ADVANCED COMMUNICATION SYSTEMS

ELEN90051 (LECTURER: MARGRETA KUIJPER)

Digital Demodulation

1st Semester 2018

Material based on Chapters 2 & 4 of "Digital Communications" by Proakis & Salehi, 2008, and Chapter 4 & 7 of "Communication Systems Engineering" by Proakis & Salehi, 2002, and Chapter 5 of "Digital Communications—a Discrete-Time Approach" by Michael Rice, 2009

Digital Communication System

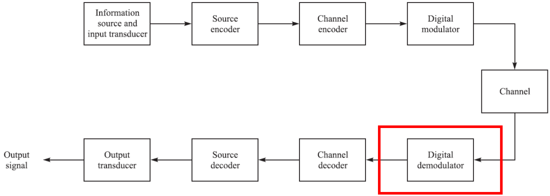
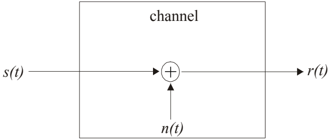


FIGURE 1.1–1
Basic elements of a digital communication system.

Digital demodulation for AWGN channels

- AWGN channels
    - Gaussian random variables
    - Gaussian random processes
- Signal demodulation and detection
    - Correlation-type demodulator
    - Matched filter-type demodulator
    - MAP and ML detector
- Non-coherent demodulation

AWGN Channel Model



- Received signal for additive white Gaussian noise (AWGN) channel is given by

$$r(t) = s_m(t) + n(t)$$

  where $n(t)$ is a random variable from an AWGN random noise process with power spectral density $\frac{N_0}{2}$ W/Hz.
- Given $r(t)$, how to decide which $s_m(t)$ was transmitted?
  $\Rightarrow$ Design an optimum receiver that minimizes the error probability (or some other criteria)

Gaussian Random Variables

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ be a Gaussian random variable (rv) with mean μ and variance σ^2, then its probability density function (PDF) is given by

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

and its cumulative distribution function (CDF) is given by

$$F(x) = \Pr[X \leq x] = \int_{-\infty}^{x} p(u)du = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} \exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right) du$$

A standard Gaussian rv is defined as $X \sim \mathcal{N}(0, 1)$ with

PDF: $p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$

CDF: $\Phi(x) = \Pr[X \leq x] = \int_{-\infty}^{x} p(u)du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{u^2}{2}\right) du$

Gaussian Random Variables

- The *Q*-function is a useful function for calculating tail probabilities of *standard Gaussian rv's* $X \sim \mathcal{N}(0, 1)$ defined as

$$Q(x) = \Pr[X > x] = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{t^2}{2}\right) dt$$

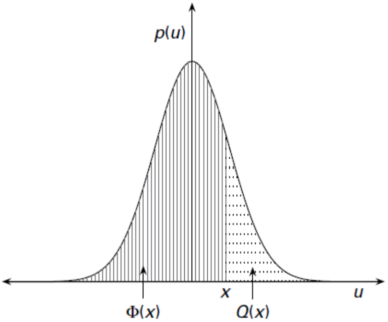
Important properties of the Q-function are

$$Q(0) = \frac{1}{2}$$

$$Q(\infty) = 0$$

$$Q(-\infty) = 1$$

$$Q(-x) = \Phi(x) = 1 - Q(x)$$



Gaussian Random Variables

- For *general Gaussian rv's* $X \sim \mathcal{N}(\mu, \sigma^2)$, we know that $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$ and therefore we can write

$$\Pr[X > x] = Q\left(\frac{x-\mu}{\sigma}\right)$$

$$\Pr[X < x] = Q\left(\frac{\mu-x}{\sigma}\right)$$

- For example, if $n \sim \mathcal{N}(0, \frac{N_0}{2})$, then

$$\Pr[n > x] = Q\left(\frac{x}{\sqrt{N_0/2}}\right).$$

Random Processes

- Continuous-time Random Process: $X(t)$, where "time' $t \in \mathbb{R}$
- the process is completely characterized by joint PDF's of the form

$$f_{X(t_1),X(t_2),\ldots,X(t_n)}(x_1, x_2, \ldots, x_n),$$

where $n \in \mathbb{Z}_+$ and $t_1, t_2, \ldots, t_n \in \mathbb{R}$

- the process $X(t)$ is strictly stationary if for all n, Δ, $(t_1, t_2, \ldots, t_n)$ we have

$$\begin{aligned} & f_{X(t_1),X(t_2),\ldots,X(t_n)}(x_1, x_2, \ldots, x_n) \\ &= f_{X(t_1+\Delta),X(t_2+\Delta),\ldots,X(t_n+\Delta)}(x_1, x_2, \ldots, x_n) \end{aligned}$$

This is a strong condition

- for fixed $t_1 \in \mathbb{R}$ we have that $X(t_1)$ is a random variable

WSS Random Processes

- Important properties of random processes

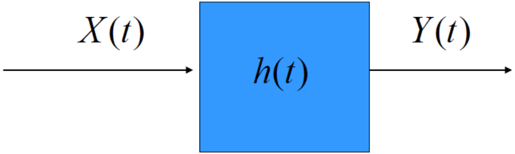
$$\text{Mean: } m_X(t) = E[X(t)]$$

$$\text{Autocorrelation: } R_X(t_1, t_2) = E[X(t_1)X^*(t_2)]$$

- Random process $X(t)$ is wide sense stationary (WSS) if
  - $m_x(t)$ is a constant $\Rightarrow m_X(t) = m_X$ for all t.
  - $R_X(t_1, t_2)$ depends only on $\tau = t_1 - t_2$, we can therefore write $R_X(\tau)$ rather than $R_X(t_1, t_2)$.
- Power spectral density (PSD) of a WSS process is the Fourier transform of autocorrelation $\mathcal{S}_X(f) = \mathscr{F}[R_X(\tau)]$. Total power content of the process is

$$P_X = E\left[|X(t)|^2\right] = R_X(0) = \int_{-\infty}^{\infty} \mathcal{S}_X(f)df$$

WSS Random Processes



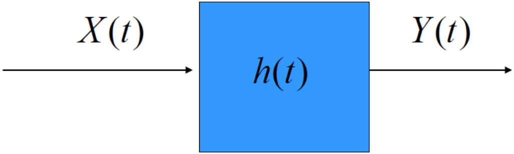
- Linear time-invariant (LTI) systems: If a WSS process $X(t)$ passes through a LTI system with impulse response $h(t)$ and frequency response $H(f)$, the output $Y(t) = \int_{-\infty}^{\infty} X(\tau)h(t-\tau)d\tau$ is also WSS with properties:

Time domain	Frequency domain
$m_Y = m_X \int_{-\infty}^{\infty} h(t)dt$	$m_Y = m_X H(0)$
$R_Y = R_X \star h \star \tilde{h}$	$\mathcal{S}_Y(f) = \mathcal{S}_X(f)\|H(f)\|^2,$

where $\tilde{h}(t) := h^*(-t)$.

