8 Optimal Detection for Additive Noise Channels: 1-D Case

We now derive the optimal demodulator. From the previous section, we have seen that instead of analyzing the waveform channel, we can convert it to an equivalent vector channel. The size of the vector is the same as the size K of the orthonormal basis for the waveforms $s_1(t), s_2(t), \ldots, s_M(t)$. In this section, we will assume $K = 1$. This is the case, for example, when we use PAM.

Definition 8.1. Detection Problem: When $K = 1$, our problem under consideration is simply that of **detecting** the scalar message S in the presence of additive noise N. The received signal R is given by

$$R = S + N.$$

- S is selected from an alphabet $\mathcal{S}$ containing M possible values $s^{(1)}$, $s^{(2)}$, …, $s^{(M)}$.
- $p_S\left(s^{(j)}\right) = P\left[S = s^{(j)}\right] \equiv p_j$.
- S and N are independent.

A detector's job is to guess the value of the channel input S from the value of the received channel output R. We denote this guessed value by $\hat{S}$. An optimal detector is the one that minimizes the (symbol) error probability $P(\mathcal{E}) = P\left[\hat{S} \neq S\right]$.

8.2. The analysis here is very similar to what we have done in Section 3. Here, for clarity, we note some important differences:

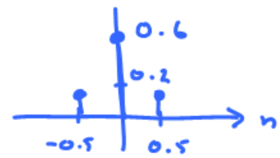
- In Section 3, The channel input and output are denoted by X and Y, respectively. Here, they are denoted by S and R.
- In Section 3, the transition probabilities are arbitrary and summarized by the matrix $\mathbf{Q}$. Here, the transition probabilities is basically controlled by the additive noise.
- In Section 3, both X and Y are discrete. Here, S is discrete. However, because noise is continuous, R will be a continuous random variable.

Even with these differences, several techniques that we used in Section 3 will be applicable here.

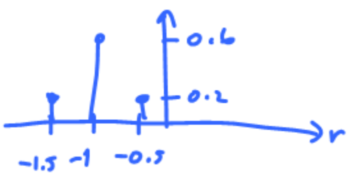
Example 8.3. Review: When the additive noise is discrete, we may attempt to write down the $\mathbf{Q}$ matrix. Suppose

$$p_S(s) = \begin{cases} 0.3, & s = -1, \\ 0.7, & s = 1, \\ 0, & \text{otherwise,} \end{cases} \quad \text{and } p_N(n) = \begin{cases} 0.2, & n \in \{-0.5, +0.5\}, \\ 0.6, & n = 0, \\ 0, & \text{otherwise.} \end{cases}$$

Because $R = S + N$, we know that



(a) given $S = -1$, we have $R = -1 + N$:



$$p_{R|S}(r|-1) = \begin{cases} 0.2, & r \in \{-1.5, -0.5\} \\ 0.6, & r = -1, \\ 0, & \text{otherwise.} \end{cases}$$

n	$-1+n$
-0.5	-1.5
0	-1
0.5	-0.5

(b) given $S = 1$, we have $R = 1 + N$:

$p_{R|S}(r|1)$

The $\mathbf{Q}$ matrix is given by

$$\mathbf{Q} \quad \begin{array}{c|cccccc} s \backslash r & -1.5 & -1 & -0.5 & 0.5 & 1 & 1.5 \\ \hline -1 & 0.2 & 0.6 & 0.2 & 0 & 0 & 0 \\ +1 & 0 & 0 & 0 & 0.2 & 0.6 & 0.2 \end{array}$$

Note that each row of the $\mathbf{Q}$ matrix is simple a shifted copy of the noise pmf. The amount of shift is the corresponding value of s for that row.

8.4. Formula-wise, when the additive noise is discrete, each row of the $\mathbf{Q}$ matrix (as in Example 8.3) is given by

$R = S + N$

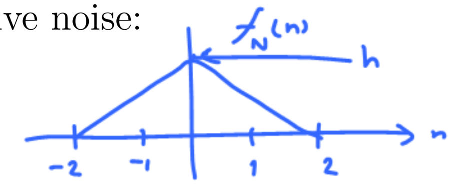
$$p_{R|S}(r|s) = p_N(r - s). \qquad r \quad s \tag{43}$$

8.5. When the additive noise is continuous, there are uncountably many possible values for the channel output R. Hence, representing conditional probabilities in the form of a matrix $\mathbf{Q}$ does not make sense here.

