Performance evaluation.

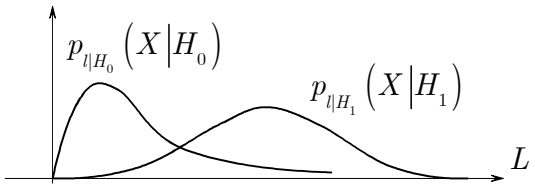
- A basic performance measures of a binary signal detection system with given decision rule are the two conditional error probabilities P_F and P_M.
- In many cases the optimum test can be derived but an exact performance is impossible to calculate.
- We can resort to bounds of the error probabilities or approximate expressions for these probabilities.
- The problem of interest is a general binary hypothesis test.
- The likelihood ratio test is:

$$l(\mathbf{R}) \triangleq \Lambda(\mathbf{R}) = \ln\left[\frac{p_{\mathrm{r}|H_1}\left(\mathbf{R} \mid H_1\right)}{p_{\mathrm{r}|H_0}\left(\mathbf{R} \mid H_0\right)}\right] \underset{H_0}{\overset{H_1}{\lessgtr}} \gamma.$$

- The variable $l(\mathrm{R})$ is a random variable whose probability density depends on which hypothesis is true.
- If two densities are known, then P_F and P_M are given by

$$P_M = \int_{-\infty}^{\gamma} p_{l|H_1}\left(L \mid H_1\right) dL$$

$$P_F = \int_{\gamma}^{\infty} p_{l|H_0}\left(L \mid H_0\right) dL$$



Typical densities

- The difficulty is that it is often hard to find $p_{l|H_i}\left(L \mid H_i\right)$, and even if it can be found it is cumbersome.
- We can estimate the error by calculating the bounds for error

<u>Gaussian multidimensional decision.</u>

The Bayes test for general M-hypothesis problem can be expressed as following:

- Compute $\beta_i \triangleq \sum_{j=0}^{M-1} C_{ij} p_{\mathrm{r}|H_j}\left(H_j \mid \mathbf{R}\right)$, $i = 0, \ldots, M-1$ and choose the smallest.
- We minimize the cost function by selecting the i that gives the smallest value for given $\mathbf{R}$.
- If $C_{ii} = 0$, $i = 0,1,2,\ldots,M-1$
  $C_{ij} = C$, $i \neq j$, $i, j = 0,1,2,\ldots,M-1$
- The equivalent test becomes:
  Compute $P\left(H_i \mid \mathbf{R}\right)$, $i = 0, \ldots, M-1$ and choose the largest.
- By choosing the largest we exclude it from contributing to the cost.

$$p_{\mathrm{r}|H_j}\left(H_j \mid \mathbf{R}\right) = \frac{P_j p_{\mathbf{r}|H_j}\left(\mathbf{R} \mid H_j\right)}{p_{\mathbf{r}}\left(\mathbf{R}\right)}$$

$$B_i \triangleq \frac{1}{p_{\mathbf{r}}\left(\mathbf{R}\right)} \sum_{j=0}^{M-1} C_{ij} P_j p_{\mathbf{r}|H_j}\left(\mathbf{R} \mid H_j\right)$$

- The Bayes test becomes: Compute

$$\mathrm{B}_i \triangleq \sum_{j=0}^{M-1} C_{ij} P_j p_{\mathbf{r}|H_j}\left(\mathbf{R} \mid H_j\right)$$

and choose the smallest.

- The error performance of a congruent-decision region receiver is invariant to the actual source statistics.
- A receiver designed to be optimal under the assumption of equally likely messages is a maximum likelihood receiver.
- When $\left\{s_i\right\}$ are completely symmetric the maximum likelihood receiver is a minimax receiver.

Symmetric signal sets.