Gaussian Random Processes

- $X(t)$ is a Gaussian random process if $\{X(t_1), X(t_2), \ldots, X(t_n)\}$ has jointly Gaussian PDF. Then $X(t_k)$ is a Gaussian rv for any fixed $t_k \in \mathbb{R}$.
- $X(t)$ is a white process if its psd $\mathcal{S}_X(f)$ is constant for all frequencies. Power content of white process $P_X = \int_{-\infty}^{\infty} \mathcal{S}_X(f) = \infty$ so not physically realizable.

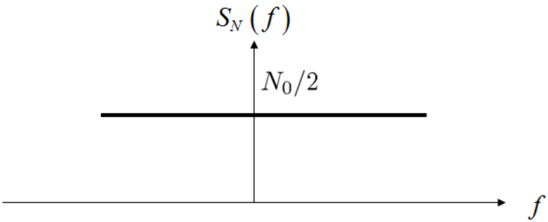


- If LTI-input $X(t)$ is a Gaussian random process, the LTI-output $Y(t)$ is also a Gaussian random process.
- If LTI-input $X(t)$ is white, LTI-output $Y(t)$ is not necessarily white (*Recall:* $\mathcal{S}_Y(f) = \mathcal{S}_X(f)|H(f)|^2$).

AWGN Random Process

Thermal noise can be modeled as a random process $N(t)$ which is

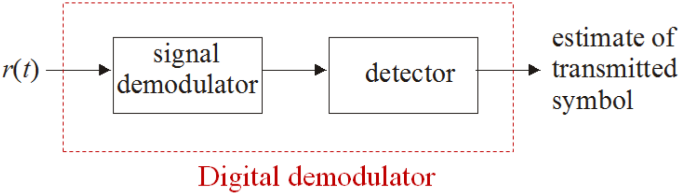
- Wide Sense Stationary: Mean $m_N(t)$ independent of time and Autocorrelation depends only on $\tau = t_1 - t_2$ with $R_N(\tau) = \mathscr{F}^{-1}[\mathcal{S}_N(f)] = \frac{N_0}{2}\delta(\tau)$
- Zero-mean: $m_N(t) = 0$
- Gaussian: $N(t) \sim \mathcal{N}(0, \frac{N_0}{2})$
- White: $\mathcal{S}_N(f) = \frac{N_0}{2}$ for all f



Demodulation

Two major steps in receiver design

- *Step 1:* The signal demodulator **projects** the received waveform $r(t)$ onto a signal $\mathbf{r} = (r_1, r_2, \ldots, r_N)$ in the signal space. Here N is the dimension of the signal space.
- *Step 2:* The detector **decides** which of the M possible signal waveforms was transmitted based on $\mathbf{r}$.



Projection

- Any signal $s_m(t)$ in the N-dimensional signal space S with orthonormal basis $\{\phi_k(t)\}_{k=1}^{N}$ can be uniquely written as

$$s_m(t) = \sum_{k=1}^{N} s_{mk}\phi_k(t),$$

  where $s_{mk} = \langle s_m(t), \phi_k(t)\rangle$, and $s_m(t)$ can be completely specified by the vector $\mathbf{s}_m = (s_{m1}, \ldots, s_{mN})$ for $m = 1, 2, \ldots, M$.

- If $s_m(t)$ is transmitted through an AWGN channel, then received signal is $r(t) = s_m(t) + n(t)$.

- How to convert $r(t)$ into a vector $\mathbf{r} = (r_1, r_2, \ldots, r_N)$?
  $\Rightarrow$ Project $r(t)$ onto the N-dimensional signal space S

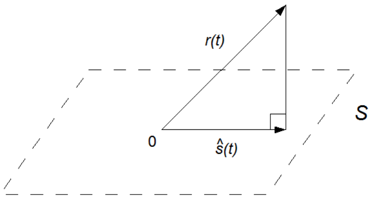
Recall: Orthonormal Bases

Question:

- Given a received signal $r(t)$ outside a subspace S and an orthonormal basis $\{\phi_k(t), k = 1, 2, \ldots, N\}$ for S, which $\widehat{s}(t)$ in S is closest to $r(t)$? Or which $\widehat{s}(t) \in S$ minimizes $||r(t) - \widehat{s}(t)||$?

Answer:

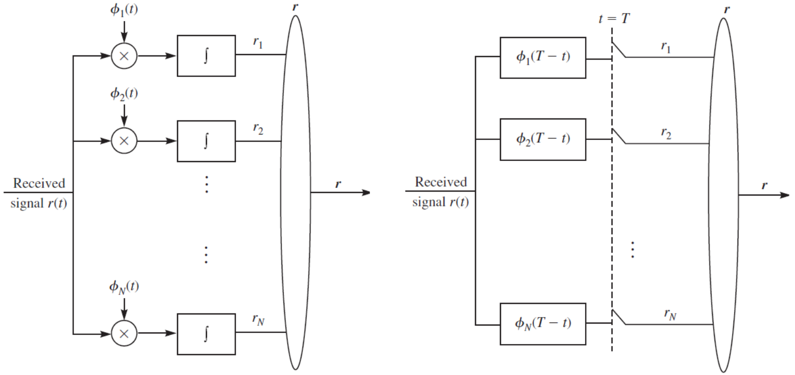
- $\widehat{s}(t)$ is the projection of $r(t)$ onto S. Use the mean square error (MSE) criterion from estimation theory to find the optimum coefficients of $\widehat{s}(t) = \sum_{k=1}^{N} s_k \phi_k(t)$ according to $s_k = \langle r(t), \phi_k(t) \rangle$, i.e., project $r(t)$ onto each basis function $\phi_k(t)$.



