EC5.203 Communication Theory (3-1-0-4):

Lecture 21:

Optimal Demodulation-4

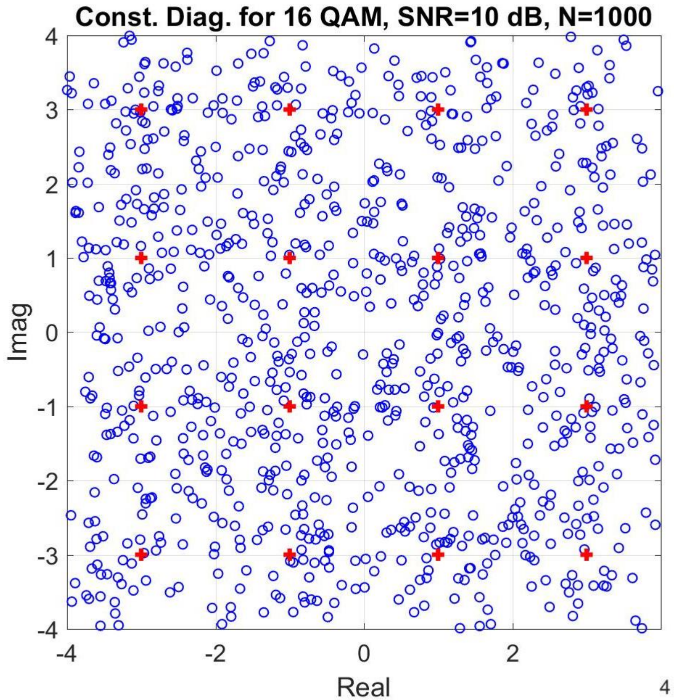
Apr. 14, 2025

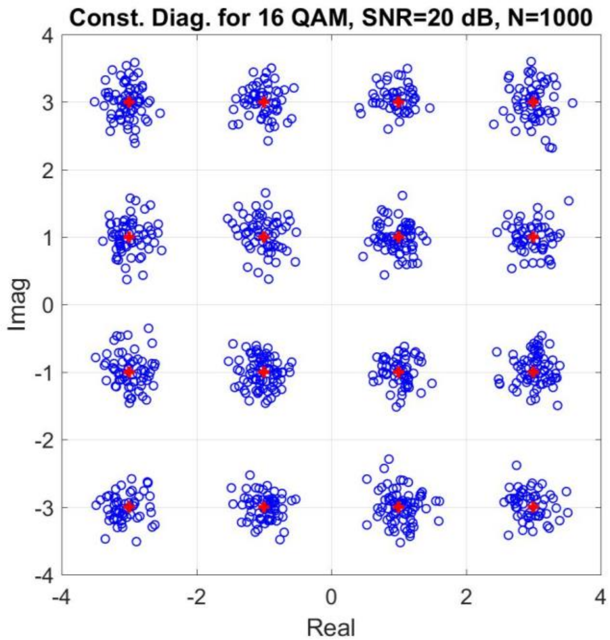
INTERNATIONAL INSTITUTE OF
INFORMATION TECHNOLOGY
HYDERABAD

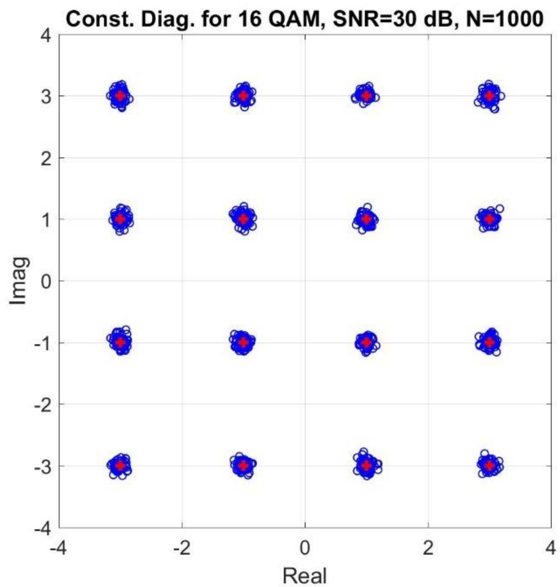
References

- Chap. 6 (Madhow)

Example: 16 QAM in AWGN







Optimal Demodulation

- In 16 QAM, one of 16 passband waveforms corresponding to 16 symbols is sent, where each passband waveform is given by

$$s_i(t) = s_{b_c,c_s} = b_c p(t)\cos(2\pi f_c t) - b_s p(t)\sin(2\pi f_c t)$$

  where b_c, b_s each takes value in $\{\pm 1, \pm 3\}$.

- At the receiver, we are faced with a hypothesis testing problem: we have M possible hypotheses about which signal was sent.

- Based on the observations

$$y(t) = s_i(t) + \boxed{n(t)} \quad \text{AWGN}$$

  we are interested in finding a decision rule to make a best guess which hypothesis was sent.

- For communications applications, performance criteria is to minimize the probability of error (i.e., the probability of making a wrong guess).

Example 5.6.3

- **Binary on-off keying in Gaussian noise**

$$Y = m + n \quad \text{if 1 is sent}$$
$$Y = n \quad \text{if 0 is sent}$$

  Here Y is the received sample, $m > 0$ is some constant and n is AWGN sample with $\mathcal{N}(0, v^2)$.

- At the receiver, the detection strategy is

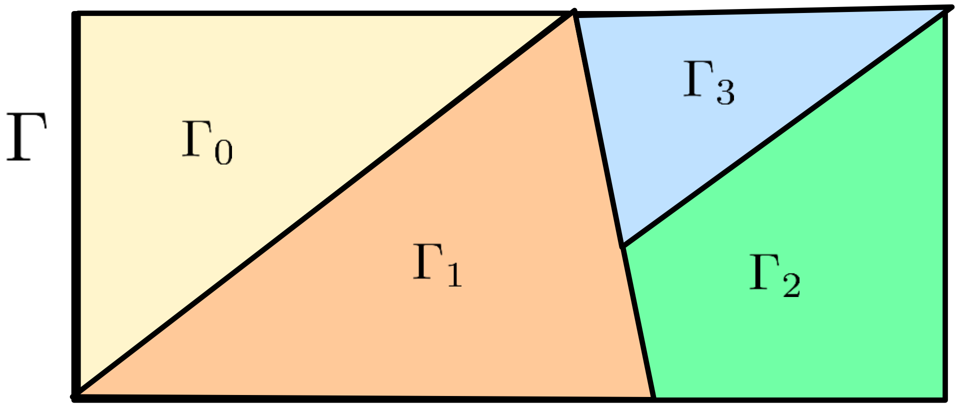
$$Y > m/2 \quad \text{Decide 1 is sent}$$
$$Y \leq m/2 \quad \text{Decide 0 is sent}$$

- Assuming that both 0 and 1 are equally likely,
  - Find the average signal power
  - Find the conditional probability of error conditioned on 0 being sent
  - Find the conditional probability of error conditioned on 1 being sent
  - Find average error probability
  - Find the probability of error for SNR of 13 dB?

