4 Optimum Reception in Additive White Gaussian Noise (AWGN)

In this chapter, we derive the optimum receiver structures for the modulation schemes introduced in Chapter 3 and analyze their performance.

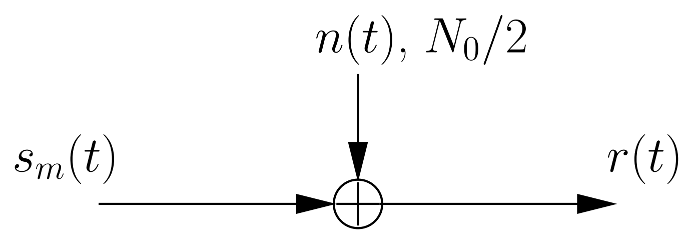
4.1 Optimum Receivers for Signals Corrupted by AWGN

■ Problem Formulation

– We first consider memoryless linear modulation formats. In symbol interval $0 \le t \le T$, information is transmitted using one of M possible waveforms $s_m(t)$, $1 \le m \le M$.

– The received passband signal $r(t)$ is corrupted by real–valued AWGN $n(t)$:

$$r(t) = s_m(t) + n(t), \qquad 0 \le t \le T.$$



– The AWGN, $n(t)$, has *power spectral density*

$$\Phi_{NN}(f) = \frac{N_0}{2} \left[\frac{\mathrm{W}}{\mathrm{Hz}}\right]$$

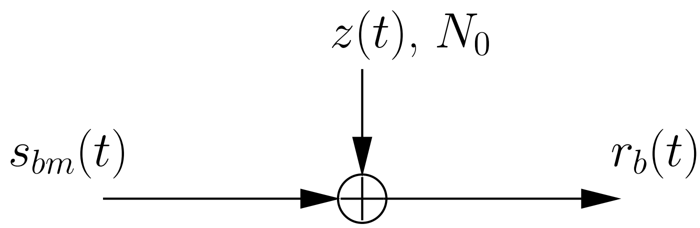
– At the receiver, we observe $r(t)$ and the question we ask is: **What is the best decision rule for determining $s_m(t)$?**

– This problem can be equivalently formulated in the complex baseband. The received baseband signal $r_b(t)$ is

$$r_b(t) = s_{bm}(t) + z(t)$$

where $z(t)$ is complex AWGN, whose real and imaginary parts are independent. $z(t)$ has a power spectral density of

$$\Phi_{ZZ}(f) = N_0 \left[\frac{\mathrm{W}}{\mathrm{Hz}}\right]$$



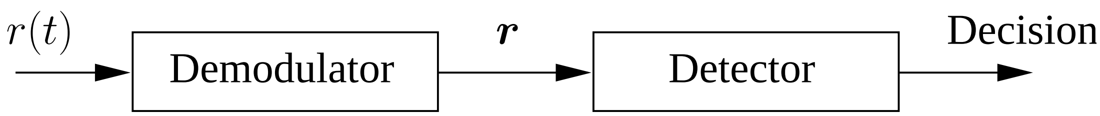
■ **Strategy:** We divide the problem into two parts:

1. First we transform the received continuous–time signal $r(t)$ (or equivalently $r_b(t)$) into an N–dimensional vector

$$\boldsymbol{r} = [r_1 \;\; r_2 \;\; \ldots \;\; r_N]^T$$

(or $\boldsymbol{r}_b$), which forms a *sufficient statistic* for the detection of $s_m(t)$ ($s_{bm}(t)$). This transformation is referred to as *demodulation*.

2. Subsequently, we determine an estimate for $s_m(t)$ (or $s_{bm}(t)$) based on vector $\boldsymbol{r}$ (or $\boldsymbol{r}_b$). This process is referred to as *detection*.



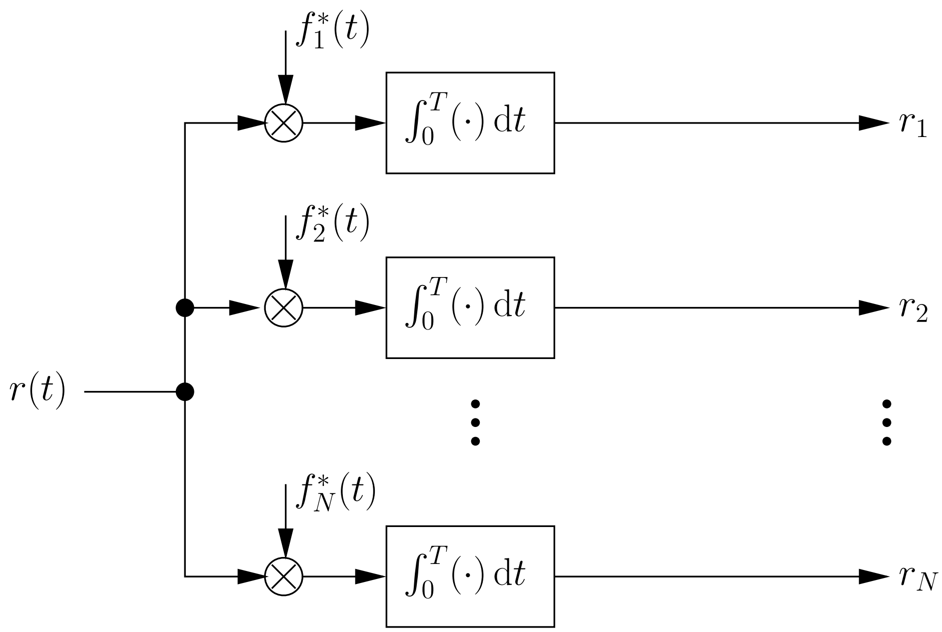
4.1.1 Demodulation

The demodulator extracts the information required for optimal detection of $s_m(t)$ and eliminates those parts of the received signal $r(t)$ that are irrelevant for the detection process.

4.1.1.1 Correlation Demodulation

- Recall that the transmit waveforms $\{s_m(t)\}$ can be represented by a set of N orthogonal basis functions $f_k(t)$, $1 \le k \le N$.
- For a complete representation of the noise $n(t)$, $0 \le t \le T$, an infinite number of basis functions are required. But fortunately, only the noise components that lie in the signal space spanned by $f_k(t)$, $1 \le k \le N$, are relevant for detection of $s_m(t)$.
- We obtain vector $\boldsymbol{r}$ by correlating $r(t)$ with $f_k(t)$, $1 \le k \le N$

$$\begin{aligned} r_k &= \int_0^T r(t) f_k^*(t)\,\mathrm{d}t = \int_0^T [s_m(t) + n(t)] f_k^*(t)\,\mathrm{d}t \\ &= \underbrace{\int_0^T s_m(t) f_k^*(t)\,\mathrm{d}t}_{s_{mk}} + \underbrace{\int_0^T n(t) f_k^*(t)\,\mathrm{d}t}_{n_k} \\ &= s_{mk} + n_k, \qquad 1 \le k \le N \end{aligned}$$



- $r(t)$ can be represented by

$$\begin{aligned} r(t) &= \sum_{k=1}^{N} s_{mk} f_k(t) + \sum_{k=1}^{N} n_k f_k(t) + n'(t) \\ &= \sum_{k=1}^{N} r_k f_k(t) + n'(t), \end{aligned}$$

where noise $n'(t)$ is given by

$$n'(t) = n(t) - \sum_{k=1}^{N} n_k f_k(t)$$

Since $n'(t)$ does not lie in the signal space spanned by the basis functions of $s_m(t)$, it is irrelevant for detection of $s_m(t)$. Therefore, without loss of optimality, we can estimate the transmitted waveform $s_m(t)$ from $\boldsymbol{r}$ instead of $r(t)$.

■ Properties of n_k

– n_k is a Gaussian random variable (RV), since $n(t)$ is Gaussian.

– Mean:

$$\begin{aligned}\mathcal{E}\{n_k\} &= \mathcal{E}\left\{\int_0^T n(t) f_k^*(t)\, \mathrm{d}t\right\} \\ &= \int_0^T \mathcal{E}\{n(t)\} f_k^*(t)\, \mathrm{d}t \\ &= 0\end{aligned}$$

– Covariance:

$$\begin{aligned}\mathcal{E}\{n_k n_m^*\} &= \mathcal{E}\left\{\left(\int_0^T n(t) f_k^*(t)\, \mathrm{d}t\right)\left(\int_0^T n(t) f_m^*(t)\, \mathrm{d}t\right)^*\right\} \\ &= \int_0^T \int_0^T \underbrace{\mathcal{E}\{n(t) n^*(\tau)\}}_{\frac{N_0}{2}\delta(t-\tau)} f_k^*(t) f_m(\tau)\, \mathrm{d}t\, \mathrm{d}\tau \\ &= \frac{N_0}{2} \int_0^T f_m(t) f_k^*(t) \mathrm{d}t \\ &= \frac{N_0}{2}\, \delta[k-m]\end{aligned}$$

where $\delta[k]$ denotes the Kronecker function

$$\delta[k] = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases}$$

We conclude that the N noise components are zero–mean, mutually uncorrelated Gaussian RVs.

■ Conditional pdf of $\boldsymbol{r}$

$\boldsymbol{r}$ can be expressed as

$$\boldsymbol{r} = \boldsymbol{s}_m + \boldsymbol{n}$$

with

$$\begin{aligned} \boldsymbol{s}_m &= [s_{m1}\ s_{m2}\ \ldots s_{mN}]^T, \\ \boldsymbol{n} &= [n_1\ n_2\ \ldots n_N]^T. \end{aligned}$$

Therefore, conditioned on $\boldsymbol{s}_m$ vector $\boldsymbol{r}$ is Gaussian distributed and we obtain

$$\begin{aligned} p(\boldsymbol{r}|\boldsymbol{s}_m) &= p_{\boldsymbol{n}}(\boldsymbol{r} - \boldsymbol{s}_m) \\ &= \prod_{k=1}^{N} p_n(r_k - s_{mk}), \end{aligned}$$

where $p_{\boldsymbol{n}}(\boldsymbol{n})$ and $p_n(n_k)$ denote the pdfs of the Gaussian noise vector $\boldsymbol{n}$ and the components n_k of $\boldsymbol{n}$, respectively. $p_n(n_k)$ is given by

$$p_n(n_k) = \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{n_k^2}{N_0}\right)$$

since n_k is a real–valued Gaussian RV with variance $\sigma_n^2 = \frac{N_0}{2}$. Therefore, $p(\boldsymbol{r}|\boldsymbol{s}_m)$ can be expressed as

$$\begin{aligned} p(\boldsymbol{r}|\boldsymbol{s}_m) &= \prod_{k=1}^{N} \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{(r_k - s_{mk})^2}{N_0}\right) \\ &= \frac{1}{(\pi N_0)^{N/2}} \exp\left(-\frac{\sum_{k=1}^{N} (r_k - s_{mk})^2}{N_0}\right), \qquad 1 \leq m \leq M \end{aligned}$$

$p(\boldsymbol{r}|\boldsymbol{s}_m)$ will be used later to find the optimum estimate for $\boldsymbol{s}_m$ (or equivalently $s_m(t)$).

- **Role of** $n'(t)$
  We consider the correlation between r_k and $n'(t)$:

$$\begin{aligned}
\mathcal{E}\{n'(t)r_k^*\} &= \mathcal{E}\{n'(t)(s_{mk}+n_k)^*\} \\
&= \underbrace{\mathcal{E}\{n'(t)\}}_{=0}\, s_{mk}^* + \mathcal{E}\{n'(t)n_k^*\} \\
&= \mathcal{E}\left\{\left(n(t) - \sum_{j=1}^{N} n_j f_j(t)\right) n_k^*\right\} \\
&= \int_0^T \underbrace{\mathcal{E}\{n(t)n^*(\tau)\}}_{\frac{N_0}{2}\delta(t-\tau)} f_k(\tau)\,\mathrm{d}\tau - \sum_{j=1}^{N} \underbrace{\mathcal{E}\{n_j n_k^*\}}_{\frac{N_0}{2}\delta[j-k]} f_j(t) \\
&= \frac{N_0}{2} f_k(t) - \frac{N_0}{2} f_k(t) \\
&= 0
\end{aligned}$$

We observe that $\boldsymbol{r}$ and $n'(t)$ are uncorrelated. Since $\boldsymbol{r}$ and $n'(t)$ are Gaussian distributed, they are also *statistically independent*. Therefore, $n'(t)$ cannot provide any useful information that is relevant for the decision, and consequently, $\boldsymbol{r}$ forms a *sufficient statistic* for detection of $s_m(t)$.

4.1.1.2 Matched–Filter Demodulation

- Instead of generating the $\{r_k\}$ using a bank of N correlators, we may use N linear filters instead.
- We define the N filter impulse responses $h_k(t)$ as

$$h_k(t) = f_k^*(T-t), \qquad 0 \le t \le T$$

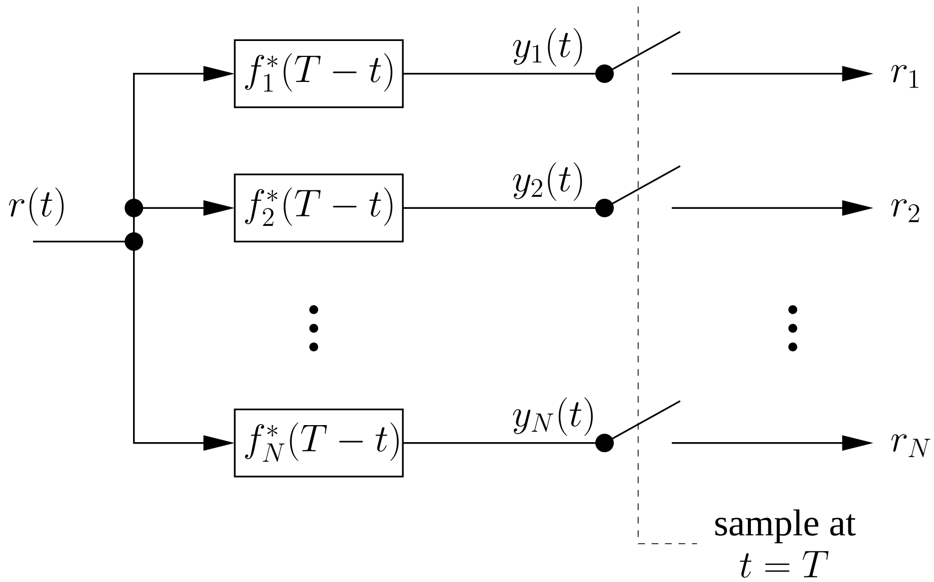
  where f_k, $1 \le k \le N$, are the N basis functions.
- The output of filter $h_k(t)$ with input $r(t)$ is

$$\begin{aligned} y_k(t) &= \int\limits_0^t r(\tau) h_k(t-\tau)\, \mathrm{d}\tau \\ &= \int\limits_0^t r(\tau) f_k^*(T-t+\tau)\, \mathrm{d}\tau \end{aligned}$$

- By sampling $y_k(t)$ at time $t = T$, we obtain

$$\begin{aligned} y_k(T) &= \int\limits_0^T r(\tau) f_k^*(\tau)\, \mathrm{d}\tau \\ &= r_k, \qquad 1 \le k \le N \end{aligned}$$

  This means the sampled output of $h_k(t)$ is r_k.



■ General Properties of Matched Filters MFs

– In general, we call a filter of the form

$$h(t) = s^*(T-t)$$

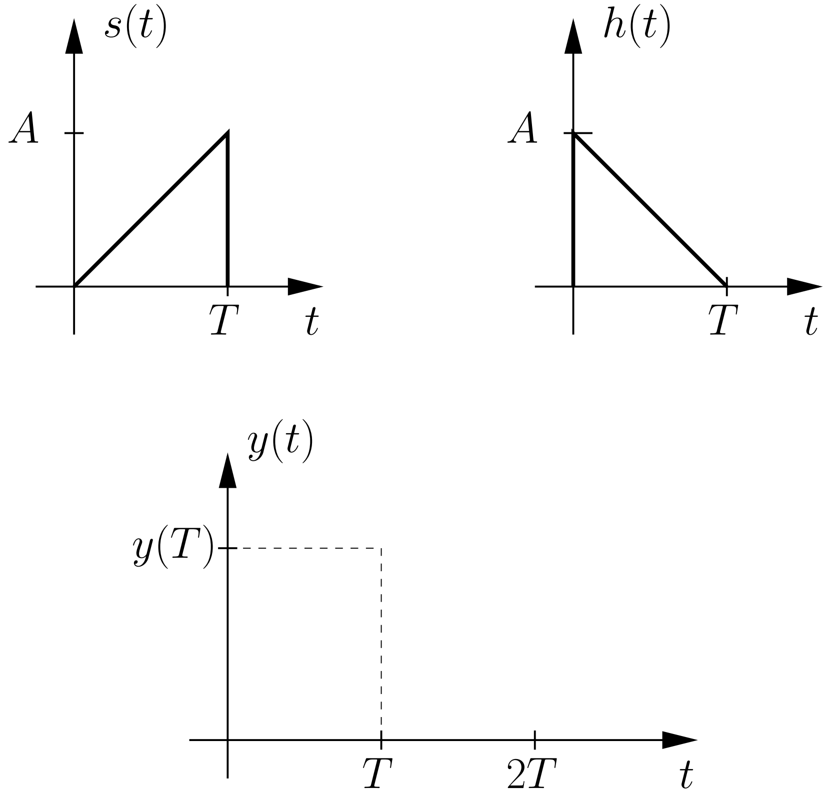
a *matched filter* for $s(t)$.

– The output

$$y(t) = \int_0^t s(\tau)s^*(T-t+\tau)\,\mathrm{d}\tau$$

is the time–shifted time–autocorrelation of $s(t)$.