- In this specific case calculation of the error probability is greatly simplified by the "completely symmetry" of the geometrical configuration of the $\left\{s_i\right\}$.
- By the complete symmetry we mean that any re-labeling of the signal points can be undone by rotation of coordinates, translation, and/or inversion of axes.
- Given:

  $P_i = \frac{1}{M}, \forall i$

  leads to a congruent decision regions and thus to a conditional probability of correct decision that is independent of the particular signal transmitted:

  $p\left(correct \mid H_i\right) =$ a constant

The union bound.

- An approximation to the error probability for any set of M equally likely signals $\left\{s_i\right\}$ in white Gaussian noise.
- Pairwise error probability

  $p\left\{s_i \rightarrow s_j\right\}$.
- An error occurs when s_i is transmitted and the received vector $\mathbf{r}$ is closer to one of signals s_j, $j \neq i$ than it is to s_i

  $p\left(e_{ij} \mid s_i\right) = p\left\{s_i \rightarrow s_j\right\}$,

  $p\left\{s_i \rightarrow s_j\right\} = \frac{1}{2} erfc\left(\frac{d_{ij}}{2\sqrt{N_0}}\right)$,

  where $d_{ij} = \left|s_i - s_j\right|$ is Euclidean distance.

- If e_{ij} is used to denote the event that $\mathbf{r}$ is closer to s_j when s_i is transmitted, we have:

$$p\left(e \mid s_i\right) = p\left(e_{i,0} \cup e_{i,1} \cup \ldots \cup e_{i,i-1} \cup e_{i,i+1} \cup \ldots \cup e_{i,M-1}\right).$$

- Since the probability of a union of events cannot exceed the sum of the individual probabilities, we have a union bound.

$$p\left(e \mid s_i\right) \le \sum_{j \ne i} p\left(e_{ij} \mid s_i\right) = \sum_{j \ne i} p\left\{s_i \to s_j\right\} \approx \sum_{j \ne i} \frac{1}{2} erfc\left(\frac{d_{ij}}{2\sqrt{N_0}}\right)$$

- By averaging over the signal set

$$p(e) \le \frac{1}{M} \sum_{i=1}^{M} \sum_{j \ne i} \frac{1}{2} erfc\left(\frac{d_{ij}}{2\sqrt{N_0}}\right)$$

- For any set of M equally likely equal-energy orthogonal signals in AWGN channel the probability of error is bounded by:

$$p(e) \le (M-1) Q\left(\sqrt{\frac{E_s}{N_0}}\right) \le MQ\left(\sqrt{\frac{E_s}{N_0}}\right)$$

- For bounding the error it suffices to know all the distances d_{ij} among signals in constellation.
- The union bound becomes tighter and tighter as N_0 decreases, i.e. when $P(e)$ decreases, so that for low enough error probabilities it provides a good approximation to their exact values.

The union Bhattacharyya bound

- A simpler form of the union bound can be obtained by using a bound to the pairwise error probability exceeds the exact value.

$$p\left\{s_i \to s_j\right\} = \frac{1}{2} erfc\left(\frac{d_{ij}}{2\sqrt{N_0}}\right) \le \exp\left\{-\frac{d_{ij}}{4N_0}\right\}$$

$$p(e) \le \frac{1}{M} \sum_{i=1}^{M} \sum_{j \ne i} \exp\left\{-\frac{d_{ij}}{4N_0}\right\}$$

Some other useful bounds

$$\frac{1}{\sqrt{2\pi}X}\left(1 - \frac{1}{X^2}\right)\exp\left(-\frac{X^2}{2}\right) < \frac{1}{2} erfc\left(\frac{d_{ij}}{2\sqrt{N_0}}\right) \le \frac{1}{\sqrt{2\pi}X} \exp\left\{-\frac{d_{ij}}{4N_0}\right\}$$