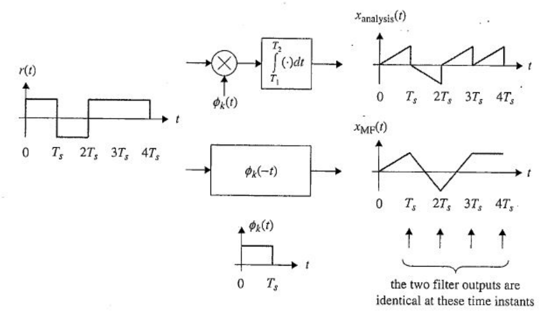
Demodulation-two main approaches

Two main approaches to demodulate $r(t)$ into N-dimensional vector $\mathbf{r} = (r_1, r_2, \ldots, r_N)$:

- Correlation-type demodulator *(see left side of figure below)*
- Matched filter-type demodulator *(see right side of figure below)*



Example: binary PAM demodulation with $r(t)$ noiseless:



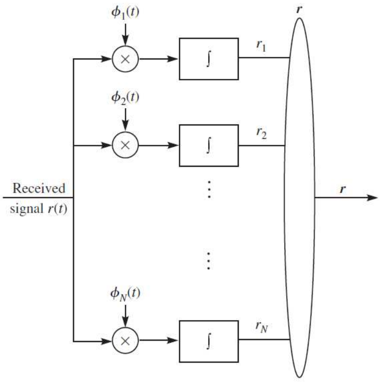
- $x_{analysis}(t)$ in the figure is produced by the correlator in a correlator-type demodulator
- $x_{MF}(t)$ in the figure is produced by the matched filter in a matched filter demodulator
- The PAM amplitudes in $r(t)$ are retrieved as the values $x(T), x(2T), \ldots$, as explained next.

Correlation-type Demodulator

- Use a parallel bank of N correlators which multiplies $r(t)$ with $\{\phi_k(t)\}_{k=1}^{N}$. Output of k-th correlator is

$$
\begin{aligned}
r_k &= \int_0^T r(t)\phi_k(t)dt \\
&= \int_0^T [s_m(t) + n(t)]\phi_k(t)dt \\
&= \int_0^T s_m(t)\phi_k(t)dt + \int_0^T n(t)\phi_k(t)dt \\
&= s_{mk} + n_k, \quad k = 1, \ldots, N.
\end{aligned}
$$

$$\Rightarrow \mathbf{r} = \mathbf{s}_m + \mathbf{n}$$



Correlation-type Demodulator

Received signal waveform can be re-expressed as

$$r(t) = \sum_{k=1}^{N} s_{mk}\phi_k(t) + \sum_{k=1}^{N} n_k\phi_k(t) + n'(t) = \sum_{k=1}^{N} r_k\phi_k(t) + n'(t).$$

We ignore $n'(t) = n(t) - \sum_{k=1}^{N} n_k\phi_k(t)$ because it is uncorrelated with r_k:

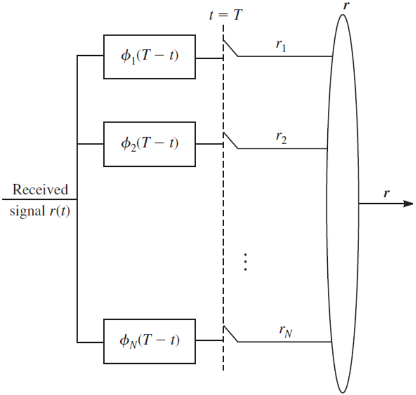
$$\begin{aligned}
E[n'(t)r_k] &= E[n'(t)s_{mk}] + E[n'(t)n_k] = E[n'(t)n_k] \\
&= E\left[\left(n(t) - \sum_{i=1}^{N} n_i\phi_i(t)\right) n_k\right] \\
&= \int_0^T E\left[n(t)n(\tau)\right]\phi_k(\tau)d\tau - \sum_{i=1}^{N} E\left[n_k n_i\right]\phi_i(t) \\
&= \frac{N_0}{2}\int_0^T \delta(t-\tau)\phi_k(\tau)d\tau - \frac{N_0}{2}\phi_k(t) \\
&= \frac{N_0}{2}\phi_k(t) - \frac{N_0}{2}\phi_k(t) = 0 \quad \Rightarrow \quad \text{Zero correlation!}
\end{aligned}$$

Matched filter-type Demodulator

- Use a parallel bank of N linear filters with impulse response $h_k(t) = \phi_k(T - t)$. Output of k-th filter is

$$\begin{aligned} r_k(t) &= (r \star h_k)(t) \\ &= \int_0^T r(\tau) h_k(t-\tau) d\tau \\ &= \int_0^T r(\tau) \phi_k(T - t + \tau) d\tau \end{aligned}$$

$\Rightarrow$ When filter outputs are sampled at $t = T$, then $r_k = \int_0^T r(\tau)\phi_k(\tau)d\tau$.



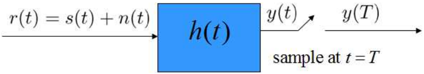
Matched filter-type Demodulator

The matched filter to a signal $s(t)$ is a filter whose impulse response is $h(t) = s(T - t)$, where $s(t)$ is confined to the time interval $0 \leq t \leq T$.

Properties of Matched Filter

- If a signal $s(t)$ is corrupted by AWGN, the filter with impulse response matched to $s(t)$ maximizes the output SNR.
- The output SNR from the matched filter depends on the energy of the waveform $s(t)$ but not on the detailed characteristics of $s(t)$.

Matched filter-type Demodulator



- Consider signal $r(t)$ passes through a filter with impulse response $h(t)$, $0 \leq t \leq T$. The output of the filter is

$$y(t) = \int_0^T (s(\tau) + n(\tau))h(t - \tau)d\tau$$

- At $t = T$, we can decompose the output into signal and noise components

$$y(t) = \underbrace{\int_0^T s(\tau)h(T - \tau)d\tau}_{y_s(T)} + \underbrace{\int_0^T n(\tau)h(T - \tau)d\tau}_{y_n(T)}$$

- Question: Which $h(t)$ maximizes the output signal-to-noise ratio (SNR) defined as $\frac{y_s^2(T)}{E[y_n^2(T)]}$?

