1 \quad \text{(one is right)} \)

and in general it is

\[ \frac{n}{2} - 1 \]

Now, consider the transmitted signal

\[ 0 \quad 1 \quad 1 \]

Each bit of this symbol can be in error 4 ways.

For example, binary one in the rightmost column can be in error because there are four zeros in that column. In general, each bit can be in error \( n-1 \) times.
\[ P(B|W) = \frac{2^{n-1}}{2^n - 1} \quad \text{or} \quad \frac{M/2}{M-1} \]

and \[ P(B) = \frac{2^{n-1}}{2^n - 1} P(e) \]

Notice that

\[ \lim_{n \to \infty} P(B) = \frac{1}{2} P_s(e) \Rightarrow P_B(e) \leq P_s(e) \]

At \( n = 1 \Rightarrow M = 2 \)

binary!

See Fig. 4-7.
2. Many Phase Shift Keying (MPSK)

Recall that for orthogonal signaling, selecting any one of the \((M-1)\) erroneous symbols is equally likely. In the case of MPSK signaling, each signal vector is not equidistant from all of the others.

![Diagram of MPSK signaling]

If symbol \((011)\) is transmitted, it is clear that should an error occur, the transmitted signal will most likely be mistaken for one of its closest neighbors, \((010)\) or \((100)\). The likelihood that \((011)\) be mistaken for \((111)\) is relatively remote. If the assignment of bits to symbols follows the ones shown above, some symbol errors will usually result in two or more bit errors!
In these situations, one often uses a binary-to-
many code such that binary sequences corresponding to adjacent symbols differ in only one bit position; thus when an error occurs, it is more likely that only one of the n input bits will be in error. A code that provides this desirable feature is the Gray Code.

Utilizing the Gray code assignment, it can be shown that

\[ P_B(e) = \frac{P_5(e)}{n} \]
This can be proved as follows:

Know

\[ P_B(e) = \frac{P(B|W) P_e(e)}{P(W+B)} \]

\[ P(B|W) \approx \frac{1}{n} \]

\[ \sum_{x} \frac{x}{n} \]
noncoherent signaling schemes:

1. NFsk (Noncoherent Fsk)

In an NFsk system the received data during the kth signaling interval are

\[ Z(t) = \sqrt{\frac{2E_s}{T_s}} \cos(w_t + \alpha) + n(t) \quad 0 \leq t \leq T_s ; \quad i = 1, 2, \ldots, M \]

\[ \text{random} \]
\[ \text{phase} \]
\[ \text{(Uniform)} \]
\[ 0, \pi \]

\[ = \sqrt{\frac{2E_s}{T_s}} \cos(w_t \sin \alpha) - \sqrt{\frac{2E_s}{T_s}} \sin(w_t \sin \alpha) + n(t) \]

\[ i = 1, M. \]

Note that, although each transmitted signal \( s_i(t) \), \( i = 1, M \), is represented by a point in \( M \) dim-space, the presence of the unknown phase \( \alpha \) makes it necessary to use \( 2M + 1 \) basis forms in order to resolve the received signal \( Z(t) \).
Let
\[
\begin{align*}
\mathbf{Q}_x(t) &= \sqrt{\frac{2}{T_s}} \cos \omega_i t & i = 1, M \\
\mathbf{Q}_y(t) &= \sqrt{\frac{2}{T_s}} \sin \omega_i t & t \in (0, T_s)
\end{align*}
\]

Given that \( s_i(t) \) was sent, the coordinates of \( \mathbf{z} \) are as follows:
\[
\mathbf{z} = (x_1, y_1, x_2, y_2, \ldots, x_M, y_M)
\]

where
\[
X_j = (\mathbf{Q}_x, \mathbf{z}) = \sum_{i=1}^{N} x_j + \sqrt{E_s} \cos \alpha + N_{x_i} \\
Y_j = (\mathbf{Q}_y, \mathbf{z}) = \sum_{i=1}^{N} y_j - \sqrt{E_s} \sin \alpha + N_{y_i}
\]

\( j = 1, M \)

\( N_{x_j} \) and \( N_{y_j} \) are uncorrelated with variance \( \frac{N_s}{2} \).
Consider a DMS source:

\[ T \]

\[ 0100101 \ldots \]

Let \( T \) be the bit interval. Thus

\[ R = \frac{1}{T} \frac{\text{bits}}{\text{sec}} = \text{bit rate} \]

In the interval \( T \), therefore, we have \( RT \) bits.

\[ \text{Pr: } \frac{1 \text{ sec}}{T} R \frac{\text{bit}}{\text{sec}} \]

Let \( K = RT \) \( \leftarrow \) block size

Wish to see that different communication systems may yield drastically different performances for the same value of \( E_b \).
Consider the following:

Contrast the results achieved when a sequence of

\[ K = RT \]

equally likely bits is communicated by
two signalling schemes.

