EE4152 / IM4152
Digital Communications

A/Prof Li Kwok Hung
EEE-S1-B1b-62
Tel: 6790-5028
Ekhli@ntu.edu.sg
EEE, NTU

Major Topics

- Pulse Detection for Binary Signaling
- Digital Carrier Systems
- Optimum Threshold Detection for Binary Signals
- Geometric Representation of Signals
- Optimum Receiver
- Decision Regions & Error Probability

Textbook & References

- **B P Lathi and Z Ding**, *Modern Digital and Analog Systems*, 4/Ed, Oxford University Press, 2010
- **S Haykin and M Moher**, *Communication Systems*, 5/Ed, John Wiley, 2010
- **J G Proakis and M Salehi**, *Communication Systems Engineering*, 2/Ed, Prentice-Hall, 2002

Pulse Detection for Binary Signaling

Ref:

Lathi and Ding, *Modern Digital and Analog Systems*

(pp. 365 - 366)

Error in Pulse Detection

- The signal received at the detector consists of the desired pulse train plus random channel noise. This results in detection error.

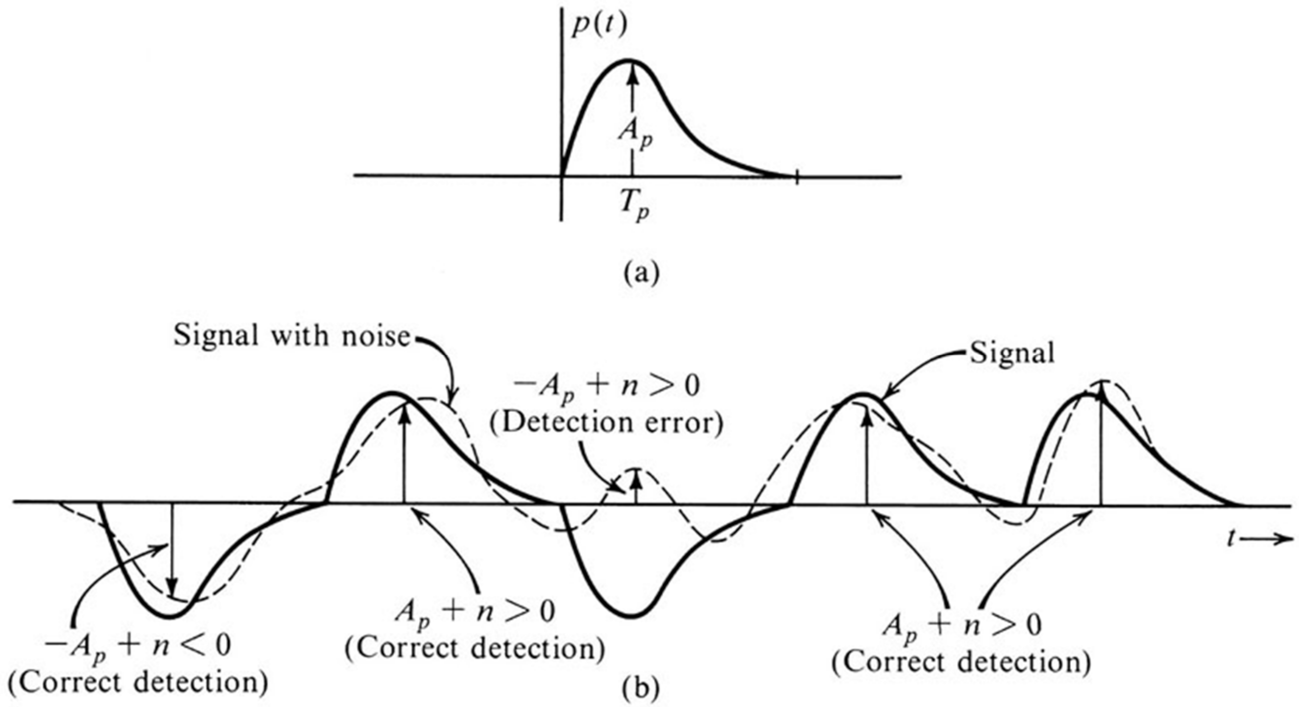


Fig. 1: Detection error in pulse detection

Polar Transmission

- Sample value = $\pm A_p + n$
- A_p is the peak amplitude of the pulse.
- n is the Gaussian noise with PDF

$$p(n) = \frac{1}{\sigma_n \sqrt{2\pi}} e^{-n^2/2\sigma_n^2} \quad (1)$$

  where σ_n^2 is the variance of the Gaussian noise.
- To distinguish between $-A_p$ and $+A_p$ for polar signaling, the detection threshold is thus $(-A_p + A_p)/2 = 0$.

$\Pr(\alpha < n \leq \beta)$

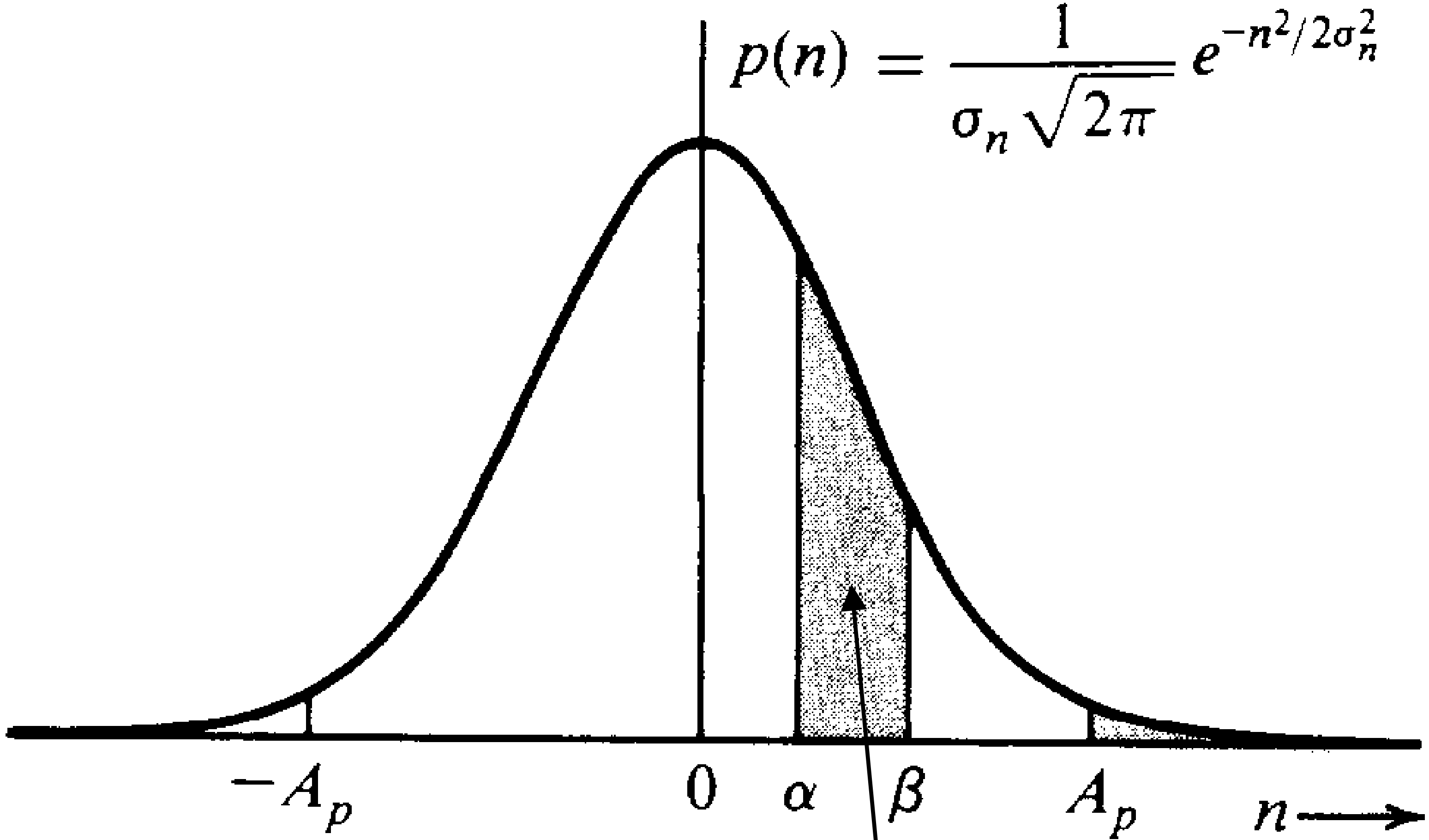


Fig. 2: $\Pr(\alpha < n \leq \beta) = \int_{\alpha}^{\beta} p(n)\,dn$

Conditional Error Prob

- Let $P(\varepsilon|0)$ be the conditional error prob given that a '0' was transmitted. Since the threshold is 0, an error occurs when

$$P(\varepsilon \,|\, 0) = \Pr(-A_p + n > 0) = \Pr\left(n > A_p\right)$$

- Let $P(\varepsilon|1)$ be the conditional error prob given that a '1' was transmitted. An error occurs when

$$P(\varepsilon \,|\, 1) = \Pr(A_p + n < 0) = \Pr\left(n < -A_p\right)$$

Derivation of $P(\varepsilon|0)$

$$P(\varepsilon\,|\,0) = \Pr\left(n > A_p\right)$$

$$= \int_{A_p}^{\infty} p(n)\,dn$$

$$= \frac{1}{\sigma_n\sqrt{2\pi}} \int_{A_p}^{\infty} \exp\left(-\frac{n^2}{2\sigma_n^2}\right) dn$$

$$= \frac{1}{\sqrt{2\pi}} \int_{A_p/\sigma_n}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx \qquad \left(\text{use } x = n/\sigma_n\right)$$

$$= Q\left(\frac{A_p}{\sigma_n}\right)$$

Error Prob for Polar Signal

- Similarly, we can show that $P(\varepsilon|1) = P(\varepsilon|0)$
- Since the bits '1' & '0' are equally likely, the average error prob is

$$P(\varepsilon) = \frac{1}{2}P(\varepsilon\,|1) + \frac{1}{2}P(\varepsilon\,|\,0) = Q\left(\frac{A_p}{\sigma_n}\right) \qquad (2)$$

- In general, if the difference between pulse amplitudes is $2A_p$, then the error prob will be

$$Q\left(\frac{2A_p/2}{\sigma_n}\right) = Q\left(A_p/\sigma_n\right) \qquad (3)$$

Error Prob for On-Off Signal

- We have to distinguish between A_p (presence of a pulse) and 0 (no pulse). The detection threshold is $(A_p + 0)/2 = A_p/2$.
- If the received sample value is greater than $A_p/2$, we decide in favor of 1. Otherwise, we decide in favor of 0.
- The derivation of error prob is similar to that of a polar signal. In particular, the difference between two amplitudes is $(A_p - 0) = A_p$. The error prob, from (3), is $Q(A_p/2\sigma_n)$.

Bipolar Signal

- We have to distinguish between $\pm A_p$ (presence of a pulse) and 0 (no pulse). The detection threshold is $\pm A_p/2$.
- If the received sample value is in $(-A_p/2, A_p/2)$, we decide in favor of 0. Otherwise, we decide in favor of 1.
- Let $p_1 = p(+ve|1)$ be the prob that the positive pulse is transmitted for '1'.

Derivation of $P(\varepsilon | 0)$

$$P(\varepsilon \,|\, 0) = \Pr\left(|n| > A_p / 2\right)$$

$$= \Pr\left(n > A_p / 2\right) + \Pr\left(n < -A_p / 2\right)$$

$$= 2 \times \Pr\left(n > A_p / 2\right)$$

$$= 2Q\left(\frac{A_p}{2\sigma_n}\right)$$

Error Prob for Bipolar Signal

$$P(\varepsilon\,|\,1) \simeq p_1 \times \Pr\left(n < -A_p/2\right) + [1 - p_1] \times \Pr\left(n > A_p/2\right)$$
$$= \Pr\left(n > A_p/2\right)$$
$$= Q\left(\frac{A_p}{2\sigma_n}\right)$$

Finally, the average error prob is

$$P(\varepsilon) = \frac{1}{2}P(\varepsilon\,|\,1) + \frac{1}{2}P(\varepsilon\,|\,0) = \frac{3}{2}Q\left(\frac{A_p}{2\sigma_n}\right) \qquad (4)$$

Observations

- The error prob for the bipolar case is 50% higher than that for the on-off case.
- The difference can be easily compensated by just a little increase in signal power.
- Thus bipolar and on-off schemes have almost the same performance.

Scheme	$P(\varepsilon)$	$P(\varepsilon) = 0.287 \times 10^{-6}$
Polar	$Q(A_p/\sigma_n)$	$A_p/\sigma_n = 5$
On-Off	$Q(A_p/2\sigma_n)$	$A_p/\sigma_n = 10$
Bipolar	$1.5 \times Q(A_p/2\sigma_n)$	$A_p/\sigma_n = 10.16$

Example 1

(a) Polar binary pulses are received with peak amplitude $A_p = 1$ mV. The channel noise rms amplitude is 192.3 μV. Threshold detection is used, and 1 and 0 are equally likely. Find the detection error prob.

(b) Find the error prob for (i) the on-off case and (ii) the bipolar case if pulses of the same shape as in part (a) are used, but their amplitudes are adjusted so that the transmitted power is the same as in part (a).