MATCHED FILTER-TYPE DEMODULATOR

- Noise variance is

$$\begin{aligned} E[y_n^2(T)] &= \int_0^T \int_0^T E[n(\tau)n(t)]h(T-\tau)h(T-t)dtd\tau \\ &= \int_0^T \int_0^T \frac{N_0}{2}\delta(t-\tau)h(T-t)h(T-\tau)dtd\tau \\ &= \frac{N_0}{2}\int_0^T h^2(T-\tau)d\tau \end{aligned}$$

- Signal energy is

$$y_s^2(T) = \left[\int_0^T s(\tau)h(T-\tau)d\tau\right]^2 \leq \int_0^T s^2(\tau)d\tau \int_0^T h^2(T-\tau)d\tau$$

with Cauchy-Schwarz inequality $|\langle x_1(t), x_2(t)\rangle| \leq ||x_1(t)||.||x_2(t)||$.
The signal energy is maximized when $h(t) = ks(T-t)$, for some constant k.
$\Rightarrow$ SNR is maximized when filter response $h(t)$ is matched to $s(t)$.

Matched filter-type Demodulator

- The output signal-to-noise ratio (SNR) is

$$\left(\frac{S}{N}\right)_o = \frac{2}{N_0}\int_0^T s^2(\tau)d\tau = \frac{2\mathcal{E}_s}{N_0}$$

$\Rightarrow$ Depends only on the energy of $s(t)$.

- The frequency response of the matched filter $h(t) = s(T - t)$ is

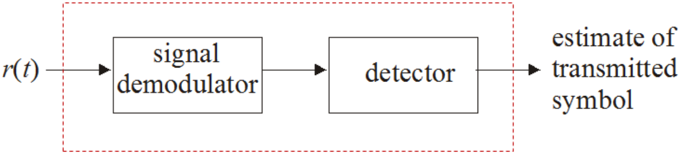
$$H(f) = S^*(f)e^{-j2\pi fT}$$

where the filter spectrum has the same magnitude as the signal spectrum $|H(f)| = |S(f)|$ and phase of $H(f)$ is negative of the phase of $S(f)$ shifted by $2\pi fT$.

RECALL

Two major steps in receiver design

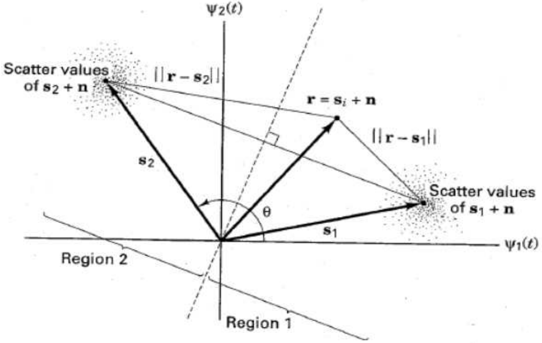
- *Step 1:* Signal demodulation – Convert the received waveform $r(t)$ into a vector $\mathbf{r} = (r_1, r_2, \ldots, r_N)$, where N is the dimension of the transmitted signal waveforms. Projection of received signal onto the signal space.
- *Step 2:* Detection – Decide which of M possible signal waveforms was transmitted based on $\mathbf{r}$.



Digital demodulator

The demodulator computes the projected received signal $\mathbf{r} = (r_1, r_2, \ldots, r_N)$ in the N-dimensional signal space of the transmit vectors $\{\mathbf{s}_m = (s_{m1}, s_{m2}, \ldots, s_{mN})\}_{m=1}^{M}$. The projected received signals

form "spherical noise clouds" around the signal points $\mathbf{s}_m$ due to the additive noise signal $\mathbf{n} = (n_1, n_2, \ldots, n_N)$ which has Gaussian components $n_k \sim N(0, N_0/2)$ for all $k = 1, \ldots, N$.



OPTIMAL DETECTION

- Objective: Minimize overall $P(\text{Error})$
- Approach:
  1. Partition N-dimensional space of signal demodulator outputs into M non-overlapping decision regions $D_1, D_2, \ldots, D_M$
  2. Decide on $\widehat{m} = \mathbf{s}_m$ iff $\mathbf{r} = (r_1, \ldots, r_N) \in D_m$, $1 \leq m \leq M$.
- Detector design $\Rightarrow$ Decide *how to partition N* dimensional space
- Some definitions:
  - Prior probability $P(\mathbf{s}_m \text{ transmitted})$
  - Posterior probability $P(\mathbf{s}_m \text{ transmitted} \mid \mathbf{r} \text{ received})$
  - Likelihood function (conditional pdf) $p(\mathbf{r} \text{ received} \mid \mathbf{s}_m \text{ transmitted})$
  - Related by Bayes' theorem

$$P(\mathbf{s}_m|\mathbf{r}) = \frac{p(\mathbf{r}|\mathbf{s}_m)P(\mathbf{s}_m)}{p(\mathbf{r})}$$

$$P(\text{No error}|\mathbf{s}_m) = P(\mathbf{r} \in D_m|\mathbf{s}_m) = \int_{D_m} p(\mathbf{r}|\mathbf{s}_m)d\mathbf{r}$$

- Minimizing P(Error) = Maximizing P(No error), in fact

$$\begin{aligned} P(\text{No error}) &= \sum_{m=1}^{M} P(\mathbf{s}_m)P(\text{No error}|\mathbf{s}_m) \\ &= \sum_{m=1}^{M} P(\mathbf{s}_m)\int_{D_m} p(\mathbf{r}|\mathbf{s}_m)d\mathbf{r} \\ &= \sum_{m=1}^{M} \int_{D_m} P(\mathbf{s}_m|\mathbf{r})p(\mathbf{r})d\mathbf{r} \quad \text{(using Bayes' theorem)} \end{aligned}$$

- Probability density $p(\mathbf{r})$ non-negative $\Rightarrow$ RHS maximised if $P(\mathbf{s}_m|\mathbf{r})$ is maximized for each $\mathbf{r}$, i.e. construct the decision regions such that

$$D_m = \{\mathbf{r} \in \mathbb{R}^N : P(\mathbf{s}_m|\mathbf{r}) \geq P(\mathbf{s}_{m'}|\mathbf{r}) \text{ for all } m' \neq m\}$$