Example:



– **MFs Maximize the SNR**

  * Consider the signal

$$r(t) = s(t) + n(t), \qquad 0 \le t \le T,$$

where $s(t)$ is some known signal with energy

$$E = \int_0^T |s(t)|^2 \, \mathrm{d}t$$

and $n(t)$ is AWGN with power spectral density

$$\Phi_{NN}(f) = \frac{N_0}{2}$$

∗ **Problem:** Which filter $h(t)$ maximizes the SNR of

$$y(T) = h(t) * r(t)\Big|_{t=T}$$

∗ **Answer:** The *matched filter* $h(t) = s^*(T-t)$!

Proof. The filter output sampled at time $t = T$ is given by

$$\begin{aligned} y(T) &= \int_0^T r(\tau)h(T-\tau)\,\mathrm{d}\tau \\ &= \underbrace{\int_0^T s(\tau)h(T-\tau)\,\mathrm{d}\tau}_{y_S(T)} + \underbrace{\int_0^T n(\tau)h(T-\tau)\,\mathrm{d}\tau}_{y_N(T)} \end{aligned}$$

Now, the SNR at the filter output can be defined as

$$\mathrm{SNR} = \frac{|y_S(T)|^2}{\mathcal{E}\{|y_N(T)|^2\}}$$

The noise power in the denominator can be calculated as

$$\begin{aligned} \mathcal{E}\left\{|y_N(T)|^2\right\} &= \mathcal{E}\left\{\left(\int_0^T n(\tau)h(T-\tau)\,\mathrm{d}\tau\right)\left(\int_0^T n(\tau)h(T-\tau)\,\mathrm{d}\tau\right)^*\right\} \\ &= \int_0^T\int_0^T \underbrace{\mathcal{E}\{n(\tau)n^*(t)\}}_{\frac{N_0}{2}\delta(\tau-t)} h(T-\tau)h^*(T-t)\,\mathrm{d}\tau\,\mathrm{d}t \\ &= \frac{N_0}{2}\int_0^T |h(T-\tau)|^2\,\mathrm{d}\tau \end{aligned}$$

Therefore, the SNR can be expressed as

$$\mathrm{SNR} = \frac{\left|\int\limits_0^T s(\tau)h(T-\tau)\,\mathrm{d}\tau\right|^2}{\frac{N_0}{2}\int\limits_0^T |h(T-\tau)|^2\,\mathrm{d}\tau}.$$

From the *Cauchy–Schwartz inequality* we know

$$\left|\int\limits_0^T s(\tau)h(T-\tau)\,\mathrm{d}\tau\right|^2 \le \int\limits_0^T |s(\tau)|^2\,\mathrm{d}\tau \cdot \int\limits_0^T |h(T-\tau)|^2\,\mathrm{d}\tau,$$

where equality holds if and only if

$$h(t) = Cs^*(T-t).$$

(C is an arbitrary non–zero constant). Therefore, the maximum output SNR is

$$\begin{aligned}\mathrm{SNR} &= \frac{\left|\int\limits_0^T |s(\tau)|^2\,\mathrm{d}\tau\right|^2}{\frac{N_0}{2}\int\limits_0^T |s(\tau)|^2\,\mathrm{d}\tau} \\ &= \frac{2}{N_0}\int\limits_0^T |s(\tau)|^2\,\mathrm{d}\tau \\ &= 2\frac{E}{N_0},\end{aligned}$$

which is achieved by the MF $h(t) = s^*(T-t)$. □

– **Frequency Domain Interpretation**

The frequency response of the MF is given by

$$\begin{aligned} H(f) &= \mathcal{F}\{h(t)\} \\ &= \int\limits_{-\infty}^{\infty} s^*(T-t)\, e^{-j2\pi f t}\, \mathrm{d}t \\ &= \int\limits_{-\infty}^{\infty} s^*(\tau)\, e^{j2\pi f \tau} e^{-j2\pi f T}\, \mathrm{d}\tau \\ &= e^{-j2\pi f T} \left(\int\limits_{-\infty}^{\infty} s(\tau)\, e^{-j2\pi f \tau}\, \mathrm{d}\tau \right)^* \\ &= e^{-j2\pi f T} S^*(f) \end{aligned}$$

Observe that $H(f)$ has the same magnitude as $S(f)$

$$|H(f)| = |S(f)|.$$

The factor $e^{-j2\pi f T}$ in the frequency response accounts for the time shift of $s^*(-t)$ by T.

4.1.2 Optimal Detection

Problem Formulation:

- The output $\boldsymbol{r}$ of the demodulator forms a sufficient statistic for detection of $s_m(t)$ ($\boldsymbol{s}_m$).
- We consider linear modulation formats without memory.
- *What is the optimal decision rule?*
- *Optimality criterion:* Probability for correct detection shall be maximized, i.e., probability of error shall be minimized.

Solution:

- The probability of error is minimized if we choose that $\boldsymbol{s}_{\tilde{m}}$ which maximizes the *posteriori probability*

  $$P(\boldsymbol{s}_{\tilde{m}}|\boldsymbol{r}), \qquad \tilde{m} = 1, 2, \ldots, M,$$

  where the "tilde" indicates that $\boldsymbol{s}_{\tilde{m}}$ is not the transmitted symbol but a *trial symbol*.

Maximum a Posteriori (MAP) Decision Rule

The resulting decision rule can be formulated as

$$\hat{m} = \underset{\tilde{m}}{\text{argmax}} \{P(\boldsymbol{s}_{\tilde{m}}|\boldsymbol{r})\}$$

where $\hat{m}$ denotes the *estimated* signal number. The above decision rule is called *maximum a posteriori (MAP)* decision rule.

- **Simplifications**

  Using Bayes rule, we can rewrite $P(\boldsymbol{s}_{\tilde{m}}|\boldsymbol{r})$ as

  $$P(\boldsymbol{s}_{\tilde{m}}|\boldsymbol{r}) = \frac{p(\boldsymbol{r}|\boldsymbol{s}_{\tilde{m}})P(\boldsymbol{s}_{\tilde{m}})}{p(\boldsymbol{r})},$$

  with

  – $p(\boldsymbol{r}|\boldsymbol{s}_{\tilde{m}})$: Conditional pdf of observed vector $\boldsymbol{r}$ given $\boldsymbol{s}_{\tilde{m}}$.

  – $P(\boldsymbol{s}_{\tilde{m}})$: *A priori probability* of transmitted symbols. Normally, we have

  $$P(\boldsymbol{s}_{\tilde{m}}) = \frac{1}{M}, \qquad 1 \leq \tilde{m} \leq M,$$

  i.e., all signals of the set are transmitted with equal probability.

  – $p(\boldsymbol{r})$: Probability density function of vector $\boldsymbol{r}$

  $$p(\boldsymbol{r}) = \sum_{m=1}^{M} p(\boldsymbol{r}|\boldsymbol{s}_m)P(\boldsymbol{s}_m).$$

  Since $p(\boldsymbol{r})$ is obviously independent of $\boldsymbol{s}_{\tilde{m}}$, we can simplify the MAP decision rule to

  $$\hat{m} = \underset{\tilde{m}}{\text{argmax}} \{p(\boldsymbol{r}|\boldsymbol{s}_{\tilde{m}})P(\boldsymbol{s}_{\tilde{m}})\}$$

Maximum–Likelihood (ML) Decision Rule

- The MAP rule requires knowledge of both $p(\boldsymbol{r}|\boldsymbol{s}_{\tilde{m}})$ and $P(\boldsymbol{s}_{\tilde{m}})$.
- In some applications $P(\boldsymbol{s}_{\tilde{m}})$ is unknown at the receiver.
- If we neglect the influence of $P(\boldsymbol{s}_{\tilde{m}})$, we get the ML decision rule

$$\boxed{\hat{m} = \underset{\tilde{m}}{\operatorname{argmax}} \, \{p(\boldsymbol{r}|\boldsymbol{s}_{\tilde{m}})\}}$$

- Note that if all $\boldsymbol{s}_m$ are equally probable, i.e., $P(\boldsymbol{s}_{\tilde{m}}) = 1/M$, $1 \leq \tilde{m} \leq M$, the MAP and the ML decision rules are *identical*.

The above MAP and ML decision rules are very general. They can be applied to any channel as long as we are able to find an expression for $p(\boldsymbol{r}|\boldsymbol{s}_{\tilde{m}})$.

ML Decision Rule for AWGN Channel

- For the AWGN channel we have

$$p(\boldsymbol{r}|\boldsymbol{s}_{\tilde{m}}) = \frac{1}{(\pi N_0)^{N/2}} \exp\left(-\frac{\sum_{k=1}^{N} |r_k - s_{\tilde{m}k}|^2}{N_0} \right), \quad 1 \le \tilde{m} \le M$$

- We note that the ML decision does not change if we maximize $\ln(p(\boldsymbol{r}|\boldsymbol{s}_{\tilde{m}}))$ instead of $p(\boldsymbol{r}|\boldsymbol{s}_{\tilde{m}})$ itself, since $\ln(\cdot)$ is a *monotonic* function.

- Therefore, the ML decision rule can be simplified as

$$\begin{aligned}
\hat{m} &= \underset{\tilde{m}}{\text{argmax}} \{p(\boldsymbol{r}|\boldsymbol{s}_{\tilde{m}})\} \\
&= \underset{\tilde{m}}{\text{argmax}} \{\ln(p(\boldsymbol{r}|\boldsymbol{s}_{\tilde{m}}))\} \\
&= \underset{\tilde{m}}{\text{argmax}} \left\{ -\frac{N}{2}\ln(\pi N_0) - \frac{1}{N_0}\sum_{k=1}^{N} |r_k - s_{\tilde{m}k}|^2 \right\} \\
&= \underset{\tilde{m}}{\text{argmin}} \left\{ \sum_{k=1}^{N} |r_k - s_{\tilde{m}k}|^2 \right\} \\
&= \underset{\tilde{m}}{\text{argmin}} \left\{ ||\boldsymbol{r} - \boldsymbol{s}_{\tilde{m}}||^2 \right\}
\end{aligned}$$

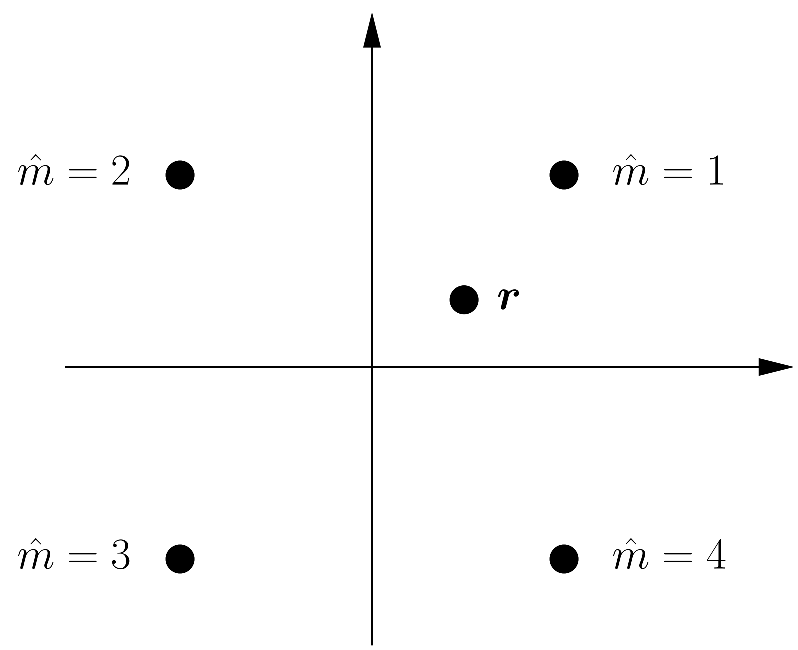
- **Interpretation:**
  We select that vector $\boldsymbol{s}_{\tilde{m}}$ which has the minimum *Euclidean distance*

$$\boxed{D(\boldsymbol{r}, \boldsymbol{s}_{\tilde{m}}) = ||\boldsymbol{r} - \boldsymbol{s}_{\tilde{m}}||}$$

from the received vector $\boldsymbol{r}$. Therefore, we can interpret the above ML decision rule graphically by dividing the signal space in *decision regions*.

Example:

4QAM



- **Alternative Representation:**
  Using the expansion

$$||\boldsymbol{r} - \boldsymbol{s}_{\tilde{m}}||^2 \;=\; ||\boldsymbol{r}||^2 - 2\mathsf{Re}\{\boldsymbol{r} \bullet \boldsymbol{s}_{\tilde{m}}\} + ||\boldsymbol{s}_{\tilde{m}}||^2,$$

we observe that $||\boldsymbol{r}||^2$ is independent of $\boldsymbol{s}_{\tilde{m}}$. Therefore, we can further simplify the ML decision rule

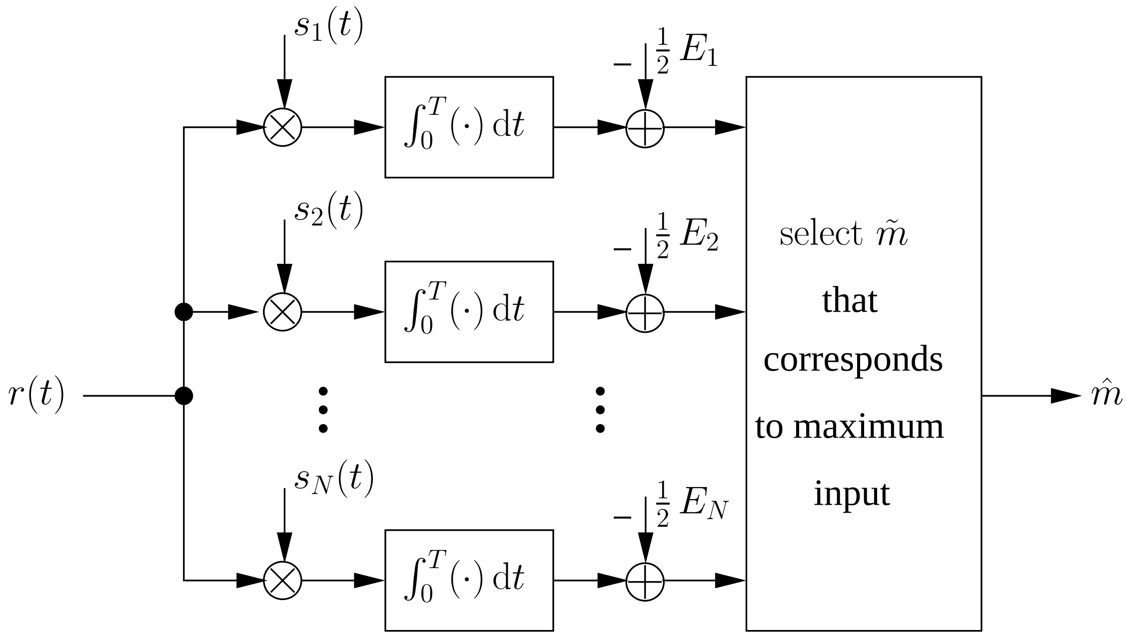
$$\begin{aligned}
\hat{m} &= \underset{\tilde{m}}{\operatorname{argmin}} \left\{||\boldsymbol{r} - \boldsymbol{s}_{\tilde{m}}||^2\right\} \\
&= \underset{\tilde{m}}{\operatorname{argmin}} \left\{-2\mathsf{Re}\{\boldsymbol{r} \bullet \boldsymbol{s}_{\tilde{m}}\} + ||\boldsymbol{s}_{\tilde{m}}||^2\right\} \\
&= \underset{\tilde{m}}{\operatorname{argmax}} \left\{2\mathsf{Re}\{\boldsymbol{r} \bullet \boldsymbol{s}_{\tilde{m}}\} - ||\boldsymbol{s}_{\tilde{m}}||^2\right\} \\
&= \underset{\tilde{m}}{\operatorname{argmax}} \left\{\mathsf{Re}\left\{\int_0^T r(t) s_{\tilde{m}}^*(t)\,\mathrm{d}t\right\} - \frac{1}{2}E_{\tilde{m}}\right\},
\end{aligned}$$

with

$$E_{\tilde{m}} = \int_0^T |s_{\tilde{m}}(t)|^2\,\mathrm{d}t.$$

If we are dealing with passband signals both $r(t)$ and $s_{\tilde{m}}^*(t)$ are real–valued, and we obtain

$$\boxed{\hat{m} = \underset{\tilde{m}}{\operatorname{argmax}} \left\{\int_0^T r(t) s_{\tilde{m}}(t)\,\mathrm{d}t - \frac{1}{2}E_{\tilde{m}}\right\}}$$



Example:

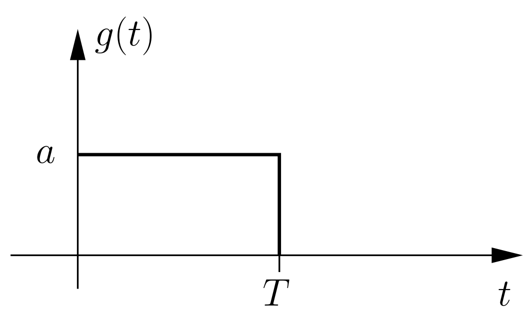
M–ary PAM transmission (baseband case)

The transmitted signals are given by

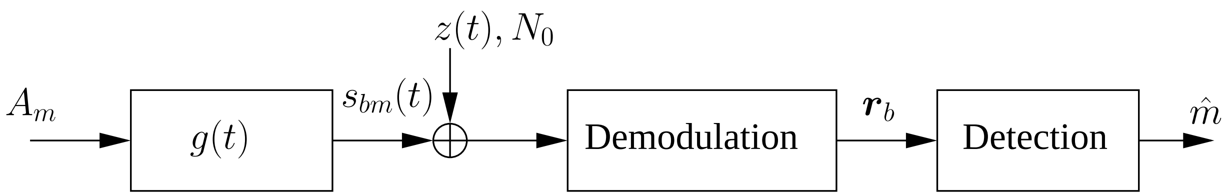
$$s_{bm}(t) = A_m g(t),$$

with $A_m = (2m - 1 - M)d$, $m = 1, 2, \ldots, M$, $2d$: distance between adjacent signal points.