Example 1 (solution)

(a) For the polar case,

$$\frac{A_p}{\sigma_n} = \frac{10^{-3}}{192.3\times 10^{-6}} = 5.2$$

From the Q-function table, we find

$$P_{polar}(\varepsilon) = Q(5.2) = 0.9964\times 10^{-7}$$

(b) Because half the bits are transmitted by no pulse, there are, on the average, only half as many pulses in the on-off case (compared to the polar). To maintain the

Example 1 (solution)

the same power, we need to double the energy of each pulse in the on-off or the bipolar case (compared to the polar). Now, doubling the pulse energy is accomplished by multiplying the pulse by $\sqrt{2}$. Thus, for on-off $\hat{A}_p$ is $\sqrt{2}$ times the A_p in the polar case, i.e., $\hat{A}_p = \sqrt{2}\times 10^{-3}$. Therefore,

$$P_{on-off}\left(\varepsilon\right) = Q\left(\frac{\hat{A}_p}{2\sigma_n}\right) = Q(3.68) = 1.166\times 10^{-4}$$

$$P_{bipolar}\left(\varepsilon\right) = 1.5Q\left(\frac{\hat{A}_p}{2\sigma_n}\right) = 1.5\times Q(3.68) = 1.749\times 10^{-4}$$

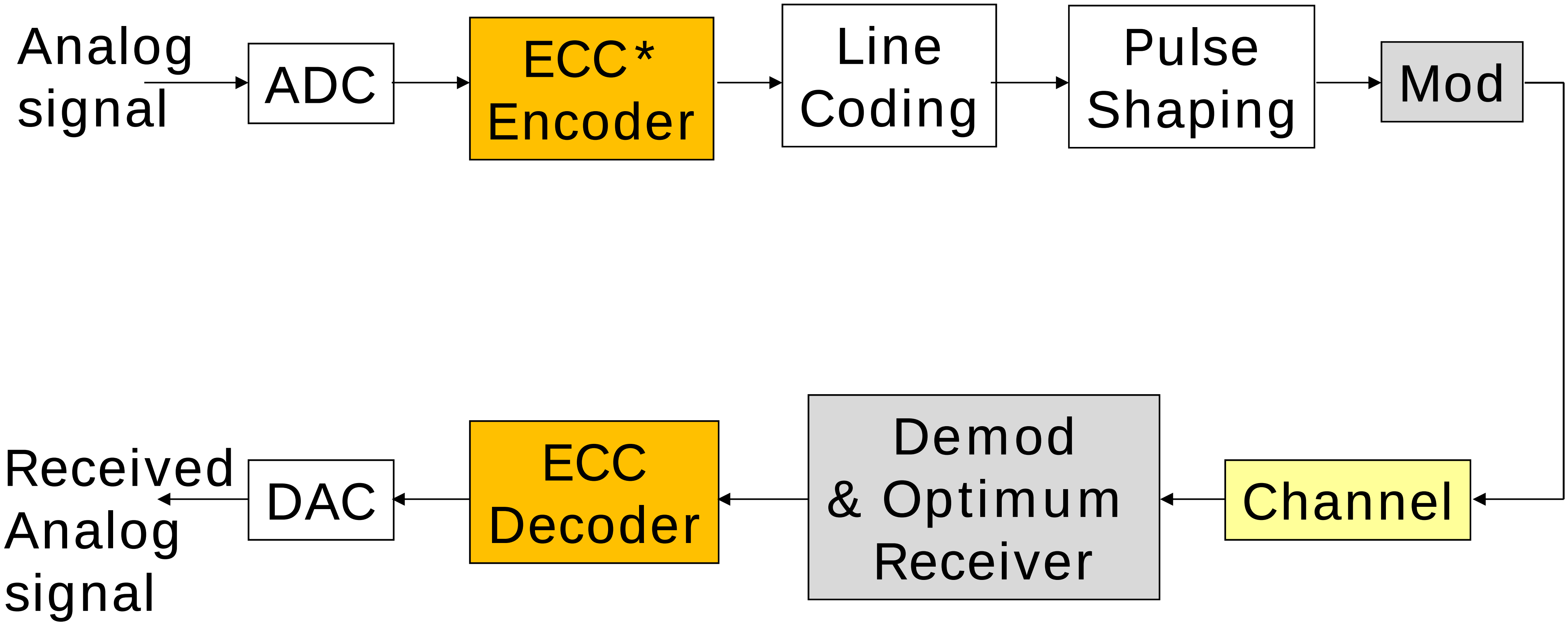
Digital Carrier Systems

Ref:

Lathi and Ding, *Modern Digital and Analog Systems*

(pp. 372 - 380)

Digital Communication System



*Note: ECC – error-control coding

Fig. 3: Overview of a digital communication system

Baseband Communication

- Baseband (BB) signals are transmitted directly without modulation
- Suitable for short-distance transmission over a pair of wires or coaxial cables
- Not suitable for transmission over a radio link as it needs a huge antenna

Digital Modulation

- The BB signal spectrum must be shifted to a higher frequency range to match (bandpass) channel characteristics
- By shifting spectra to non-overlapping bands, several BB signals can share the available BW simultaneously through frequency division multiplexing (FDM)
- Two basic forms: *amplitude modulation* and *angle modulation*

Amplitude Modulation (AM)

- In AM, the carrier amplitude is varied in proportion to the modulating signal.
- Note that the modulated signal is still an on-off signal.
- This modulation scheme of transmitting binary data is known as *amplitude-shift keying* (ASK) or *on-off keying* (OOK).

Amplitude Modulation (...)

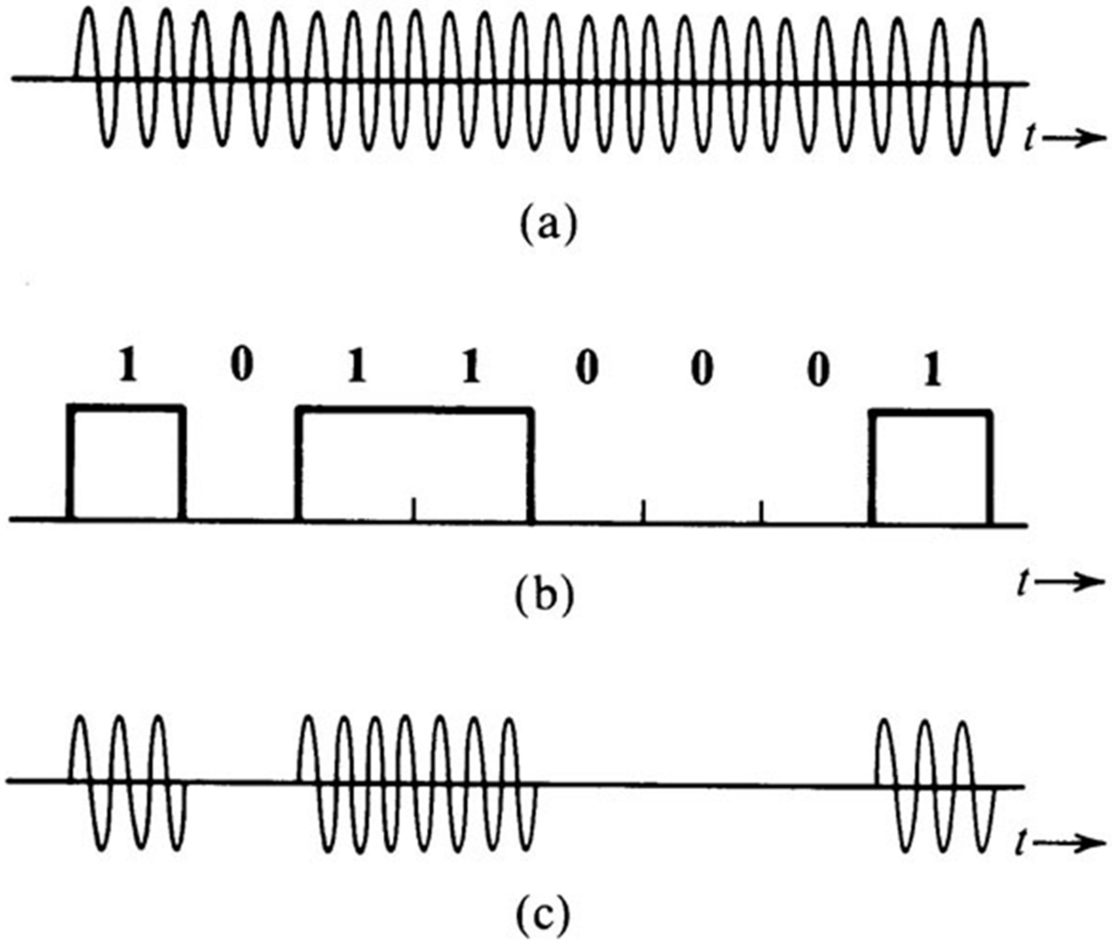


Fig. 4: (a) The carrier cos $\omega_c t$. (b) The modulating signal $m(t)$. (c) ASK: the modulated signal $m(t)$ cos $\omega_c t$

Angle Modulation

- *Phase-shift keying* (PSK): Using the phase of the modulated signal to convey binary info
  ‘1’ $\rightarrow p(t)\cos\left(2\pi f_c t\right)$
  ‘0’ $\rightarrow p(t)\cos\left(2\pi f_c t + \pi\right) = -p(t)\cos\left(2\pi f_c t\right)$

- *Frequency-shift keying* (FSK): Using the freq of the modulated signal to convey binary info
  ‘1’ $\rightarrow \cos\left(2\pi f_{c_1} t\right)$
  ‘0’ $\rightarrow \cos\left(2\pi f_{c_0} t\right)$

Angle Modulation (...)

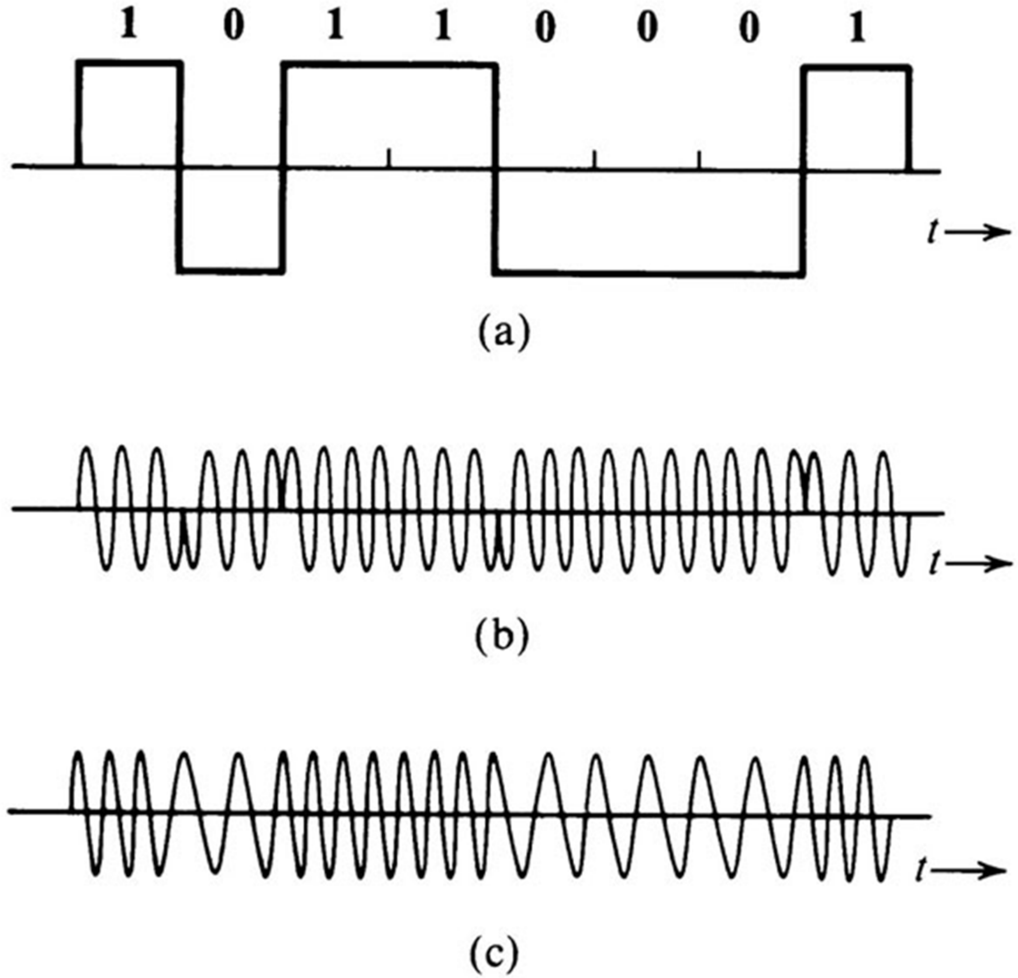


Fig. 5: (a) The modulating signal m(t). (b) PSK: the modulated signal $m(t)\cos\omega_c t$. (c) FSK: the modulated signal

Demodulation

- Demodulation of digital-modulated signals is similar to that of analog-modulated signals
- **ASK**: Can be demodulated coherently (synchronous) or noncoherently (envelope detection). Noncoherent detection is commonly used due to its simplicity
- **PSK**: Only coherent detection can be employed. Noncoherent detection can only be used for differential PSK (DPSK)
- **FSK**: Both coherent detection & noncoherent detection (envelope detection) are possible

PSD of ASK, PSK & FSK

- **ASK**: Same as on-off signal shifted to $\pm f_c$, with discrete components. For full-width rectangular pulse, BW = $2R_b$.
- **PSK**: Same as polar signal shifted to $\pm f_c$, without discrete components. For full-width rectangular pulse, BW = $2R_b$.
- **FSK**: Same as two on-off signals shifted to $\pm f_{c_0}$ and $\pm f_{c_1}$, without discrete components. For full-width rectangular pulse, BW > $2R_b$.
- Being polar, PSK requires 3 dB *less* power than ASK/FSK for the same noise immunity.

PSD of ASK, PSK & FSK (...)

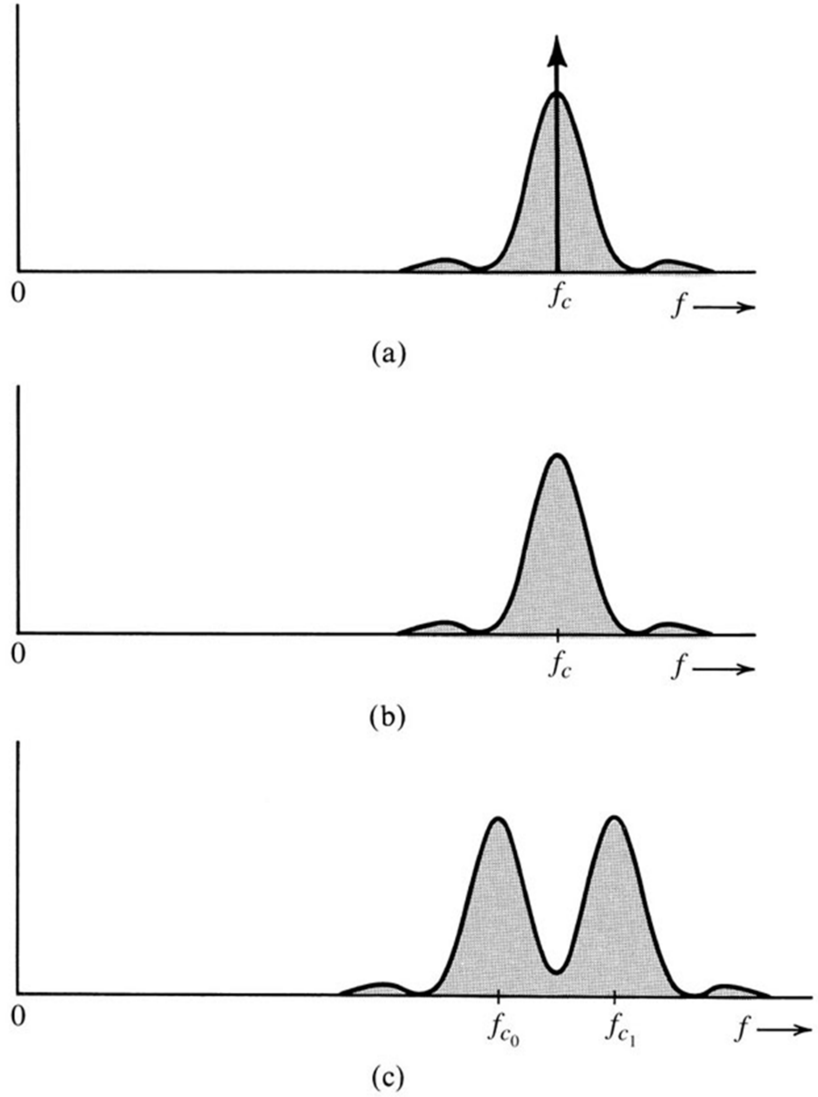


Fig. 6: PSD of (a) ASK, (b) PSK & (c) FSK

Optimum Threshold Detection for Binary Signals

Ref:

Lathi and Ding, *Modern Digital and Analog Systems*

(pp. 506 - 512)

Minimizing Error Prob

- For binary polar signaling, the sampled value of the received pulse is $\pm A_p + n$, where A_p is its peak amplitude and n is the Gaussian noise with mean zero and variance σ_n^2
- From (3), we know the error prob $P_e = Q(\rho)$, where $\rho = A_p/\sigma_n$
- To minimize P_e, we need to maximize ρ because the Q-function decreases monotonically with ρ
- Can we do better?