Ingredients of Hypothesis Testing Framework

- Hypotheses $H_0, H_1, \ldots, H_{M-1}$
- Observation $Y \in \Gamma$
- Conditional densities $p(y|i)$ for $i = 0, 1, \ldots, M-1$
- Prior probabilities $\pi_i = P(H_i)$ with $\sum_i \pi_i = 1$
- Decision rule $\delta : \Gamma \to \{0, 1, M-1\}$
- Decision region $\Gamma_i : \{y \in \Gamma_i : \delta(y) = i\}$ for $i = 0, 1, M-1$

Example of Decision regions for M=4



Error Probabilities

- Conditional error probabilities, conditioned on H_i, is

$$\begin{aligned} P_{e|i} &= P(\text{decide } j \text{ for some } j \neq i | H_i \text{ is true}) \\ &= \sum_{j \neq i} P(Y \in \Gamma_j | H_i) \\ &= 1 - P(Y \in \Gamma_i | H_i) \end{aligned}$$

- Conditional probabilities of correct detection, conditioned on H_i, is

$$\begin{aligned} P_{c|i} &= P(Y \in \Gamma_i | H_i) \\ &= 1 - P_{e|i} \end{aligned}$$

- Average error probability

$$P_e = \sum_{i=1}^{M} \pi_i P_{e|i}$$

- Average probability of correct detection

$$P_c = \sum_{i=1}^{M} \pi_i P_{c|i}$$

Ingredients of Hypothesis Testing Framework

- Hypotheses $H_0, H_1, \ldots, H_{M-1}$
- Observation $Y \in \Gamma$
- Conditional densities $p(y|i)$ for $i = 0, 1, \ldots, M-1$
- Prior probabilities $\pi_i = P(H_i)$ with $\sum_i \pi_i = 1$
- Decision rule $\delta : \Gamma \to \{0, 1, M-1\}$
- Decision region $\Gamma_i : \{y \in \Gamma_i : \delta(y) = i\}$ for $i = 0, 1, M-1$

- In earlier example
  - Hypotheses H_0, H_1, Observation $Y \in \Gamma = \mathcal{R}$
  - Conditional densities $p(y|0)$ and $p(y|1)$
  - Prior probabilities π_0 and π_1
  - Decision rule δ:
  $$\delta(y) = \begin{cases} 0, & y \leq m/2 \\ 1, & y > m/2 \end{cases}$$
  - Decision regions: $\Gamma_0 = (-\infty, m/2]$ and $\Gamma_1 = (m/2, \infty)$

MAP rule

- Definitions:
  - *A priori* probability: Before the data is observed : $P(H_i) = \pi_i$
  - *A posteriori* probability: After the data is observed: $P(H_i|y)$
- Maximum *a posteriori* probability (MAP) rule:

$$\delta_{\mathrm{MAP}}(y) = \arg\max_i P(H_i|Y = y)$$

  where $i = 0, 1, \ldots, M-1$

- Using Bayes rule $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$, MAP rule can be rewritten as

$$\begin{aligned}
\delta_{\mathrm{MAP}}(y) &= \arg\max_i \frac{P(Y = y|H_i)P(H_i)}{P(Y = y)} \\
&= \arg\max_i \frac{p(y|i)\pi_i}{p(y)} \\
&= \arg\max_i p(y|i)\pi_i \\
&= \arg\max_i \log \pi_i + \log p(y|i)
\end{aligned}$$

Optimality of MAP (or MPE) rule

- Optimality of MAP rule: The MAP rule minimizes the probability of error. Proof!

ML rule

- Definitions:
  - Likelihood function: $p(y|i) = P(Y = y|H_i)$
- Maximum likelihood (ML) rule

$$\delta_{\mathrm{ML}}(y) = \arg\max_i p(y|i)$$

  where $i = 0, 1, \ldots, M-1$

- Equivalently,

$$\delta_{\mathrm{ML}}(y) = \arg\max_i \log p(y|i)$$

- ML is equivalent of MAP for equal prior probabilities, i.e., $\pi_i = \frac{1}{M}$, we have

$$\begin{aligned}\delta_{\mathrm{MAP}}(y) &= \arg\max_i \log \pi_i + \log p(y|i) \\ &= \arg\max_i \log \frac{1}{M} + \log p(y|i) \\ &= \arg\max_i \log p(y|i)\end{aligned}$$

Binary Hypothesis Testing Problem

- For two hypotheses case, ML decision rule is

$$\begin{aligned}\delta_{\mathrm{ML}}(y) &= \arg\max_i p(y|i) \\ &= \arg\max_i \{p(y|0), p(y|1)\}\end{aligned}$$

- Equivalently,

$$\begin{aligned} p(y|0) > p(y|1) \quad &\rightarrow \quad \delta_{\mathrm{ML}}(y) = 0 \\ p(y|1) > p(y|0) \quad &\rightarrow \quad \delta_{\mathrm{ML}}(y) = 1 \end{aligned}$$

- This can be written as

$$p(y|1) \underset{H_0}{\overset{H_1}{\gtrless}} p(y|0)$$

- Similarly, MAP or MPE rule for binary hypothesis testing problem can be written as

$$\pi_1 p(y|1) \underset{H_0}{\overset{H_1}{\gtrless}} \pi_0 p(y|0)$$

$$P(H_1|Y=y) \underset{H_0}{\overset{H_1}{\gtrless}} P(H_0|Y=y)$$

Likelihood Ratio

- For two hypotheses case, ML decision rule can be written as

$$p(y|1) \underset{H_0}{\overset{H_1}{\gtrless}} p(y|0)$$

- Equivalently,

$$\frac{p(y|1)}{p(y|0)} \underset{H_0}{\overset{H_1}{\gtrless}} 1$$

  This ratio of likelihood functions is called likelihood ratio(LR) and denoted by $L(y)$ and the test is called as likelihood ratio test (LRT)

- Taking log of both sides

$$\log \frac{p(y|1)}{p(y|0)} \underset{H_0}{\overset{H_1}{\gtrless}} 0$$

  The test statistic in this case is Log of LR and is called log-likelihood ratio (LLR) while the test is called as LLRT

Likelihood Ratio: MAP

- For two hypotheses case, MAP decision rule is

$$\pi_1 p(y|1) \underset{H_0}{\overset{H_1}{\gtrless}} \pi_0 p(y|0)$$

- In terms of LR, the LRT is

$$L(y) = \frac{p(y|1)}{p(y|0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\pi_0}{\pi_1}$$