1. **Bit-by-bit Signalling**

   the first transmits a signal consisting of a
   sequence of \( K \) nonoverlapping pulse trains,
   as shown (message sequence is 1010)

   \[ s(t) \]

   \[ 0 \quad \frac{T_0}{T} \quad \frac{2T_0}{T} \quad \frac{3T_0}{T} \quad \frac{4T_0}{T} \quad \frac{5T_0}{T} \quad t \]

   \[ K = 5 \text{ in this case!} \]

   There are \( 2^K \) of these.

   This is called bit-by-bit signalling. But

   \[ x_0(t) \]

   \[ E_b = \int_{-\infty}^{\infty} x_0^2(t) \, dt \]
\[ s(t) = \sum_{j=1}^{K} s_j x_0(t-jT_0) \]

where
\[ s_j = \begin{cases} +1 & \text{bit 1} \\ -1 & \text{bit 0} \end{cases} \]

Let
\[ q_j(t) = \frac{x_0(t-jT_0)}{\sqrt{E_b}} \quad ; \quad j = 1, 2, \ldots, K \]

\[ q_1(t) \quad q_2(t) \quad \ldots \quad q_K(t) \]

Know \( M = 2^K \). Choosing \( q_j(t) \) as given, means that we associate the \( M \) possible signals with \( M \) vertices of a \( K \)-dim hypercube.
\[ K = 2 \quad (M = 2^2 = 4) \]

\[ \Phi (t) = \frac{x_0(t-T_0)}{\sqrt{E_b}} \]

\[ \Phi_r(t) = \frac{x_0(t-2T_0)}{\sqrt{E_b}} \]

\[ S_1(t) = -\sqrt{E_b} \Phi(t) \sqrt{E_b} \Phi_r(t) = 0 \]

\[ S_1 = (-\sqrt{E_b}, -\sqrt{E_b}) \]

\[ S_2 = (-\sqrt{E_b}, \sqrt{E_b}) \]

\[ S_3 = (\sqrt{E_b}, -\sqrt{E_b}) \]

\[ S_4 = (\sqrt{E_b}, \sqrt{E_b}) \]
Now consider one of these sequences:

\[ X \text{ bits} \]

\[ X X X \ldots X \]

the prob of at least one error with such a
signal set is

\[
P_w(e) = 1 - P_c(c) = 1 - \left(1 - P_b(e)\right)^K
\]

Recall that for binary antipodal signals of energy \( E_b \) in AWGN

\[
P_b(e) = Q\left(\sqrt{\frac{2E_b}{N_0}}\right)
\]

\[
\therefore P_w(e) = 1 - \left[1 - Q\left(\sqrt{\frac{2E_b}{N_0}}\right)\right]^K
\]

\[
\lim_{K \to \infty} P_w(e) = 1
\]

Thus, no matter how small \( P_b(e) \) is, \( P_w(e) \to 1 \).

\( P_w(e) \) can be made small only if \( E_b \uparrow \)
Indeed, for many years communications assumed that decreased $P_w(e)$ could be achieved only by increasing $E_b$. To see that this assumption is false, we consider the second approach.

2. **Block-Orthogonal Signaling**

This PPM has been considered before. Recall that in this case we have $M$-orthogonal signals $S_i(t)$, $i = 0, M-1$.

\[
\therefore \quad P_i(t) = \frac{S_i(t)}{\sqrt{E_s}}
\]

**Aside**: An example of this case is PPM.
Recall that
\[ P(e) \leq (M-1) Q(\sqrt{\frac{E_b}{N_0}}) < M e^{-\frac{E_b}{2N_0}} \]

\[ \therefore \quad P(e) < 2^k e^{-\frac{KE_b}{2N_0}} \]

But
\[ 2^k = e^{k \ln 2} = e^{k \ln 2} \]

\[ \therefore \quad P(e) < e^{k \ln 2} e^{-\frac{KE_b}{2N_0}} \]

\[ = e^{-k \left[ \frac{E_b}{2N_0} - \ln 2 \right]} \]