Source rate

- Given a message source that produces a sequence of discrete symbols, we are interested in characterizing how much transmission capability is required to communicate the source output to a distant terminal.
- This can be characterized how many hypotheses (messages) in the time interval can correctly be separated at the receiver.
- For a set of M equally likely hypothesis (messages) in any time interval T we define the source rate R as:

$$R = \frac{1}{T} \log_2 M.$$

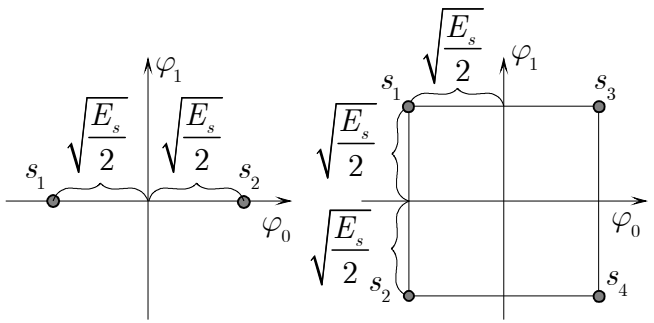
- The maximal rate that the communication system can satisfy satisfactory when constrained in power is called channel capacity.

Bit-by-Bit and Block-Orthogonal signaling

- For the same value of E_b different communication systems yield different performance.
- Assume $K = RT$ equally likely binary digits is communicated by two different schemes:
  - Bit-by-Bit
    - System transmits K nonoverlapping pulses.
    - Each pulse has the same waveshape and is positive when input bit is 1 and negative when input bit is 0.
  - Block-Orthogonal
    - The signaling scheme uses a signal set of 2^K orthogonal pulses, each having energy $E_s = KE_b$.
    - The choice of transmitted signal is made by observing the entire input sequence at once and transmitting the i-th pulse when the binary number specified by this sequence is i.

Bit-by-Bit signaling

- We associate the $M = 2^K$ possible signals with the 2^K vertices of a K-dimensional hypercube.
- The probability of at least error with such a signal set is
  $p(e) = 1 - (1-p)^K = 1 - (1-p)^{RT}$
- For additive white Gaussian noise (AWGS) channel the probability of error for binary decision between two antipodal signals of energy E_b is
  $$p = Q\left(\sqrt{\frac{2E_b}{N_0}}\right)$$
- For any choice of R and E_b the probability of error tends to 1 when T becomes large.
- For fixed T the probability of error can be made small only by increasing energy expended per bit.
- The distance between nearest neighbor remains fixed as K increases.
- The number of nearest neighbors and the number of dimensions occupied by the signal set increases linearly with K.



Block-Orthogonal signaling

- For example one out of 2^K orthogonal pulses is transmitted every T second.
- Recall that for any set of M equally likely equal-energy orthogonal signals in AWGN channel the probability of error is bounded by

$$p(e) \le MQ\left(\sqrt{\frac{E_s}{N_0}}\right) \le Me^{-E_s/2N_0}$$

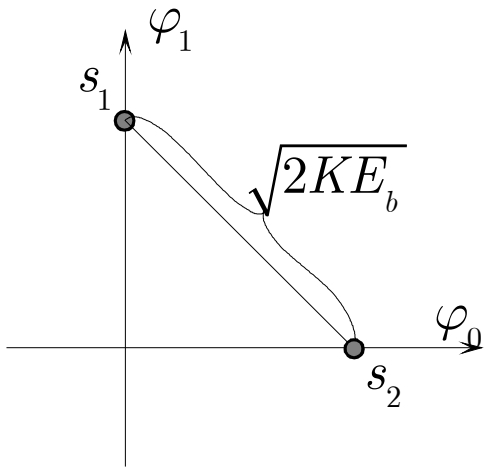
- By substituting $M = 2^K = 2^{RT}$ and $E_s = KE_b = TP_s$.

$$p(e) \le 2^{RT} e^{-TP_s/2N_0} = \exp\left(-T\left(\frac{P_s}{2N_0} - R\ln 2\right)\right) = \exp\left(-K\left(\frac{E_b}{2N_0} - \ln 2\right)\right)$$

- The probability of error approaches zero exponentially with increasing T, as long as the rate R satisfies the bound