We assume the transmit pulse $g(t)$ is as shown below.



In the interval $0 \leq t \leq T$, the transmission scheme is modeled as



1. **Demodulator**

   – Energy of transmit pulse

$$E_g = \int_0^T |g(t)|^2 \, \mathrm{d}t = a^2 T$$

   – Basis function $f(t)$

$$\begin{aligned} f(t) &= \frac{1}{\sqrt{E_g}} g(t) \\ &= \begin{cases} \frac{1}{\sqrt{T}}, & 0 \leq t \leq T \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

– Correlation demodulator

$$\begin{aligned}
\boldsymbol{r}_b &= \int_0^T r_b(t) f^*(t)\,\mathrm{d}t \\
&= \frac{1}{\sqrt{T}} \int_0^T r_b(t)\,\mathrm{d}t \\
&= \underbrace{\frac{1}{\sqrt{T}} \int_0^T s_{bm}(t)\,\mathrm{d}t}_{s_{bm}} + \underbrace{\frac{1}{\sqrt{T}} \int_0^T z(t)\,\mathrm{d}t}_{z} \\
&= s_{bm} + z
\end{aligned}$$

s_{bm} is given by

$$\begin{aligned}
s_{bm} &= \frac{1}{\sqrt{T}} \int_0^T A_m g(t)\,\mathrm{d}t \\
&= \frac{1}{\sqrt{T}} \int_0^T A_m a\,\mathrm{d}t \\
&= a\sqrt{T} A_m = \sqrt{E_g} A_m.
\end{aligned}$$

On the other hand, the noise variance is

$$\begin{aligned}
\sigma_z^2 &= \mathcal{E}\{|z|^2\} \\
&= \frac{1}{T} \int_0^T \int_0^T \underbrace{\mathcal{E}\{z(t) z^*(\tau)\}}_{N_0 \delta(t-\tau)}\,\mathrm{d}t\,\mathrm{d}\tau \\
&= N_0
\end{aligned}$$

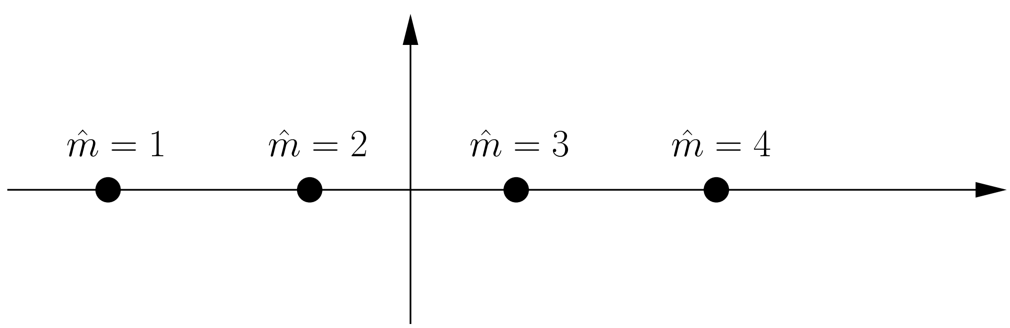
– pdf $p(r_b|A_m)$:

$$p(r_b|A_m) = \frac{1}{\pi N_0} \exp\left(-\frac{|r_b - \sqrt{E_g}A_m|^2}{N_0}\right)$$

2. **Optimum Detector**
   The ML decision rule is given by

$$\begin{aligned}\hat{m} &= \underset{\tilde{m}}{\operatorname{argmax}}\{\ln(p(r_b|A_{\tilde{m}}))\} \\ &= \underset{\tilde{m}}{\operatorname{argmax}}\{-|r_b - \sqrt{E_g}A_{\tilde{m}}|^2\} \\ &= \underset{\tilde{m}}{\operatorname{argmin}}\{|r_b - \sqrt{E_g}A_{\tilde{m}}|^2\}\end{aligned}$$

Illustration in the Signal Space



4.2 Performance of Optimum Receivers

In this section, we evaluate the performance of the optimum receivers introduced in the previous section. We assume again memoryless modulation. We adopt the *symbol error probability (SEP)* (also referred to as symbol error rate (SER)) and the *bit error probability (BEP)* (also referred to as bit error rate (BER)) as performance criteria.

4.2.1 Binary Modulation

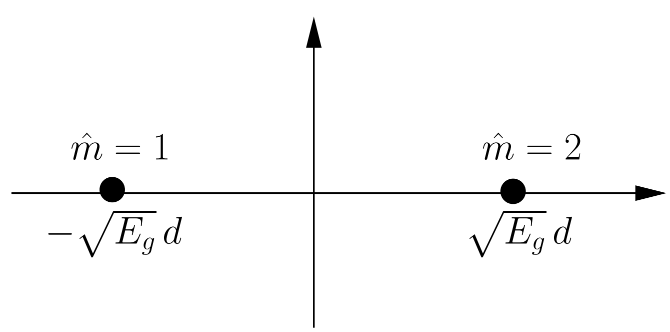
1. **Binary PAM** ($M = 2$)

   - From the example in the previous section we know that the detector input signal in this case is

$$r_b = \sqrt{E_g} A_m + z, \qquad m = 1,\, 2,$$

   where the noise variance of the complex baseband noise is $\sigma_z^2 = N_0$.

   - Decision Regions



- Assuming s_1 has been transmitted, the received signal is

$$r_b = -\sqrt{E_g} d + z$$

and a correct decision is made if

$$r_R < 0,$$

whereas an error is made if

$$r_R > 0,$$

where $r_R = \mathsf{Re}\{r_b\}$ denotes the real part of r. r_R is given by

$$r_R = -\sqrt{E_g} d + z_R$$

where $z_R = \mathsf{Re}\{z\}$ is real Gaussian noise with variance $\sigma_{z_R}^2 = N_0/2$.

- Consequently, the (conditional) error probability is

$$P(e|s_1) = \int_0^\infty p_{r_R}(r_R|s_1)\,\mathrm{d}r_R.$$

Therefore, we get

$$\begin{aligned} P(e|s_1) &= \int_0^\infty \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{(r_R - (-\sqrt{E_g} d))^2}{N_0}\right) \mathrm{d}r_R \\ &= \frac{1}{\sqrt{2\pi}} \int_{\sqrt{\frac{2E_g}{N_0}}d}^\infty \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x \\ &= Q\left(\sqrt{\frac{2E_g}{N_0}} d\right) \end{aligned}$$

where we have used the substitution $x = \sqrt{2}(r_R+\sqrt{E_g}d)/\sqrt{N_0}$ and the Q–function is defined as

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int\limits_x^\infty \exp\left(-\frac{t^2}{2}\right) \,\mathrm{d}t$$

- The BEP, which is equal to the SEP for binary modulation, is given by

$$P_b = P(s_1)P(e|s_1) + P(s_2)P(e|s_2)$$

For the usual case, $P(s_1) = P(s_2) = \frac{1}{2}$, we get

$$\begin{aligned} P_b &= \frac{1}{2}P(e|s_1) + \frac{1}{2}P(e|s_2) \\ &= P(e|s_1), \end{aligned}$$

since $P(e|s_1) = P(e|s_2)$ is true because of the symmetry of the signal constellation.

- In general, the BEP is expressed as a function of the *received energy per bit* E_b. Here, E_b is given by

$$\begin{aligned} E_b &= \mathcal{E}\{|\sqrt{E_g}A_m|^2\} \\ &= E_g\left(\frac{1}{2}(-d)^2 + \frac{1}{2}(d)^2\right) \\ &= E_g d^2. \end{aligned}$$

Therefore, the BEP can be expressed as

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

- Note that binary PSK (BPSK) yields the same BEP as 2PAM.

2. **Binary Orthogonal Modulation**

- For binary orthogonal modulation, the transmitted signals can be represented as

$$\boldsymbol{s}_1 = \begin{bmatrix} \sqrt{E_b} \\ 0 \end{bmatrix}$$
$$\boldsymbol{s}_2 = \begin{bmatrix} 0 \\ \sqrt{E_b} \end{bmatrix}$$

- The demodulated received signal is given by

$$\boldsymbol{r} = \begin{bmatrix} \sqrt{E_b} + n_1 \\ n_2 \end{bmatrix}$$

and

$$\boldsymbol{r} = \begin{bmatrix} n_1 \\ \sqrt{E_b} + n_2 \end{bmatrix}$$

if $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ were sent, respectively. The noise variances are given by

$$\sigma_{n_1}^2 = \mathcal{E}\{n_1^2\} = \sigma_{n_2}^2 = \mathcal{E}\{n_2^2\} = \frac{N_0}{2},$$

and n_1 and n_2 are mutually independent Gaussian RVs.

- **Decision Rule**

  The ML decision rule is given by

$$\begin{aligned}\hat{m} &= \underset{\tilde{m}}{\operatorname{argmax}} \left\{2\boldsymbol{r} \bullet \boldsymbol{s}_{\tilde{m}} - ||\boldsymbol{s}_{\tilde{m}}||^2\right\} \\ &= \underset{\tilde{m}}{\operatorname{argmax}} \left\{\boldsymbol{r} \bullet \boldsymbol{s}_{\tilde{m}}\right\},\end{aligned}$$

  where we have used the fact that $||\boldsymbol{s}_{\tilde{m}}||^2 = E_b$ is independent of $\tilde{m}$.

- **Error Probability**

  – Let us assume that $m = 1$ has been transmitted.

  – From the above decision rule we conclude that an error is made if

$$\boldsymbol{r} \bullet \boldsymbol{s}_1 < \boldsymbol{r} \bullet \boldsymbol{s}_2$$

  – Therefore, the conditional BEP is given by

$$\begin{aligned}P(e|\boldsymbol{s}_1) &= P(\boldsymbol{r} \bullet \boldsymbol{s}_2 > \boldsymbol{r} \bullet \boldsymbol{s}_1) \\ &= P(\sqrt{E_b}n_2 > E_b + \sqrt{E_b}n_1) \\ &= P(\underbrace{n_2 - n_1}_{X} > \sqrt{E_b})\end{aligned}$$

  Note that X is a Gaussian RV with variance

$$\begin{aligned}\sigma_X^2 &= \mathcal{E}\left\{|n_2 - n_1|^2\right\} \\ &= \mathcal{E}\left\{|n_2|^2\right\} - 2\mathcal{E}\left\{n_1 n_2\right\} + \mathcal{E}\left\{|n_1|^2\right\} \\ &= N_0\end{aligned}$$

Therefore, $P(e|\boldsymbol{s}_1)$ can be calculated to

$$\begin{aligned} P(e|\boldsymbol{s}_1) &= \frac{1}{\sqrt{2\pi N_0}} \int\limits_{\sqrt{E_b}}^{\infty} \exp\left(-\frac{x^2}{2N_0}\right) \mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \int\limits_{\sqrt{\frac{E_b}{N_0}}}^{\infty} \exp\left(-\frac{u^2}{2}\right) \mathrm{d}u \\ &= Q\left(\sqrt{\frac{E_b}{N_0}}\right) \end{aligned}$$

Finally, because of the symmetry of the signal constellation we obtain $P_b = P(e|\boldsymbol{s}_1) = P(e|\boldsymbol{s}_2)$ or

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

■ Comparison of 2PAM and Binary Orthogonal Modulation

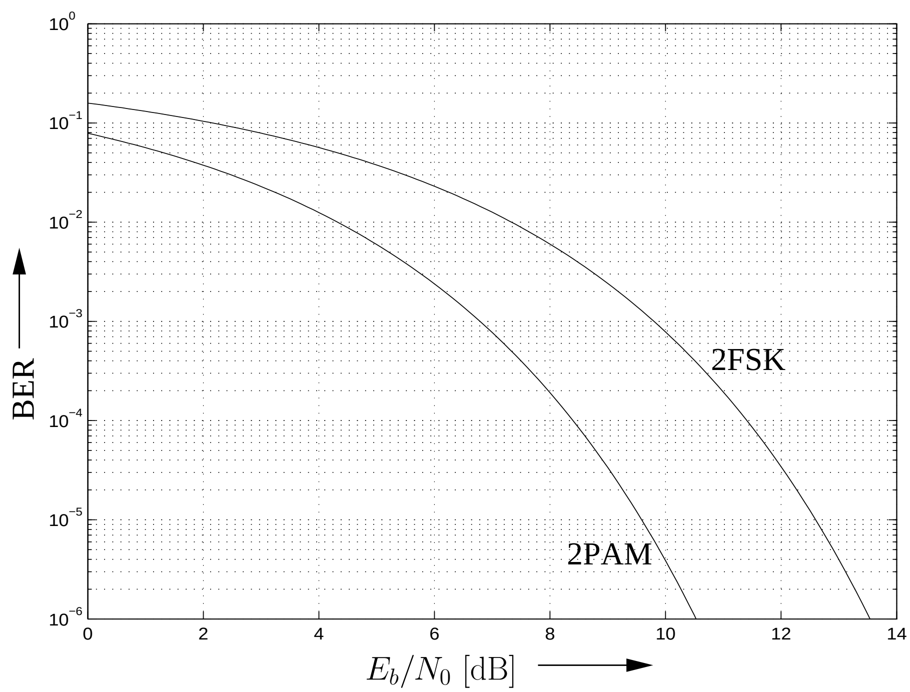
– PAM:

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

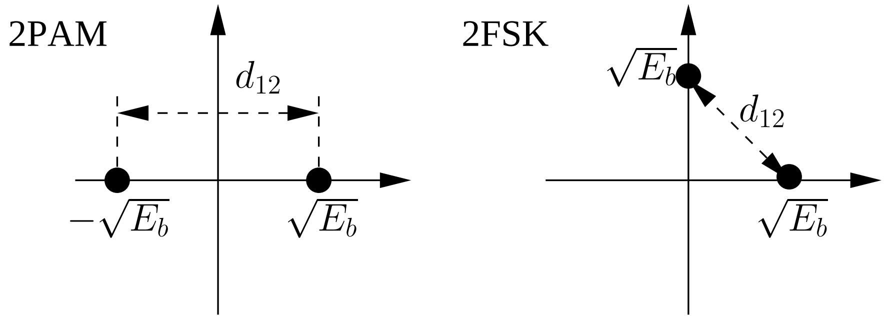
– Orthogonal Signaling (e.g. FSK)

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

– We observe that in order to achieve the same BEP the E_b–to–N_0 ratio (SNR) has to be 3 dB higher for orthogonal signaling than for PAM. Therefore, orthogonal signaling (FSK) is less *power efficient* than antipodal signaling (PAM).



– Signal Space



We observe that the (minimum) Euclidean distance between signal points is given by

$$d_{12}^{\mathrm{PAM}} = 2\sqrt{E_b}$$

and

$$d_{12}^{\mathrm{FSK}} = \sqrt{2E_b}$$

for 2PAM and 2FSK, respectively. The ratio of the squared Euclidean distances is given by

$$\left(\frac{d_{12}^{\mathrm{PAM}}}{d_{12}^{\mathrm{FSK}}}\right)^2 = 2.$$

Since the average energy of the signal points is identical for both constellations, the higher power efficiency of 2PAM can be directly deduced from the higher minimum Euclidean distance of the signal points in the signal space. Note that the BEP for both 2PAM and 2FSK can also be expressed as

$$\boxed{P_b = Q\left(\sqrt{\frac{d_{12}^2}{2N_0}}\right)}$$

– **Rule of Thumb:**
In general, for a given average energy of the signal points, the BEP of a linear modulation scheme is larger if the minimum Euclidean distance of the signals in the signal space is smaller.

4.2.2 M–ary PAM

- The transmitted signal points are given by

$$s_{bm} = \sqrt{E_g} A_m, \qquad 1 \le m \le M$$

with pulse energy E_g and amplitude

$$A_m = (2m - 1 - M)d, \qquad 1 \le m \le M.$$

- **Average Energy of Signal Points**

$$\begin{aligned}
E_S &= \frac{1}{M} \sum_{m=1}^{M} E_m \\
&= \frac{1}{M} E_g d^2 \sum_{m=1}^{M} (2m - 1 - M)^2 \\
&= \frac{E_g d^2}{M} \Bigg[4 \underbrace{\sum_{m=1}^{M} m^2}_{\frac{1}{6}M(M+1)(2M+1)} - 4(M+1) \underbrace{\sum_{m=1}^{M} m}_{\frac{1}{2}M(M+1)} + \underbrace{\sum_{m=1}^{M} (M+1)^2}_{M(M+1)^2} \Bigg] \\
&= \frac{M^2 - 1}{3} d^2 E_g
\end{aligned}$$

■ **Received Baseband Signal**

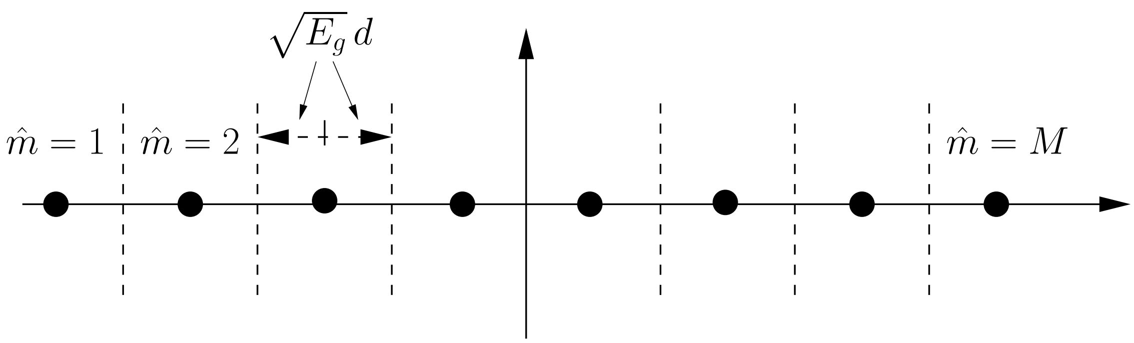
$$r_b = s_{bm} + z,$$

with $\sigma_z^2 = \mathcal{E}\{|z|^2\} = N_0$. Again only the real part of the received signal is relevant for detection and we get

$$r_R = s_{bm} + z_R,$$

with noise variance $\sigma_{z_R}^2 = \mathcal{E}\{z_R^2\} = N_0/2$.