Notice that
\[ P(e) \rightarrow 0 \quad \text{as} \quad k \rightarrow \infty \quad \text{as long as} \]
\[ \frac{E_b}{2N_0} - \ln 2 > 0 \quad \Rightarrow \quad \frac{E_b}{N_0} > 2 \ln 2 \approx 1.39 \]

this can be shown in another fashion

as follows:
\[ P(e) \leq (M-1) Q(\sqrt{\frac{E_s}{N_0}}) < M e^{-\frac{E_s}{2N_0}} \]

Know

\[ M = \frac{K}{2} = \frac{RT}{2} \]

\[ P_s = \frac{E_s}{T} \]

\[ E_s = KE_b = T P_s \]

\[ P(e) \leq \frac{RT}{2} T P_s \]

\[ \frac{e^{\frac{P_s}{2N_0} - RL_n^2}}{e^{\frac{RT}{2N_0}}} = e^{-T \left[ \frac{P_s}{2N_0} - RL_n^2 \right]} \]

We see that as \( K \to \infty \) \( (T \to 0) \) the term

\[ e^{-T \left[ \frac{P_s}{2N_0} - RL_n^2 \right]} \to 0 \] as long as

\[ \frac{P_s}{2N_0} - RL_n^2 > 0 \implies 0 < R < \frac{P_s}{2N_0} \frac{1}{L_n^2} \approx 0.72 \frac{P_s}{N_0} \]
the contrast between the results of the two schemes is dramatic. In the bit-by-bit scheme, as \( K \to \infty \), \( P(e) \to 1 \) regardless of how large \( \frac{E_b}{N_0} \) is. In the second, by letting \( K \to \infty \), we can force \( P(e) \to 0 \) as we wish, provided that \( \frac{E_b}{N_0} > 1.39 \). An alternative statement is that \( \frac{P_s}{N_0} \) implies a bound on the max rate of communications; at rates below this max \( P(e) \) can be made as small as possible by choosing \( T \).
Geometric Interpretation

the geometry of the signal-vector constellations for bit-by-bit and block-orthogonal signaling provides insight into the contrast between their performances.

bit-by-bit

\[ k = 1 \]

Note that the distance between nearest neighbors remains fixed as \( k \) grows, whereas the \( k \) of nearest neighbors and the \( k \) of dimensions occupied by the signal set increase linearly with \( k \). The problem that at least one of the \( k \) relevant noise components will carry the received signal vector closer to a neighbor than to the true vector becomes very large as \( k \uparrow \); there are \( k \) chances for this to happen!
Notice that in this case, the distance between nearest neighbors grows linearly with $\sqrt{K}$ as shown above. (All signals are nearest neighbor here!)

It might seem that the block-orthogonal PPM signaling scheme provides a pulse $\xi$ to the general probe of accurate, efficient communications over AWGN. Unfortunately, such is not the case: for $R$ close to $0.72 \frac{P_0}{N_0}$, a very large value of $T$ is required to obtain a large negative exponent in the bound given above. This can be obtained at the expense of increasing $BW$. (Note that $M = 2^RT$, $T \uparrow \to 0 \; M \downarrow$)
To emphasize the dependence of $BW$ on $K = RT$, the dimensionality theorem is stated:

**Dimensionality Theorem**

Let $\{f_j(t)\}$ denote any set of orthogonal functions of duration $T$ and $BW$ $W$. More precisely, require that each $f_j(t)$

1. be identically zero outside a time interval $T$ and
2. have no more than $\frac{1}{T}$ of its energy outside the frequency interval $-W < f < W$.

Then the number of different waveforms in the set $\{f_j(t)\}$ is overbounded by $2.4TW$ when $TW$ is large.

Q.E.D.

For the case of orthogonal signaling, we have

$$M = 2^{RT} \leq 2.4TW \Rightarrow W > \frac{R^2RT}{2.4TR} = \frac{R^2}{2.4K}$$

which goes to infinity with $K$ or $T \to \infty$. Thus, error-free transmission is achieved with block-orthogonal signaling at the expense of infinite $BW$. 
Remark on dimensionality. The text says

\[ N \leq 2.4 TW \]

Thus for "bandlimited" channels (W fixed), N can grow no faster than linearly with time, T.

The converse statement is that for bandlimited channels

\[ N = DT \]

\[ \uparrow \]

no. of dim \# available per sec.

Where D varies linearly with W but is relatively insensitive to T. Note that for bandlimited channels, D (which is related to W) is fixed and N grows linearly with T.
Exa. (A way to obtain D)

\[ x(t) \xrightarrow{\mathcal{F}} X(f) \]

\[ N = \frac{T}{\gamma} = \frac{1}{\gamma T} = D T \]

\[ \therefore D = \frac{1}{T} = \# f \text{ Dim Per sec.} \]

Exa. (Relationship between D and B.W. ?)