MAP Detection

- MAP criterion:

$$
\begin{aligned}
& P(\mathbf{s}_m|\mathbf{r}) \geq P(\mathbf{s}_{m'}|\mathbf{r}) \text{ for all } m' \neq m \\
\iff \ & \widehat{m} = \arg\max_m \underbrace{P(\mathbf{s}_m|\mathbf{r})}_{\text{Posterior probability}} \\
& = \arg\max_m \frac{p(\mathbf{r}|\mathbf{s}_m)P(\mathbf{s}_m)}{p(\mathbf{r})} \\
& = \arg\max_m p(\mathbf{r}|\mathbf{s}_m)P(\mathbf{s}_m)
\end{aligned} \tag{1}
$$

- Important result: Maximum a posteriori probability (MAP) detector is an optimal detector for minimizing the probability of error.

ML Detection

- Another detection rule is the maximum-likelihood (ML) criterion:

$$\widehat{m} = \arg\max_{m} \underbrace{p(\mathbf{r}|\mathbf{s}_m)}_{\text{Likelihood function}}$$

- It follows from (1) on the previous page that, if all symbols are equally likely to be transmitted, i.e., $P(\mathbf{s}_m) = \frac{1}{M}$ for all $1 \leq m \leq M$, then the MAP criterion simplifies to the ML criterion.

- Important result: **A Maximum-Likelihood (ML) detector is optimal when all symbols are equiprobable.**

- In AWGN channels, the received vector components of $\mathbf{r} = (r_1, r_2, \ldots, r_N)$ are

$$r_k = s_{mk} + n_k, \quad k = 1, \ldots, N$$

where $n_k \sim \mathcal{N}(0, N_0/2)$, so that $r_i \sim \mathcal{N}(s_{mk}, N_0/2)$

- The likelihood function $p(\mathbf{r}|\mathbf{s}_m)$ can be calculated as

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

where $||(x_1, x_2, \ldots, x_N)|| = \sqrt{\sum_{i=1}^{N} x_i^2}$ is the Euclidean norm.

ML Detection

- The ML detection criterion for AWGN channels given by

$$\widehat{m} = \arg\max_{m} \frac{1}{(\pi N_0)^{\frac{N}{2}}} \exp\left(-\frac{||\mathbf{r} - \mathbf{s}_m||^2}{N_0}\right)$$

can be simplified to

$$\widehat{m} = \arg\min_{m} ||\mathbf{r} - \mathbf{s}_m||$$

since the exponential function $\exp(-x)$ is a decreasing function of x.

- Decide on $\mathbf{s}_m$ that is *closest* to $\mathbf{r}$ $\Rightarrow$ minimum distance detection
- This means that we can determine decision regions $D_1, D_2, \ldots D_M$ graphically.

Decision Regions

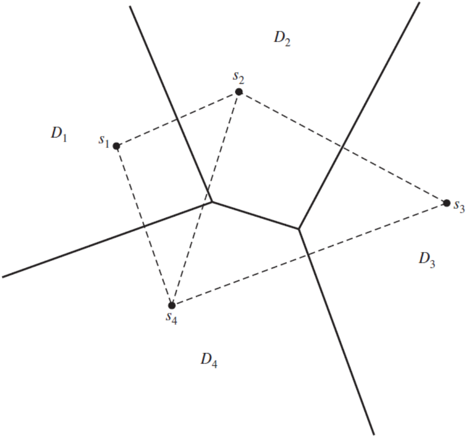
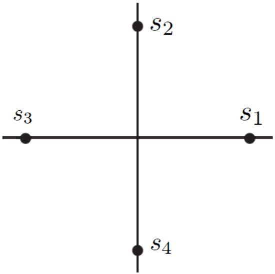


FIGURE 4.2–1
The decision regions for equiprobable signaling.

Decision Regions

Assume we use an ML detector and assume that signals are equally likely to be transmitted.

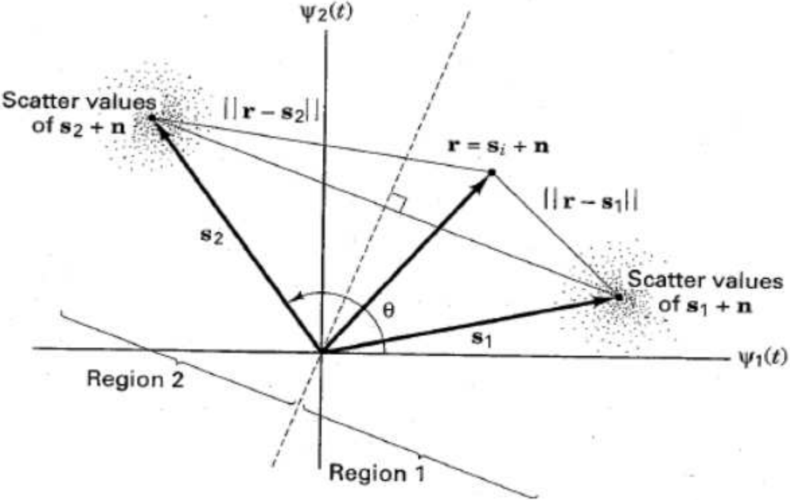
Question 1: What are the decision regions for 4-PSK modulation?



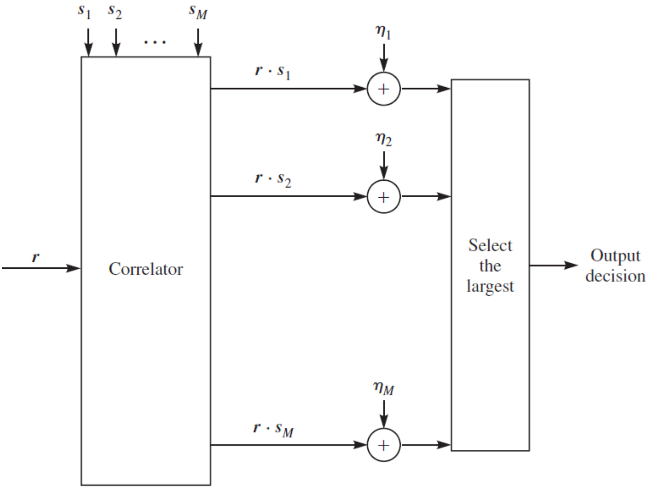
Question 2: Draw the signal space for 2-FSK modulation. What are the decision regions?