■ **Decision Regions for ML Detection**



■ We observe that there are two different types of signal points:

1. Outer Signal Points
   We refer to the signal points with $\hat{m} = 1$ and $\hat{m} = M$ as *outer signal points* since they have only one neighboring signal point. In this case, we make on average 1/2 symbol errors if

$$|r_R - s_{bm}| > d\sqrt{E_g}$$

2. Inner Signal Points
   Signal points with $2 \leq \hat{m} \leq M - 1$ are referred to as *inner signal points* since they have two neighbors. Here, we make on average 1 symbol error if $|r_R - s_{bm}| > d\sqrt{E_g}$.

- **Symbol Error Probability (SEP)**
  The SEP can be calculated to

$$
\begin{aligned}
P_M &= \frac{1}{M}\left[(M-2)+\frac{1}{2}\cdot 2\right] P\left(|r_R - s_{bm}| > d\sqrt{E_g}\right) \\
&= \frac{M-1}{M} P\left([r_R - s_{bm} > d\sqrt{E_g}] \vee [r_R - s_{bm} < -d\sqrt{E_g}]\right) \\
&= \frac{M-1}{M}\left(P\left(r_R - s_{bm} > d\sqrt{E_g}\right) + P\left(r_R - s_{bm} < -d\sqrt{E_g}\right)\right) \\
&= \frac{M-1}{M} 2P\left(r_R - s_{bm} > d\sqrt{E_g}\right) \\
&= 2\,\frac{M-1}{M}\,\frac{1}{\sqrt{\pi N_0}} \int\limits_{d\sqrt{E_g}}^{\infty} \exp\left(-\frac{x^2}{N_0}\right)\, \mathrm{d}x \\
&= 2\,\frac{M-1}{M}\,\frac{1}{\sqrt{2\pi}} \int\limits_{\sqrt{2d^2\frac{E_g}{N_0}}}^{\infty} \exp\left(-\frac{y^2}{2}\right)\, \mathrm{d}y \\
&= 2\,\frac{M-1}{M}\, Q\left(\sqrt{2d^2\frac{E_g}{N_0}}\right)
\end{aligned}
$$

- Using the identity

$$d^2 E_g = 3\frac{E_S}{M^2-1},$$

  we obtain

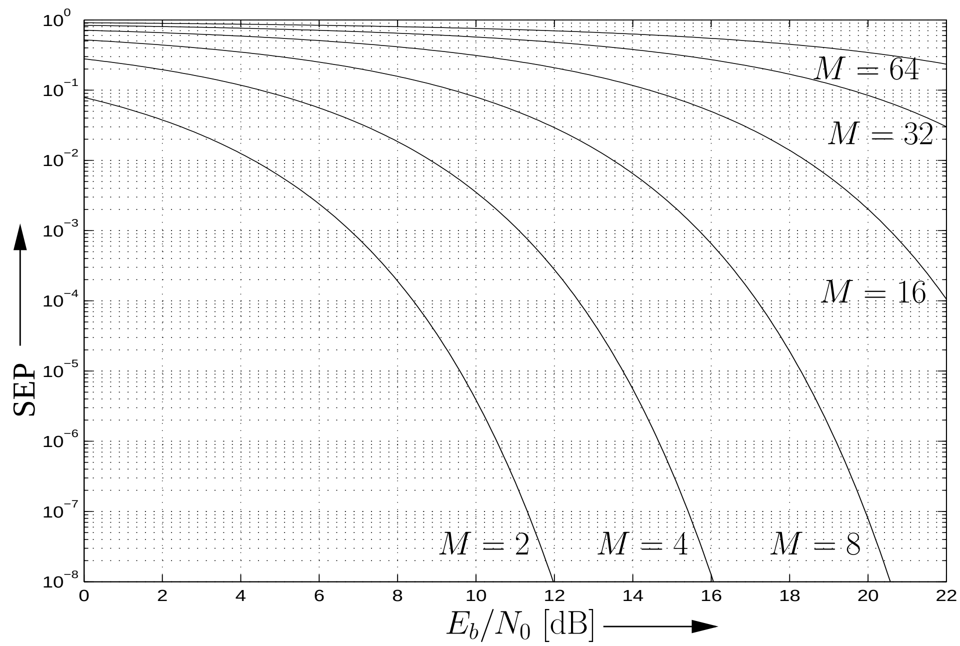
$$\boxed{P_M = 2\,\frac{M-1}{M}\, Q\left(\sqrt{\frac{6E_S}{(M^2-1)N_0}}\right).}$$

We make the following observations

1. For constant E_S the error probability increases with increasing M.
2. For a given SEP the required E_S/N_0 increases as
$$10\log_{10}(M^2-1) \approx 20\log_{10} M.$$
This means if we double the number of signal points, i.e., $M = 2^k$ is increased to $M = 2^{k+1}$, the required E_S/N_0 increases (approximately) as
$$20\log_{10}\left(2^{k+1}/2^k\right) = 20\log_{10} 2 \approx 6\,\text{dB}.$$

- Alternatively, we may express P_M as a function of the average energy per bit E_b, which is given by
$$E_b = \frac{E_S}{k} = \frac{E_S}{\log_2 M}$$
Therefore, the resulting expression for P_M is
$$\boxed{P_M = 2\,\frac{M-1}{M}\,Q\left(\sqrt{\frac{6\log_2(M)E_b}{(M^2-1)N_0}}\right).}$$

- An exact expression for the *bit error probability (BEP)* is more difficult to derive than the expression for the SEP. However, for high E_S/N_0 ratios most errors only involve neighboring signal points. Therefore, if we use *Gray labeling* we make approximately one bit error per symbol error. Since there are $\log_2 M$ bits per symbol, the PEP can be approximated by
$$\boxed{P_b \approx \frac{1}{\log_2 M}\,P_M.}$$



4.2.3 M–ary PSK

- For 2PSK the same SEP as for 2PAM results.

$$P_2 = Q\left(\sqrt{\frac{2E_b}{N_0}}\right).$$

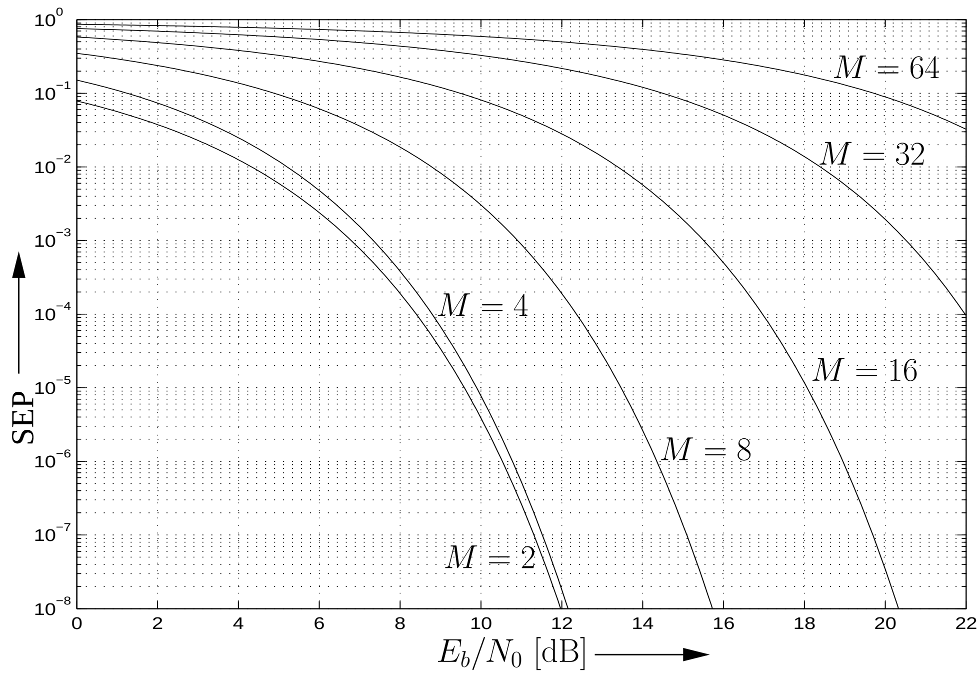
- For 4PSK the SEP is given by

$$P_4 = 2\,Q\left(\sqrt{\frac{2E_b}{N_0}}\right)\left[1 - \frac{1}{2}Q\left(\sqrt{\frac{2E_b}{N_0}}\right)\right].$$

- For optimum detection of M–ary PSK the SEP can be tightly approximated as

$$P_M \approx 2\,Q\left(\sqrt{\frac{2\log_2(M)E_b}{N_0}}\sin\frac{\pi}{M}\right).$$

- The approximate SEP is illustrated below for several values of M. For $M = 2$ and $M = 4$ the exact SEP is shown.

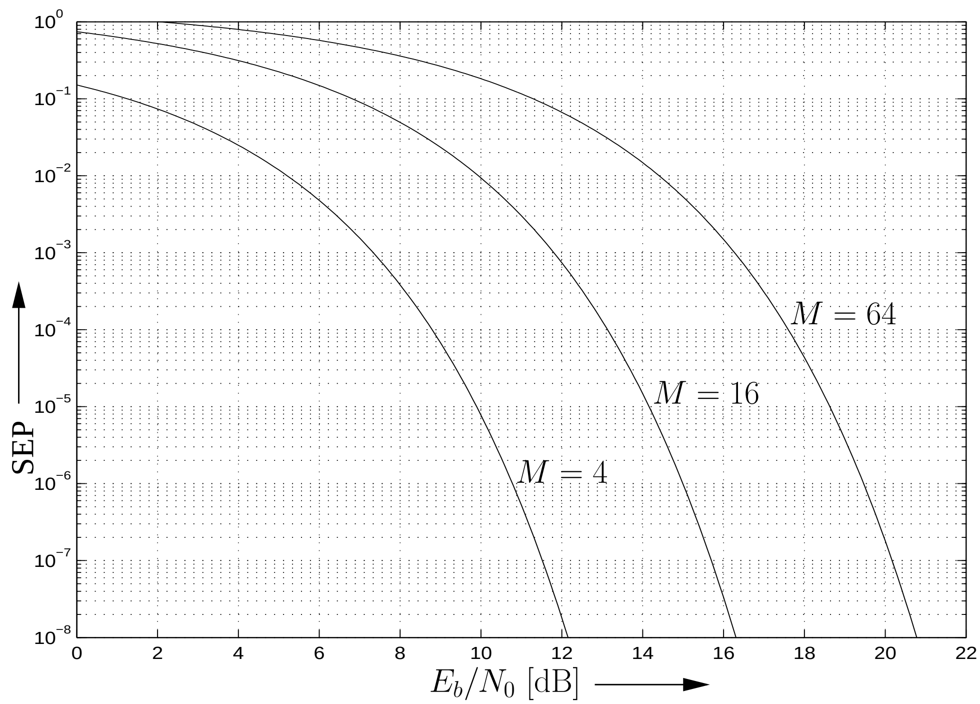


4.2.4 M–ary QAM

- For $M = 4$ the SEP of QAM is identical to that of PSK.
- In general, the SEP can be tightly upper bounded by

$$P_M \leq 4Q\left(\sqrt{\frac{3\log_2(M)E_b}{(M-1)N_0}}\right).$$

- The bound on SEP is shown below. For $M = 4$ the exact SEP is shown.



4.2.5 Upper Bound for Arbitrary Linear Modulation Schemes

Although exact (and complicated) expressions for the SEP and BEP of most regular linear modulation formats exist, it is sometimes more convenient to employ simple bounds and approximation. In this sections, we derive the union upper bound valid for arbitrary signal constellations and a related approximation for the SEP.

- We consider an M–ary modulation scheme with M signal points $\boldsymbol{s}_m$, $1 \leq m \leq M$, in the signal space.
- We denote the *pairwise* error probability of two signal points $\boldsymbol{s}_\mu$ and $\boldsymbol{s}_\nu$, $\mu \neq \nu$ by

$$\text{PEP}(\boldsymbol{s}_\mu \rightarrow \boldsymbol{s}_\nu) = P(\boldsymbol{s}_\mu \,\text{transmitted},\, \boldsymbol{s}_\nu,\, \text{detected})$$

- The union bound for the SEP can be expressed as

$$P_M \leq \frac{1}{M} \sum_{\mu=1}^{M} \sum_{\substack{\nu=1 \\ \nu \neq \mu}}^{M} \mathrm{PEP}(\boldsymbol{s}_\mu \to \boldsymbol{s}_\nu)$$

which is an upper bound since some regions of the signal space may be included in multiple PEPs.

- The advantage of the union bound is that the PEP can be usually easily obtained. Assuming Gaussian noise and an Euclidean distance of $d_{\mu\nu} = ||\boldsymbol{s}_\mu - \boldsymbol{s}_\nu||$ between signal points $\boldsymbol{s}_\mu$ and $\boldsymbol{s}_\nu$, we obtain

$$P_M \leq \frac{1}{M} \sum_{\mu=1}^{M} \sum_{\substack{\nu=1 \\ \nu \neq \mu}}^{M} Q\left(\sqrt{\frac{d_{\mu\nu}^2}{2N_0}}\right)$$

- Assuming that each signal point has on average C_M nearest neighbor signal points with minimum distance $d_{\min} = \min_{\mu \neq \nu}\{d_{\mu\nu}\}$ and exploiting the fact that for high SNR the minimum distance terms will dominate the union bound, we obtain the approximation

$$P_M \approx C_M Q\left(\sqrt{\frac{d_{\min}^2}{2N_0}}\right)$$

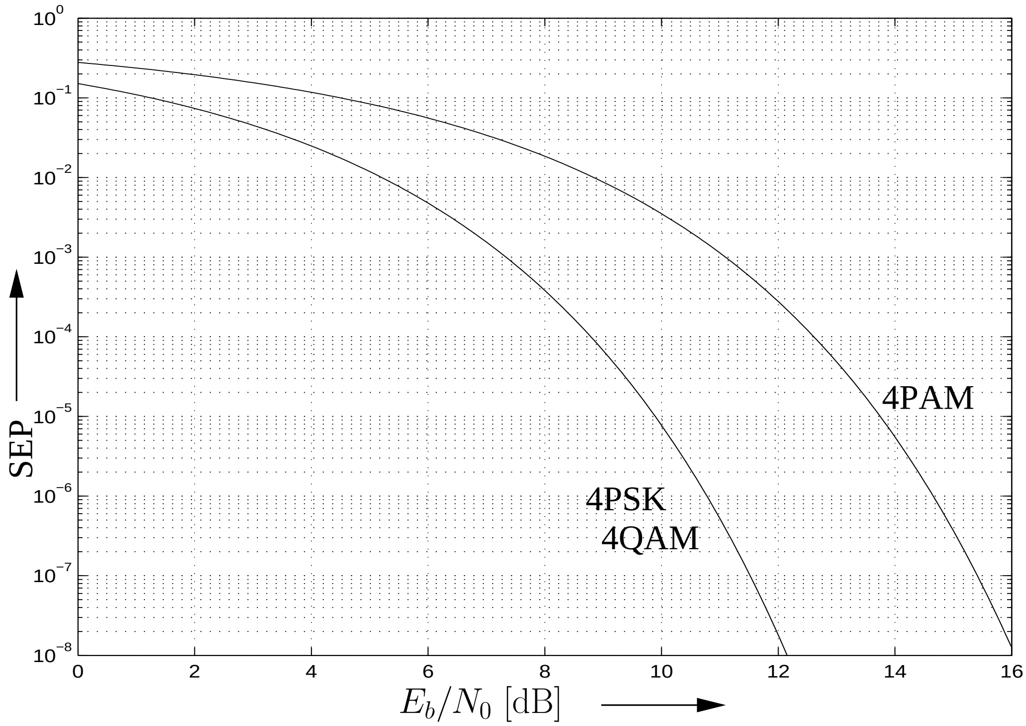
Note: The SEP approximation given above for MPSK can be obtained with this approximation.

- For Gray labeling, approximations for the BEP are obtained from the above equations with $P_b \approx P_M / \log_2(M)$

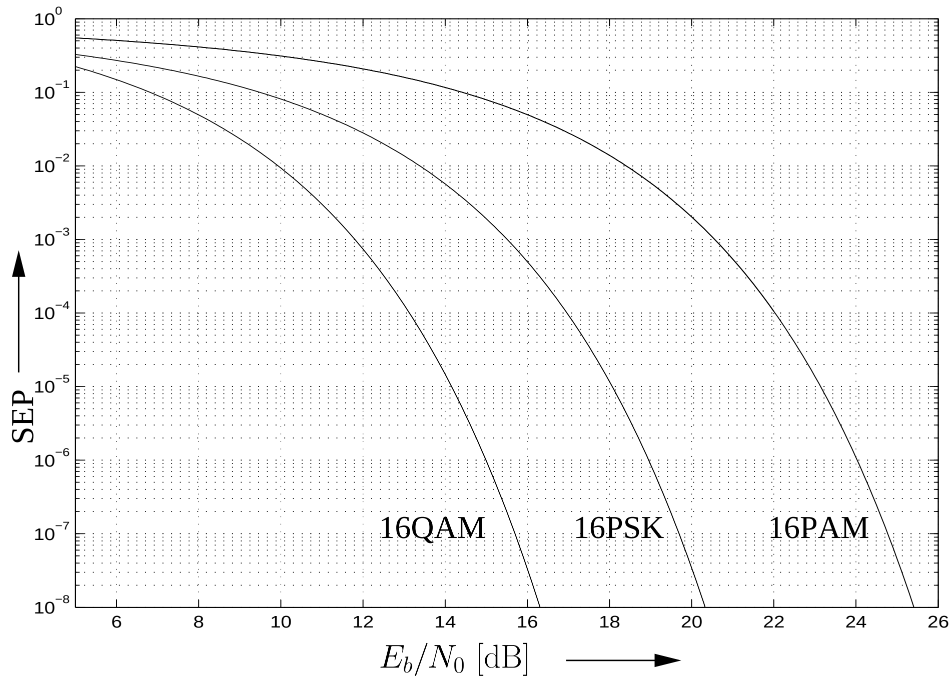
4.2.6 Comparison of Different Linear Modulations

We compare PAM, PSK, and QAM for different M.