- Taking log of both sides, the test is given in terms of LLR

$$\log L(y) \underset{H_0}{\overset{H_1}{\gtrless}} \log \frac{\pi_0}{\pi_1}$$

Receiver design as hypothesis testing

- Consider the multiple hypothesis testing

$$H_i \;:\; y(t) = s_i(t) + n(t) \quad i = 0, 1, \ldots, M-1$$

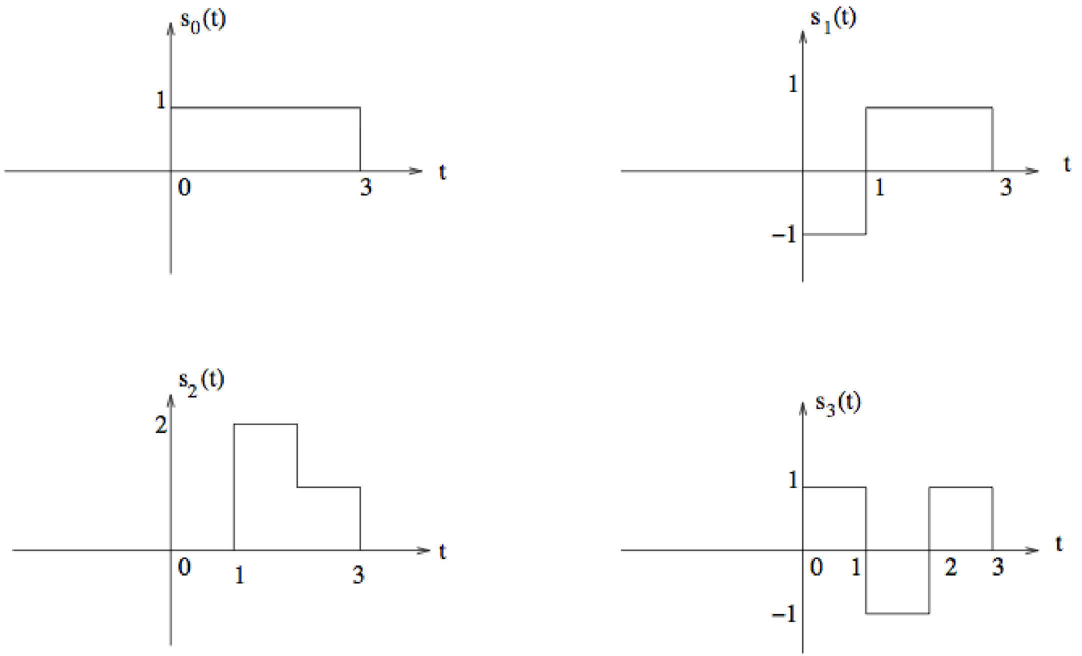
where $n(t)$ is WGN with $S_n(f) = \frac{N_0}{2} = \sigma^2$

- Strategy:
  - Show that we can reduce the continuous time received signal to a finite-dimensional vector without losing information
  - Derive the optimal receiver based on the finite-dimensional vector observation
  - Map the optimal receiver back to continuous time

- This approach is based on signal space concepts: even though the received signal lives in an infinite-dimensional space, we can restrict attention to the subspace spanned by the signals that could have been transmitted

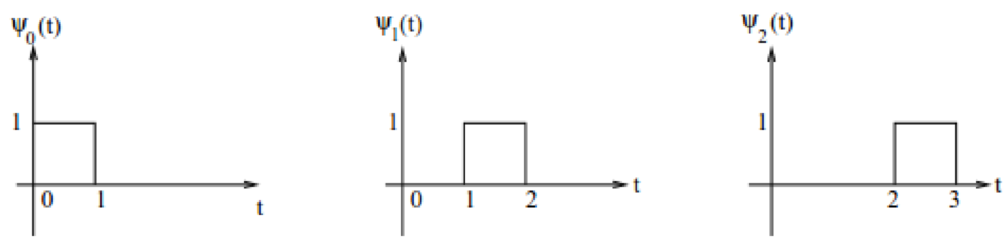
Signal space concepts: an example

- Four possible transmitted signals living in a 3-dimensional space



- Orthonormal basis (by inspection)

$$\psi_0(t) = I_{[0,1]}(t),\ \psi_1(t) = I_{[1,2]}(t),\ \psi_2(t) = I_{[2,3]}(t),$$



From Signals to Vectors

- Signal space consists of all possible linear combinations of $s_0(t), s_1(t), \ldots, s_{M-1}(t)$.
- We can always find an orthonormal basis for the signal space, i.e., $\psi_0(t)$, $\psi_1(t)$, …, $\psi_{n-1}(t)$ where $n \leq M$
- We can express each signal as a vector of basic coefficients

$$s_i(t) = \sum_{k=0}^{n-1} s_i[k]\psi_k(t)$$

  where $s_i[k] =< s_i, \psi_k >$

- Using $s_i[k]$, we can now express $s_i(t)$ as a $n \times 1$ vector $\mathbf{s}_i = [s_i[0] \; s_i[1] \; \ldots s_i[n-1]]^T$
- Note that finite-dimensional basis always exists
- If it is not possible to find natural basis by inspection, then we can always use Gram-Schmidt orthoginalization procedure

Gram-Schmidt Orthogonalization Process

- **Step 0 (Initialization)**: Let $\phi_0 = s_0$. If $\phi_0 \neq 0$, then set $\psi_0 = \frac{\phi_0}{||\phi_0||}$. Note that ψ_0 provides a basis function for $\mathcal{S}_0$.

- **Step k**: Suppose that we have constructed an orthornormal basis $\mathcal{B}_{k-1} = \{\psi_0, \ldots, \psi_{m-1}\}$ for subspace $\mathcal{S}_{k-1}$ spanned by the first k signals $s_0, \ldots, s_{k-1}$. Note that $m \leq k$. Define

$$\phi_k(t) = s_k(t) - \sum_{i=0}^{m-1} < s_k, \psi_i > \psi_i(t)$$

- The signal $\phi_k(t)$ is the component of $s_k(t)$ orthogonal to the subspace $\mathcal{S}_{k-1}$. If $\phi_k \neq 0$, define a new basis function $\psi_m = \frac{\phi_k(t)}{||\phi_k||}$ and the basis as $\mathcal{B}_k = \{\psi_0, \ldots, \psi_m\}$. If $\phi_k = 0$, then $s_k \in \mathcal{S}_{k-1}$ and it is not necessary to update the basis. In this case $\mathcal{B}_k = \mathcal{B}_{k-1} = \{\psi_0, \ldots, \psi_{m-1}\}$.