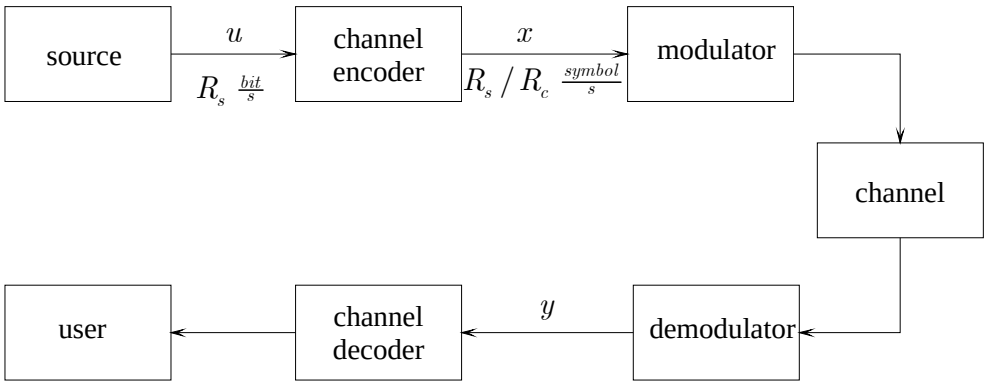
$$R < \frac{P_s}{2N_0}\frac{1}{\ln 2} \approx 0.72\frac{P_s}{N_0},\ \frac{E_b}{N_0} > 2\ln 2 \approx 1.39$$

- By increasing K we can force the probability of error to be as close to zero as we wish, provided that $\dfrac{E_b}{N_0} > 1.39$
- The distance between the nearest neighbors grows linearly with $\sqrt{K}$.
- The increase of distance is achieved by introducing a new dimension for each of the additional signals and re-scaling the amplitude.



Error control codes.

- Techniques to control the error probability are based on the addition of redundancy to the information sequence.



- Source producing a binary sequence – data stream, R_s, formed from independent identically distributed random variables.
- Encoder maps the data stream into a code stream R_c.
- Data stream may be segmented into data words.
- Segmented code streams form code words. (n,k) block code consist the code works of n bits and data word of k bits.
- A channel code is the set of 2^k n-tuples of bits.
- Encoder is the set of the 2^k pairs (u,x) where u is a data word.
- The redundancy of a code is $r = n - k$.

Types of codes.

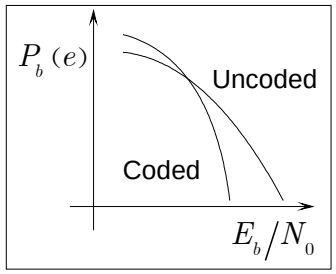
- The block code is an (n,k) code and the ratio $R_c \triangleq \frac{k}{n}$is the rate of the code.
- If the n bits of the code word depends not only on the k bits of the data word, but also on some previous data words, the code is a tree code.
- Tree codes with a special memory and linearity structure are convolutional codes.

Benefits due to channel encoding.

- Benefits of the channel coding compared with the uncoded schemes.
- The source emits binary digits at a rate of R_s $bits/s$ and the encoder represents each data word of k source bits using $n = k/R_c$ $bits$.
- To keep the pace of the source the transmission speed of the channel must be increased to the values R_s/R_c binary symbols per second. (the required bandwidth must be increased by the same factor.)
- The use of channel coding decreases the bandwidth efficiency with respect to the uncoded transmission by factor $1/R_c$.
- Coding decreases the energy of channel symbols to the value E_b/R_c.
- More channel symbols will be incorrectly demodulated than with uncoded transmission.

- Bandwidth efficiency is decreased and more errors in the demodulated sequence are to be expected.
- At receiver we attempt to compensate the large number of errors at the demodulator output by the error-correcting capabilities of the decoder.
- A coded transmission should trade bandwidth efficiency for a better overall error performance.
- The decrease in the required power for the coded system is referred to as coding gain.
- The overall error performance of the coded scheme depends on the implementation of efficient algorithms for error detection and correction.