Maximizing A_p / σ_n

- Let the received pulse $p(t)$ be time limited to T_o. The maximization of ρ is possible by passing the received pulse through a filter that enhances the pulse amplitude at some time instant t_m and simultaneously reduces the noise power
- Let $p_o(t)$ be the output of $p(t)$ after the filter
- We are interested in finding the filter $H(f)$ that maximizes

$$\rho^2 = p_o^2(t_m) / \sigma_n^2 \qquad (5)$$

Maximizing A_p / σ_n (...)

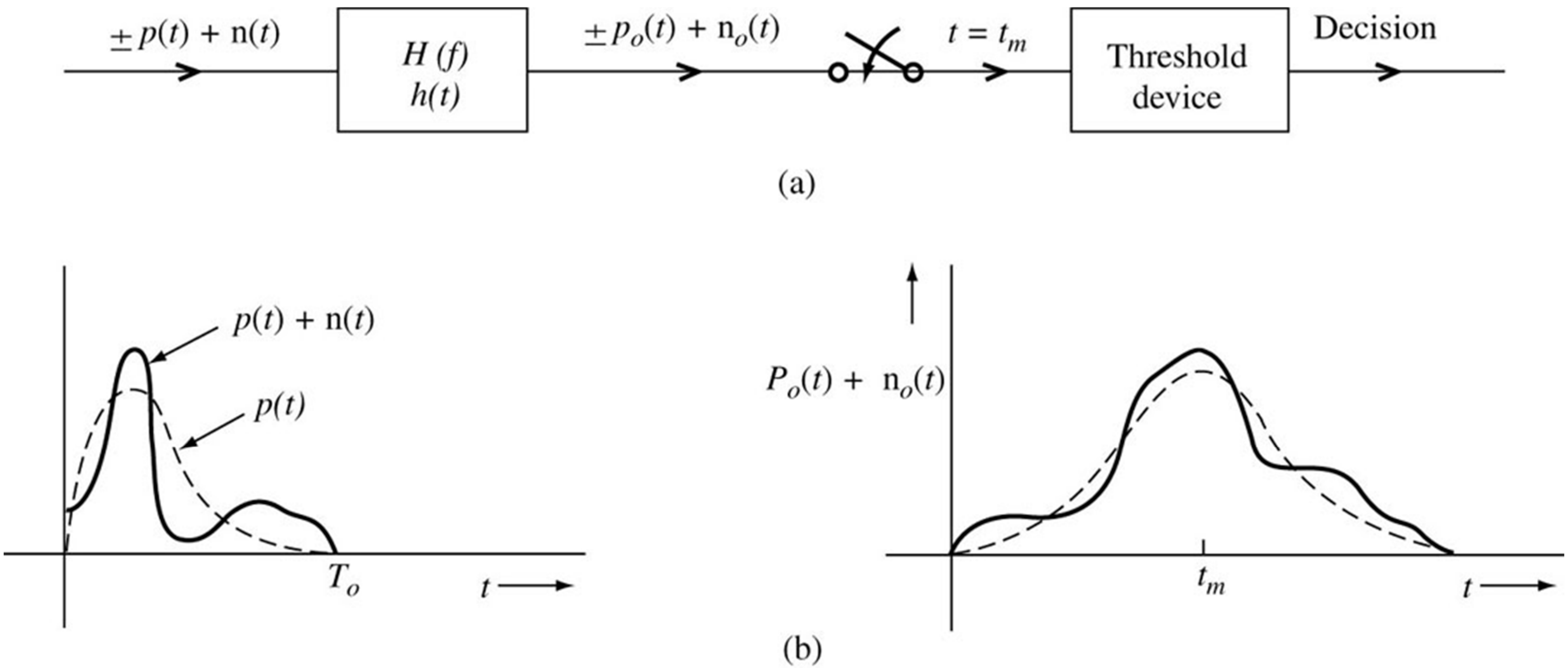


Fig. 7: Typical binary polar signaling and linear receiver

Maximizing A_p / σ_n (...)

- Since $P_o(f) = P(f)\, H(f)$, we have

$$p_o(t) = F^{-1}\left[P_0(f)\right] = \int_{-\infty}^{\infty} P(f) H(f) e^{j2\pi ft}\, df$$

$$p_o(t_m) = \int_{-\infty}^{\infty} P(f) H(f) e^{j2\pi f t_m}\, df \qquad (6)$$

Also,

$$\sigma_n^2 = \int_{-\infty}^{\infty} S_n(f) \left|H(f)\right|^2 df \qquad (7)$$

Hence,

$$\rho^2 = \frac{\left[\int_{-\infty}^{\infty} P(f) H(f) e^{j2\pi f t_m}\, df\right]^2}{\int_{-\infty}^{\infty} S_n(f) \left|H(f)\right|^2 df} \qquad (8)$$

Maximizing A_p / σ_n (...)

- Using the Schwarz inequality,

$$\left| \int_{-\infty}^{\infty} G_1(f) G_2(f)\, df \right|^2 \le \int_{-\infty}^{\infty} \left| G_1(f) \right|^2 df \int_{-\infty}^{\infty} \left| G_2(f) \right|^2 df$$

  with equality if and only if $G_1(f) = k[G_2(f)]^*$, where k is an arbitrary constant and $(\bullet)^*$ denotes complex conjugation.

- Choose

$$G_1(f) = H(f)\sqrt{S_n(f)} \qquad (9)$$

$$G_2(f) = P(f)e^{j2\pi f t_m} / \sqrt{S_n(f)} \qquad (10)$$

Maximizing A_p / σ_n (...)

- Then (8) can be written as

$$\rho^2 \le \frac{\int_{-\infty}^{\infty} \left|H(f)\right|^2 S_n(f)\,df \int_{-\infty}^{\infty} \frac{\left|P(f)\right|^2}{S_n(f)}\,df}{\int_{-\infty}^{\infty} S_n(f)\left|H(f)\right|^2 df}$$

$$\rho^2 \le \int_{-\infty}^{\infty} \frac{\left|P(f)\right|^2}{S_n(f)}\,df \tag{11}$$

AWGN Channels

- For AWGN channels, $S_n(f) = N_0/2$. Hence, (11) becomes

$$\rho_{\max}^2 = \int_{-\infty}^{\infty} \frac{|P(f)|^2}{N_0/2} df = \frac{2}{N_0} \int_{-\infty}^{\infty} |P(f)|^2 df = \frac{2E_p}{N_0} \qquad (12)$$

where

$$E_p = \int_{-\infty}^{\infty} |P(f)|^2 df = \int_{-\infty}^{\infty} |p(t)|^2 dt \qquad (13)$$

is the energy of $p(t)$.

Matched Filter

- For max ρ^2, $G_1(f) = k[G_2(f)]^*$. Hence,

$$H(f)\sqrt{S_n(f)} = k \times \left[\frac{P(f)e^{j2\pi f t_m}}{\sqrt{S_n(f)}}\right]^*$$

$$H(f)\sqrt{S_n(f)} = k \times \frac{P(-f)e^{-j2\pi f t_m}}{\sqrt{S_n(f)}}$$

$$H(f) = k \times \frac{P(-f)e^{-j2\pi f t_m}}{S_n(f)} \qquad (14)$$

Matched Filter (...)

- For AWGN channels, $S_n(f) = N_0/2$. Eqn. (14) becomes

$$H(f) = k \times \frac{P(-f)e^{-j2\pi f t_m}}{N_0/2}$$

$$= k_1 \times P(-f)e^{-j2\pi f t_m} \qquad (15)$$

where $k_1 = 2k/N_0$ is an arbitrary constant

- The impulse response $h(t)$ of the optimum filter is thus

$$h(t) = k_1 \times p(t_m - t) \qquad (16)$$

Sampling Instant

- If $t_m < T_o$, the filter is noncausal and physically unrealizable. If $t_m \geq T_o$, the filter is realizable. Hence, $t_m = T_o$ gives the min delay for decision making using a realizable filter.
- The constant k_1 will affect both the signal and the noise by the same factor. We may conveniently choose $k_1 = 1$ and get

$$h(t) = p(T_o - t) \tag{17}$$

$$H(f) = P(-f)e^{-j2\pi fT_o} \tag{18}$$

Sampling Instant (...)

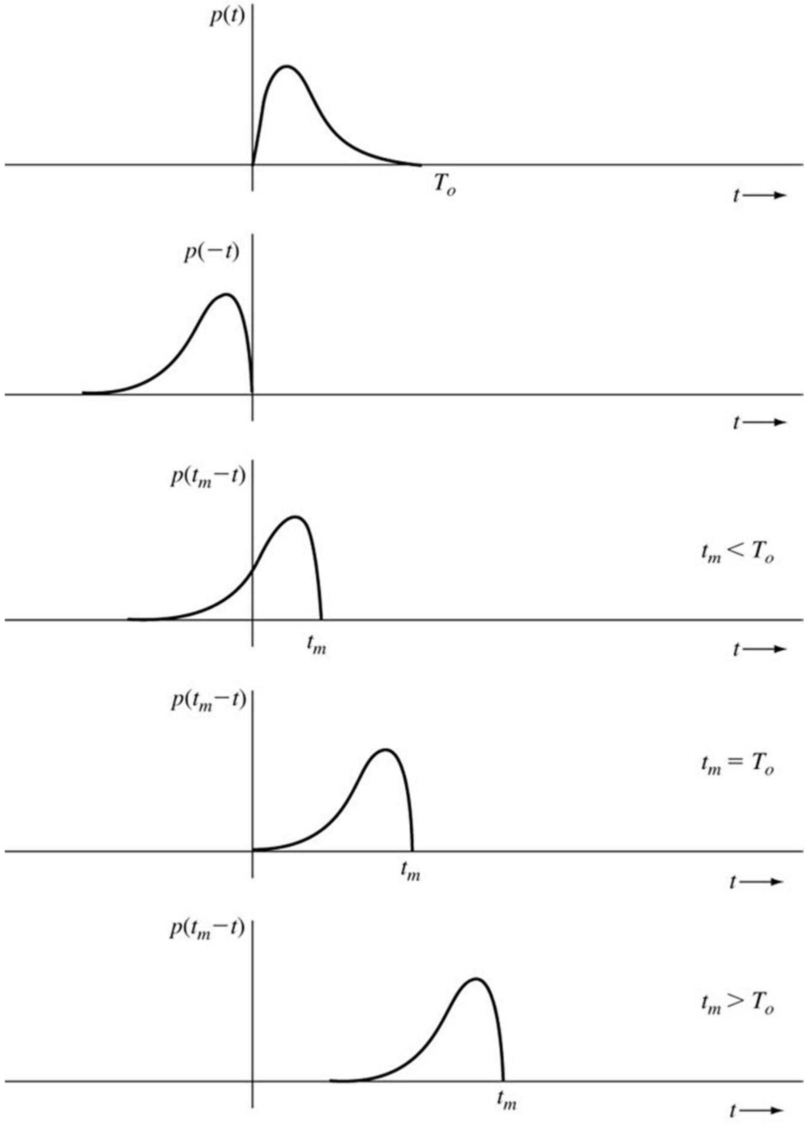


Fig. 8: Optimum choice for sampling instant

Sampling Instant (...)

- The optimum filter in (17) or (18) is known as the *matched filter* (MF) for the AWGN channel. At the output of the filter, the signal to rms noise amplitude ratio, ρ, is max at the decision-making instant $t = T_o$.
- Observe that both $p(t)$ and $h(t)$ have a width of T_o seconds. The filter output, which is the convolution of $p(t)$ and $h(t)$, has a width of $2T_o$ seconds. The peak occurs at $t = T_o$.

Error Prob for MF

- It can be shown that when the noise is white Gaussian, the MF receiver is indeed the optimum filter that minimizes the detection error prob.
- From (12), the *min* detection error prob is

$$P_e = Q(\rho_{\max}) = Q\left(\sqrt{\frac{2E_p}{N_0}}\right) \quad (19)$$

- The error prob depends on $p(t)$ only through its energy E_p, not its detailed waveform.

Correlation Receiver

- Note that the MF output is

$$r(t) = \int_{-\infty}^{\infty} y(x)\, h(t-x)\, dx \qquad (20)$$

where

$$h(t) = p(T_o - t)$$

$$h(t-x) = p(T_o - (t-x)) = p(x + T_o - t)$$

- Hence, at the decision-making instant $t = T_o$, we have

$$r(T_o) = \int_{-\infty}^{\infty} y(x)\, p(x + T_o - T_o)\, dx = \int_0^{T_o} y(x)\, p(x)\, dx \qquad (21)$$

Correlation Receiver (…)

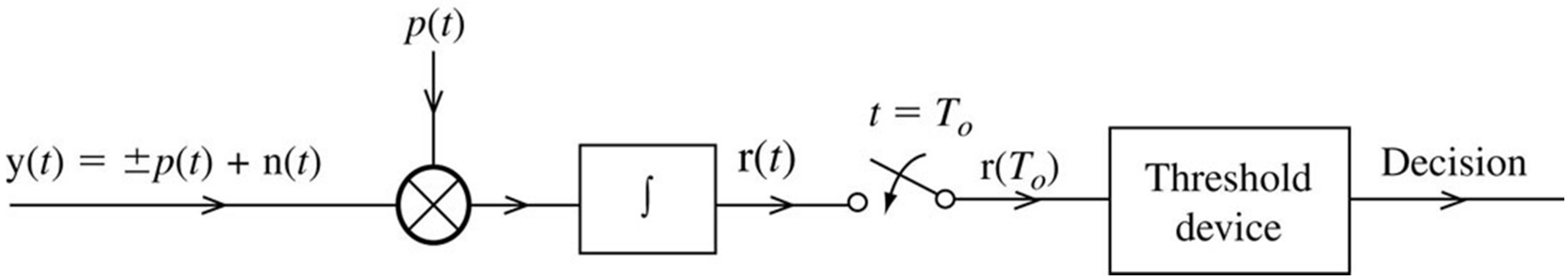


Fig. 9: Correlation detector

Geometric Representation of Signals

Ref:

Lathi and Ding, *Modern Digital and Analog Systems*

(pp. 525 - 536)

Vector Space

- In order to determine the optimum receiver for general *M*-ary signaling over AWGN channels, we need to introduce the concept of signal space.
- We first consider 3 unit vectors as
  $\varphi_1 = (1,0,0)$,
  $\varphi_2 = (0,1,0)$,
  $\varphi_3 = (0,0,1)$.
- These unit vectors are *orthonormal* (orthogonal and unit length) because their dot products (e.g., $\varphi_1 . \varphi_2$) are zero.

Vector Space(...)