- $M = 4$

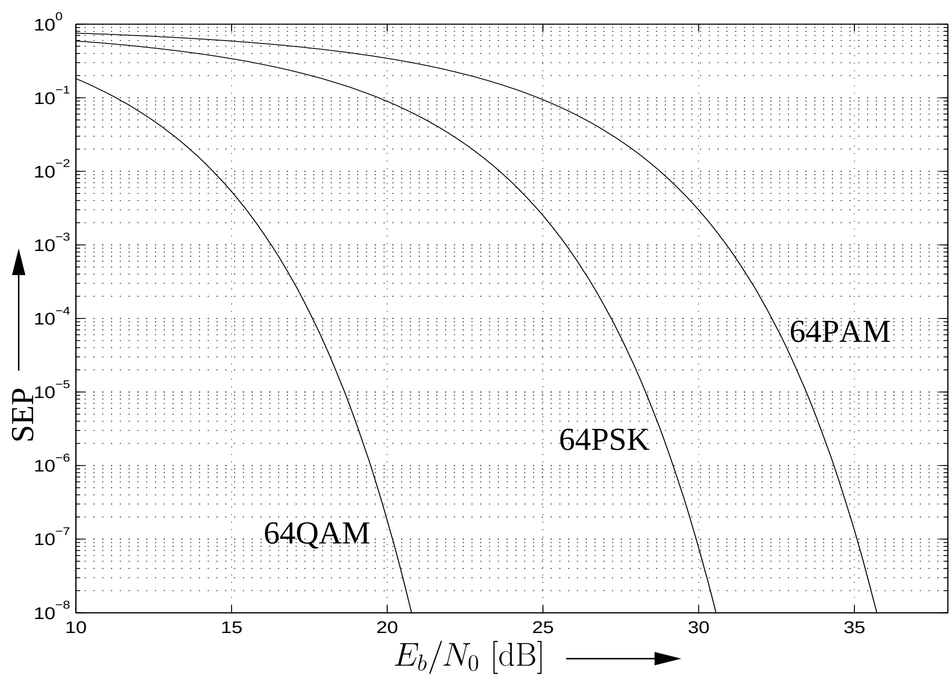


- $M = 16$



10^0
10^-1
10^-2
10^-3
10^-4
10^-5
10^-6
10^-7
10^-8
SEP
6
8
10
12
14
16
18
20
22
24
26
16QAM
16PSK
16PAM
E_b/N_0 [dB]

- $M = 64$



10^0
10^-1
10^-2
10^-3
10^-4
10^-5
10^-6
10^-7
10^-8
SEP
10
15
20
25
30
35
64PAM
64PSK
64QAM
E_b/N_0 [dB]

- Obviously, as M increases PAM and PSK become less favorable and the gap to QAM increases. The reason for this behavior is the smaller *minimum Euclidean distance* $d_{\min}$ of PAM and PSK. For a given transmit energy $d_{\min}$ of PAM and PSK is smaller since the signal points are confined to a line and a circle, respectively. For QAM on the other hand, the signal points are on a rectangular grid, which guarantees a comparatively large $d_{\min}$.

4.3 Receivers for Signals with Random Phase in AWGN

4.3.1 Channel Model

- **Passband Signal**

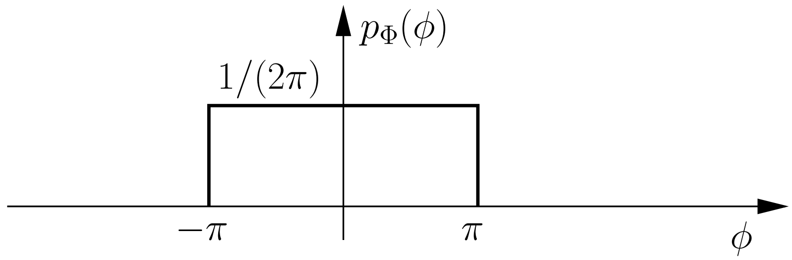
  We assume that the received passband signal can be modeled as

$$r(t) = \sqrt{2}\mathsf{Re}\left\{ \left(e^{j\phi} s_{bm}(t) + z(t) \right) e^{j2\pi f_c t} \right\},$$

  where $z(t)$ is complex AWGN with power spectral density $\Phi_{zz}(f) = N_0$, and ϕ is an unknown, random but constant phase.

  - ϕ may originate from the local oscillator or the transmission channel.
  - ϕ is often modeled as uniformly distributed, i.e.,

$$p_\Phi(\phi) = \begin{cases} \frac{1}{2\pi}, & -\pi \le \phi < \pi \\ 0, & \text{otherwise} \end{cases}$$



- **Baseband Signal**

  The received baseband signal is given by

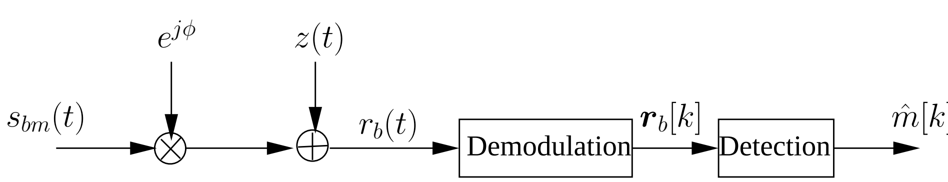
$$r_b(t) = e^{j\phi} s_{bm}(t) + z(t).$$

- **Optimal Demodulation**

  Since the unknown phase results just in the multiplication of the transmitted baseband waveform $s_{bm}(t)$ by a constant factor $e^{j\phi}$, demodulators that are optimum for $\phi = 0$ are also optimum for $\phi \neq 0$. Therefore, both correlation demodulation and matched–filter demodulation are also optimum if the channel phase is unknown. The demodulated signal in interval $kT \leq t \leq (k+1)T$ can be written as

$$\boldsymbol{r}_b[k] = e^{j\phi} \boldsymbol{s}_{bm}[k] + \boldsymbol{z}[k],$$

  where the components of the AWGN vector $\boldsymbol{z}[k]$ are mutually independent, zero–mean complex Gaussian processes with variance $\sigma_z^2 = N_0$.

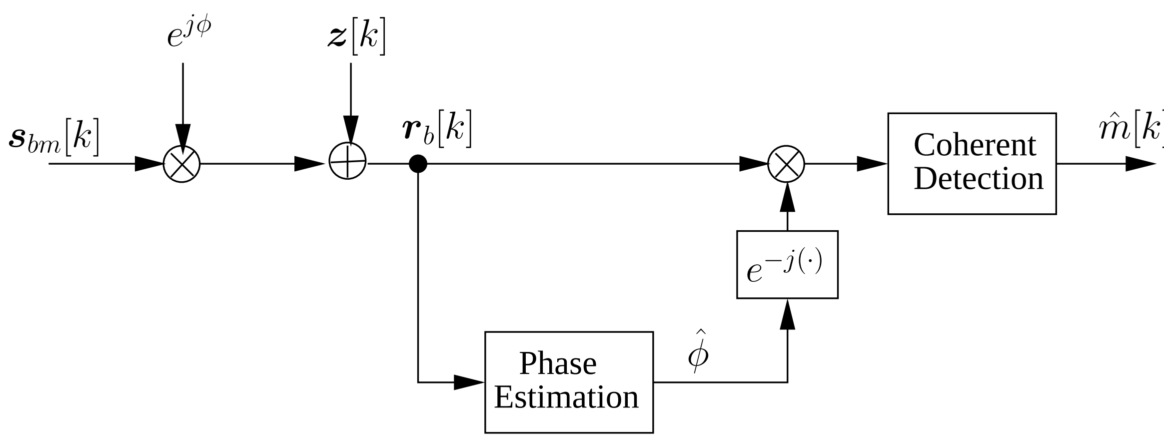


4.3.2 Noncoherent Detectors

In general, we distinguish between *coherent* and *noncoherent* detection.

1. **Coherent Detection**

- Coherent detectors first estimate the unknown phase ϕ using e.g. known pilot symbols introduced into the transmitted signal stream.
- For detection it is assumed that the phase estimate $\hat{\phi}$ is *perfect*, i.e., $\hat{\phi} = \phi$, and the same detectors as for the pure AWGN channel can be used.



- The performance of ideal coherent detection with $\hat{\phi} = \phi$ constitutes an *upper* bound for any realizable non–ideal coherent or noncoherent detection scheme.
- *Disadvantages*
  - In practice, ideal coherent detection is not possible and the *ad hoc* separation of phase estimation and detection is *suboptimum*.

- Phase estimation is often complicated and may require pilot symbols.

2. **Noncoherent Detection**

- In noncoherent detectors no attempt is made to explicitly estimate the phase ϕ.
- *Advantages*
  - For many modulation schemes simple noncoherent receiver structures exist.
  - More complex *optimal* noncoherent receivers can be derived.

4.3.2.1 A Simple Noncoherent Detector for PSK with Differential Encoding (DPSK)

- As an example for a simple, suboptimum noncoherent detector we derive the so–called *differential detector* for DPSK.
- **Transmit Signal**
  The transmitted complex baseband waveform in the interval $kT \leq t \leq (k+1)T$ is given by

$$s_{bm}(t) = b[k]g(t - kT),$$

where $g(t)$ denotes the transmit pulse of length T and $b[k]$ is the transmitted PSK signal which is given by

$$b[k] = e^{j\Theta[k]}.$$

The PSK symbols are generated from the differential PSK (DPSK) symbols $a[k]$ as

$$b[k] = a[k]b[k-1],$$

where $a[k]$ is given by

$$a[k] = e^{j\Delta\Theta[k]}, \qquad \Delta\Theta[k] = 2\pi(m-1)/M,\, m \in \{1, 2, \ldots, M\}.$$

For simplicity, we have dropped the symbol index m in $\Delta\Theta[k]$ and $a[k]$. Note that the *absolute phase* $\Theta[k]$ is related to the *differential phase* $\Delta\Theta[k]$ by

$$\Theta[k] = \Theta[k-1] + \Delta\Theta[k].$$

- For demodulation, we use the basis function

$$f(t) = \frac{1}{\sqrt{E_g}}\, g(t)$$

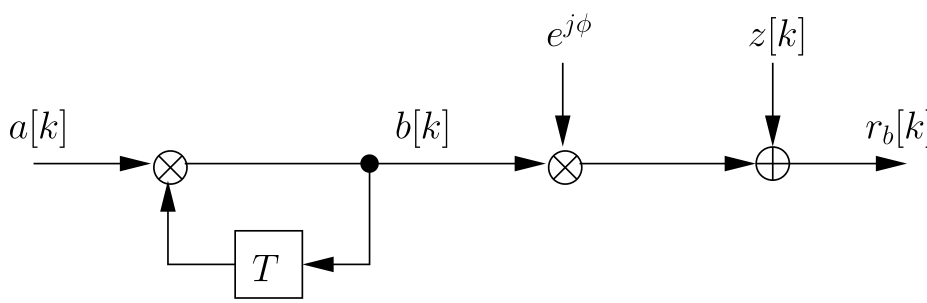
- The demodulated signal in the kth interval can be represented as

$$r'_b[k] = e^{j\phi}\sqrt{E_g}\, b[k] + z'[k],$$

where $z'[k]$ is an AWGN process with variance $\sigma^2_{z'} = N_0$. It is convenient to define the new signal

$$\begin{aligned} r_b[k] &= \frac{1}{\sqrt{E_g}}\, r'_b[k] \\ &= e^{j\phi}\, b[k] + z[k], \end{aligned}$$

where $z[k]$ has variance $\sigma^2_z = N_0/E_g$.



■ Differential Detection (DD)

– The received signal in the kth and the $(k-1)$st symbol intervals are given by

$$r_b[k] = e^{j\phi}a[k]b[k-1] + z[k]$$

and

$$r_b[k-1] = e^{j\phi}b[k-1] + z[k-1],$$

respectively.

– If we assume $z[k] \approx 0$ and $z[k-1] \approx 0$, the variable

$$d[k] = r_b[k]r_b^*[k-1]$$

can be simplified to

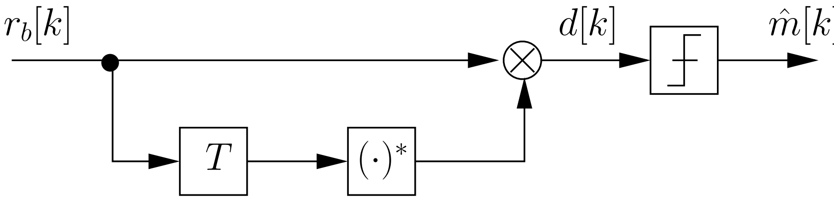
$$\begin{aligned} d[k] &= r_b[k]r_b^*[k-1] \\ &\approx e^{j\phi}a[k]b[k-1](e^{j\phi}a[k]b[k-1])^* \\ &= a[k]\,|b[k]|^2 \\ &= a[k]. \end{aligned}$$

This means $d[k]$ is independent of the phase ϕ and is suitable for detecting $a[k]$.

– A detector based on $d[k]$ is referred to as *(conventional) differential detector*. The resulting decision rule is

$$\hat{m}[k] = \underset{\tilde{m}}{\text{argmin}} \left\{ |d[k] - a_{\tilde{m}}[k]|^2 \right\},$$

where $a_{\tilde{m}}[k] = e^{j2\pi(\tilde{m}-1)/M}$, $1 \leq \tilde{m} \leq M$. Alternatively, we may base our decision on the location of $d[k]$ in the signal space, as usual.



■ Comparison with Coherent Detection

– *Coherent Detection (CD)*

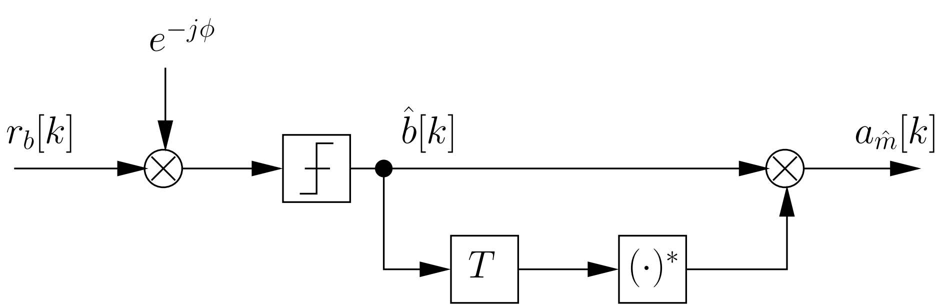
An upper bound on the performance of DD can be obtained by CD with ideal knowledge of ϕ. In that case, we can use the decision variable $r_b[k]$ and directly make a decision on the absolute phase symbols $b[k] = e^{j\Theta[k]} \in \{e^{j2\pi(m-1)/M} | m = 1, 2, \ldots, M\}$.

$$\hat{b}[k] = \underset{\tilde{b}[k]}{\text{argmin}} \left\{ |\tilde{b}[k] - e^{-j\phi} r_b[k]|^2 \right\},$$

and obtain an estimate for the differential symbol from

$$a_{\hat{m}}[k] = \hat{b}[k] \cdot \hat{b}^*[k-1].$$

$\hat{b}[k]$ and $\tilde{b}[k] = e^{j\tilde{\Theta}[k]} \in \{e^{j2\pi(m-1)/M} | m = 1, 2, \ldots, M\}$ denote the estimated transmitted symbol and a trial symbol, respectively. The above decision rule is identical to that for PSK except for the inversion of the differential encoding operation.



Since two successive absolute symbols $\hat{b}[k]$ are necessary for estimation of the differential symbol $a_{\hat{m}}[k]$ (or equivalently $\hat{m}$), isolated single errors in the absolute phase symbols will lead to *two* symbol errors in the differential phase symbols, i.e., if all absolute phase symbols but that at time k_0 are correct, then all differential phase symbols but the ones at times k_0 and $k_0 + 1$ will be correct. Since at high SNRs single errors in the absolute phase symbols dominate, the SEP $\mathrm{SEP}_{\mathrm{DPSK}}^{\mathrm{CD}}$ of DPSK with CD is approximately by a factor of two higher than that of PSK with CD, which we refer to as $\mathrm{SEP}_{\mathrm{PSK}}^{\mathrm{CD}}$.

$$\mathrm{SEP}_{\mathrm{DPSK}}^{\mathrm{CD}} \approx 2\mathrm{SEP}_{\mathrm{PSK}}^{\mathrm{CD}}$$

Note that at high SNRs this factor of two difference in SEP corresponds to a negligible difference in required E_b/N_0 ratio to achieve a certain BER, since the SEP decays approximately exponentially.