- In Gaussian channel the coding gain, which depends on value of the bit error probability, increases with the signal-to-noise ration and tends to an asymptotic value.



- For low values of the signal-to-noise ratio, there can be a crossing between the uncoded and coded curves, meaning that the coding gain becomes negative. In other words there is a limit to what a code can do in terms of improving a bad channel.

Transmission errors.

- Error detection. Determination by the decoder whether errors are present in a received word.
- Undetectable errors. The error pattern causes the received word to be a valid word other than the transmitted word.
- Error correction. The arithmetic of algebraic structure of the code is used to determine which of the valid code words is most likely to have been sent, given the erroneous received word.
- Decoder error. The decoder selects a code word other than that which was actually transmitted.

Decoder types.

- Assume a binary transmission.
- The goal of forward error correction (FEC) system is to minimize the probability of decoder error given a received word r.

Hard decision decoding

- The demodulator output is quantized to two levels [1 0].
- The decoder attempts to recover the information sequence by using the code word redundancy for either detecting or correcting the errors that ate present at the demodulator output.

  (In this model a binary coherent modulation and AWGN channel, the combination of modulator, channel, and demodulator is equivalent to a binary symmetric channel (BSC)).

Unquantized soft decision decoding

- Binary input continuous output channel.
- The unquantized output of the demodulator.
- The receiver derives the sufficient statistics and supplies it to the decoder, which performs the estimation of the information sequence.
- Decoder stores the n outputs corresponding to each sequence of n binary waveforms and builds 2^kdecision variables.
- The decoder can take advantage of the additional information contained in the unquantized samples that represent each individual binary transmitted waveform.

Quantized soft-decision decoding.

- Binary input q -ary output discrete channel.
- The demodulator output is quantized to q levels with $q > 2$.

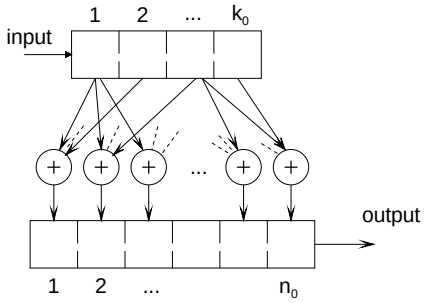
Maximum posteriori decoder. Identifies the code word c_i that maximizes $p\left(c = c_i \mid r\right)$.

Maximum likelihood decoder. Identifies the code word c_i that maximizes $p\left(r \mid c = c_i\right)$.

Bayes rule $p\left(s \mid r\right) = \dfrac{p_S\left(s\right) p\left(r \mid s\right)}{p_R\left(r\right)}$

Linear block codes

- The block of n digits generated by the encoder depends only on the corresponding block of k digits generated by the source.
- From the 2^n possible code words in a binary block code of length n we select $M = 2^k$ code words to form a code.
- Coder maps a block of k information bits into a code work of length n
- The encoding and decoding functions involve the arithmetic operations of addition and multiplication performed on the code words.



Properties of Linear Codes

- The linear combination of any set of code words is a codeword.
- The minimum distance of a linear code is equal to weight of a code word with minimal weight.
- The undetectable error patterns for a linear code word are independent of the code word transmitted and always consist of the set of all nonzero code words.

Example: Hamming code $(7, 4)$

$x_i = u_i,\ i = 1, 2, 3, 4$

$x_5 = u_1 + u_2 + u_3$

$x_6 = u_2 + u_3 + u_4$

$x_7 = u_1 + u_2 + u_4$

- An encoder is called systematic when the first k digits in the code word are a replica of the information digits in the data word, and the remaining $(n - k)$ digits are parity checks on the k information digits.
- The encoding rule can be represented by the generator $k \times n$ matrix $\mathbf{G}$.
- The code word x can be calculated by multiplying the data word u with the generator matrix.