ML Detection

ML detection decides on closest signal $\Rightarrow \widehat{m} = \mathbf{s}_1$



ML DETECTION



This figure is explained on the next page.

ML Detection

- We can further expand the distance metrics for the ML criterion in AWGN channels as

$$\begin{aligned}||\mathbf{r}-\mathbf{s}_m||^2 &= \langle \mathbf{r}-\mathbf{s}_m, \mathbf{r}-\mathbf{s}_m\rangle \\ &= \langle \mathbf{r},\mathbf{r}\rangle - \langle \mathbf{r},\mathbf{s}_m\rangle - \langle \mathbf{s}_m,\mathbf{r}\rangle + \langle \mathbf{s}_m,\mathbf{s}_m\rangle \\ &= \underbrace{||\mathbf{r}||^2}_{\text{Independent of } m} - 2\langle \mathbf{r},\mathbf{s}_m\rangle + \underbrace{||\mathbf{s}_m||^2}_{\text{Energy of } m\text{-th signal, } \mathcal{E}_m}\end{aligned}$$

- ML criterion can be re-expressed as

$$\begin{aligned}\widehat{m} &= \arg\min_m ||\mathbf{r}-\mathbf{s}_m|| \\ &= \arg\max_m \underbrace{[\langle \mathbf{r},\mathbf{s}_m\rangle + \eta_m]}_{\text{Correlation metrics}}\end{aligned}$$

where $\eta_m = -\frac{1}{2}\mathcal{E}_m$ is a bias term that compensates for signal sets that have unequal energies such as PAM.

Rectangular 16-QAM example

Suppose that 16-QAM modulation is used, with

$$s(t) = I(t)\sqrt{2}\cos\omega_0 t - Q(t)\sqrt{2}\sin\omega_0 t, \text{ where}$$

- $I(t) = \sum_n a_0(n)p(t-nT)$ is the inphase component of $s(t)$
- $Q(t) = \sum_n a_1(n)p(t-nT)$ is the quadrature component of $s(t)$
- ω_0 is the carrier frequency; $p(t)$ is the pulse signal.

A matched filter detector is given by the next figure; note that the demodulator consists of a mixer, a matched filter and a sampler:

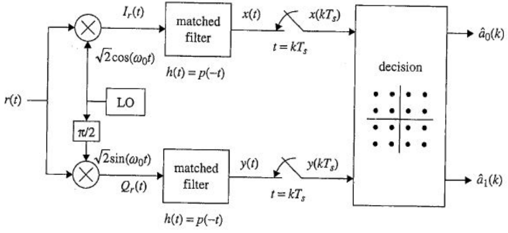


Figure 5.3.11 The matched filter detector for M-ary QAM using continuous-time processing.

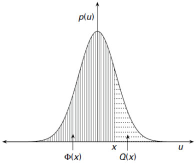
Summary so far

AWGN channel model

- Received signals from AWGN channels can be modelled as Gaussian random process due to addition of noise which is Gaussian distributed with zero mean and variance $N_0/2$, i.e., $n(t) \sim \mathcal{N}(0, N_0/2)$.

- Q-function is useful function for calculating tail probabilities of $X \sim \mathcal{N}(0, 1)$ defined as

$$Q(x) = \Pr[X > x] = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{v^2}{2}\right) dv$$



Signal demodulation

- Signal demodulator converts the received waveform $r(t)$ into an N-dimensional vector $\mathbf{r} = (r_1, r_2, \ldots, r_N)$ by taking the inner product of $r(t)$ and the N basis functions of the transmitted signal space $\{\phi_1(t), \ldots, \phi_N(t)\}$.
- Correlation-type demodulator multiplies $r(t)$ with $\{\phi_1(t), \ldots, \phi_N(t)\}$ and integrates the N product terms from 0 to T.
- Matched filter-type demodulator convolves $r(t)$ with matched impulse responses $\{\phi_1(T-t), \ldots, \phi_N(T-t)\}$ and samples the N filter outputs at time T.
- The matched filter to a transmitted signal $s(t)$ is a filter whose impulse response is $h(t) = s(T-t)$. In AWGN channels, the matched filter output sampled at time T results in the maximum SNR.

Detection

- Detection decides which of M possible symbols $\mathbf{s}_m$ was transmitted based on the demodulated vector $\mathbf{r} = (r_1, r_2, \ldots, r_N)$.
- The optimal detector that minimizes the error probability is the MAP detector

$$\widehat{m} = \arg\max_m P(\mathbf{s}_m|\mathbf{r})$$

  which maximizes the posterior probabilities $P(\mathbf{s}_m|\mathbf{r})$, or equivalently $p(\mathbf{r}|\mathbf{s}_m)P(\mathbf{s}_m)$.
- When symbols are equiprobable, the MAP detector is the same as the ML detector

$$\widehat{m} = \arg\max_m p(\mathbf{r}|\mathbf{s}_m).$$

- In AWGN channels, the ML detector simply finds the minimum distance

$$\widehat{m} = \arg\min_m ||\mathbf{r} - \mathbf{s}_m||.$$

Practical concerns

- Implicit assumptions made so far:
  - the demodulator knows the symbol times exactly
  - the demodulator knows the phase of the carrier signal exactly
- However, this is often not true.
- How to deal with this?
  - see reference book by M. Rice, Chapter 8 on Symbol Timing Synchronization (*not covered here*), for example polyphase filterbank interpolation
  - see reference book by M. Rice, Chapter 7 on Carrier Phase Synchronization

Carrier phase mismatch

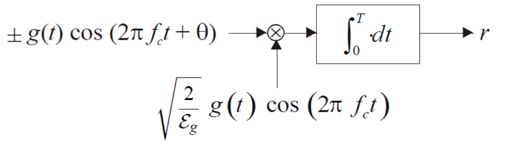
- How does carrier phase mismatch arise? For example, consider a BPSK transmitted signal

$$s(t) = \pm\cos(2\pi f_c t), \quad 0 \le t < T.$$

  The received signal after AWGN channel with imperfect carrier synchronisation is

$$r(t) = \pm\cos(2\pi f_c (t - \tau)) + n(t) = \pm\cos(2\pi f_c t - \theta) + n(t).$$