– *Differential Detection (DD)*
The decision variable $d[k]$ can be rewritten as

$$\begin{aligned} d[k] &= r_b[k] r_b^*[k-1] \\ &= (e^{j\phi} a[k] b[k-1] + z[k])(e^{j\phi} b[k-1] + z[k-1])^* \\ &= a[k] + \underbrace{e^{-j\phi} b^*[k-1] z[k] + e^{j\phi} b[k] z^*[k-1] + z[k] z^*[k-1]}_{z_{\mathrm{eff}}[k]}, \end{aligned}$$

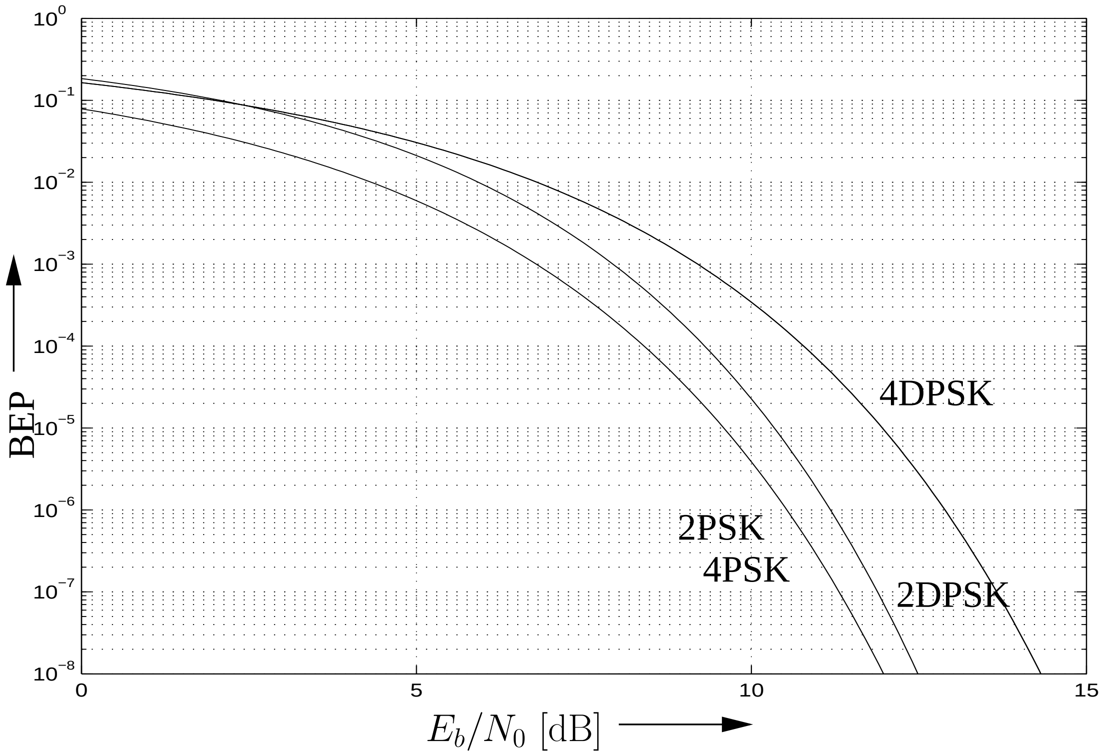
where $z_{\mathrm{eff}}[k]$ denotes the *effective noise* in the decision variable $d[k]$. It can be shown that $z_{\mathrm{eff}}[k]$ is a white process with variance

$$\sigma^2_{z_{\text{eff}}} = 2\sigma_z^2 + \sigma_z^4$$
$$= 2\frac{N_0}{E_g} + \left(\frac{N_0}{E_g}\right)^2$$

For high SNRs we can approximate $\sigma^2_{z_{\text{eff}}}$ as

$$\sigma^2_{z_{\text{eff}}} \approx 2\sigma_z^2$$
$$= 2\frac{N_0}{E_g}.$$

– *Comparison*
We observe that the variance $\sigma^2_{z_{\text{eff}}}$ of the effective noise in the decision variable for DD is twice as high as that for CD. However, for small M the distribution of $z_{\text{eff}}[k]$ is significantly different from a Gaussian distribution. Therefore, for small M a direct comparison of DD and CD is difficult and requires a detailed BEP or SEP analysis. On the other hand, for large $M \geq 8$ the distribution of $z_{\text{eff}}[k]$ can be well approximated as Gaussian. Therefore, we expect that at high SNRs DPSK with DD requires approximately a 3 dB higher E_b/N_0 ratio to achieve the same BEP as CD. For $M = 2$ and $M = 4$ this difference is smaller. At a BEP of 10^{-5} the loss of DD compared to CD is only about 0.8 dB and 2.3 dB for $M = 2$ and $M = 4$, respectively.



4.3.2.2 Optimum Noncoherent Detection

- The demodulated received baseband signal is given by

$$\boldsymbol{r}_b = e^{j\phi}\, \boldsymbol{s}_{bm} + \boldsymbol{z}.$$

- We assume that all possible symbols $\boldsymbol{s}_{bm}$ are transmitted with equal probability, i.e., ML detection is optimum.

- We already know that the ML decision rule is given by

$$\hat{m} = \underset{\tilde{m}}{\operatorname{argmax}} \{p(\boldsymbol{r}|\boldsymbol{s}_{b\tilde{m}})\}.$$

  Thus, we only have to find an analytic expression for $p(\boldsymbol{r}|\boldsymbol{s}_{b\tilde{m}})$. The problem we encounter here is that, in contrast to coherent detection, there is an additional random variable, namely the unknown phase ϕ, involved.

- We know the pdf of $\boldsymbol{r}_b$ conditioned on both ϕ and $\boldsymbol{s}_{b\tilde{m}}$. It is given by

$$p(\boldsymbol{r}_b|\boldsymbol{s}_{b\tilde{m}},\,\phi) = \frac{1}{(\pi N_0)^N}\exp\left(-\frac{||\boldsymbol{r}_b - e^{j\phi}\,\boldsymbol{s}_{b\tilde{m}}||^2}{N_0}\right)$$

- We obtain $p(\boldsymbol{r}_b|\boldsymbol{s}_{b\tilde{m}})$ from $p(\boldsymbol{r}_b|\boldsymbol{s}_{b\tilde{m}},\,\phi)$ as

$$p(\boldsymbol{r}_b|\boldsymbol{s}_{b\tilde{m}}) = \int\limits_{-\infty}^{\infty} p(\boldsymbol{r}_b|\boldsymbol{s}_{b\tilde{m}},\,\phi)\,p_\Phi(\phi)\,\mathrm{d}\phi.$$

Since we assume for the distribution of ϕ

$$p_\Phi(\phi) = \begin{cases} \frac{1}{2\pi}, & -\pi \le \phi < \pi \\ 0, & \text{otherwise} \end{cases}$$

we get

$$\begin{aligned}
p(\boldsymbol{r}_b|\boldsymbol{s}_{b\tilde{m}}) &= \frac{1}{2\pi}\int\limits_{-\pi}^{\pi} p(\boldsymbol{r}_b|\boldsymbol{s}_{b\tilde{m}},\,\phi)\,\mathrm{d}\phi \\
&= \frac{1}{(\pi N_0)^N}\,\frac{1}{2\pi}\int\limits_{-\pi}^{\pi}\exp\left(-\frac{||\boldsymbol{r}_b - e^{j\phi}\,\boldsymbol{s}_{b\tilde{m}}||^2}{N_0}\right)\mathrm{d}\phi \\
&= \frac{1}{(\pi N_0)^N}\,\frac{1}{2\pi}\int\limits_{-\pi}^{\pi}\exp\left(-\frac{||\boldsymbol{r}_b||^2 + ||\boldsymbol{s}_{b\tilde{m}}||^2 - 2\mathsf{Re}\{\boldsymbol{r}_b \bullet e^{j\phi}\boldsymbol{s}_{b\tilde{m}}\}}{N_0}\right)\mathrm{d}\phi
\end{aligned}$$

Now, we use the relations

$$\begin{aligned}
\mathsf{Re}\{\boldsymbol{r}_b \bullet e^{j\phi}\boldsymbol{s}_{b\tilde{m}}\} &= \mathsf{Re}\{|\boldsymbol{r}_b \bullet \boldsymbol{s}_{b\tilde{m}}|\,e^{j\alpha}e^{j\phi}\} \\
&= |\boldsymbol{r}_b \bullet \boldsymbol{s}_{b\tilde{m}}|\mathsf{cos}(\phi+\alpha),
\end{aligned}$$

where α is the argument of $\boldsymbol{r}_b \bullet \boldsymbol{s}_{b\tilde{m}}$. With this simplification $p(\boldsymbol{r}_b|\boldsymbol{s}_{b\tilde{m}})$ can be rewritten as

$$\begin{aligned} p(\boldsymbol{r}_b|\boldsymbol{s}_{b\tilde{m}}) &= \frac{1}{(\pi N_0)^N} \exp\left(-\frac{||\boldsymbol{r}_b||^2 + ||\boldsymbol{s}_{b\tilde{m}}||^2}{N_0}\right) \\ &\quad \cdot \frac{1}{2\pi} \int\limits_{-\pi}^{\pi} \exp\left(\frac{2}{N_0}|\boldsymbol{r}_b \bullet \boldsymbol{s}_{b\tilde{m}}| \cos(\phi + \alpha)\right) \, \mathrm{d}\phi \end{aligned}$$

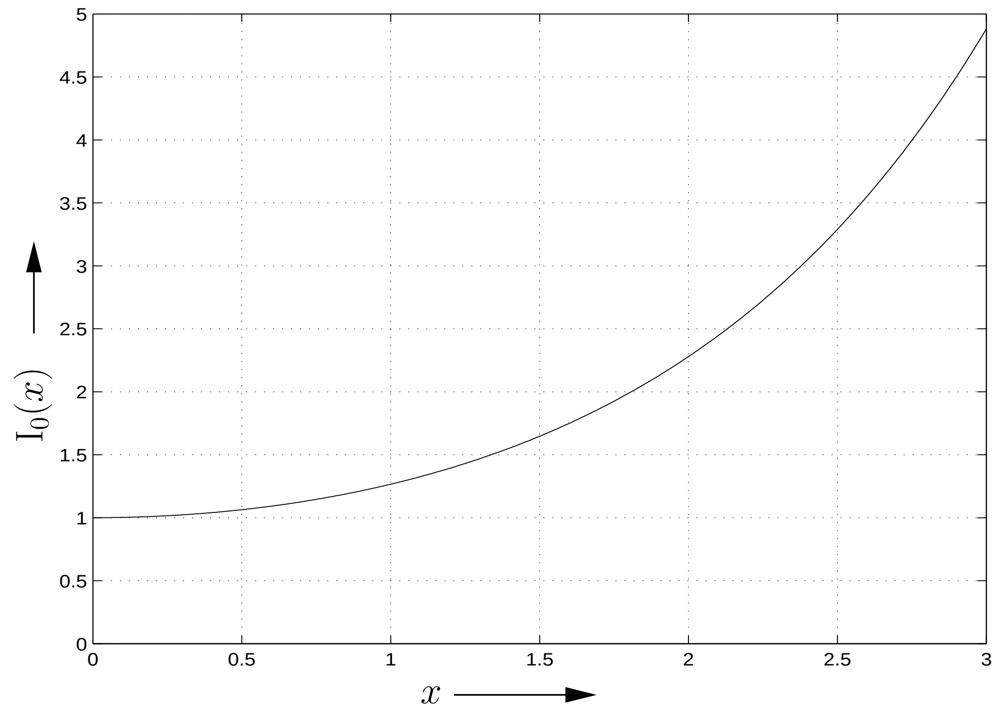
Note that the above integral is *independent* of α since α is independent of ϕ and we integrate over an entire period of the cosine function. Therefore, using the definition

$$\boxed{\mathrm{I}_0(x) = \frac{1}{2\pi} \int\limits_{-\pi}^{\pi} \exp(x \cos\phi) \, \mathrm{d}\phi}$$

we finally obtain

$$\boxed{p(\boldsymbol{r}_b|\boldsymbol{s}_{b\tilde{m}}) = \frac{1}{(\pi N_0)^N} \exp\left(-\frac{||\boldsymbol{r}_b||^2 + ||\boldsymbol{s}_{b\tilde{m}}||^2}{N_0}\right) \cdot \mathrm{I}_0\left(\frac{2}{N_0}|\boldsymbol{r}_b \bullet \boldsymbol{s}_{b\tilde{m}}|\right)}$$

Note that $\mathrm{I}_0(x)$ is the *zeroth order modified Bessel function of the first kind.*



- **ML Detection**

  With the above expression for $p(\boldsymbol{r}_b|\boldsymbol{s}_{b\tilde{m}})$ the noncoherent ML decision rule becomes

$$\begin{aligned}\hat{m} &= \operatorname*{argmax}_{\tilde{m}}\{p(\boldsymbol{r}|\boldsymbol{s}_{b\tilde{m}})\}. \\ &= \operatorname*{argmax}_{\tilde{m}}\{\ln[p(\boldsymbol{r}|\boldsymbol{s}_{b\tilde{m}})]\}. \\ &= \operatorname*{argmax}_{\tilde{m}}\left\{-\frac{||\boldsymbol{s}_{b\tilde{m}}||^2}{N_0}+\ln\left[\mathrm{I}_0\left(\frac{2}{N_0}|\boldsymbol{r}_b\bullet\boldsymbol{s}_{b\tilde{m}}|\right)\right]\right\}\end{aligned}$$

- **Simplification**

The above ML decision rule depends on N_0, which may not be desirable in practice, since N_0 (or equivalently the SNR) has to be estimated. However, for $x \gg 1$ the approximation

$$\ln[\mathrm{I}_0(x)] \approx x$$

holds. Therefore, at high SNR (or small N_0) the above ML metric can be simplified to

$$\hat{m} = \underset{\tilde{m}}{\operatorname{argmax}} \left\{ -||\boldsymbol{s}_{b\tilde{m}}||^2 + 2|\boldsymbol{r}_b \bullet \boldsymbol{s}_{b\tilde{m}}| \right\},$$

which is independent of N_0. In practice, the above simplification has a negligible impact on performance even for small arguments (corresponding to low SNRs) of the Bessel function.

4.3.2.3 Optimum Noncoherent Detection of Binary Orthogonal Modulation

- **Transmitted Waveform**

We assume that the signal space representation of the *complex baseband* transmit waveforms is

$$\begin{aligned} \boldsymbol{s}_{b1} &= [\sqrt{E_b} \quad 0]^T \\ \boldsymbol{s}_{b2} &= [0 \quad \sqrt{E_b}]^T \end{aligned}$$

- **Received Signal**

The corresponding demodulated received signal is

$$\boldsymbol{r}_b = [e^{j\phi}\sqrt{E_b} + z_1 \qquad z_2]^T$$

and

$$\boldsymbol{r}_b = [z_1 \qquad e^{j\phi}\sqrt{E_b} + z_2]^T$$

if $\boldsymbol{s}_{b1}$ and $\boldsymbol{s}_{b2}$ have been transmitted, respectively. z_1 and z_2 are mutually independent complex Gaussian noise processes with identical variances $\sigma_z^2 = N_0$.

- **ML Detection**

  For the simplification of the general ML decision rule, we exploit the fact that for binary orthogonal modulation the relation

$$||\boldsymbol{s}_{b1}||^2 = ||\boldsymbol{s}_{b2}||^2 = E_b$$

  holds. The ML decision rule can be simplified as

$$\begin{aligned}
\hat{m} &= \underset{\tilde{m}}{\operatorname{argmax}} \left\{ -\frac{||\boldsymbol{s}_{b\tilde{m}}||^2}{N_0} + \ln\left[\mathrm{I}_0\left(\frac{2}{N_0} |\boldsymbol{r}_b \bullet \boldsymbol{s}_{b\tilde{m}}| \right) \right] \right\} \\
&= \underset{\tilde{m}}{\operatorname{argmax}} \left\{ -\frac{E_b}{N_0} + \ln\left[\mathrm{I}_0\left(\frac{2}{N_0} |\boldsymbol{r}_b \bullet \boldsymbol{s}_{b\tilde{m}}| \right) \right] \right\} \\
&= \underset{\tilde{m}}{\operatorname{argmax}} \left\{ \ln\left[\mathrm{I}_0\left(\frac{2}{N_0} |\boldsymbol{r}_b \bullet \boldsymbol{s}_{b\tilde{m}}| \right) \right] \right\} \\
&= \underset{\tilde{m}}{\operatorname{argmax}} \left\{ |\boldsymbol{r}_b \bullet \boldsymbol{s}_{b\tilde{m}}| \right\} \\
&= \underset{\tilde{m}}{\operatorname{argmax}} \left\{ |\boldsymbol{r}_b \bullet \boldsymbol{s}_{b\tilde{m}}|^2 \right\}
\end{aligned}$$

  We decide in favor of that signal point which has a larger correlation with the received signal.