From signals to vectors

- Signal space consists of all possible linear combinations of $s_0(t), s_1(t), \ldots, s_{M-1}(t)$.
- We can always find an orthonormal basis for the signal space, i.e., $\psi_0(t), \psi_1(t), \ldots, \psi_{n-1}(t)$ where $n \leq M$

We can express each signal as a vector of basis coefficients

$$s_i(t) = \sum_{l=0}^{n-1} s_i[k]\psi_k(t), \text{ where } s_i[k] = \langle s_i, \psi_k \rangle \quad \leftrightarrow \mathbf{s}_i = \begin{pmatrix} s_i[0] \\ s_i[1] \\ . \\ . \\ . \\ s_i[n-1] \end{pmatrix}$$

Inner products are preserved

- Performance of M-ary signaling in AWGN depends only on the inner products between the signals if the noise PSD is fixed.
- When mapping the CT hypothesis testing problem to DT, it is important to check that these inner products are preserved when projecting them to signal space.
- Check the following

$$\langle s_i, s_j \rangle = \int s_i(t) s_j(t) dt = \sum_{k=0}^{n-1} s_i[k] s_j[k] = \langle \mathbf{s}_i, \mathbf{s}_j \rangle$$

Inner products are preserved from CT to DT and vice versa
➔ norms, energies, distances are preserved

Modeling WGN in signal space

- What about projection of noise on the signal subspace?
- The noise projection onto the i^{th} basis function as

$$N[i] = \langle \boxed{n}, \psi_i \rangle = \int n(t)\psi_i(t)$$

Noise

for $i = 0, 1, \ldots, \boxed{n} - 1$

Dimension of signal space

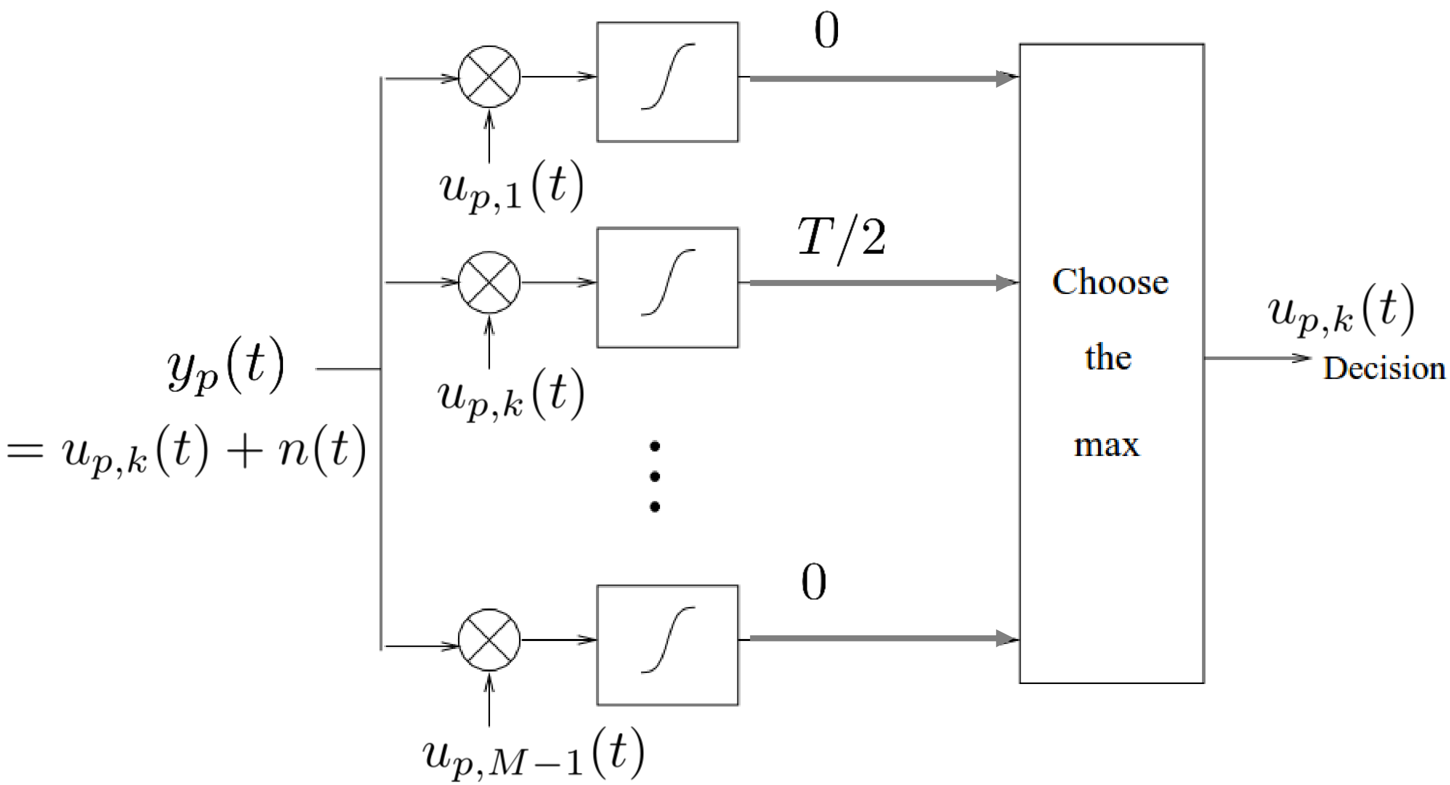
- We can write noise $n(t)$ as follows

$$n(t) = \boxed{\sum_{i=0}^{n-1} N[i]\psi_i(t)} + \boxed{n^{\perp}(t)}$$

where $n^{\perp}(t)$ is the projection of noise orthogonal noise subspace.

Today’s Class

Recap: *FSK coherent demodulation*



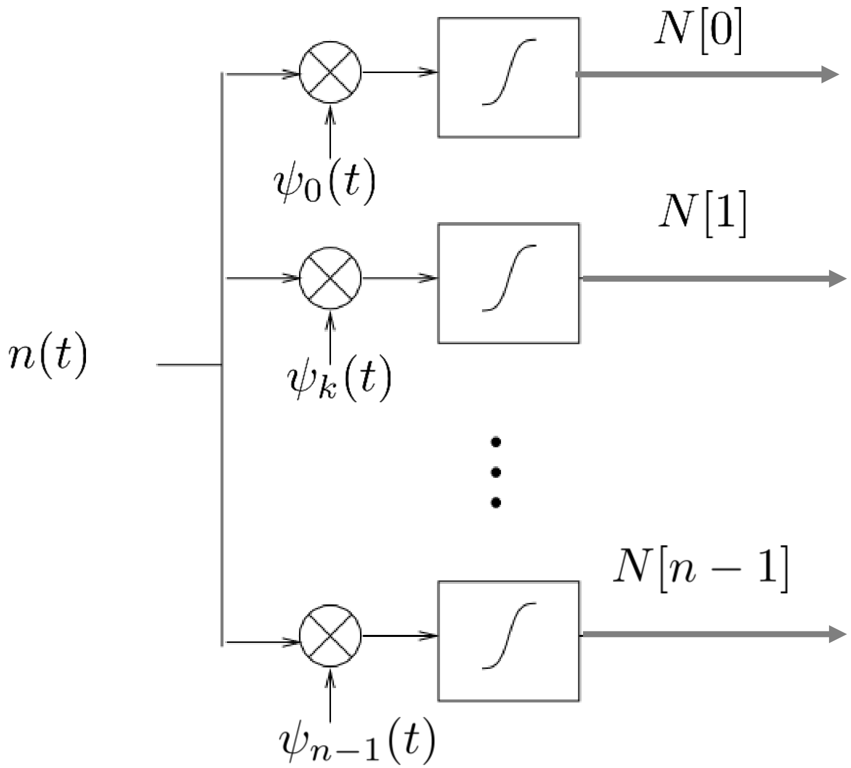
- Correlate the incoming signal with all M reference sinusoidal tones (passband signal) given by

$$u_{p,k} = \cos(2\pi(f_c + k\Delta f)t) \quad 0 \leq t \leq T$$

for $k = 0, 1, \ldots, M-1$.