- Any vector $\mathbf{x} = (x_1, x_2, x_3)$ in the 3-dimensional space can be expressed as

$$\mathbf{x} = x_1\varphi_1 + x_2\varphi_2 + x_3\varphi_3$$

- $x_k = \mathbf{x}\cdot\varphi_k, \quad k = 1, 2, 3$ (22)
- The length of $\mathbf{x}$ is $|\mathbf{x}|$, defined by

$$|\mathbf{x}|^2 = \mathbf{x}\cdot\mathbf{x} = x_1^2 + x_2^2 + x_3^2 \tag{23}$$

Signal Space

- We define *K orthonormal* (orthogonal and unit length) signals, or waveforms, $\varphi_1(t)$, $\varphi_2(t)$, ..., $\varphi_K(t)$.
- These *K* signals form the *basis signals* of a *K*-dimensional signal space.
- The basis signals are *mutually orthogonal* if

$$\varphi_j(t) \cdot \varphi_k(t) = \begin{cases} 0 & j \neq k \\ A & j = k \end{cases} \tag{24}$$

Signal Space (...)

- The basis signals are *orthonormal* if they are orthogonal and their lengths $A = 1$
- A signal $x(t)$ can be represented by

$$x(t) = x_1\varphi_1(t) + x_2\varphi_2(t) + \cdots + x_K\varphi_K(t)$$

$$= \sum_{k=1}^{K} x_k \varphi_k(t) \qquad (25)$$

- With respect to $\{\varphi_m(t), m = 1, 2, ..., K\}$, $x(t)$ and $(x_1, x_2, ..., x_K)$ are equivalent.

Signal Space (...)

- The coefficient x_k is given by

$$x_k = x(t)\cdot\varphi_k(t) = \int_{-\infty}^{\infty} x(t)\,\varphi_k(t)\,dt$$

- We define the *scalar product* $x(t)\cdot y(t)$ as

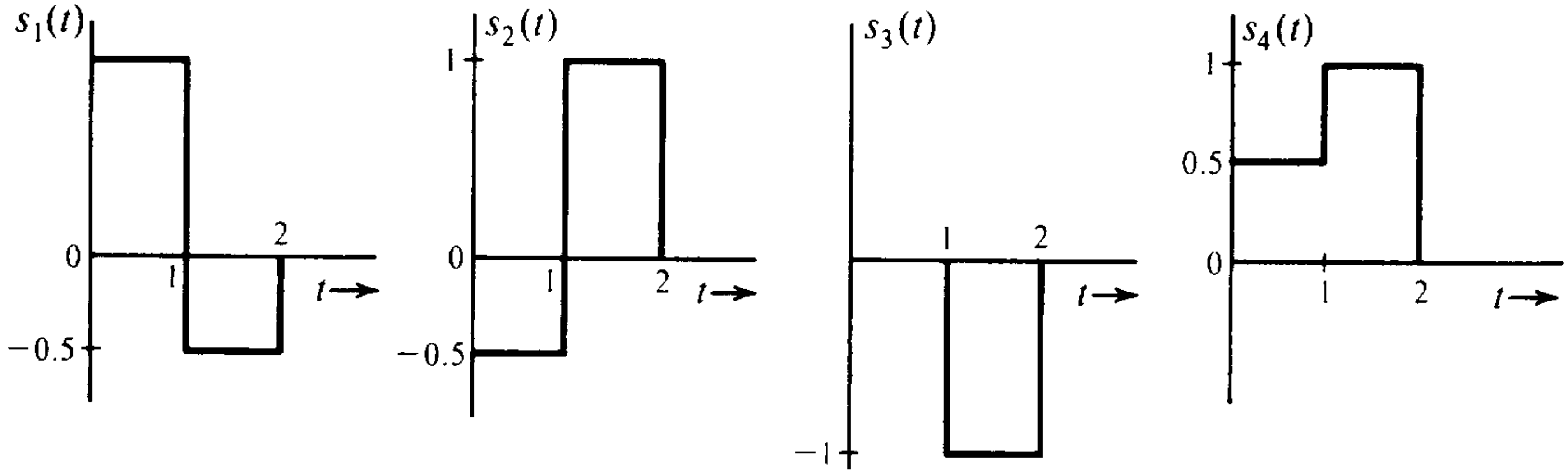
$$x(t)\cdot y(t) = \int_{-\infty}^{\infty} x(t)\,y(t)\,dt = \sum_{k=1}^{K} x_k y_k \qquad (26)$$

- $x(t)$ and $y(t)$ are said to be *orthogonal* if

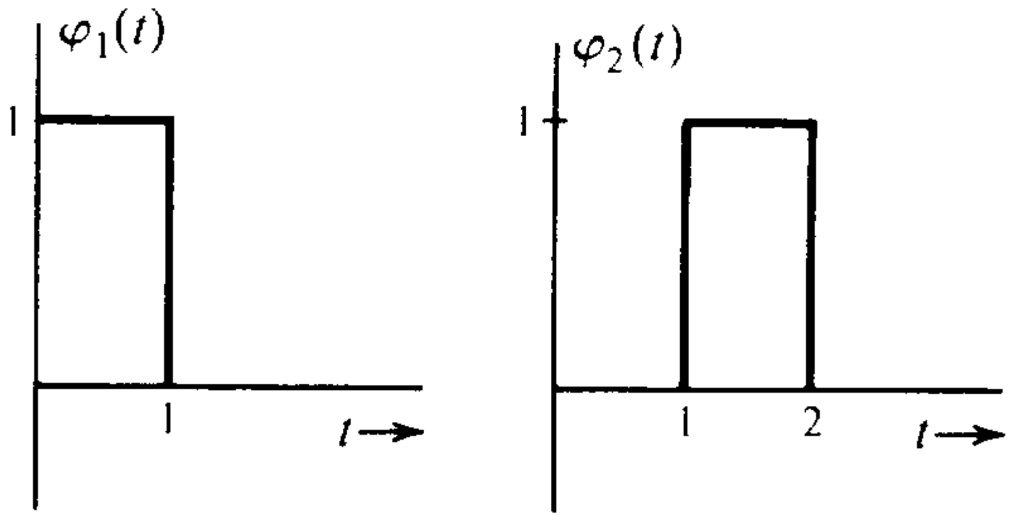
$$x(t)\cdot y(t) = 0 \quad \text{or} \quad \sum_{k=1}^{K} x_k y_k = 0 \qquad (27)$$

Example 2

A signal space consists of four signals $s_1(t)$, $s_2(t)$, $s_3(t)$, and $s_4(t)$. Choose a suitable set of basis signals. Represent these signals geometrically in the signal space.



Example 2 (Solution)

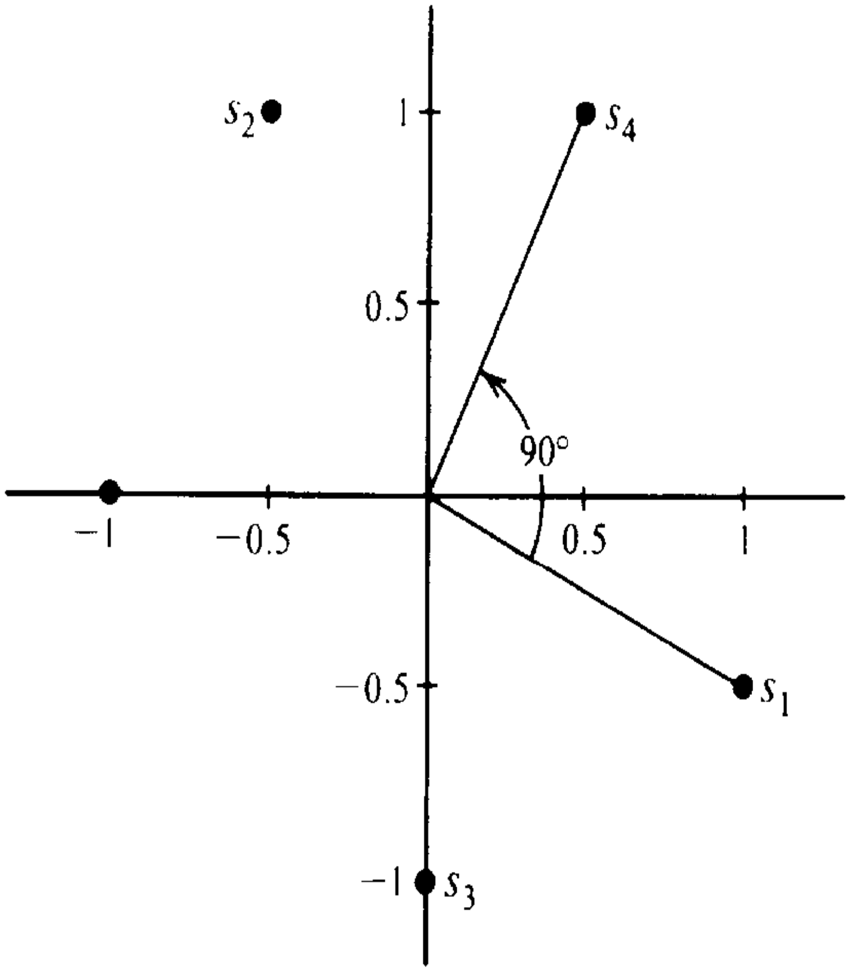


- We may choose $\varphi_1(t)$ and $\varphi_2(t)$ to be the basis signals.
- The signals can be represented as

$\mathbf{s}_1 = (1, -0.5)$ $\mathbf{s}_2 = (-0.5, 1)$

$\mathbf{s}_3 = (0, -1)$ $\mathbf{s}_4 = (0.5, 1)$

Example 2 (Solution)



- Note that $\mathbf{s}_1 \bullet \mathbf{s}_4 = 0.5 - 0.5 = 0$. Hence, they are orthogonal.
- We may also verify this through

$$\int_{-\infty}^{\infty} s_1(t) s_4(t)\, dt = 0$$

Random Process

- A *deterministic signal* can be represented by a point in a signal space.
- A *random process* consists of an ensemble of points in a signal space.
- Let $n(t)$ be a random process given by

$$n(t) = \sum_k n_k \varphi_k(t)$$

where $\{\varphi_k(t)\}$ is a complete set of *orthonormal* basis signals and

Random Process (...)

$$n_k = \int_{-\infty}^{\infty} n(t)\,\varphi_k(t)\,dt$$

- Note that n_k's are random variables (RVs)
- If the RVs are independent, then the joint PDF can be represented as the *product* of marginal PDFs. That is,

$$p_{\mathbf{n}}(\mathbf{n}) = p_{n_1, n_2, \ldots, n_K}(n_1, n_2, \ldots, n_K)$$

$$= p_{n_1}(n_1) \times p_{n_2}(n_2) \times \cdots \times p_{n_K}(n_K)$$

Gaussian PDF

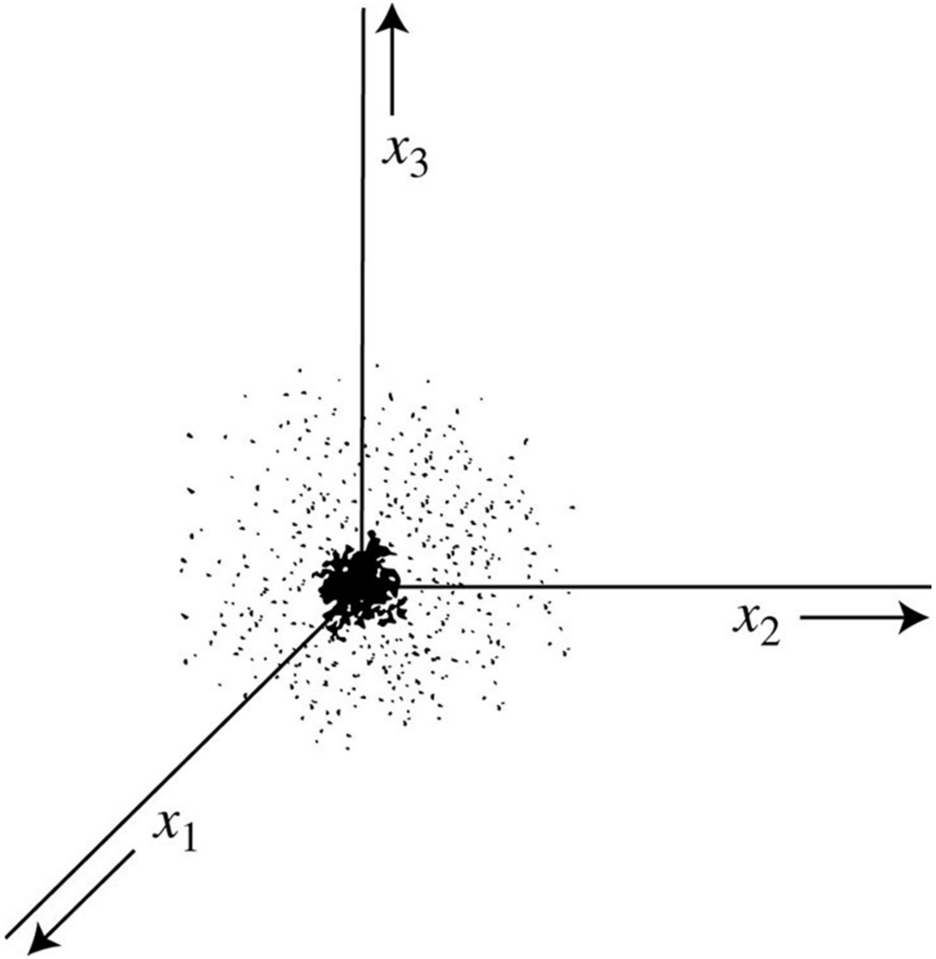
- If n_k's are Gaussian random variables with zero mean and variance $\sigma_n^2 = N_0/2$, then

$$p_{\mathbf{n}}(\mathbf{n}) = p_{n_1}(n_1) \times p_{n_2}(n_2) \times \cdots \times p_{n_K}(n_K)$$

$$= \prod_{k=1}^{K} \frac{1}{\sigma_n \sqrt{2\pi}} e^{-n_k^2/2\sigma_n^2} = \prod_{k=1}^{K} \frac{1}{(\pi N_0)^{1/2}} e^{-n_k^2/N_0}$$

$$= \frac{1}{(\pi N_0)^{K/2}} e^{-(n_1^2+n_2^2+\ldots+n_K^2)/N_0} \tag{28}$$

Geometrical Representation



- According to (28), each sample function will have a specific point **n** and will map into one point in the signal space.