- We decide for $\hat{m} = 1$ if

$$|\boldsymbol{r}_b \bullet \boldsymbol{s}_{b1}| > |\boldsymbol{r}_b \bullet \boldsymbol{s}_{b2}|.$$

  Using the definition

$$\boldsymbol{r}_b = [r_{b1} \qquad r_{b2}]^T,$$

we obtain

$$\left|\begin{bmatrix} r_{b1} \\ r_{b2} \end{bmatrix} \bullet \begin{bmatrix} \sqrt{E_b} \\ 0 \end{bmatrix}\right|^2 \;>\; \left|\begin{bmatrix} r_{b1} \\ r_{b2} \end{bmatrix} \bullet \begin{bmatrix} 0 \\ \sqrt{E_b} \end{bmatrix}\right|^2$$

$$|r_1|^2 \;>\; |r_2|^2.$$

In other words, the ML decision rule can be simplified to

$$\begin{array}{lll} \hat{m} = 1 & \text{if} & |r_1|^2 > |r_2|^2 \\ \hat{m} = 2 & \text{if} & |r_1|^2 < |r_2|^2 \end{array}$$

Example:

Binary FSK

– In this case, in the interval $0 \le t \le T$ we have

$$s_{b1}(t) \;=\; \sqrt{\frac{E_b}{T}}$$

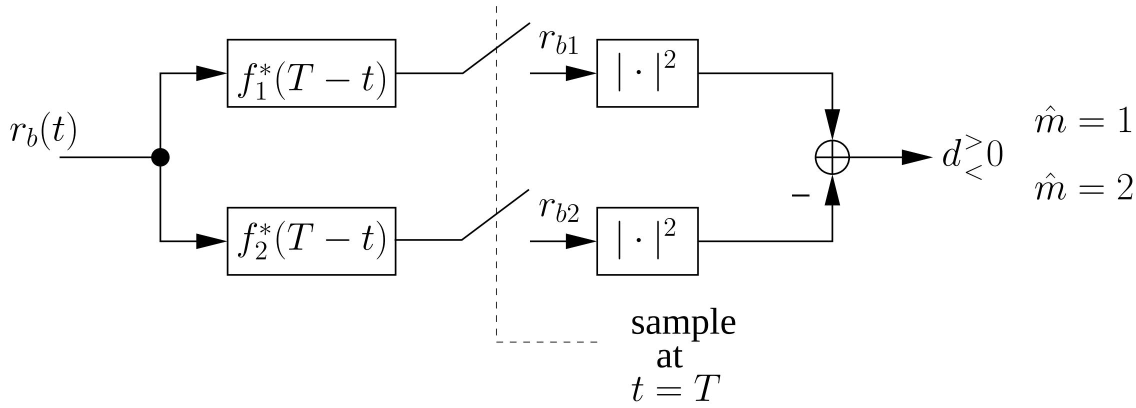
$$s_{b2}(t) \;=\; \sqrt{\frac{E_b}{T}} e^{j2\pi\Delta f t}$$

– If $s_{b1}(t)$ and $s_{b2}(t)$ are orthogonal, the basis functions are

$$f_1(t) \;=\; \begin{cases} \frac{1}{\sqrt{T}}, & 0 \le t \le T \\ 0, & \text{otherwise} \end{cases}$$

$$f_2(t) \;=\; \begin{cases} \frac{1}{\sqrt{T}} e^{j2\pi\Delta f t}, & 0 \le t \le T \\ 0, & \text{otherwise} \end{cases}$$

– **Receiver Structure**



– If $s_{b2}(t)$ was transmitted, we get for r_{b1}

$$
\begin{aligned}
r_{b1} &= \int_0^T r_b(t) f_1^*(t)\,\mathrm{d}t \\
&= \int_0^T e^{j\phi} s_{b2}(t) f_1^*(t)\,\mathrm{d}t + z_1 \\
&= e^{j\phi}\frac{1}{\sqrt{T}} \int_0^T s_{b2}(t)\,\mathrm{d}t + z_1 \\
&= \sqrt{E_b} e^{j\phi} \frac{1}{T} \int_0^T e^{j2\pi\Delta f t}\,\mathrm{d}t + z_1 \\
&= \sqrt{E_b} e^{j\phi} \frac{1}{j2\pi\Delta f T}\, e^{j2\pi\Delta f t}\Big|_0^T + z_1 \\
&= \sqrt{E_b} e^{j\phi} \frac{1}{j2\pi\Delta f T}\, e^{j\pi\Delta f T}\left(e^{j\pi\Delta f T} - e^{-j\pi\Delta f T}\right) + z_1 \\
&= \sqrt{E_b} e^{j(\pi\Delta f T + \phi)}\, \mathrm{sinc}(\pi\Delta f T) + z_1
\end{aligned}
$$

For the proposed detector to be optimum we require orthogonality, i.e., $r_{b1} = z_1$ has to be valid. Therefore,

$$\mathrm{sinc}(\pi \Delta f T) = 0$$

is necessary, which implies

$$\Delta f T = \frac{1}{T} + \frac{k}{T}, \qquad k \in \{\ldots, -1, 0, 1, \ldots\}.$$

This means that for noncoherent detection of FSK signals the minimum required frequency separation for orthogonality is

$$\Delta f = \frac{1}{T}.$$

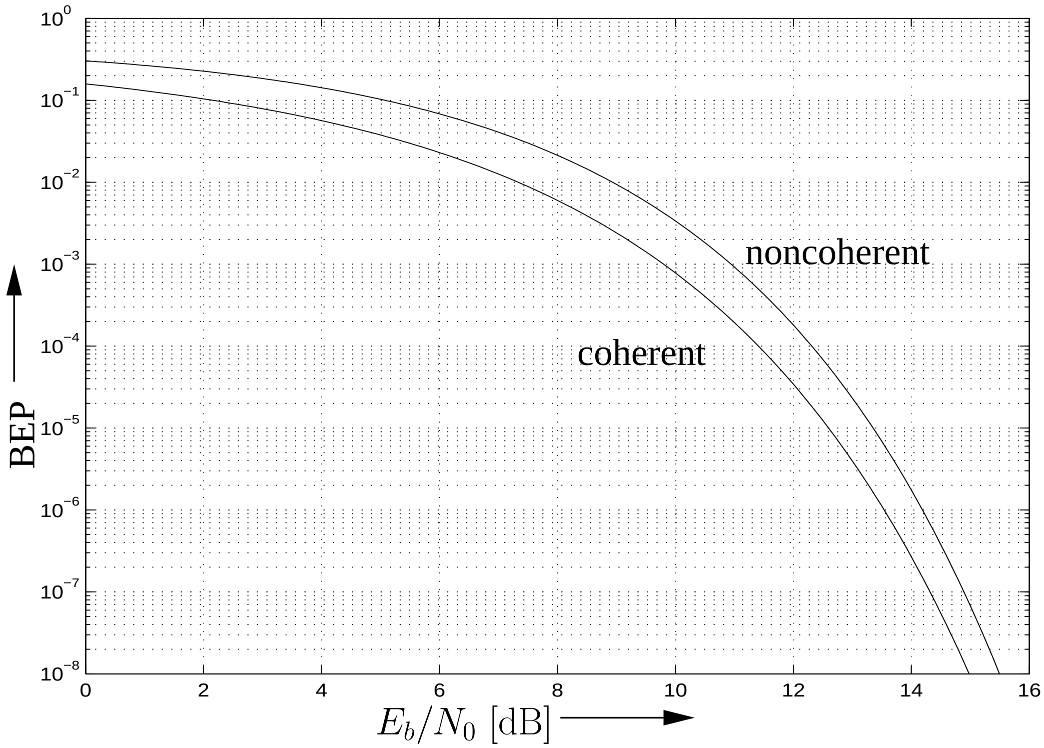
Recall that the transmitted passband signals are orthogonal for

$$\Delta f = \frac{1}{2T},$$

which is also the minimum required separation for *coherent detection* of FSK.

– The BEP (or SEP) for binary FSK with noncoherent detection can be calculated to

$$\boxed{P_b = \frac{1}{2} \exp\left(-\frac{E_b}{2N_0}\right)}$$



4.3.2.4 Optimum Noncoherent Detection of On–Off Keying

- On–off keying (OOK) is a binary modulation format, and the transmitted signal points in complex baseband representation are given by

$$\begin{aligned} s_{b1} &= \sqrt{2E_b} \\ s_{b2} &= 0 \end{aligned}$$

- The demodulated baseband signal is given by

$$r_b = \begin{cases} e^{j\phi}\sqrt{2E_b} + z & \text{if} \quad m = 1 \\ z & \text{if} \quad m = 2 \end{cases}$$

where z is complex Gaussian noise with variance $\sigma_z^2 = N_0$.

- **ML Detection**

  The ML decision rule is

$$\hat{m} \;=\; \underset{\tilde{m}}{\operatorname{argmax}} \left\{ -\frac{|s_{b\tilde{m}}|^2}{N_0} + \ln\left[\mathrm{I}_0\left(\frac{2}{N_0} |r_b s^*_{b\tilde{m}}| \right) \right] \right\}$$

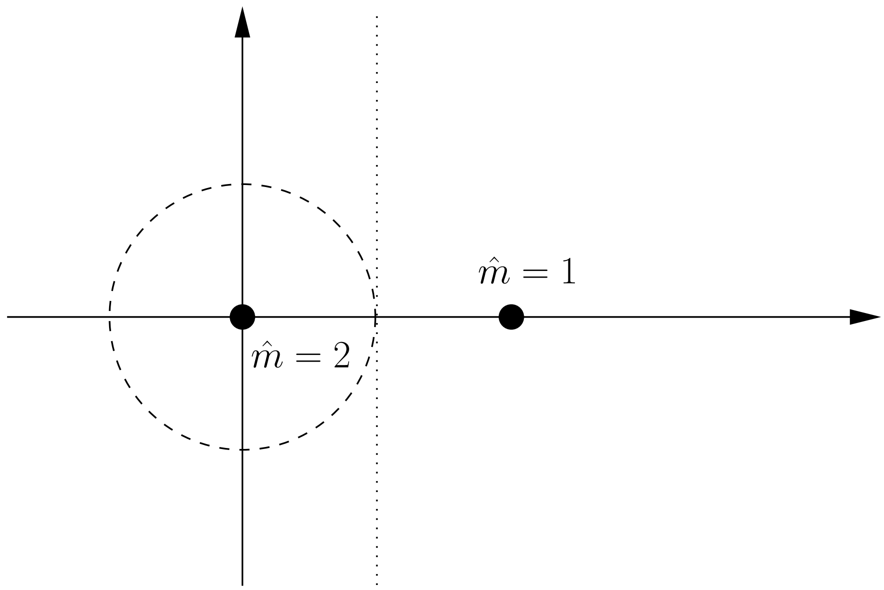
- We decide for $\hat{m} = 1$ if

$$\begin{aligned} -\frac{2E_b}{N_0} + \ln\left[\mathrm{I}_0\left(\frac{2}{N_0} |\sqrt{2E_b} r_b| \right) \right] \;&>\; \ln(\mathrm{I}_0(0)) \\ \underbrace{\ln\left[\mathrm{I}_0\left(\frac{2}{N_0} \sqrt{2E_b} |r_b| \right) \right]}_{\approx \frac{2}{N_0}\sqrt{2E_b}|r_b|} \;&>\; \frac{2E_b}{N_0} \\ |r_b| \;&>\; \sqrt{\frac{E_b}{2}} \end{aligned}$$

  Therefore, the (approximate) ML decision rule can be simplified to

$$\begin{aligned} \hat{m} = 1 \qquad &\text{if} \qquad |r_b| > \sqrt{\tfrac{E_b}{2}} \\ \hat{m} = 2 \qquad &\text{if} \qquad |r_b| < \sqrt{\tfrac{E_b}{2}} \end{aligned}$$

- **Signal Space Decision Regions**



4.3.2.5 Multiple–Symbol Differential Detection (MSDD) of DPSK

- Noncoherent detection of *PSK* is not possible, since for PSK the information is represented in the *absolute* phase of the transmitted signal.
- In order to enable noncoherent detection *differential encoding* is necessary. The resulting scheme is referred to as *differential* PSK (DPSK).

- The (normalized) demodulated DPSK signal in symbol interval k is given by

$$r_b[k] = e^{j\phi} b[k] + z[k]$$

with $b[k] = a[k]b[k-1]$ and noise variance $\sigma_z^2 = N_0/E_S$, where E_S denotes the received energy per DPSK symbol.

- Since the differential encoding operation introduces memory, it is advantageous to detect the transmitted symbols in a block–wise manner.

- In the following, we consider vectors (blocks) of length N

$$\boldsymbol{r}_b = e^{j\phi} \boldsymbol{b} + \boldsymbol{z},$$

where

$$\begin{aligned} \boldsymbol{r}_b &= [r_b[k] \;\; r_b[k-1] \;\; \ldots \;\; r_b[k-(N-1)]^T \\ \boldsymbol{b} &= [b[k] \;\; b[k-1] \;\; \ldots \;\; b[k-(N-1)]^T \\ \boldsymbol{z} &= [z[k] \;\; z[k-1] \;\; \ldots \;\; z[k-(N-1)]^T \end{aligned}$$

- Since $\boldsymbol{z}$ is a Gaussian noise vector, whose components are mutually independent and have variances of $\sigma_z^2 = N_0$, respectively, the general optimum noncoherent ML decision rule can also be applied in this case.

- In the following, we interpret $\boldsymbol{b}$ as the transmitted baseband signal points, i.e., $\boldsymbol{s}_b = \boldsymbol{b}$, and introduce the vector of $N-1$ differential *information bearing* symbols

$$\boldsymbol{a} = [a[k] \;\; a[k-1] \;\; \ldots \;\; a[k-(N-2)]]^T.$$

Note that the vector $\boldsymbol{b}$ of absolute phase symbols corresponds to just the $N-1$ differential symbol contained in $\boldsymbol{a}$.

- The estimate for $\boldsymbol{a}$ is denoted by $\hat{\boldsymbol{a}}$, and we also introduce the trial vectors $\tilde{\boldsymbol{a}}$ and $\tilde{\boldsymbol{b}}$. With these definitions the ML decision rule becomes

$$\begin{aligned}
\hat{\boldsymbol{a}} &= \underset{\tilde{\boldsymbol{a}}}{\operatorname{argmax}} \left\{ -\frac{||\tilde{\boldsymbol{b}}||^2}{N_0} + \ln\left[\mathrm{I}_0\left(\frac{2}{N_0} |\boldsymbol{r}_b \bullet \tilde{\boldsymbol{b}}| \right) \right] \right\} \\
&= \underset{\tilde{\boldsymbol{a}}}{\operatorname{argmax}} \left\{ -\frac{N}{N_0} + \ln\left[\mathrm{I}_0\left(\frac{2}{N_0} |\boldsymbol{r}_b \bullet \tilde{\boldsymbol{b}}| \right) \right] \right\} \\
&= \underset{\tilde{\boldsymbol{a}}}{\operatorname{argmax}} \left\{ |\boldsymbol{r}_b \bullet \tilde{\boldsymbol{b}}| \right\} \\
&= \underset{\tilde{\boldsymbol{a}}}{\operatorname{argmax}} \left\{ \left| \sum_{\nu=0}^{N-1} r_b[k-\nu] \tilde{b}^*[k-\nu] \right| \right\}
\end{aligned}$$

In order to further simplify this result we make use of the relation

$$\tilde{b}[k-\nu] = \prod_{\mu=\nu}^{N-2} \tilde{a}[k-\mu] \tilde{b}[k-(N-1)]$$

and obtain the final ML decision rule

$$\begin{aligned}
\hat{\boldsymbol{a}} &= \underset{\tilde{\boldsymbol{a}}}{\operatorname{argmax}} \left\{ \left| \sum_{\nu=0}^{N-1} r_b[k-\nu] \prod_{\mu=\nu}^{N-2} \tilde{a}^*[k-\mu] \tilde{b}^*[k-(N-1)] \right| \right\} \\
&= \underset{\tilde{\boldsymbol{a}}}{\operatorname{argmax}} \left\{ \left| \sum_{\nu=0}^{N-1} r_b[k-\nu] \prod_{\mu=\nu}^{N-2} \tilde{a}^*[k-\mu] \right| \right\}
\end{aligned}$$

Note that this final result is independent of $\tilde{b}[k-(N-1)]$.

- We make a *joint* decision on $N-1$ differential symbols $a[k-\nu]$, $0 \leq \nu \leq N-2$, based on the observation of N received signal points. N is also referred to as the *observation window size.*

- The above ML decision rule is known as *multiple symbol differential detection (MSDD)* and was reported first by Divsalar & Simon in 1990.
- In general, the performance of MSDD increases with increasing N. For $N \to \infty$ the performance of ideal coherent detection is approached.
- $N = 2$
  For the special case $N = 2$, we obtain

$$
\begin{aligned}
\hat{a}[k] &= \underset{\tilde{a}[k]}{\operatorname{argmax}} \left\{ \left| \sum_{\nu=0}^{1} r_b[k-\nu] \prod_{\mu=\nu}^{0} \tilde{a}^*[k-\mu] \right| \right\} \\
&= \underset{\tilde{a}[k]}{\operatorname{argmax}} \left\{ |r_b[k]\tilde{a}^*[k] + r_b[k-1]| \right\} \\
&= \underset{\tilde{a}[k]}{\operatorname{argmax}} \left\{ |r_b[k]\tilde{a}^*[k] + r_b[k-1]|^2 \right\} \\
&= \underset{\tilde{a}[k]}{\operatorname{argmax}} \left\{ |r_b[k]|^2 + |r_b[k-1]|^2 + 2\mathsf{Re}\{r_b[k]\tilde{a}^*[k]r_b^*[k-1]\} \right\} \\
&= \underset{\tilde{a}[k]}{\operatorname{argmin}} \left\{ -2\mathsf{Re}\{r_b[k]\tilde{a}^*[k]r_b^*[k-1]\} \right\} \\
&= \underset{\tilde{a}[k]}{\operatorname{argmin}} \left\{ |r_b[k]r_b^*[k-1]|^2 + |\tilde{a}[k]|^2 - 2\mathsf{Re}\{r_b[k]\tilde{a}^*[k]r_b^*[k-1]\} \right\} \\
&= \underset{\tilde{a}[k]}{\operatorname{argmin}} \left\{ |d[k] - \tilde{a}[k]|^2 \right\}
\end{aligned}
$$

  with

$$
d[k] = r_b[k]r_b^*[k-1].
$$

  Obviously, for $N = 2$ the ML (MSDD) decision rule is identical to the heuristically derived conventional differential detection decision rule. For $N > 2$ the gap to coherent detection becomes smaller.