- Choose $u_{p,m}$ for which the output is maximum among all the correlated outputs.

Noise projection on Signal Subspace



- Prove that noise projection on signal space is discrete-time WGN.
- For this we will need theorem on the next slide!

Noise projection on Signal Subspace

- **WGN through correlators** The random variables $Z_1 = \langle n, u_1 \rangle$ and $Z_2 = \langle n, u_2 \rangle$ are zero-mean, jointly Gaussian, with

$$\text{cov}(Z_1, Z_2) = \sigma^2 \langle u_1, u_2 \rangle$$

so that $\mathbf{Z} = (Z_1, Z_2)^T \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ with covariance matrix

$$\mathbf{C} = \begin{pmatrix} \sigma^2 ||u_1||^2 & \sigma^2 \langle u_1, u_2 \rangle \\ \sigma^2 \langle u_1, u_2 \rangle & \sigma^2 ||u_2||^2 \end{pmatrix}$$

Proof!

- Prove that noise projection on signal space is discrete-time WGN.

Geometric interpretation of WGN

- The projection of WGN in any direction in signal space is an $\mathcal{N}(0, \sigma^2)$ random variable
- Projections in orthogonal directions are independent
- Projections along an orthonormal basis are iid $\mathcal{N}(0, \sigma^2)$ random variables

Hypothesis testing in signal space

- For the H_i hypothesis, the received signal is given by

$$y(t) = s_i(t) + n(t)$$

for $i = 0, 1, \ldots, M-1$

- By projecting this on signal space, we get

$$Y[k] = \langle y, \psi_k \rangle = s_i[k] + N[k] \quad k = 0, 1, \ldots, n-1$$

- On collecting these into an $n-$dimensional vector, we get

$$\mathbf{Y} = \mathbf{s}_i + \mathbf{N}$$

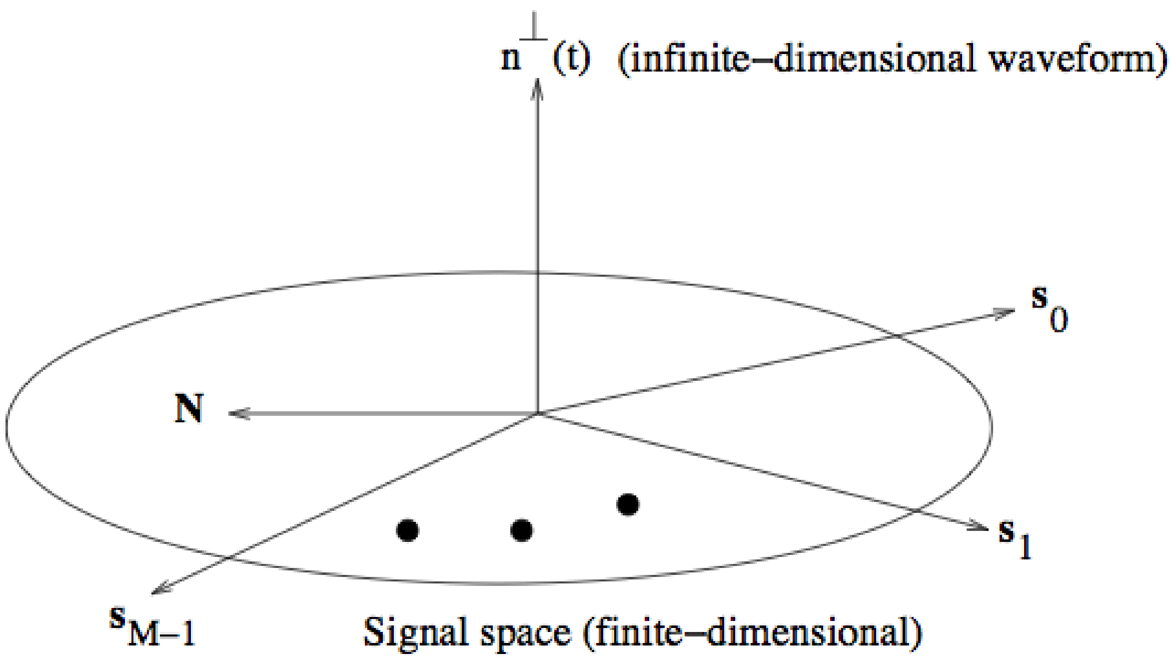
with $\mathbf{N} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$

- Note that the vector $\mathbf{Y}$ completely describes the component of the received signal $y(t)$ in the signal space given by

$$y_{\mathcal{S}} = \sum_{j=0}^{n-1} \langle y, \psi_j \rangle \psi_j(t) = \sum_{j=0}^{n-1} Y[j] \psi_j(t)$$

Irrelevance of orthogonal-noise component

- The noise component orthogonal to signal space, denoted by $n^{\perp}$,
  - does not give any information about signal
  - does not give any information about noise components in the signal space



Optimal Reception in AWGN

- Prove that the MAP/MPE rule for optimum demodulation in AWGN is

$$\delta_{\mathrm{MAP}}(\mathbf{y}) = \arg\max_i ||\mathbf{y} - \mathbf{s}_i||^2 - 2\sigma^2 \log \pi_i$$