Fig. 10: Geometrical representation of a 3-dimensional Gaussian Random process

Optimum Receiver

Ref:

Lathi and Ding, *Modern Digital and Analog Systems*

(pp. 536 - 545)

Minimum-Error Receiver

- **Problem statement**: *What receiver will yield the minimum-error probability for M-ary communications over AWGN channels?*

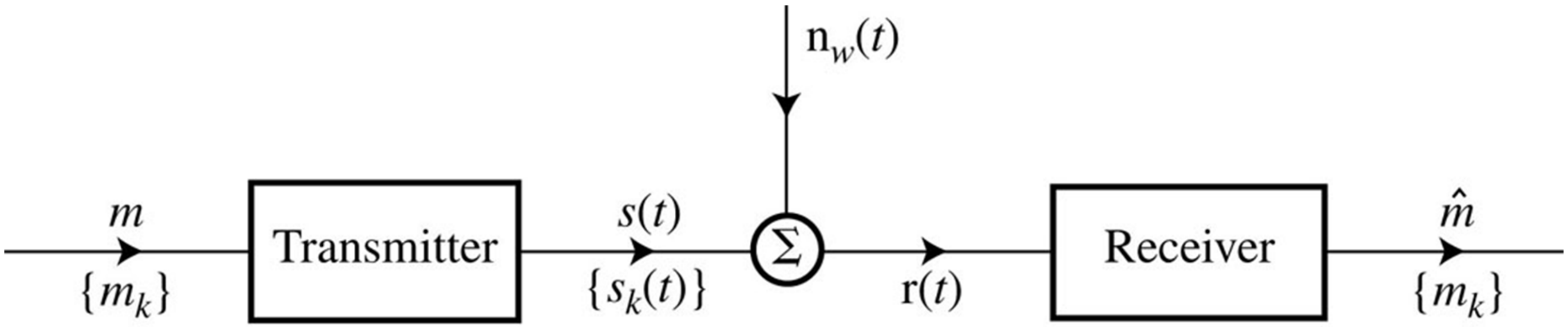


Fig. 11: M-ary communication system

Signals & Noise

- The comprehension of the signal-detection problem is greatly simplified by geometrical representation of signals.
- A signal can be represented by a fixed point (or a vector) in the signal space.
- In the M-ary scheme, we use m_1, m_2, ..., m_M to represent M symbols. The corresponding transmitted waveforms are $s_1(t)$, $s_2(t)$, ..., $s_M(t)$, respectively.

Signals & Noise (...)

- These waveforms are corrupted by AWGN $n_w(t)$ with PSD

$$S_{n_w}(f) = N_0/2$$

- At the receiver, the received signal is

$$r(t) = s_k(t) + n_w(t) \tag{29}$$

  or in vector representation

$$\mathbf{r} = \mathbf{s}_k + \mathbf{n}_w \tag{30}$$

Signals & Noise (...)

- The vector $\mathbf{s}_k$ is a fixed point whereas the vector $\mathbf{n}_w$ is random. Hence, $\mathbf{r}$ is also random.
- Because $\mathbf{n}_w$ is white Gaussian noise, the distribution of $\mathbf{r}$ is a spherical distribution centered at $\mathbf{s}_k$.

Effect of Noise

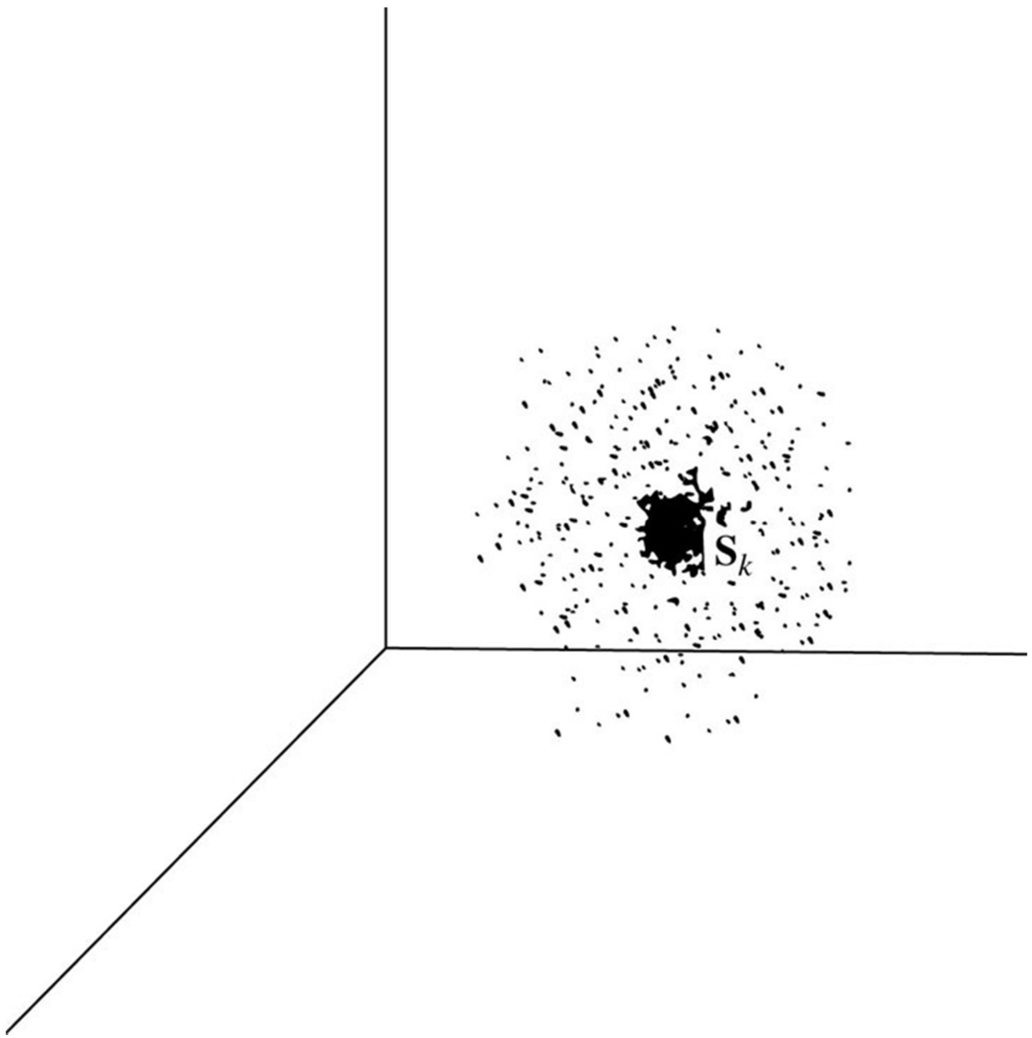


Fig. 12: Effect of Gaussian channel noise on the received signal

In Binary Communications

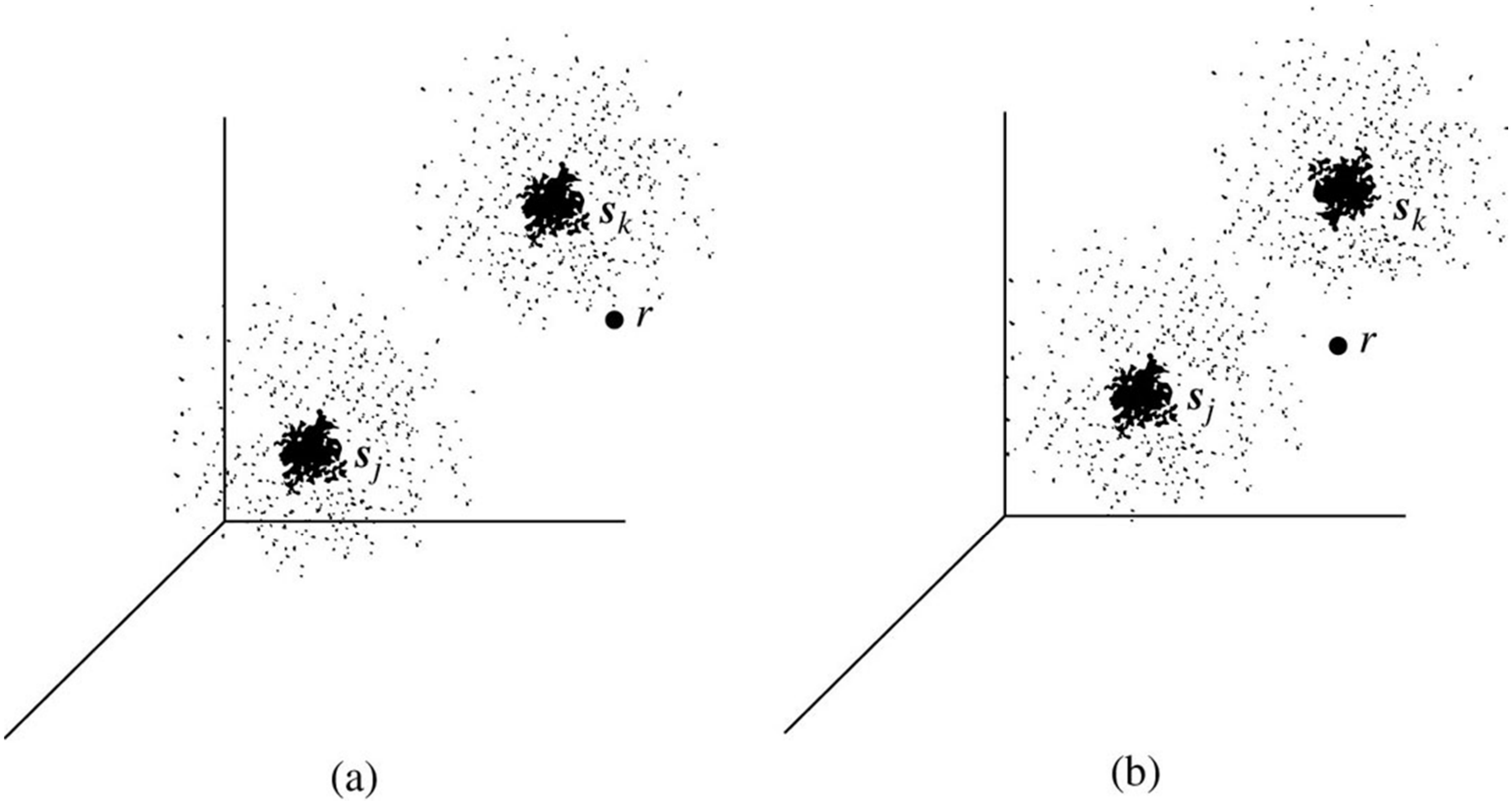


Fig. 13: Binary communication in the presence of noise for (a) high SNR and (b) low SNR

Analysis

- The optimum receiver must decide which message has been sent from the knowledge of $\mathbf{r}$.
- The whole space must be divided into M nonoverlapping regions $R_1, R_2, \ldots, R_M$.
- The receiver must choose proper boundaries such that the error prob is min.
- We assume that the signal space has K dimensions, where $K \leq M$.

Analysis (...)

- Let $\varphi_1(t)$, $\varphi_2(t)$, ..., $\varphi_K(t)$ be the orthonormal basis set for the signal space. Then

$$s_i(t) = \sum_{k=1}^{K} s_{ik}\varphi_k(t), \quad i = 1, 2, \ldots, M \qquad (31)$$

where

$$s_{ik} = s_i(t) \cdot \varphi_k(t) = \int_{-\infty}^{\infty} s_i(t)\varphi_k(t)\,dt \qquad (32)$$

- For the noise $n_w(t)$, the BW is infinite and its dimension is also infinite.

Analysis (…)

- Split $n_w(t)$ into two terms: $n(t)$ is the projection of $n_w(t)$ on the K-dimensional signal space and $n_0(t)$ is the remaining component. That is,

$$n_w(t) = n(t) + n_0(t)$$

$$n(t) = \sum_{k=1}^{K} n_k \varphi_k(t) \qquad n_k = \int_{-\infty}^{\infty} n_w(t)\varphi_k(t)\,dt,$$

$$k = 1, 2, \cdots, K$$

$$n_0(t) = n_w(t) - n(t)$$

(33)

Analysis (...)

- The received signal can be expressed as

$$
\begin{aligned}
r(t) &= s_i(t) + n_w(t) \\
&= s_i(t) + n(t) + n_0(t) \\
&= q(t) + n_0(t)
\end{aligned}
\tag{34}
$$

where $q(t)$ is the portion of $r(t)$ in the K-dimensional signal space given by

$$
q(t) = \sum_{k=1}^{K} (s_{ik} + n_k)\,\varphi_k(t) = \sum_{k=1}^{K} q_k\ \varphi_k(t)
$$

Analysis (...)

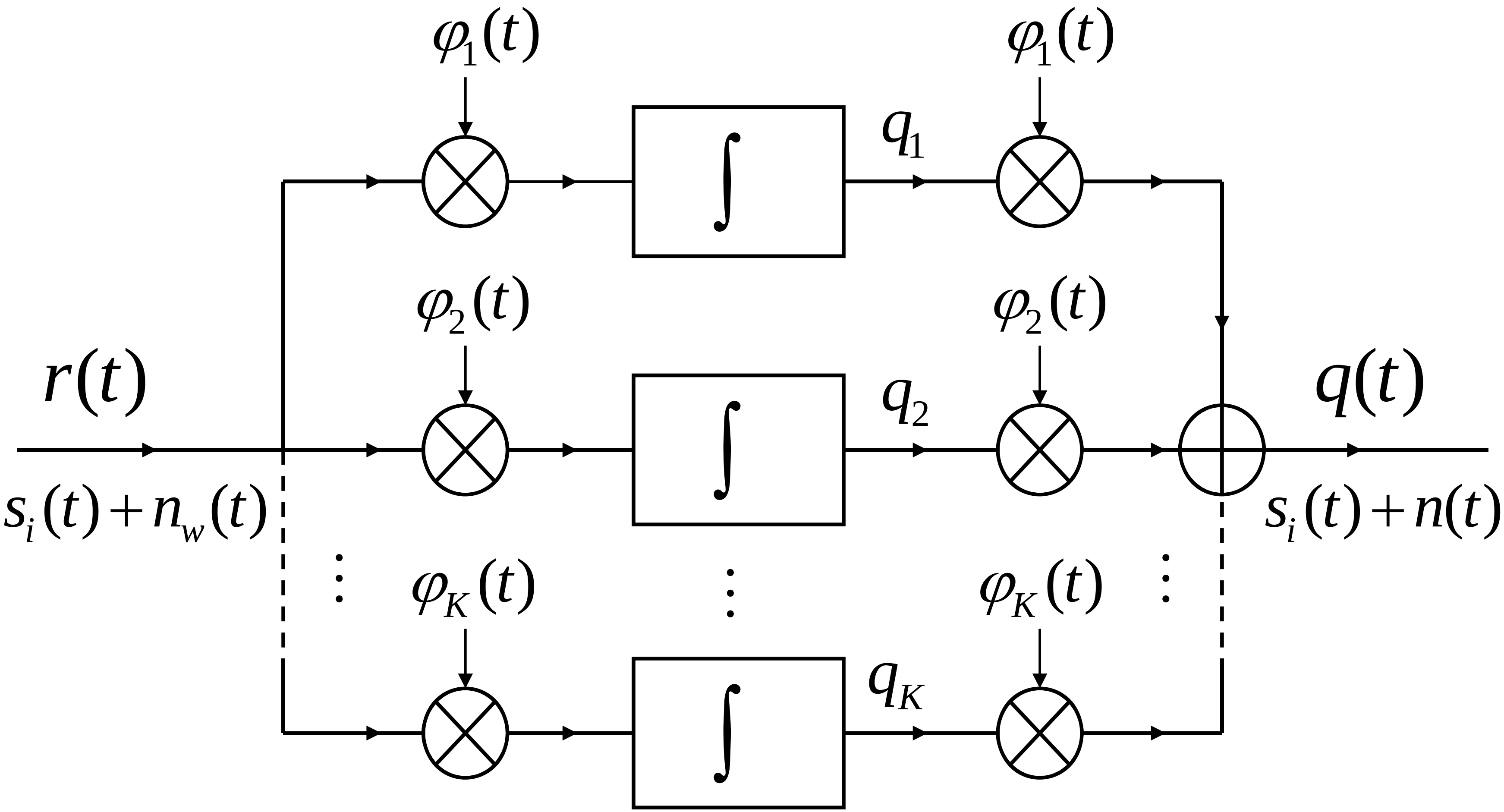


Fig. 14: to generate q(t) from r(t)

Analysis (...)