- Reasons for carrier phase mismatch:
  - Phase $\theta = 2\pi f_c \tau$ not known exactly
  - Carrier and receiver not phase-synchronised
  - Carrier frequency value only approximately known, not exactly



Effect of mismatch: Lump all phase uncertainty into added carrier angle θ. Output of correlator for BPSK is

$$\begin{aligned} r &= \pm\sqrt{\frac{2}{\mathcal{E}_g}}\int_0^T g^2(t)\cos(2\pi f_c t+\theta)\cos(2\pi f_c t)dt + n \\ &= \pm\sqrt{\frac{2}{\mathcal{E}_g}}\int_0^T g^2(t)\left[\frac{1}{2}\cos(\theta)+\frac{1}{2}\cos(4\pi f_c t+\theta)\right]dt + n \\ &\approx \pm\sqrt{\frac{\mathcal{E}_g}{2}}\cos(\theta)+n \Rightarrow \text{Loss of information when } \cos(\theta)<1 \end{aligned}$$

The phase mismatch causes a rotation that may lead to the projection lying in the wrong decision region. Even in the noiseless case.

Dealing with phase mismatch:

- Approach 1: If the phase mismatch is *unknown* and *changing rapidly* (e.g., radio channels), then if *small*, treat it as random noise and design an optimal detector. If *large*, it is better to use modulation schemes whose detectors ignore phase, e.g., envelope detectors used for orthogonal signalling (e.g. **FSK**) and on-off keying (**OOK**) modulations and then use so-called **noncoherent demodulators** (see next pages). Note that PAM is then not suitable because its detector is highly phase-dependent, e.g., for binary PAM, 180° phase shift results in a flip to the other symbol
- Approach 2: If the phase mismatch is *unknown* but *fixed or slowly varying*, then modulate the phase differences, rather than the absolute phase, e.g., differential phase-shift keying (**DPSK**) (see later pages).
- Other approaches: see reference book by M. Rice, Chapter 7 on Carrier Phase Synchronization

Noncoherent OOK Demodulation

Approach 1: Phase mismatch is large, *unknown* and *changing rapidly*

- Transmit signal for OOK is

$$s_m(t) = A_m g(t)\cos(2\pi f_c t), \quad \text{for } m = 1, 2, \quad A_1 = 1, A_2 = 0.$$

and received signal after AWGN channel with imperfect carrier synchronisation is

$$\begin{aligned} r(t) &= A_m g(t)\cos(2\pi f_c t + \theta) + n(t) \\ &= A_m g(t)\cos(\theta)\cos(2\pi f_c t) - A_m g(t)\sin(\theta)\sin(2\pi f_c t) + n(t) \end{aligned}$$

- Dimension of transmitted signal space is 1, but dimension of received signal space is 2 due to the phase mismatch (even when there is no noise!)
  $\Rightarrow$ Signal demodulator must have 2 correlators, otherwise some symbol info will be lost.

- Signal demodulator output is $\mathbf{r} = (r_1, r_2)$ where

$$\begin{aligned} r_1 &= \left\langle A_m g(t)\cos(2\pi f_c t + \theta) + n(t), \sqrt{\frac{2}{\mathcal{E}_g}} g(t)\cos(2\pi f_c t) \right\rangle \\ &= \sqrt{\frac{2}{\mathcal{E}_g}} \int_0^\infty A_m g^2(t)\cos(2\pi f_c t + \theta)\cos(2\pi f_c t) dt + n_1 \\ &\approx \sqrt{\frac{\mathcal{E}_g}{2}} A_m \cos(\theta) + n_1, \end{aligned}$$

and

$$\begin{aligned} r_2 &= \left\langle A_m g(t)\cos(2\pi f_c t + \theta) + n(t), \sqrt{\frac{2}{\mathcal{E}_g}} g(t)\sin(2\pi f_c t) \right\rangle \\ &\approx \sqrt{\frac{\mathcal{E}_g}{2}} A_m \sin(\theta) + n_2. \end{aligned}$$

- Optimal detector uses $\mathbf{r} = (r_1, r_2)$ to obtain MAP estimate of $\widehat{m}$ as

$$\widehat{m} = \arg\max_{m=1,2} P(\mathbf{s}_m)p(\mathbf{r}|\mathbf{s}_m) = \arg\max_{m=1,2} P(\mathbf{s}_m)\int_0^{2\pi} p(\mathbf{r}|\mathbf{s}_m, \theta)p(\theta)d\theta$$

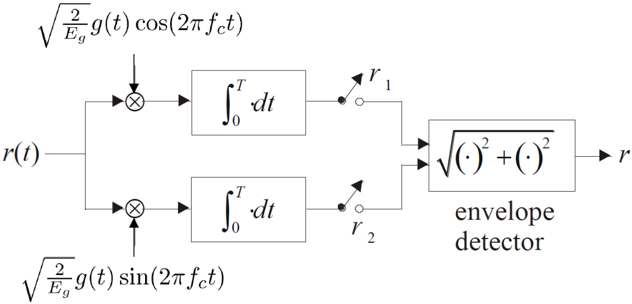
- Worst case: phase uncertainty is uniformly distributed over $[0, 2\pi)$.
- Assuming equiprobable symbols,

$$\widehat{m} = \arg\max_{m=1,2}\int_0^{2\pi} p(\mathbf{r}|\mathbf{s}_m, \theta)p(\theta)d\theta = \begin{cases} 1, & r_1^2 + r_2^2 > V_T \\ 2, & r_1^2 + r_2^2 < V_T \end{cases}$$

  which depends only on envelope of projected received signal (V_T in terms of Bessel function, see textbook pp. 212-214 for details).
- At high SNRs (i.e., $n_1, n_2 \approx 0$),

$$r_1^2 + r_2^2 \approx \frac{1}{2}A_m^2\mathcal{E}_g(\cos^2(\theta) + \sin^2(\theta)) = \begin{cases} \frac{1}{2}\mathcal{E}_g, & m = 1 \\ 0, & m = 2. \end{cases}$$

  which is independent of θ.