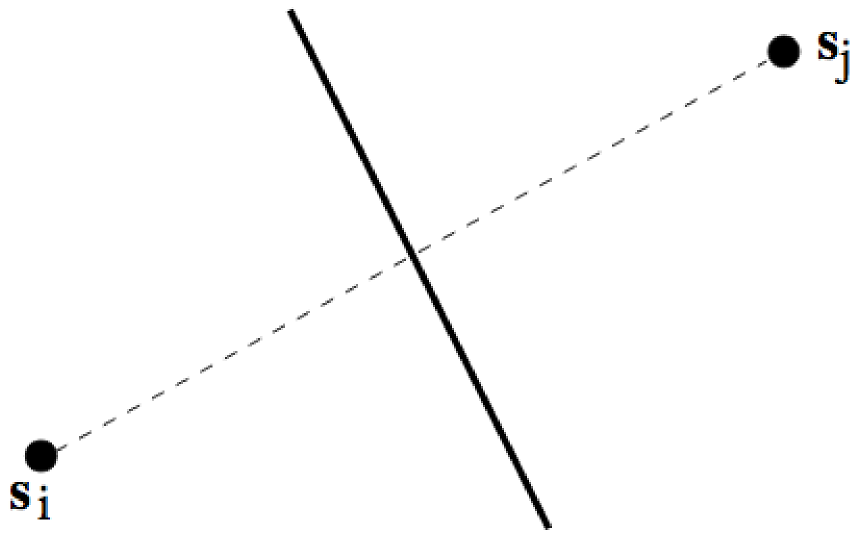
- For equiprobable prior probabilities, MAP is equivalent to ML

$$\begin{aligned}
\delta_{\mathrm{ML}}(\mathbf{y}) &= \delta_{\mathrm{MAP}}(\mathbf{y}) \\
&= \arg\min_i ||\mathbf{y} - \mathbf{s}_i||^2 - 2\sigma^2 \log \frac{1}{M} \\
&= \arg\min_i ||\mathbf{y} - \mathbf{s}_i||^2
\end{aligned}$$

This is **minimum distance rule**!

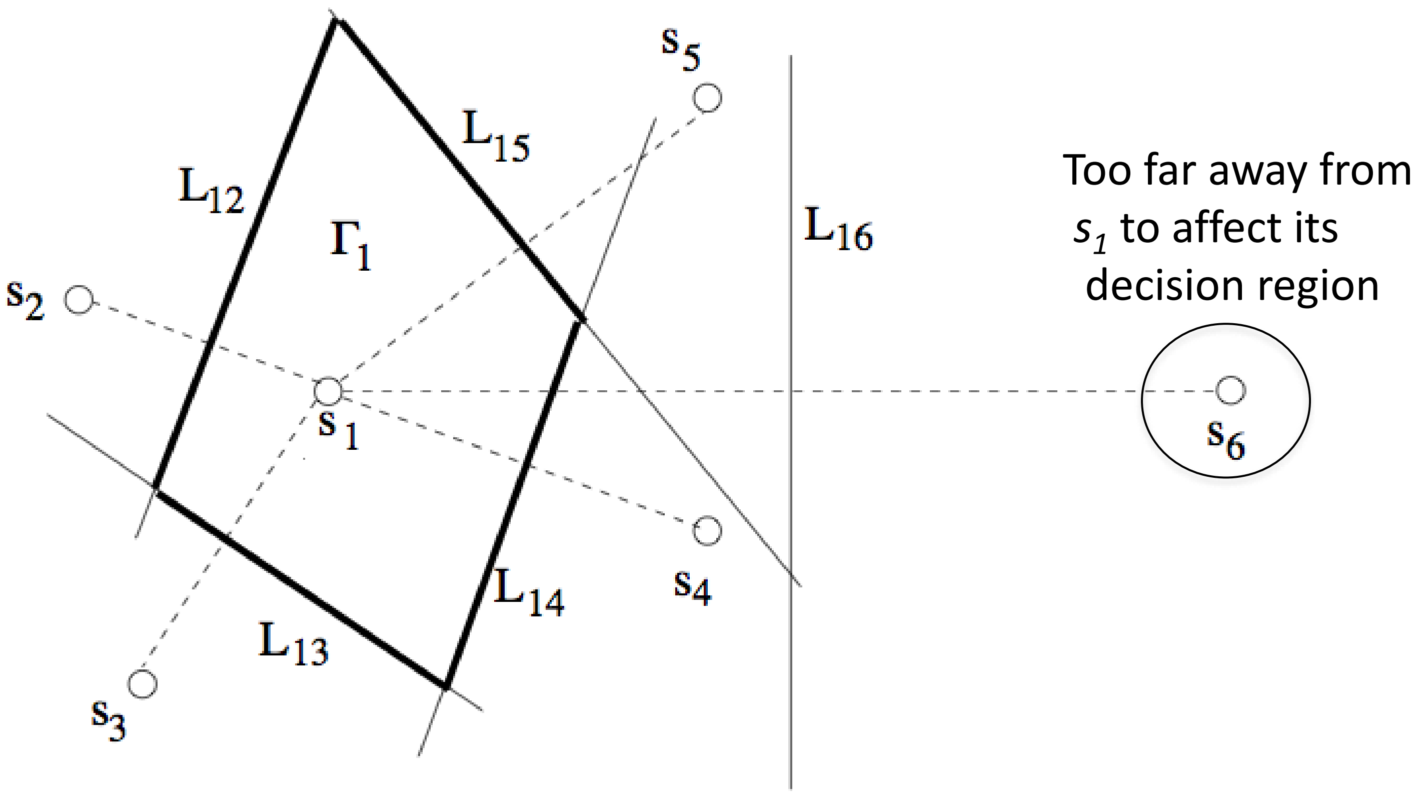
Min distance rule between 2 signals

- Draw line between signal points.
- The ML decision boundary between them is the line bisecting this in 2D
- For two signal points in n dimensions, the ML decision boundary is $(n-1)$ dimensional hyperplane bisecting the line connecting the two points



ML decision boundary
is an (n–1) dimensional hyperplane

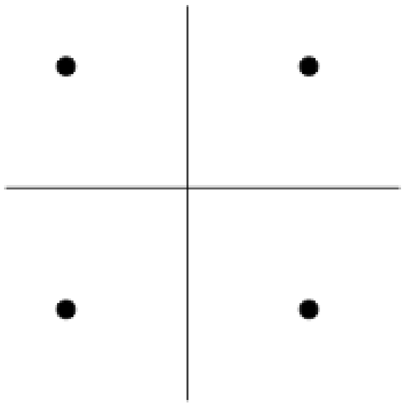
Geometry of min distance rule



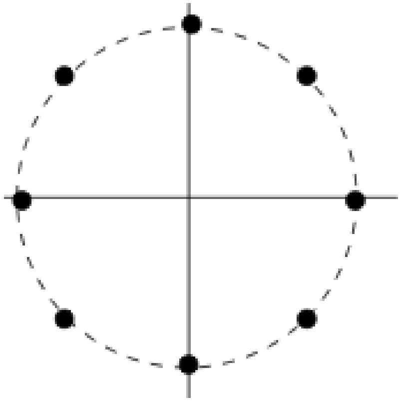
- Decision region for a signal is where it wins compared to all other signals
- Intersect the half-planes corresponding to all the pairwise comparisons
- Note that the decision region is typically determined by nearby neighbors