Know \[ x(\alpha t) \xrightarrow{\mathcal{F}} \frac{1}{\alpha} X\left(\frac{f}{\alpha}\right) \]

\[ B.W \]

\[ N = \frac{T}{\gamma_{1/\alpha}} = \frac{\alpha}{T} \]

\[ \therefore D = \frac{\alpha}{T} = \# f \text{ Dim} \frac{\text{sec}}{\text{sec}} \]

\[ BV = \alpha W \text{ and } D = \frac{\alpha}{T} = \frac{1}{W T} \text{ B.W.} \]
Efficient Signal Selection

We observed that block-orthogonal signaling over AWGN channel would yield a $P_e \to 0$ as $T \to \infty$ for $R < 0.72 \frac{P_t}{N_0}$.

The drawback was that it requires $BW \to \infty$.

Wish to show that it is possible to achieve a $P_e$ similar to block-orthogonal signaling while simultaneously meeting the bandlimited channel constraint that the dimensionality of the signal space grows only linearly with $T$.

Note that

$$K = RT$$

$$N = DT$$

Let us consider a case in which

$$N > K \implies D > R$$
For simplicity, we restrict the signals to lie on the vertices of a hypercube.

\[ N = 3 \Rightarrow 2^3 = 8 \text{ vertices,} \]
\[ k = 2 \Rightarrow M = 2^2 = 4. \]

Since the number of vertices on a hypercube of dim \( N \) is \( 2^N \) and the # of signals required is \( M = 2^k \), not all of the vertices need be used. (Recall that \( k \leq N \)).
In fact, the fraction of vertices used is

\[
\frac{Z^K}{Z^N} = \frac{Z^{RT}}{Z^{DT}} = e^{- (D - R)T} \quad \rightarrow 0
\]

\[T \rightarrow \infty\]

there is a possibility that we can avoid convergence of \( P(s) \rightarrow 1 \) for bit-by-bit case considered before when \( D = R \) ( \( N = K \)).

Restricting the symbols \( \mathbf{\{s_i(t)\}} \) to lie on the vertices of a hypersphere implies that

\[
s_i(t) = \sum_{j=1}^{N} s_{ij} c_j(t) \quad ; \quad i = 0, 1, \ldots, M-1
\]

where

\[
s_{ij} = \pm \sqrt{E_N}
\]

\( E_N = \text{available signal energy/dimension} \)
the problem of signal selection for this example is to assign the vectors (code words)

$$S_i = (s_{i1}, s_{i2}, \ldots, s_{in}) \quad ; \quad i = 0, 1, \ldots, M-1$$

Notice that a good specific assignment is hard to find. The way around this is by bounding the

$$P_s(e)$$

averaged over all possible code selections.

there are $$2^N = 2^{DT}$$ possible vertices and

$$M = 2^{RT}$$ signals $$\{S_i\}$$ to be assigned to one of them.

It follows that there are $$\left(2^N\right)^M = 2^{NM}$$ distinct ways to assign the $$M$$ signals. We assume that each of these $$2^{NM}$$ signal sets is used by one (and only one) of the communication systems in our collection and that each system uses a receiver that is optimum for its signal set.
It is clear that each system in our collection has a definite prob of error, say $P_l$ for the $l$th system, $l = 1, 2, \ldots, 2^{NM}$. Some of the systems — for example, those with codes in which all $M$ of the vectors $\mathbf{s}_i$ are assigned to the same vertex — have a very large prob of error. On the other hand, most of the systems have a prob of error that is quite small, a fact that we shall prove by calculating a bound on the arithmetic avg, denoted $\overline{P(e)}$, over the entire collection:

$$\overline{P(e)} = \frac{1}{2^{NM}} \sum_{l=1}^{NM} P_l$$

Clearly, not all of the $P_l$ can be greater than $\overline{P(e)}$.

Note that the prob of the $i$th code, $s_j$, being used is

$$P(\{s_j\}) = \frac{1}{2^{NM}}$$
When the message $m_k$ is transmitted using a specific code $S_j$, the probability of error is

$$P(e | m_k, S_j) = \frac{1}{2^{Nm}} \sum_{\text{all codes}} P(e | m_k, S_j)$$

But, using the union bound on $P(e | m_k, S_j)$ gives

$$P(e | m_k, S_j) \leq \sum_{j=1}^{M} P(S_k \rightarrow S_j)$$

$$\therefore \quad P(e | m_k) \leq \frac{1}{2^{Nm}} \sum_{\text{all codes}} \sum_{j=1}^{M} P(S_k \rightarrow S_j)$$