- We can eliminate $n_0(t)$, which is not relevant to the decision, from $r(t)$ and obtain $q(t)$.
- In vector representation, $q(t)$ is equiv to $\mathbf{q} = \mathbf{s} + \mathbf{n}$, where $\mathbf{s}$ may be any one of the vectors $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_M$.
- The joint PDF of $\mathbf{n}$ has been given in (28).

Decision Procedure

- After obtaining $\mathbf{q}$, the receiver evaluates all the M *a posteriori* probabilities. The receiver decides $\hat{m} = m_i$ if

$$P(m_i \mid \mathbf{q}) > P(m_j \mid \mathbf{q}), \quad j \neq i$$

  where $P(m_j|\mathbf{q})$ is the conditional prob that m_j was transmitted given that $\mathbf{q}$ was received.

- This detector is called the maximum *a posteriori probability* (MAP) detector.

Bayes' Rule

- Based on the Bayes' rule, we have

$$P(m_j \mid \mathbf{q}) = \frac{P\left(m_j\right) p_{\mathbf{q}}(\mathbf{q} \mid m_j)}{p_{\mathbf{q}}(\mathbf{q})}$$

  Note that the denominator is common for all conditional prob's and can be ignored.
- Hence, the receiver decides $\hat{m} = m_i$ if

$$P\left(m_j\right) p_{\mathbf{q}}(\mathbf{q} \mid m_j) < P\left(m_i\right) p_{\mathbf{q}}(\mathbf{q} \mid m_i), \quad j \neq i \tag{35}$$

Decision Function

- Note that $\mathbf{q} = \mathbf{s}_j + \mathbf{n}$ or $\mathbf{n} = \mathbf{q} - \mathbf{s}_j$

$$p_{\mathbf{q}}(\mathbf{q} \mid m_j) = p_{\mathbf{n}}(\mathbf{q} - \mathbf{s}_j) = \frac{1}{(\pi N_0)^{K/2}} e^{-\frac{|\mathbf{q}-\mathbf{s}_j|^2}{N_0}}$$

- The decision function in (35) becomes

$$\frac{P(m_j)}{(\pi N_0)^{K/2}} e^{-\frac{|\mathbf{q}-\mathbf{s}_j|^2}{N_0}}, \quad j = 1, 2, \ldots, M \tag{36}$$

It is always nonnegative. We may just compare the corresponding logarithms.

Decision Function (...)

- By ignoring the common factor $(\pi N_0)^{K/2}$, the decision function becomes

$$
\begin{aligned}
&\ln P(m_j) - \frac{\left|\mathbf{q}-\mathbf{s}_j\right|^2}{N_0} \\
&= \ln P(m_j) - \frac{(\mathbf{q}-\mathbf{s}_j)\cdot(\mathbf{q}-\mathbf{s}_j)}{N_0} \qquad (37) \\
&= \ln P(m_j) - \frac{\left|\mathbf{q}\right|^2 + \left|\mathbf{s}_j\right|^2 - 2\mathbf{q}\cdot\mathbf{s}_j}{N_0}
\end{aligned}
$$

Decision Function (...)

- After multiplying throughout by $N_0/2$, the decision function becomes

$$\frac{N_0}{2}\ln P\left(m_j\right)-\frac{1}{2}\left[\left|\mathbf{q}\right|^2+\left|\mathbf{s}_j\right|^2-2\mathbf{q}\cdot\mathbf{s}_j\right] \qquad (38)$$

- Note that $|\mathbf{s}_j|^2$ is the energy of signal $s_j(t)$. Ignoring the common term $|\mathbf{q}|^2$ and let

$$a_j=\frac{N_0}{2}\ln P\left(m_j\right)-\frac{1}{2}\left|\mathbf{s}_j\right|^2 \qquad (39)$$

Then the new decision function b_j is

$$b_j=a_j+\mathbf{q}\cdot\mathbf{s}_j \qquad j=1,2,\ldots,M \qquad (40)$$

Decision Function (...)

- We compute b_j for $j = 1, 2, ..., M$, and the receiver decides in favor of the maximum b_j
- Note that a_j's can be first computed and stored in a table. We may compute

$$\mathbf{q}\cdot\mathbf{s}_j = \int_{-\infty}^{\infty} q(t)s_j(t)\,dt$$

  using the correlation detector.
- Alternatively, it is the output at $t = T_o$ of a MF with $h(t) = s_j(T_o - t)$ when $q(t)$ is the input.

Decision Function (...)

- Note that

$$\int_{-\infty}^{\infty} r(t) s_j(t)\, dt = \int_{-\infty}^{\infty} [q(t) + n_0(t)] s_j(t)\, dt$$

$$= \int_{-\infty}^{\infty} q(t) s_j(t)\, dt + \int_{-\infty}^{\infty} n_0(t) s_j(t)\, dt$$

$$= \int_{-\infty}^{\infty} q(t) s_j(t)\, dt \qquad (41)$$

- Hence, we may use $r(t)$ or $q(t)$ at the input of the MF or the correlator.

Optimum *M*-ary receiver (1)

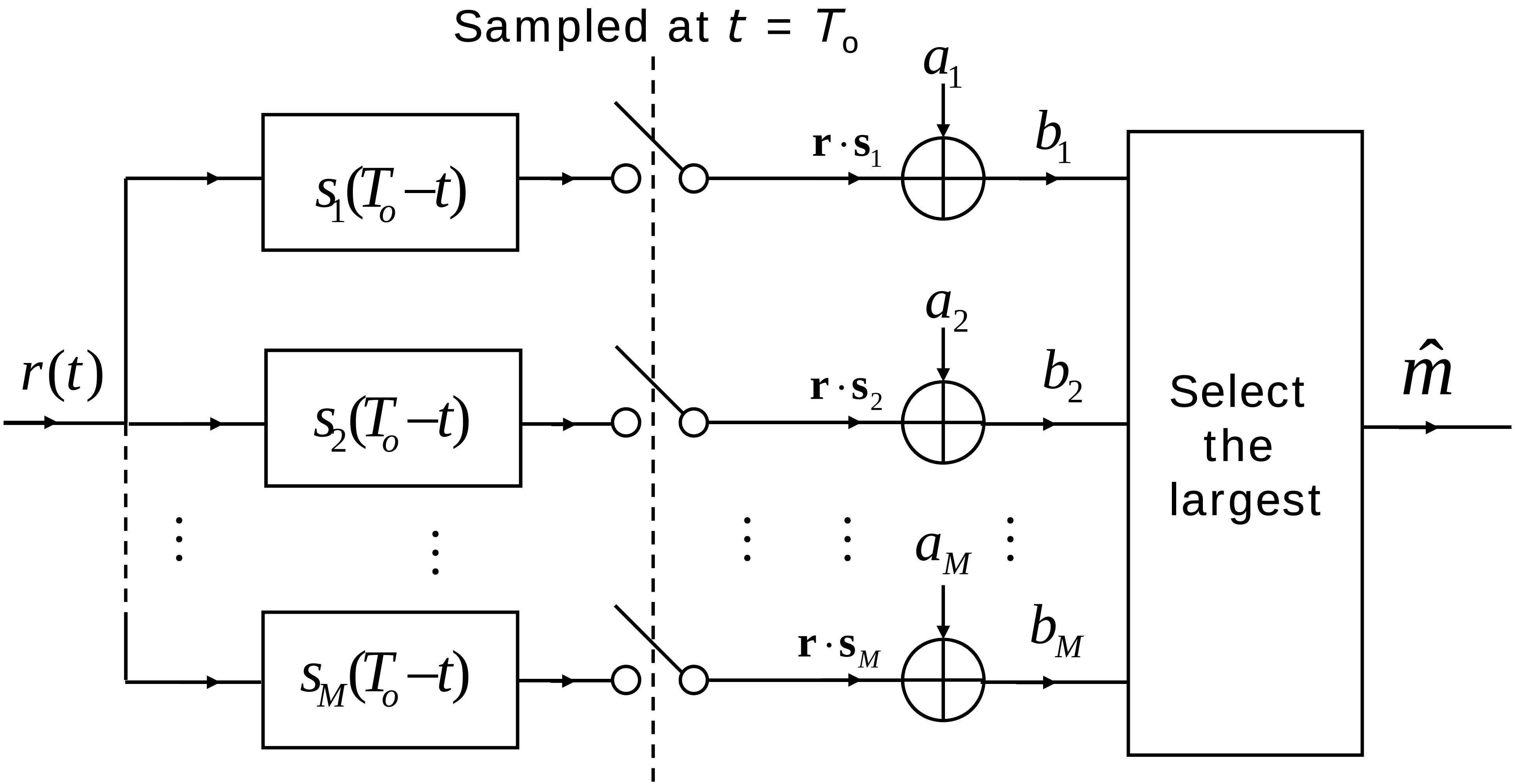


Fig. 15: Matched-filter detector

Optimum *M*-ary receiver (2)

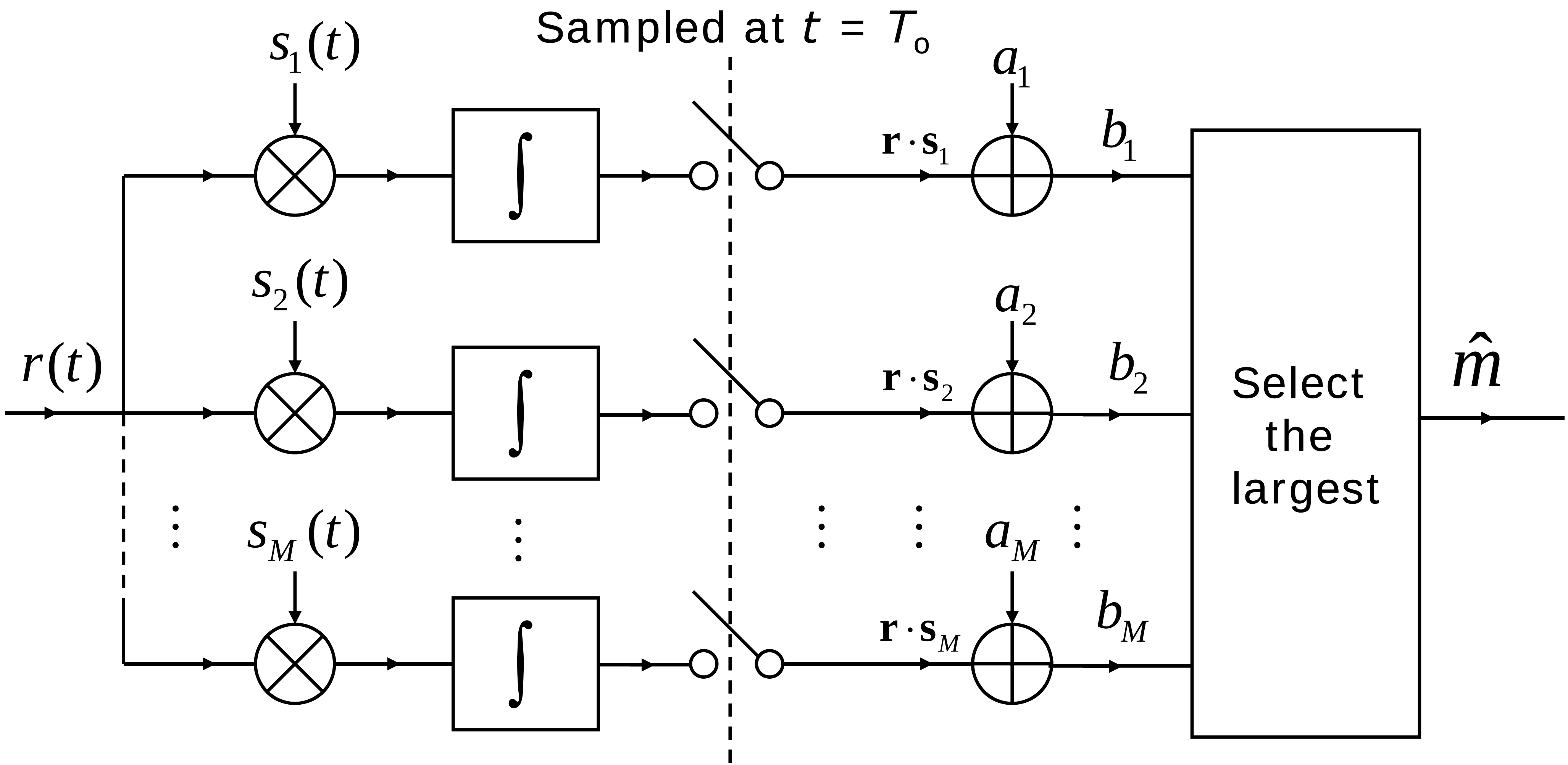


Fig. 16: Correlation detector

Decision Function (...)

- The optimum detector can be implemented in another way as

$$\mathbf{q}\cdot\mathbf{s}_j = \mathbf{r}\cdot\mathbf{s}_j = \sum_{k=1}^{K} r_k s_{jk}$$

- We first generate r_k and then compute $r_k\, s_{jk}$. In this way, we can change the number of MF's from M to K.
- Can be implemented using K correlators.
- All optimum receivers perform identically and the choice depends on the values of M and K.