2 Dimensional Examples

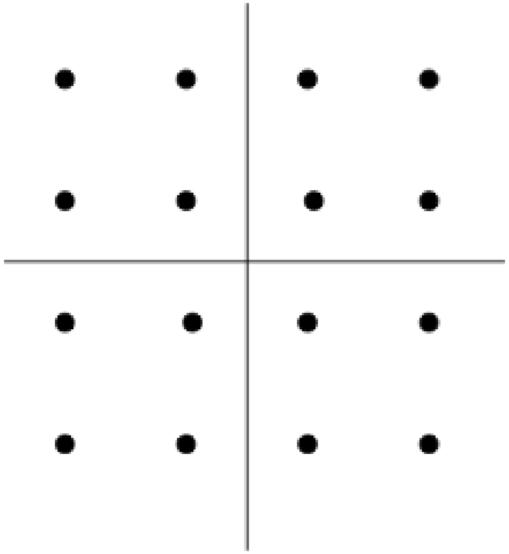
QPSK/4PSK/4QAM



8PSK

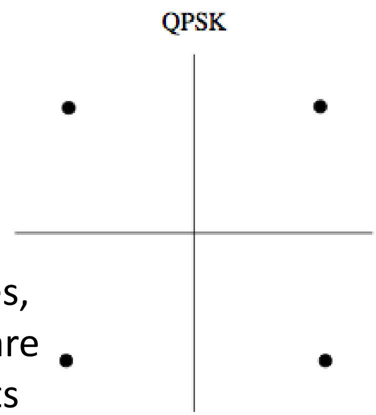


16QAM

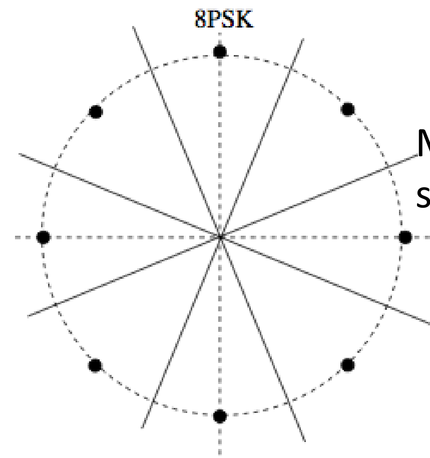


Solve: Find decision regions for these constellations for ML demodulation

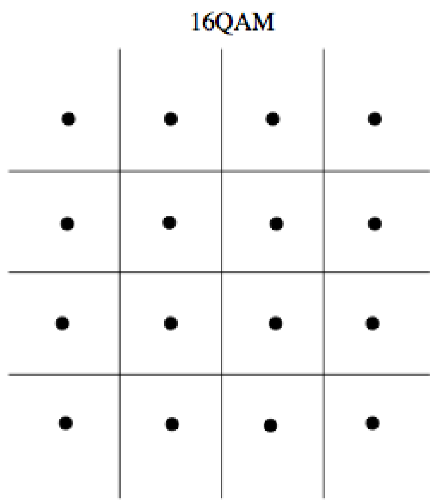
ML decisions for common 2D constellations



ML boundaries parallel to I/Q axes, decision regions are the four quadrants



ML regions are sectors



ML boundaries parallel to I/Q axes

MAP Demodulation In CT

- Remember that optimal modulation for signaling in DT AWGN is the MAP/MPE rule given by

DT(Finite Dimensional)

$$\begin{aligned}\delta_{\text{MAP}}(\mathbf{y}) &= \arg\min_i ||\mathbf{y} - \mathbf{s}_i||^2 - 2\sigma^2 \log \pi_i \\ &= \arg\max_i \langle \mathbf{y}, \mathbf{s}_i \rangle - ||\mathbf{s}_i||^2/2 + 2\sigma^2 \log \pi_i\end{aligned}$$

- Equivalently, optimal demodulation for signaling in CT AWGN is

$$\begin{aligned}\delta_{\text{MAP}}(\mathbf{y}) &= \arg\max_i \langle y, s_i \rangle - ||s_i||^2/2 + \sigma^2 \log \pi_i \\ &= \arg\max_i \langle y_{\mathcal{S}}, s_i \rangle - ||s_i||^2/2 + \sigma^2 \log \pi_i\end{aligned}$$

CT(Infinite Dimensional)

Recap: Receiver design as hypothesis testing

- Consider the multiple hypothesis testing

$$H_i \; : \; y(t) = s_i(t) + n(t) \quad i = 0, 1, \ldots, M-1$$

where $n(t)$ is WGN with $S_n(f) = \frac{N_0}{2} = \sigma^2$

- Strategy:
  - Show that we can reduce the continuous time received signal to a finite-dimensional vector without losing information
  - Derive the optimal receiver based on the finite-dimensional vector observation
  - Map the optimal receiver back to continuous time

- This approach is based on signal space concepts: even though the received signal lives in an infinite-dimensional space, we can restrict attention to the subspace spanned by the signals that could have been transmitted

ML Demodulation In CT

- For equiprobable prior probabilities, MAP is equivalent to ML

$$\delta_{\mathrm{ML}}(\mathbf{y}) = \arg\min_i ||\mathbf{y} - \mathbf{s}_i||^2$$

DT(Finite Dimensional)

$$= \arg\max_i \langle y, s_i \rangle - ||s_i||^2/2$$

CT(Infinite Dimensional)

- Here we have used

$$||\mathbf{s}_i||^2 = ||s_i||^2$$

$$\begin{aligned}
\langle y, s_i \rangle &= \langle y_{\mathcal{S}} + y^{\perp}, s_i \rangle \\
&= \langle y_{\mathcal{S}}, s_i \rangle + \langle y^{\perp}, s_i \rangle \\
&= \langle y_{\mathcal{S}}, s_i \rangle \\
&= \langle \mathbf{y}, \mathbf{s}_i \rangle
\end{aligned}$$

- Note that we don't have minimum distance rule in CT as the squares of distance in CT will be

$$\begin{aligned}
||y - s_i||^2 &= ||y_{\mathcal{S}} - s_i||^2 + ||y^{\perp}||^2 \\
&= ||y_{\mathcal{S}} - s_i||^2 + \boxed{||n^{\perp}||^2}
\end{aligned}$$