$\sqrt{\frac{2}{E_g}} g(t) \cos(2\pi f_c t)$
$r(t)$
$\int_0^T \cdot dt$
$\int_0^T \cdot dt$
r_1
r_2
$\sqrt{(\cdot)^2 + (\cdot)^2}$
r
envelope detector
$\sqrt{\frac{2}{E_g}} g(t) \sin(2\pi f_c t)$

Noncoherent Binary FSK Demodulation

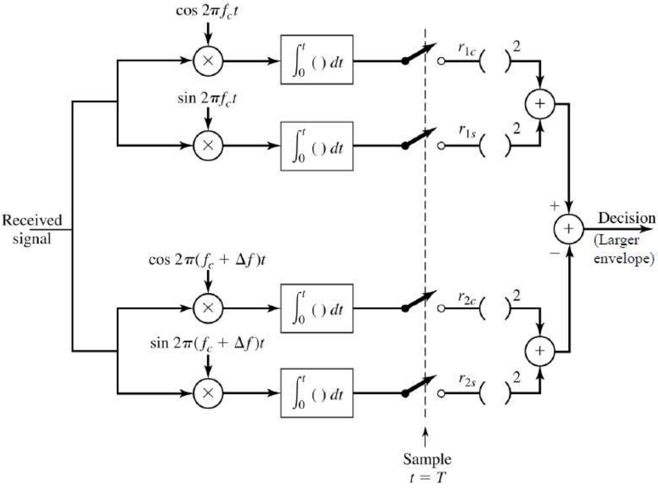


Figure 7.49

Demodulation and square-law detection of binary FSK signals.

Noncoherent M-ary FSK Demodulation

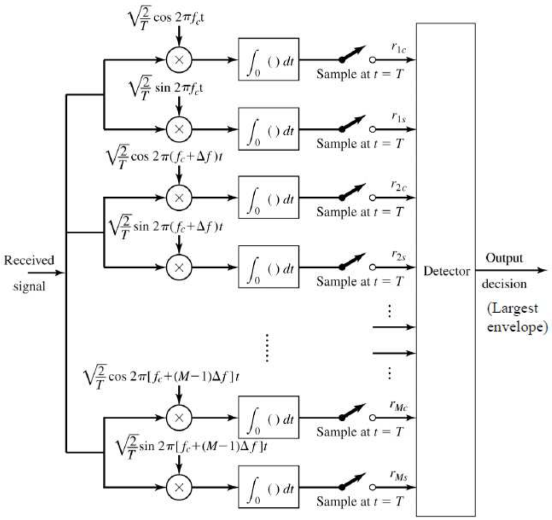


Figure 7.47

RECALL: DIFFERENTIAL MODULATION

Differential MPSK (DMPSK) modulation involves precoding of the information symbol sequence $b(n)$ into a symbol sequence $\delta(n)$ that is then input to a MPSK modulator. The DMPSK demodulator then needs to invert the precoding, see later figures.

- in MPSK each symbol value determines the actual value of the phase, but in DMPSK it determines the phase change from the previous signalling interval's symbol
- thus DMPSK is modulation with memory
- differential modulation is used in situations where the MPSK demodulator's errors are always of the same type, determined by a fixed (or slowly varying) phase mismatch ϕ, the value of which we do not know (it may even be zero).
- differential modulation leads to a better symbol error performance since it is robust against a fixed unknown phase mismatch.

Example: DBPSK modulation

- Example: a symbol value of 0 causes 0° phase change and a symbol value of 1 causes 180° phase change

Data Bit		1	0	1	0	0	1
BPSK phase		0°	180°	0°	180°	180°	0°
DBPSK phase	180°	0°	0°	180°	180°	180°	0°

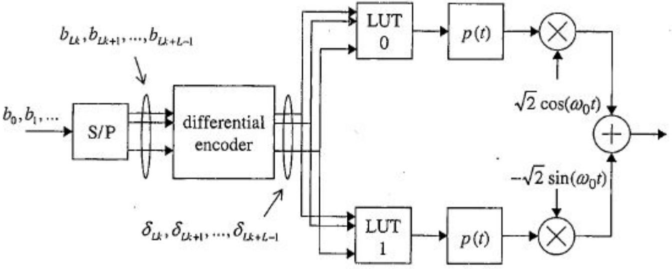


Figure 7.7.3 A block diagram of an MPSK modulator with a differential encoder. Note that the input bits b_k are encoded to produce differentially encoded bits δ_k. The differentially encoded bits are used to select the constellation points from the look-up tables.

Example: DBPSK Demodulation

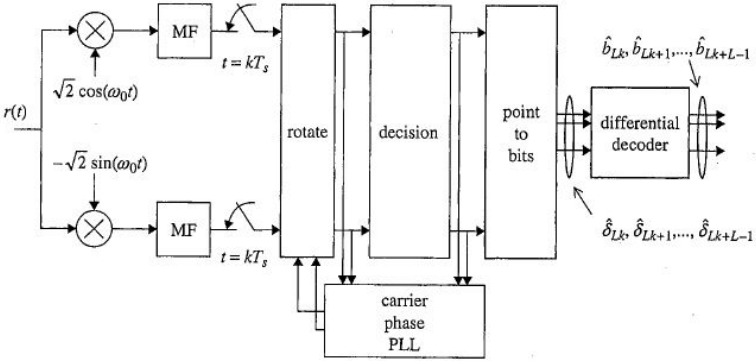


Figure 7.7.4 A block diagram of an MPSK detector with a differential decoder. Note that the detector decisions are estimates of the differentially encoded bits $\hat{\delta}_k$. The estimates of these encoded bits are used by the differential decoder to produce estimates of the original bits.

Example: DBPSK Demodulation—continued

- Question: What happens if the demodulator makes an error?

Data Bit		1	0	1	0	0	1
DPSK phase	180°	0°	0°	180°	180°	180°	0°
Estimated phase	180°	0°	180°	180°	180°	180°	0°
Demodulated Bit		1	1	0	0	0	1

$\Rightarrow$ 1 phase error causes bit errors over 2 consecutive intervals

- Easy to extend to more bits, e.g., for $M = 4$

Data Bits	DQPSK Phase Shift
00	0°
01	90°
11	180°
10	270°