Optimum *M*-ary receiver (3)

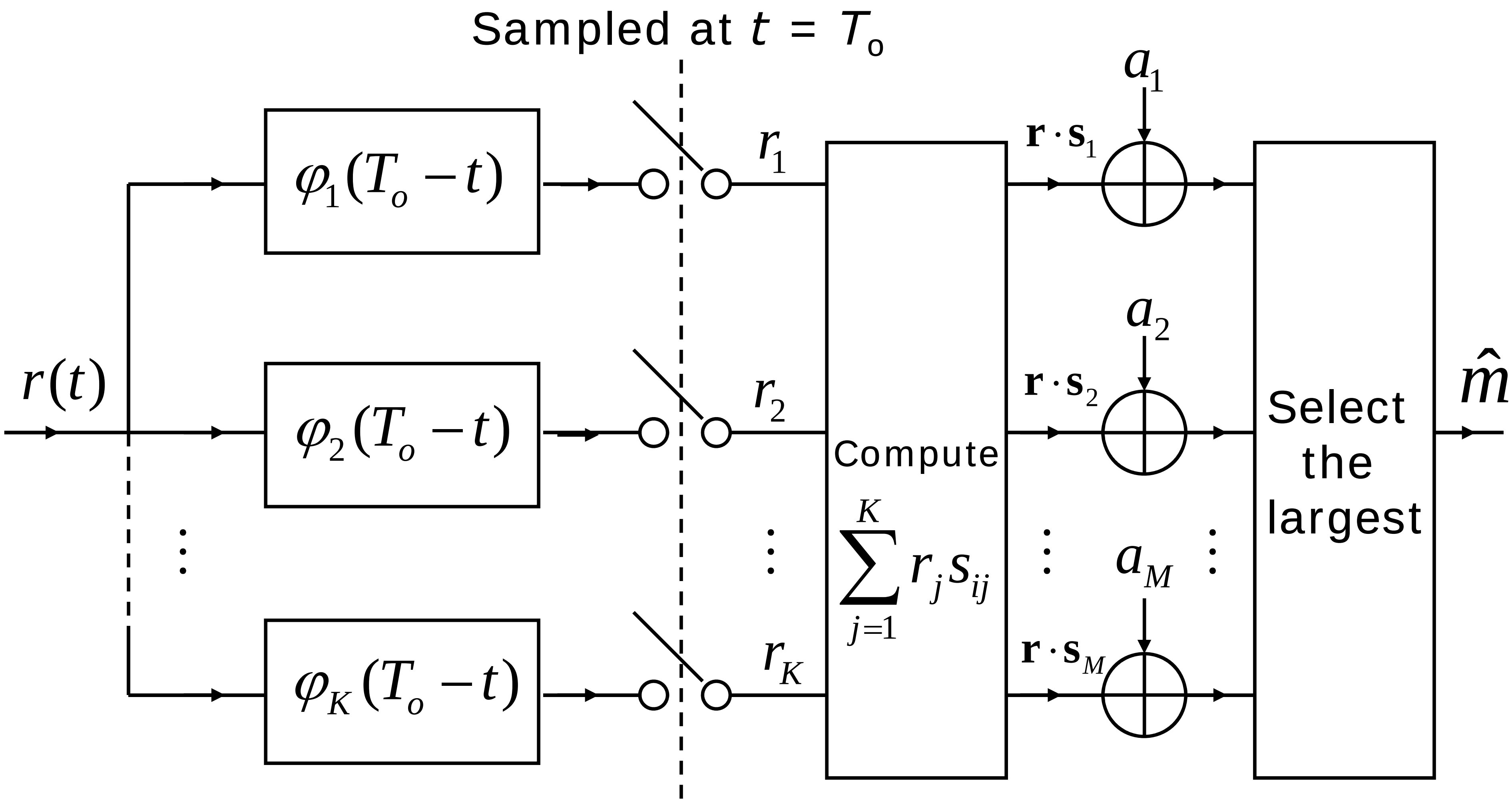


Fig. 17: Matched-filter detector

Optimum *M*-ary receiver (4)

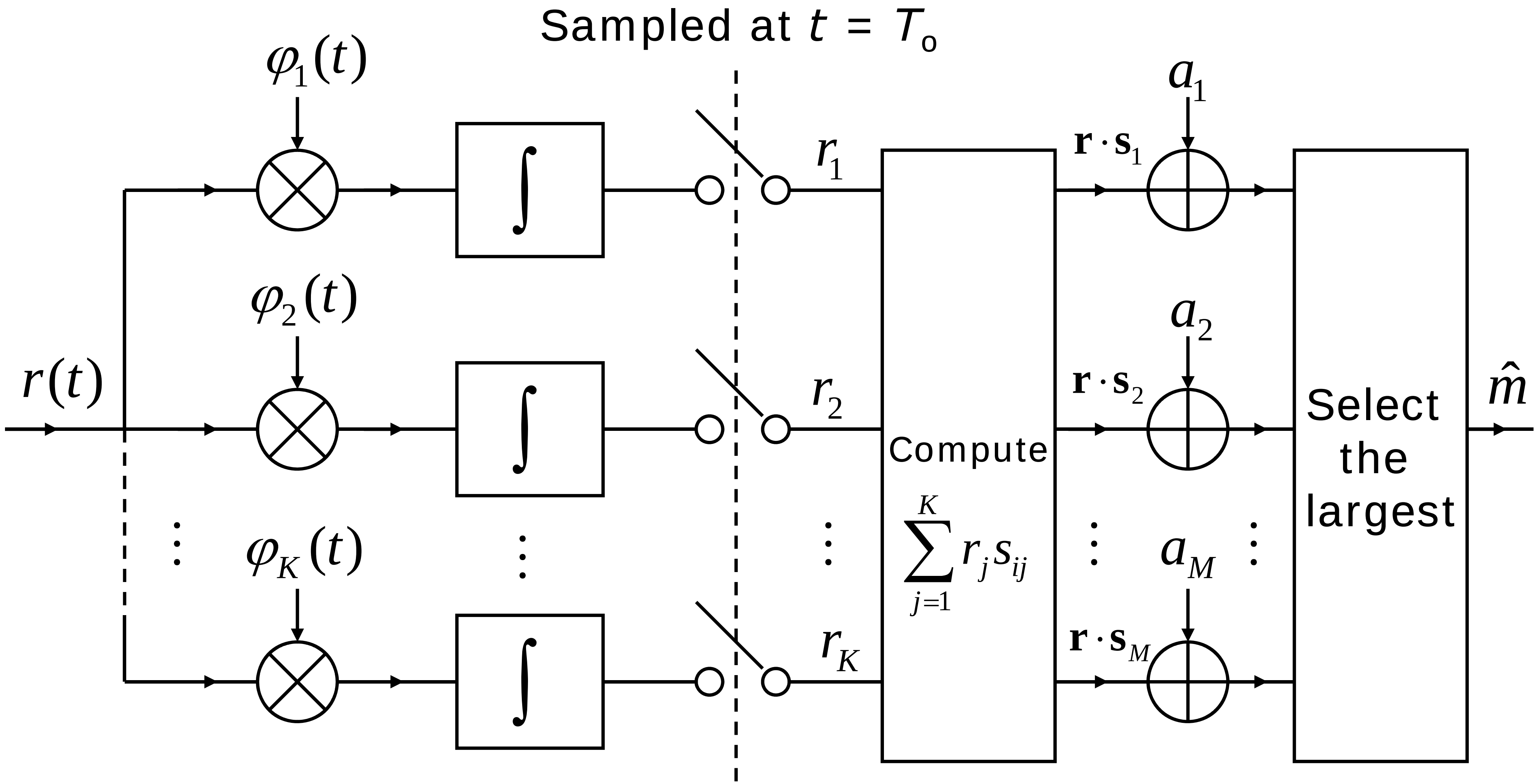


Fig. 18: Correlation detector

Decision Regions and Error probability

Ref:

Lathi and Ding, *Modern Digital and Analog Systems* (pp. 545 - 557)

Decision Regions

- In order to compute the error prob, we need to determine decision regions in the signal space. That is, the signal space is divided into M disjoint regions R_1, R_2, ..., R_M, corresponding to M messages.
- If $\mathbf{q}$ falls in the region R_j, the decision is that m_j was transmitted.
- The decision function given in (37) can be rewritten as

$$N_0 \ln P\left(m_j\right) - \left(\mathbf{q} - \mathbf{s}_j\right) \cdot \left(\mathbf{q} - \mathbf{s}_j\right) \qquad (42)$$

Decision Regions (...)

- If all the M messages are equally likely, i.e., $P(m_j) = 1/M$ for all j, the first term in (42) is common to all and can be ignored
- Maximizing $-|\mathbf{q} - \mathbf{s}_j|^2$ is equiv to minimizing $|\mathbf{q} - \mathbf{s}_j|^2$
- The decision is made in favor of that signal which is closest to $\mathbf{q}$

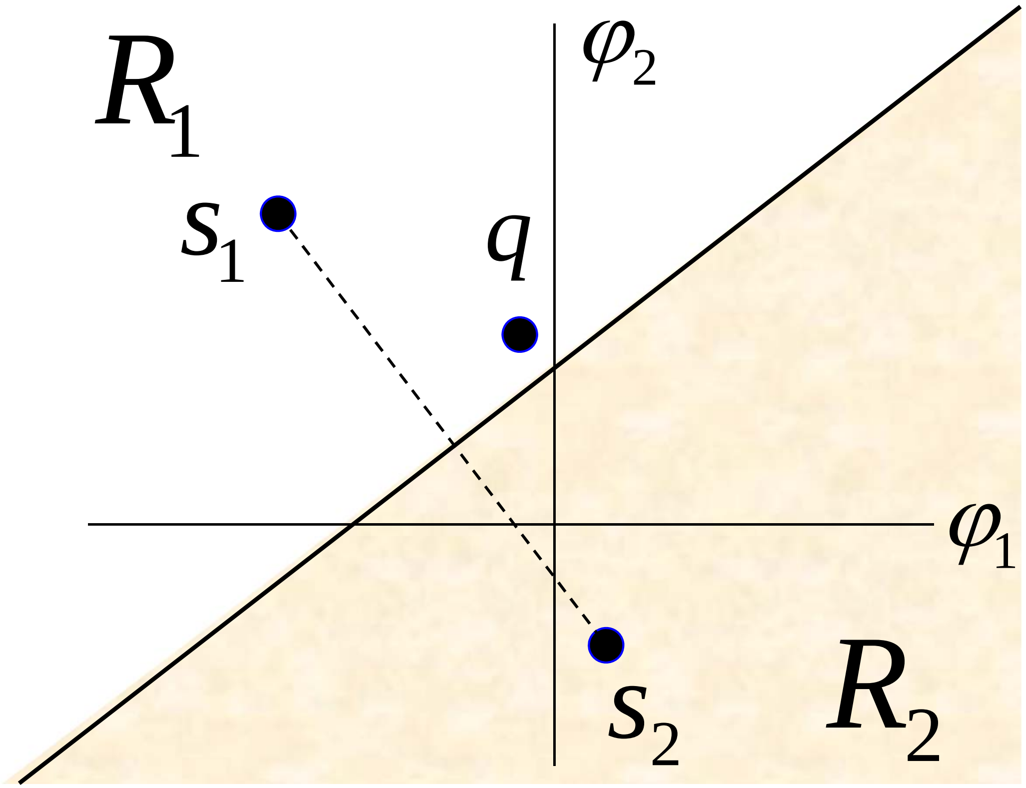


Fig. 19: Optimum decision region (equally likely)

Decision Regions (…)

- If the messages are not equally likely, the decision will be based on (42).
- $d = |\mathbf{s}_1 - \mathbf{s}_2|$
- Boundary line is perpendicular to line $\mathbf{s}_1\mathbf{s}_2$.
- $\mathbf{s}_1$ is at a distance μ from the boundary line.

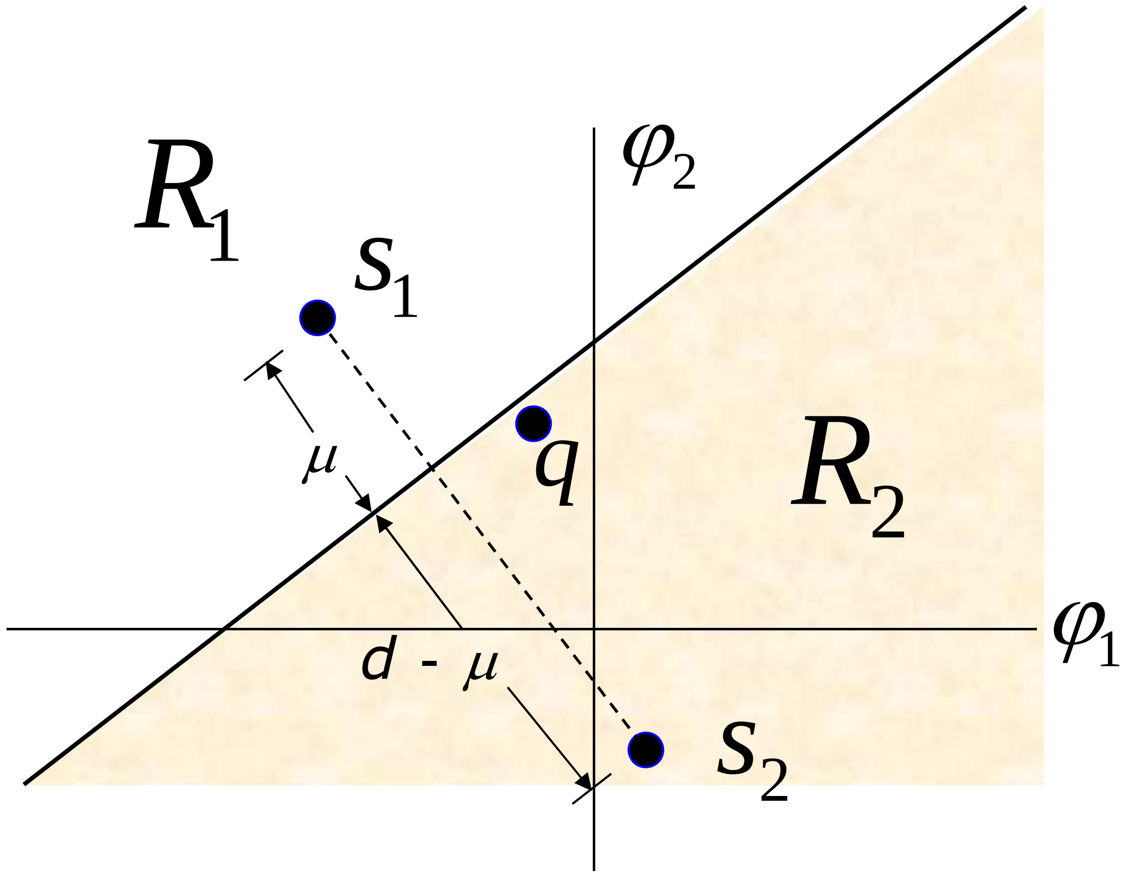


Fig. 20: Optimum decision region (unequal)

Error Prob for Optimum Receiver

- If there are M messages m_1, m_2, ..., m_M with decision regions R_1, R_2, ..., R_M, respectively, then the prob of correct decision, given that m_i was transmitted, is given by

$$P(C \mid m_i) = \Pr\left(\mathbf{q} \in R_i\right)$$

- The average prob of correct decision is

$$P(C) = \sum_{i=1}^{M} P(m_i) \times P(C \mid m_i) \tag{43}$$

- The corresponding prob of error is given by

$$P_{eM} = 1 - P(C) \tag{44}$$

Example 3

Binary data is transmitted using polar signaling over AWGN channel with PSD $N_0/2$. The two signals used are

$$s_1(t) = -s_2(t) = \sqrt{E}\varphi(t)$$

The symbol probabilities $P(m_1)$ and $P(m_2)$ are unequal. Design the optimum receiver and determine the corresponding error prob.