$\mathbf{x} = \mathbf{uG}$

- For the $(7,4)$ Hamming code

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{I} \mid \mathbf{P} \end{bmatrix}$$

$$\mathbf{u} = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{uG} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

Example

For the $(7,4)$ Hamming code

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

Parity check matrix

- Associated with any linear (n,k) code is the dual code of dimension $(n-k)$.
- The dual code is a linear $(n, n-k)$ code with 2^{n-k} code vectors.
- The code vectors of the dual code are in the null space of the (n,k) code.
- The dual code generator matrix $\mathbf{H}$ consists of $(n-k)$ linearly independent code vectors selected from the null space of $\mathbf{G}$ matrix.
- The code word of the (n,k) code is orthogonal to every row of the matrix $\mathbf{H}$:

  $\mathbf{xH}' = \mathbf{uGH}' = 0.$
- If $\mathbf{G}$ is systematic then $\mathbf{H} = \begin{bmatrix} \mathbf{P}' \mid \mathbf{I}_{n-k} \end{bmatrix}$.
- $\mathbf{H}$ is called parity check matrix.

Error detection and error correction capabilities of a block code.

- The decoder recomputes the $(n-k)$ parity checks using the first k received bits, and compares them with the $(n-k)$ received parity checks. If they match the received sequence is a code word otherwise, and error is detected.
- An error pattern is detected whenever at least one of the $(n-k)$ controls on parity checks fails.
- Multiplication of the parity check matrix with received symbols gives a parity check vector $\mathbf{S}$ called syndrome.

  $\mathbf{s} = \mathbf{yH}'$
- The received symbols can be expressed in the form

  $\mathbf{y} = \mathbf{x} + \mathbf{e}$, where $\mathrm{e} = \begin{bmatrix} e_1, e_2, \ldots, e_n \end{bmatrix}$ is an error vector.
- The syndrome associated with a sequence y is a zero vector if and only if $\mathbf{y}$ is a code word.

- Decoder can detect all channel errors represented by vectors $\mathbf{e}$ that are not code words.
- Which of the error vectors $\mathbf{e}$ occurred can not be decides based on the syndrome.
- ML decoding is achieved with minimum Hamming distance decision.
- The " best" decoding algorithm decides for the code word $\mathbf{x}_i$ which is closest to $\mathbf{y}$.
- The decoding algorithm is implemented by assigning to each code word a decision region containing the subset of all the received sequences that are closer to it than to any other.
- An error vector with no more than $t = \lfloor (d_{\min} - 1)/2 \rfloor$ errors produces a received sequence lying inside the correct decision region. Error correction is therefore possible.

Distance of a code.

- Hamming weight of a word s, denoted $w_H(s)$, is the number of nonzero coordinates in s.
- Hamming distance between two words, $v = (v_0, v_1, \ldots, v_{n-1})$ and $w = (w_0, w_1, \ldots, w_{n-1})$, is the number of coordinates in which they differ.
- The Eucledian distance between $v = (v_0, v_1, \ldots, v_{n-1})$ and $w = (w_0, w_1, \ldots, w_{n-1})$, is
  $d_E(v, w) = \sqrt{(v_0 - w_0)^2 + (v_1 - w_1)^2 + \ldots + (v_{n-1} - w_{n-1})^2}$
- The minimum distance of a block code is the minimum Hamming distance between all distinct pairs of code words in the codeset.

Minimum distance and error

- Let $d_{\min}$ be minimum distance of the code.
- A linear block code (n, k) with minimum distance $d_{\min}$ can detect all error vectors of weight not greater than $(d_{\min} - 1)$.
- Decoder can detect all channel errors represented by vectors e that are not code words.
- A linear block code (n, k) with minimum distance $d_{\min}$ can correct all error vectors containing no more than $t = \lfloor (d_{\min} - 1)/2 \rfloor$ errors, where $\lfloor a \rfloor$ denotes the largest integer contained in a. The code is a t-error correcting code and is often denoted as a (n, k, t) code.