Infinite Power

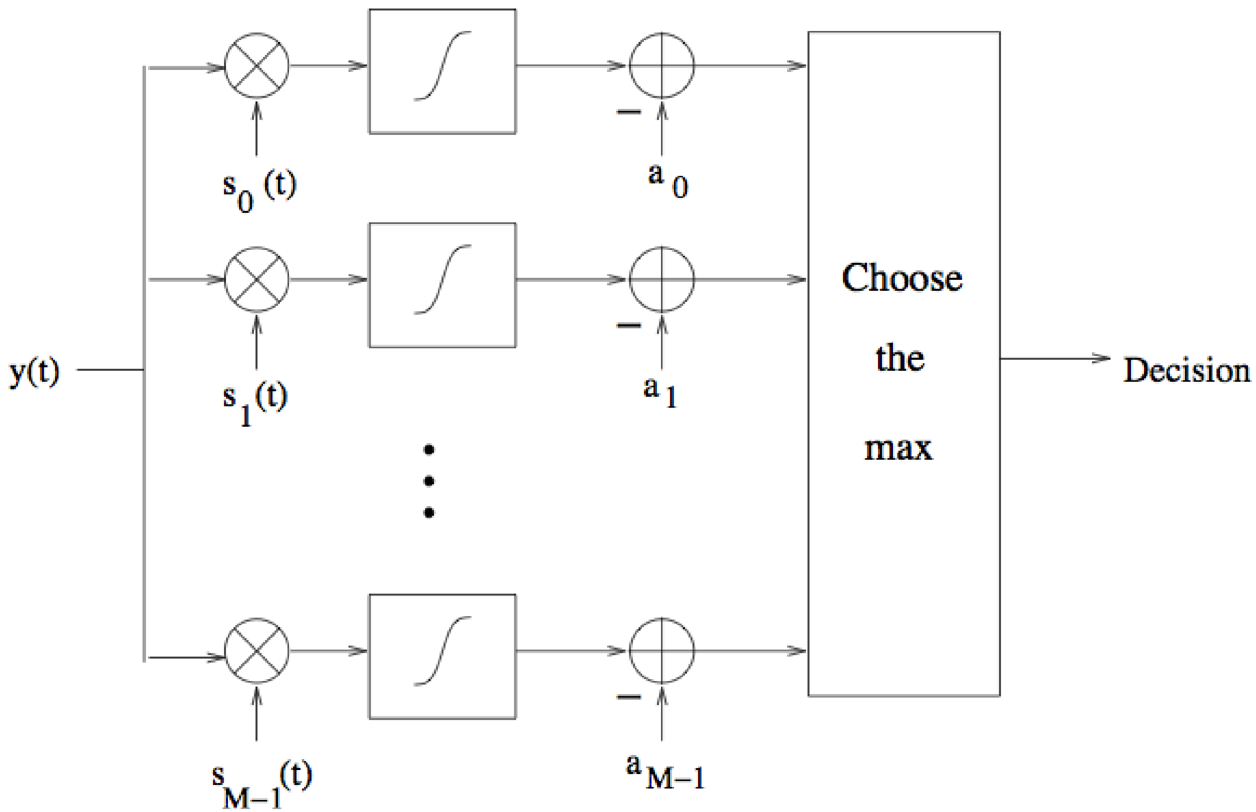
Implementation using correlation

- For equiprobable prior probabilities, MAP is equivalent to ML

$$\delta_{\mathrm{ML}}(\mathbf{y}) = \arg\min_i ||\mathbf{y} - \mathbf{s}_i||^2$$ DT(Finite Dimensional)

$$= \arg\max_i \langle y, s_i \rangle - ||s_i||^2/2$$

CT(Infinite Dimensional)



Equivalent Match Filter Implementation

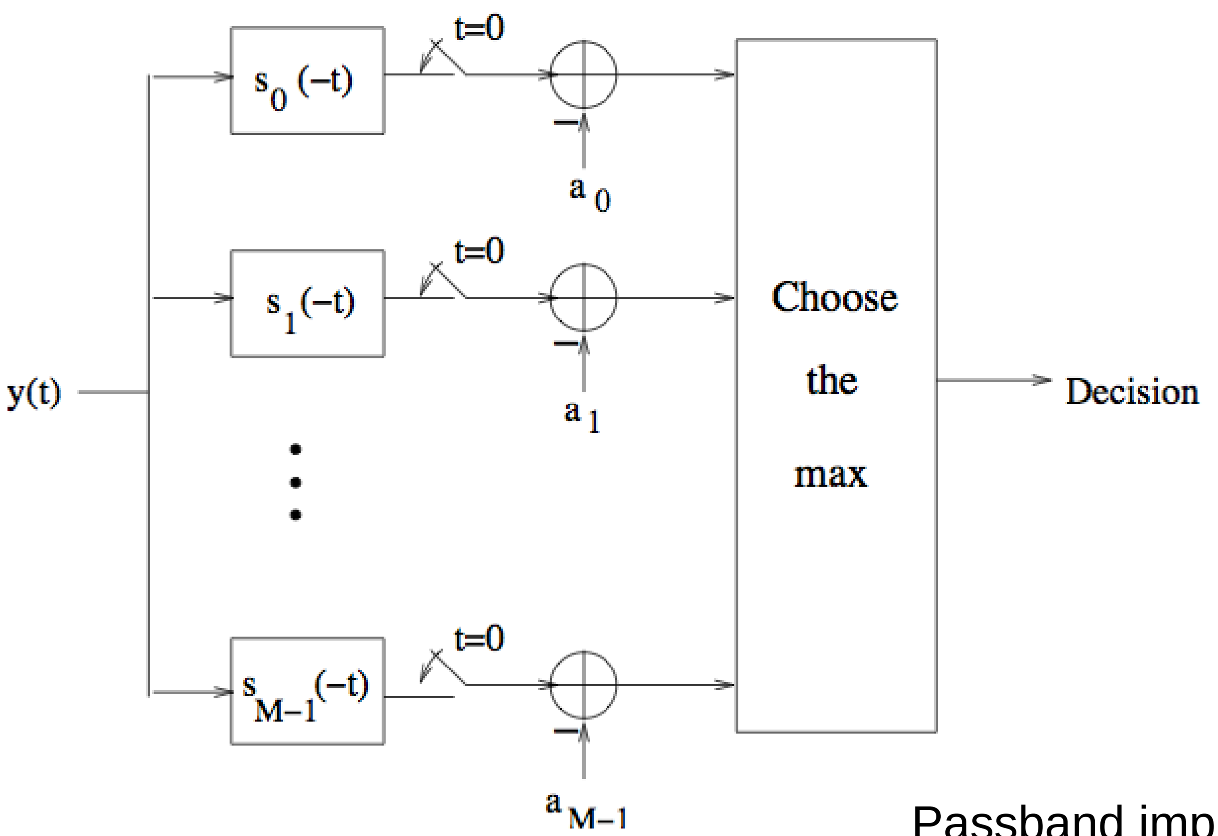
- For equiprobable prior probabilities, MAP is equivalent to ML

$$\delta_{\mathrm{ML}}(\mathbf{y}) = \arg\min_i ||\mathbf{y} - \mathbf{s}_i||^2$$

DT(Finite Dimensional)

$$= \arg\max_i \langle y, s_i \rangle - ||s_i||^2/2$$

CT(Infinite Dimensional)



Passband implementation

Implementation in Complex Baseband

- For equiprobable prior probabilities, MAP is equivalent to ML

$$\delta_{\mathrm{ML}}(\mathbf{y}) = \arg\min_i ||\mathbf{y} - \mathbf{s}_i||^2$$
$$= \arg\max_i \langle y, s_i \rangle - ||s_i||^2/2$$

- Note the relation between correlation in passband and baseband is given by

$$\langle u_p, v_p \rangle = \frac{1}{2}\mathrm{Re}\langle u, v \rangle = \frac{1}{2}(\langle u_c, v_c \rangle + \langle u_s, v_s \rangle)$$