Example 3 (solution)

- The distance between two signals is $d = 2\sqrt{E}$ Based on (42), we have

$$N_0 \ln P(m_1) - \mu^2 = N_0 \ln P(m_2) - (d - \mu)^2$$

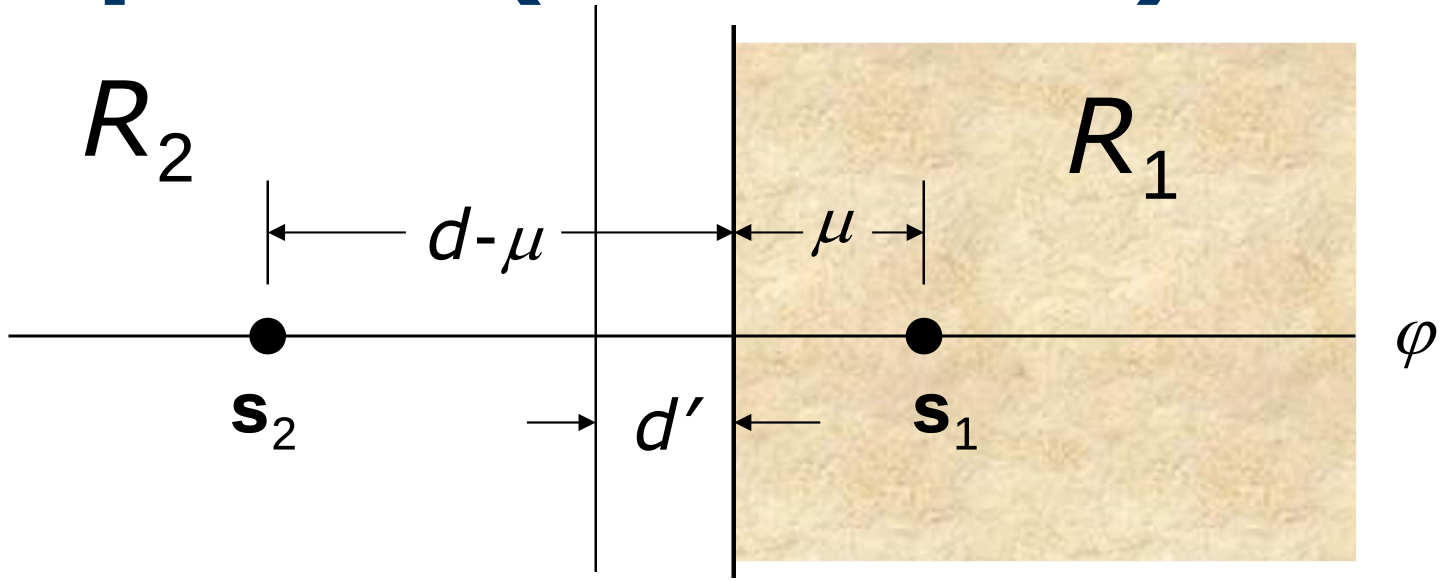
The decision regions are determined by

$$\mu = \frac{N_0}{4\sqrt{E}} \times \ln\left[\frac{P(m_1)}{P(m_2)}\right] + \sqrt{E}$$

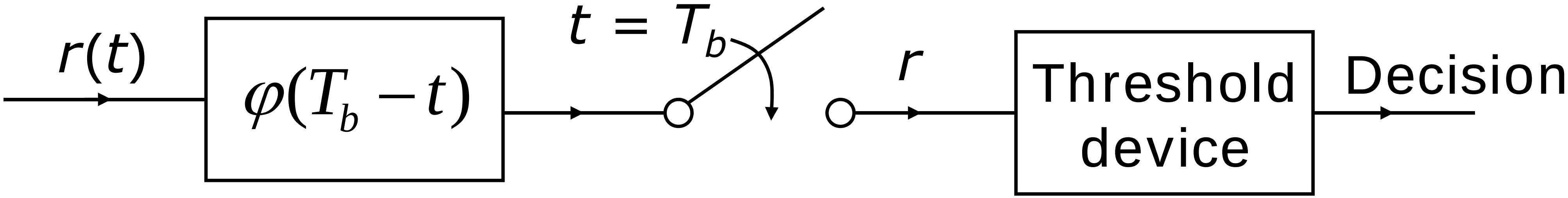
- The conditional correct decision prob is

$$P(C \mid m_1) = \Pr\left(n > -\mu\right) = 1 - Q\left(\frac{\mu}{\sigma_n}\right) = 1 - Q\left(\frac{\mu}{\sqrt{N_0 / 2}}\right)$$

Example 3 (solution)



(a) Optimum decision region



(b) Optimum receiver

Fig. 21: Decision region for a binary case

Example 3 (...)

- Similarly, $P(C\,|\,m_2) = 1 - Q\left(\dfrac{d-\mu}{\sqrt{N_0/2}}\right)$

- The correct decision prob is

$$P(C) = \sum_{i=1}^{2} P(m_i) \times P(C\,|\,m_i)$$

$$= P(m_1)\left[1 - Q\left(\frac{\mu}{\sqrt{N_0/2}}\right)\right] + P(m_2)\left[1 - Q\left(\frac{d-\mu}{\sqrt{N_0/2}}\right)\right]$$

$$= 1 - P(m_1)\,Q\left(\frac{\mu}{\sqrt{N_0/2}}\right) - P(m_2)\,Q\left(\frac{d-\mu}{\sqrt{N_0/2}}\right) \tag{45}$$

Example 3 (...)

- The average error prob $P_e = 1 - P(C)$ is

$$P_e = P(m_1)Q\left(\frac{\mu}{\sqrt{N_0/2}}\right) + P(m_2)Q\left(\frac{d-\mu}{\sqrt{N_0/2}}\right) \tag{46}$$

- If the messages are equally likely, then

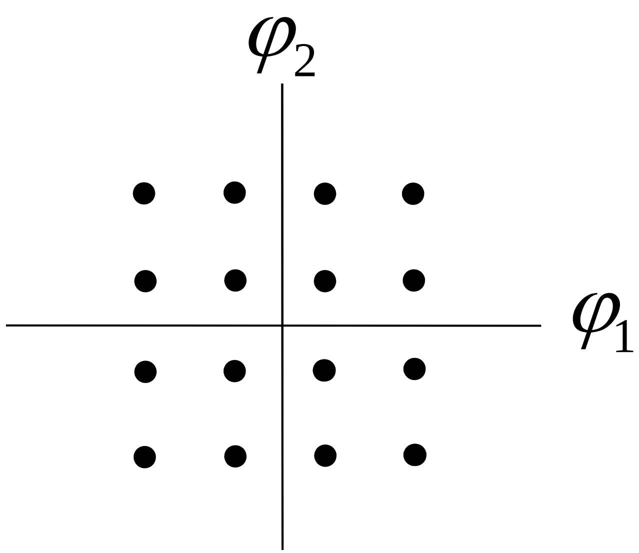
$$P_e = Q\left(\frac{\mu}{\sqrt{N_0/2}}\right) = Q\left(\frac{d/2}{\sqrt{N_0/2}}\right) = Q\left(\sqrt{\frac{2E}{N_0}}\right) \tag{47}$$

- The decision threshold is

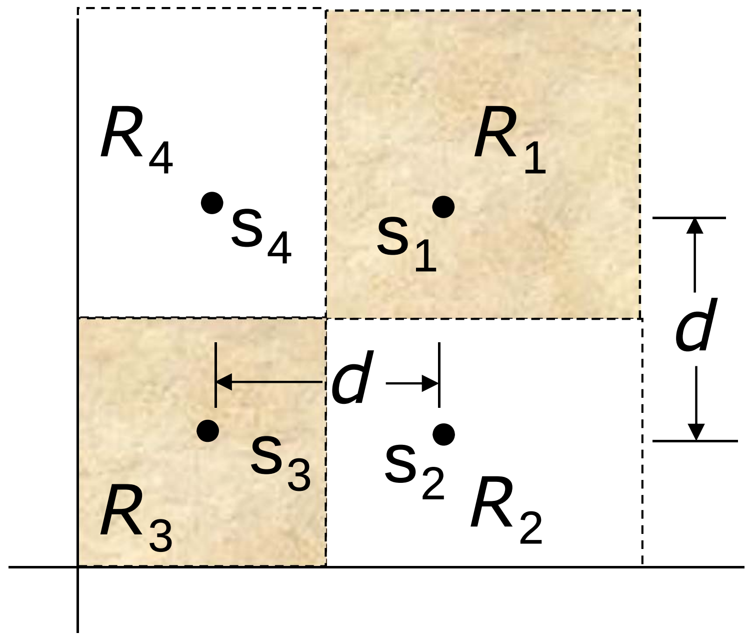
$$d' = \sqrt{E} - \mu = \frac{N_0}{4\sqrt{E}} \times \ln\left[\frac{P(m_2)}{P(m_1)}\right] \tag{48}$$

Example 4

Design the optimum receiver and compute the corresponding error prob for the 16-pt QAM with equiprobable signals over an AWGN channel.



(a) 16-point QAM



(b) Signal space of 1st quadrant

Fig. 22: 16-point QAM

Example 4 (solution)

- Because all points are equiprobable, the decision boundaries will be perpendicular bisectors joining various signals
- Hence,

$$P(C \mid m_1) = \Pr\left(\text{received point in } R_1\right)$$

$$= \Pr\left(n_1 > -d/2 \text{ and } n_2 > -d/2\right)$$

$$= \Pr\left(n_1 > -d/2\right) \times \Pr\left(n_2 > -d/2\right)$$

$$= \left[1 - Q\left(\frac{d/2}{\sigma_n}\right)\right]^2 = \left[1 - Q\left(\frac{d}{\sqrt{2N_0}}\right)\right]^2 = (1-\varepsilon)^2$$

Example 4 (...)

- Using similar arguments, we have

$$P(C \mid m_2) = P(C \mid m_4)$$

$$= \left[1 - Q\left(\frac{d}{\sqrt{2N_0}}\right)\right] \times \left[1 - 2Q\left(\frac{d}{\sqrt{2N_0}}\right)\right]$$

$$= (1-\varepsilon)(1-2\varepsilon)$$

and

$$P(C \mid m_3) = (1-2\varepsilon)^2$$

Example 4 (...)

- Because of the symmetry of the signal points in all 4 quadrants, we get similar results for each quadrant. Hence, the correct-decision prob is

$$P(C)=\sum_{i=1}^{16}P(m_i)\times P\left(C\,|\,m_i\right)=\frac{1}{16}\sum_{i=1}^{16}P\left(C\,|\,m_i\right)$$

$$=\frac{1}{16}\left[4(1-\varepsilon)^2+8(1-\varepsilon)(1-2\varepsilon)+4(1-2\varepsilon)^2\right]$$

$$=\frac{1}{4}\left[4-12\varepsilon+9\varepsilon^2\right] \qquad (49)$$

Example 4 (...)

- The error prob is

$$
\begin{aligned}
P_{eM} &= 1 - P(C) \\
&= 1 - \frac{1}{4}\left[4 - 12\varepsilon + 9\varepsilon^2\right] \\
&= 3\varepsilon - \frac{9}{4}\varepsilon^2
\end{aligned} \tag{50}
$$

- In practice, $d/\sqrt{N_0} >> 1 \Rightarrow \varepsilon$ is very small and

$$
P_{eM} \approx 3\varepsilon = 3Q\left(\frac{d}{\sqrt{2N_0}}\right) \tag{51}
$$

Example 4 (...)

- In this case, $K = 2$ and $M = 16$. Hence, the preferred receiver uses $K = 2$ correlators
- From (40), we have

$$b_i = a_i + \mathbf{q} \cdot \mathbf{s}_i = a_i + \mathbf{r} \cdot \mathbf{s}_i$$

where $a_i = N_0 \ln P(m_i)/2 - |\mathbf{s}_i|^2/2$

- Note that the first term in a_i is common and can be ignored. Hence, we may just use

$$a_i = -|\mathbf{s}_i|^2/2 = -E_i/2$$

Example 4 (...)

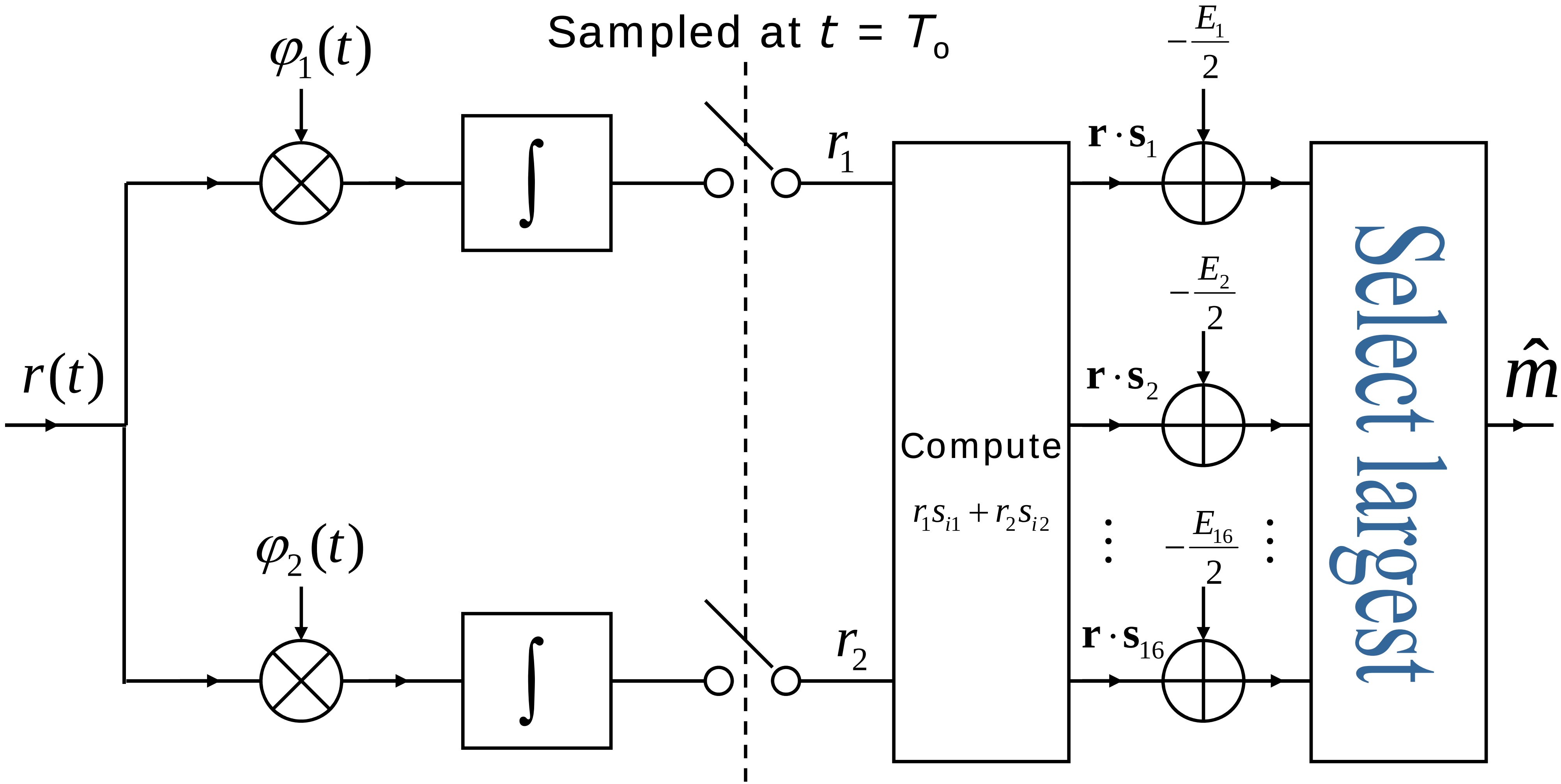


Fig. 23: Implementation of optimum receiver

Example 4 (...)

- For QAM with signal duration T_o seconds, we use

$$\varphi_1(t) = \sqrt{\frac{2}{T_o}} \cos 2\pi f_0 t \qquad \varphi_2(t) = \sqrt{\frac{2}{T_o}} \sin 2\pi f_0 t$$

where $\varphi_1(t)$ and $\varphi_2(t)$ are orthonormal basis signals.

Observations

- Whenever the optimum receiver is used, the error probability does not depend on specific signal waveforms but depends only on their geometrical configuration in the signal space.
- Error probability depends on signal waveform only through the average energy of the set.
- The average signal energy emerges as a fundamental parameter that determines the error probability.