$$= \sum_{j=1}^{M} \sum_{j \neq k} P(S_k \rightarrow S_j)$$

$$= \sum_{j=1}^{M} P(S_k \rightarrow S_j)$$
For AWGN

\[ P(s_k \rightarrow s_j) = Q \left( \frac{\| s_j - s_k \|}{\sqrt{t N_0}} \right) \]

and it can be shown that if \( s_j \) and \( s_k \) differ in \( P \) coordinates

\[ \| s_j - s_k \|^2 = 4 PN_0 \]

and

\[ P(s_k \rightarrow s_j) = \sum_{P=0}^{N} \binom{N}{P} 2^{-N} \frac{1}{\sqrt{2 PN_0}} Q \left( \frac{\sqrt{2 PN_0}}{N_0} \right) \]

\[ \text{where } i, j, \text{ and } k \]

\[ P(e \mid m_k) \leq \sum_{j=1, j \neq k}^{M} P(s_k \rightarrow s_j) = (M-1) \overline{P(s_k \rightarrow s_j)} \]

\[ \leq M \overline{P(s_k \rightarrow s_j)} \]

But

\[ P(e) = \sum_{K=1}^{M} P(e \mid m_k) P(m_k) \]

\[ \leq M \overline{P(s_k \rightarrow s_j)} \sum_{K=1}^{M} P(m_k) \]
Recall

\[ Q(\alpha) \leq \frac{1}{2} e^{-\frac{\alpha^2}{2}} < e^{-\frac{\alpha^2}{2}} \quad \alpha > 1 \]

\[ P(\mathbf{s}_j \rightarrow \mathbf{s}_k) < z^{-N} \left(1 + e^{-\frac{E_N}{N_0}}\right)^N \]

and

\[ P(e) < M z^{-NR_0} \]

where

\[ R_0 = \log_2 \frac{z}{1 + \exp(-E_N/N_0)} \]

let us define

\[ R_N = \frac{R}{D} = 0 \quad M = 2^{RT} = 2^{NR_0} \]

\[ M = 2^{NR_N} \]

so that

\[ P(e) < 2^{NR_N} - NR_0 = 2^{-N(R_0 - R_N)} \]
where $R_0$ is called its exponential bound.

\[
\begin{align*}
R_0 \rightarrow 1 \\
a = \frac{EN}{N_0} \\
10 \log \left( \frac{EN}{N_0} \right) \\
(fy \left( \frac{4.20}{237} + \text{txt} \right)
\end{align*}
\]

It is clear that as long as $R_0 > R_N$, $P(\text{er})$ can be made arbitrarily small by choosing $N \uparrow$.

Recall that our original assumption was

\[ R < D \rightarrow R_N < 1 \]

So let $R_0 \approx 1$ to ensure $R_0 > R_N$.!
Summary

We showed that

\[ P(e) < 2^{-N[R_0 - R_N]} \quad (*) \]

which guarantees that signal sets exist which afford communication through AWGN at any rate

\[ R_N < R_0 \]

with arbitrary prob of error. The value of \( N \) obtained by setting the right-hand side equal to the desired error prob is an upper bound on the # of dimensions necessary to achieve such performance.

On the other hand, (*) do not imply that an arbitrarily slow prob of error cannot be obtained for \( R_N > R_0 \).
the discussion so far can be stated as follows:

there exists a constant, \( C_N \), given by

\[
C_N = \frac{1}{2} \log_2 (1 + \frac{2EN}{N_0}) \quad \text{bits/dim}
\]

and called the Gaussian channel capacity, with the following properties:

**Negative statement:** If \( R_N > C_N \) and the number of equally likely messages, \( M = 2^{NR_N} \), is large, the \( P(e) \) is close to one for any possible set of \( M \) transmitter signals.

**Positive statement:** If \( R_N < C_N \) and \( M \) is sufficiently large, there exist sets of \( M \) transmitter signals such that \( P(e) \) achieved with optimum receivers is arbitrarily small.

**Note:** \( R_N < C_N \Rightarrow \frac{R}{D} < C_N \Rightarrow R < DC_N \) \quad \text{bits/sec/Hz}

But it can be shown that for band-limited channels

\[
D \leq 2W = 0 \quad \text{for} \quad D = 2W \quad \text{and} \quad E_N = \frac{P_s}{D} = \frac{P_s}{2W}
\]

\[
R < DC_N = 2W \frac{1}{2} \log_2 (1 + \frac{2P_s}{2WN_0}) = W \log_2 (1 + \frac{P_s}{WN_0}) = C
\]