- **Disadvantage of MSDD**

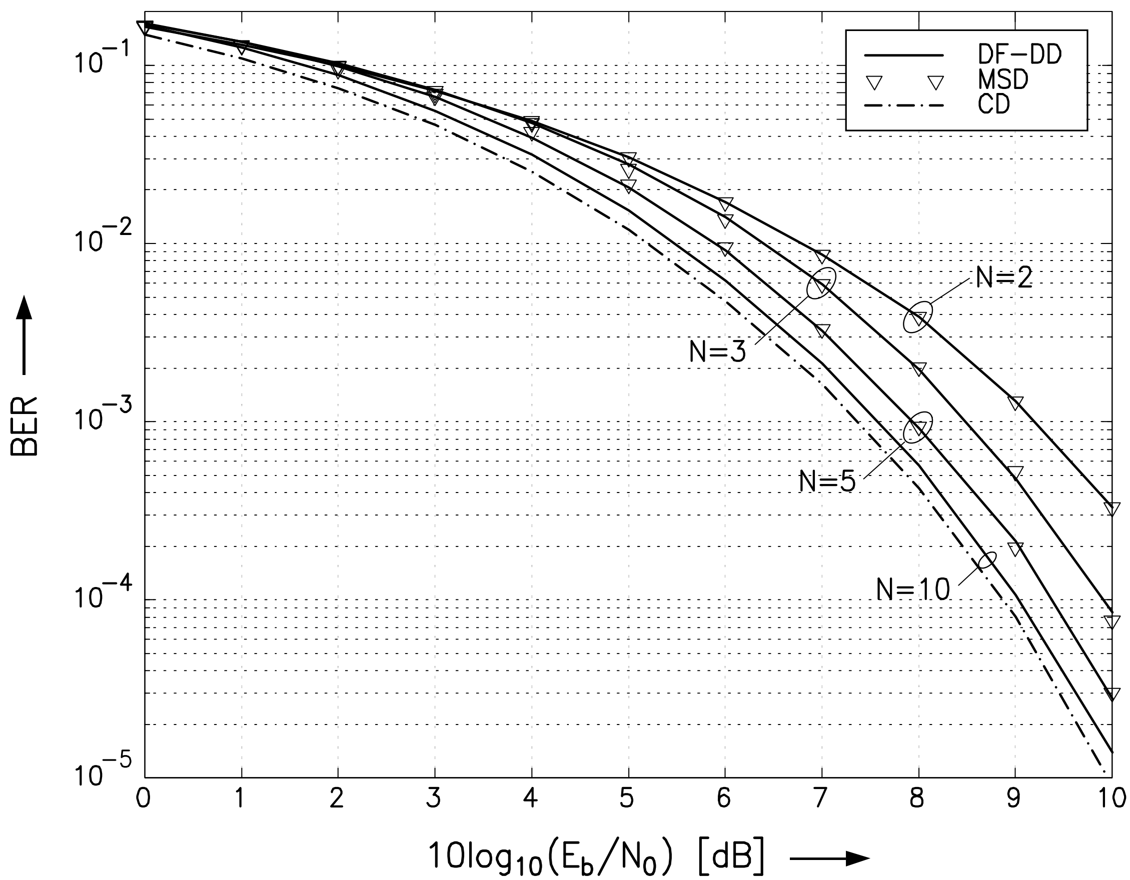
  The trial vector $\tilde{\boldsymbol{a}}$ has $N-1$ elements and each element has M possible values. Therefore, there are M^{N-1} different trial vectors $\tilde{\boldsymbol{a}}$. This means for MSDD we have to make $M^{N-1}/(N-1)$ tests per (scalar) symbol decision. This means the complexity of MSDD grows exponentially with N. For example, for $M = 4$ we have to make 4 and 8192 tests for $N = 2$ and $N = 9$, respectively.

- **Alternatives**

  In order to avoid the exponential complexity of MSDD there are two different approaches:

  1. The first approach is to replace the above *brute–force* search with a smarter approach. In this smarter approach the trial vectors $\tilde{\boldsymbol{a}}$ are sorted first. The resulting fast decoding algorithm still performs optimum MSDD but has only a complexity of $N \log(N)$ (Mackenthun, 1994). For fading channels a similar approach based on *sphere decoding* exists (Pauli et al.).
  2. The second approach is suboptimum but achieves a similar performance as MSDD. Complexity is reduced by using *decision feedback* of previously decided symbols. The resulting scheme is referred to as decision–feedback differential detection (DF–DD). The complexity of this approach is only linear in N (Edbauer, 1992).

- Comparison of BEP (BER) of MSDD (MSD) and DF–DD for 4DPSK



4.4 Optimum Coherent Detection of Continuous Phase Modulation (CPM)

■ CPM Modulation

– *CPM Transmit Signal*

Recall that the CPM transmit signal in complex baseband representation is given by

$$s_b(t,\,\boldsymbol{I}) = \sqrt{\frac{E}{T}}\,\exp(j[\phi(t,\,\boldsymbol{I}) + \phi_0]),$$

where $\boldsymbol{I}$ is the sequence $\{I[k]\}$ of information bearing symbols $I[k] \in \{\pm 1,\, \pm 3,\, \ldots,\, \pm(M-1)\}$, $\phi(t, \boldsymbol{I})$ is the information carrying phase, and ϕ_0 is the initial carrier phase. Without loss of generality, we assume $\phi_0 = 0$ in the following.

– *Information Carrying Phase*

In the interval $kT \leq t \leq (k+1)T$ the phase $\phi(t,\,\boldsymbol{I})$ can be written as

$$\begin{aligned}\phi(t,\,\boldsymbol{I}) \;=\; & \Theta[k] + 2\pi h \sum_{\nu=1}^{L-1} I[k-\nu]q(t-[k-\nu]T) \\ & +2\pi h I[k] q(t-kT),\end{aligned}$$

where $\Theta[k]$ represents the accumulated phase up to time kT, h is the so–called modulation index, and $q(t)$ is the phase shaping pulse with

$$q(t) = \begin{cases} 0, & t < 0 \\ \text{monotonic}, & 0 \leq t \leq LT \\ 1/2, & t > LT \end{cases}$$

– *Trellis Diagram*
If $h = q/p$ is a rational number with relative prime integers q and p, CPM can be described by a trellis diagram whose number of states S is given by

$$S = \begin{cases} pM^{L-1}, & \text{even}\, q \\ 2pM^{L-1}, & \text{odd}\, q \end{cases}$$

- **Received Signal**
The received complex baseband signal is given by

$$r_b(t) = s_b(t, \boldsymbol{I}) + z(t)$$

with complex AWGN $z(t)$, which has a power spectral density of $\Phi_{ZZ}(f) = N_0$.

- **ML Detection**

  – Since the CPM signal has memory, ideally we have to observe the entire received signal $r_b(t)$, $-\infty \leq t \leq \infty$, in order to make a decision on *any* $I[k]$ in the sequence of transmitted signals.

  – The conditional pdf $p(r_b(t)|s_b(t,\, \tilde{\boldsymbol{I}}))$ is given by

$$p(r_b(t)|s_b(t,\, \tilde{\boldsymbol{I}})) \propto \exp\left(-\frac{1}{N_0} \int_{-\infty}^{\infty} |r_b(t) - s_b(t,\, \tilde{\boldsymbol{I}})|^2 \,\mathrm{d}t \right),$$

where $\tilde{\boldsymbol{I}} \in \{\pm 1,\, \pm 3,\, \ldots,\, \pm(M-1)\}$ is a trial sequence. For

ML detection we have the decision rule

$$\begin{aligned}
\hat{\boldsymbol{I}} &= \underset{\tilde{\boldsymbol{I}}}{\operatorname{argmax}} \left\{ p(r_b(t)|s_b(t,\,\tilde{\boldsymbol{I}})) \right\} \\
&= \underset{\tilde{\boldsymbol{I}}}{\operatorname{argmax}} \left\{ \ln[p(r_b(t)|s_b(t,\,\tilde{\boldsymbol{I}}))] \right\} \\
&= \underset{\tilde{\boldsymbol{I}}}{\operatorname{argmax}} \left\{ -\int\limits_{-\infty}^{\infty} |r_b(t) - s_b(t,\,\tilde{\boldsymbol{I}})|^2 \, \mathrm{d}t \right\} \\
&= \underset{\tilde{\boldsymbol{I}}}{\operatorname{argmax}} \left\{ -\int\limits_{-\infty}^{\infty} |r_b(t)|^2 \, \mathrm{d}t - \int\limits_{-\infty}^{\infty} |s_b(t,\,\tilde{\boldsymbol{I}})|^2 \, \mathrm{d}t \right. \\
&\qquad \left. + \int\limits_{-\infty}^{\infty} 2\mathsf{Re}\{r_b(t) s_b^*(t,\,\tilde{\boldsymbol{I}})\} \, \mathrm{d}t \right\} \\
&= \underset{\tilde{\boldsymbol{I}}}{\operatorname{argmax}} \left\{ \int\limits_{-\infty}^{\infty} \mathsf{Re}\{r_b(t) s_b^*(t,\,\tilde{\boldsymbol{I}})\} \, \mathrm{d}t, \right\}
\end{aligned}$$

where $\hat{\boldsymbol{I}}$ refers to the ML decision. If $\boldsymbol{I}$ is a sequence of length K, there are M^K different sequences $\tilde{\boldsymbol{I}}$. Since we have to calculate the function $\int_{-\infty}^{\infty} \mathsf{Re}\{r_b(t) s_b^*(t,\,\tilde{\boldsymbol{I}})\} \, \mathrm{d}t$ for each of these sequences, the complexity of ML detection with *brute–force* search grows exponentially with the sequence length K, which is prohibitive for a practical implementation.

■ Viterbi Algorithm

– The exponential complexity of brute–force search can be avoided using the so–called *Viterbi algorithm (VA)*.

– Introducing the definition

$$\Lambda[k] = \int\limits_{-\infty}^{(k+1)T} \mathsf{Re}\{r_b(t)s_b^*(t,\, \tilde{\boldsymbol{I}})\}\,\mathrm{d}t,$$

we observe that the function to be maximized for ML detection is $\Lambda[\infty]$. On the other hand, $\Lambda[k]$ may be calculated recursively as

$$\begin{aligned}
\Lambda[k] &= \int\limits_{-\infty}^{kT} \mathsf{Re}\{r_b(t)s_b^*(t,\, \tilde{\boldsymbol{I}})\}\,\mathrm{d}t + \int\limits_{kT}^{(k+1)T} \mathsf{Re}\{r_b(t)s_b^*(t,\, \tilde{\boldsymbol{I}})\}\,\mathrm{d}t \\
\Lambda[k] &= \Lambda[k-1] + \int\limits_{kT}^{(k+1)T} \mathsf{Re}\{r_b(t)s_b^*(t,\, \tilde{\boldsymbol{I}})\}\,\mathrm{d}t \\
\Lambda[k] &= \Lambda[k-1] + \lambda[k],
\end{aligned}$$

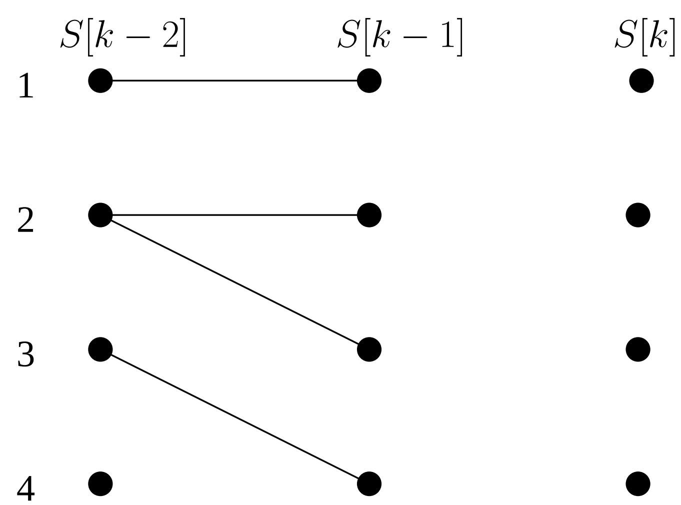
where we use the definition

$$\begin{aligned}
\lambda[k] &= \int\limits_{kT}^{(k+1)T} \mathsf{Re}\{r_b(t)s_b^*(t,\, \tilde{\boldsymbol{I}})\}\,\mathrm{d}t. \\
&= \int\limits_{kT}^{(k+1)T} \mathsf{Re}\Bigg\{ r_b(t)\exp\Bigg(-j\Bigg[\tilde{\Theta}[k] + 2\pi h \sum_{\nu=1}^{L-1} \tilde{I}[k-\nu] \\
&\qquad \cdot q(t-[k-\nu]T) + 2\pi h \tilde{I}[k] q(t-kT)\Bigg]\Bigg)\Bigg\}\,\mathrm{d}t.
\end{aligned}$$

$\Lambda[k]$ and $\lambda[k]$ are referred to as the *accumulated metric* and the *branch metric* of the VA, respectively.

– Since CPM can be described by a trellis with a finite number of states $S = pM^{L-1}$ ($S = 2pM^{L-1}$), at time kT we have to consider only S different $\Lambda[\boldsymbol{S}[k-1], k-1]$. Each $\Lambda[\boldsymbol{S}[k-1], k-1]$ corresponds to exactly one state $\boldsymbol{S}[k-1]$ which is defined by

$$\boldsymbol{S}[k-1] = [\tilde{\Theta}[k-1],\ \tilde{I}[k-(L-1)],\ \ldots,\ \tilde{I}[k-1]].$$



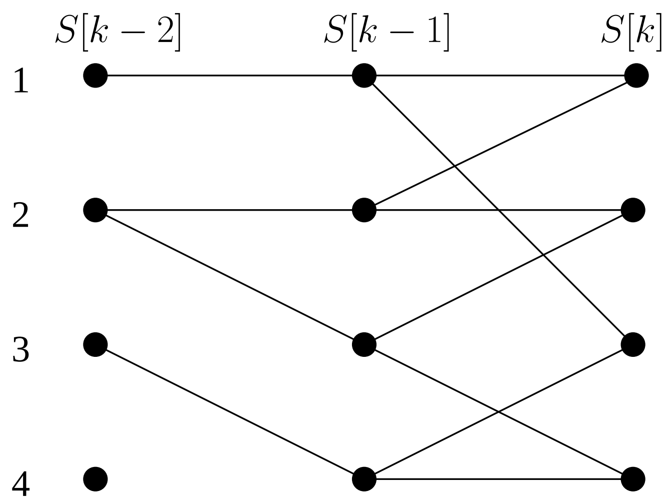
For the interval $kT \leq t \leq (k+1)T$ we have to calculate M branch metrics $\lambda[\boldsymbol{S}[k-1], \tilde{I}[k], k]$ for each state $\boldsymbol{S}[k-1]$ corresponding to the M different $\tilde{I}[k]$. Then at time $(k+1)T$ the new states are defined by

$$\boldsymbol{S}[k] = [\tilde{\Theta}[k],\ \tilde{I}[k-(L-2)],\ \ldots,\ \tilde{I}[k]],$$

with $\tilde{\Theta}[k] = \tilde{\Theta}[k-1] + \pi h \tilde{I}[k-(L-1)]$. M branches defined by $\boldsymbol{S}[k-1]$ and $\tilde{I}[k]$ emanate in each state $\boldsymbol{S}[k]$. We calculate the M accumulated metrics

$$\Lambda[\boldsymbol{S}[k-1],\, \tilde{I}[k],\, k] = \Lambda[\boldsymbol{S}[k-1],\, k-1] + \lambda[\boldsymbol{S}[k-1],\, \tilde{I}[k],\, k]$$

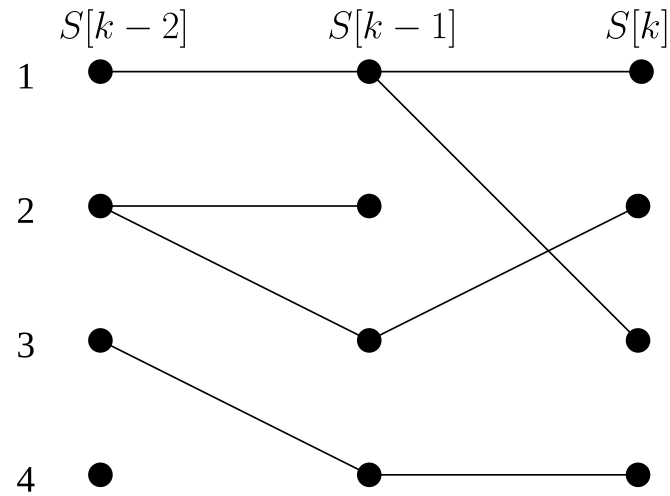
for each state $\boldsymbol{S}[k]$.



From all the paths (partial sequences) that emanate in a state, we have to retain only that one with the largest accumulated metric denoted by

$$\Lambda[\boldsymbol{S}[k],\, k] = \max_{\tilde{I}[k]} \left\{ \Lambda[\boldsymbol{S}[k-1],\, \tilde{I}[k],\, k] \right\},$$

since any other path with a smaller accumulated metric at time $(k+1)T$ cannot have a larger metric $\Lambda[\infty]$. Therefore, at time $(k+1)T$ there will be again only S so–called surviving paths with corresponding accumulated metrics.



– The above steps are carried out for all symbol intervals and at the end of the transmission, a decision is made on the transmitted sequence. Alternatively, at time kT we may use the symbol $\tilde{I}[k-k_0]$ corresponding to the surviving path with the largest accumulated metric as estimate for $I[k-k_0]$. It has been shown that as long as the decision delay k_0 is large enough (e.g. $k_0 \geq 5\log_2(S)$), this method yields the same performance as true ML detection.

– The complexity of the VA is *exponential* in the number of states, but only *linear* in the length of the transmitted sequence.

- **Remarks:**

  – At the expense of a certain loss in performance the complexity of the VA can be further reduced by reducing the number of states.

  – Alternative implementations receivers for CPM are based on Laurent's decomposition of the CPM signal into a sum of PAM signals