When R is continuous, the conditional pmf $p_{R|S}(r|s)$ is replaced by the conditional pdf $f_{R|S}(r|s)$. For additive noise N with pdf $f_N(n)$, we have

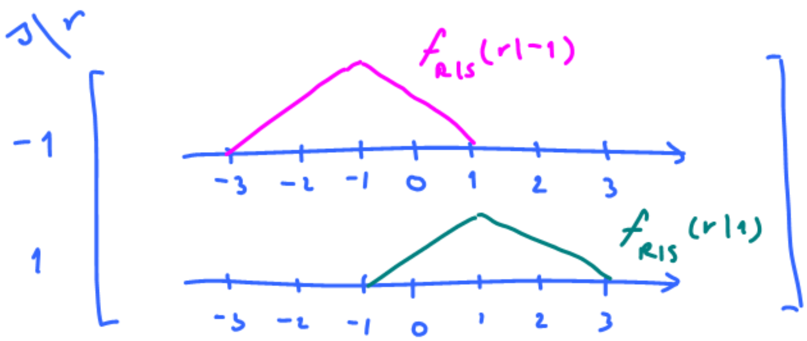
$$f_{R|S}(r|s) = f_N(r-s). \tag{44}$$

Example 8.6. Suppose the discrete additive noise in Example 8.3 is replaced by a continuous additive noise:



$\frac{1}{2} \times h \times 4 = 1$

$h = \frac{1}{2}$



8.7. The optimal detector, which minimize the error probability, is the **MAP detector**:

$$\hat{s}_{\mathrm{MAP}}(r) = \arg\max_{s \in \mathcal{S}} p_S(s) f_{R|S}(r\,|s) = \arg\max_{s \in \mathcal{S}} p_S(s) f_N(r-s). \tag{45}$$

Because event $[W = j]$ is the same as event $[S = s^{(j)}]$, we also have

$$\hat{w}_{\mathrm{MAP}}(r) = \arg\max_{j \in \{1,2,\ldots,M\}} p_j f_N\left(r - s^{(j)}\right). \tag{46}$$

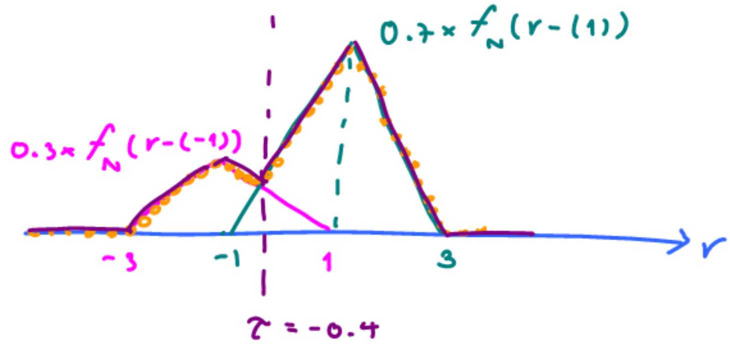
When the prior probabilities are ignored, we have the (sub-optimal) **ML detector**:

$$\hat{s}_{\mathrm{ML}}(r) = \arg\max_{s \in \mathcal{S}} f_{R|S}(r\,|s) = \arg\max_{s \in \mathcal{S}} f_N(r-s). \tag{47}$$

and

$$\hat{w}_{\mathrm{ML}}(r) = \arg\max_{j \in \{1,2,\ldots,M\}} f_N\left(r - s^{(j)}\right). \tag{48}$$

Example 8.8. Back to Example 8.6.



$$\hat{s}_{MAP}(r) = \begin{cases} -1, & r < \tau \\ 1, & r \geq \tau \\ \cancel{\text{any}}, & \cancel{r = \tau} \end{cases}$$

$$\hat{s}_{ML}(r) = \begin{cases} -1, & r < 0 \\ 1, & r \geq 0 \end{cases}$$

detection

decoding

1, 2, ..., M

Definition 8.9. The ith **decision "region"**, denoted by $\mathcal{D}_i$ for a decoder $\hat{s}(r)$ is defined as the collection of all the r values at which r is decoded as $s^{(i)}$.

- The collection $\mathcal{D}_1$, $\mathcal{D}_2$, ...,$\mathcal{D}_M$ should partition the whole observable values (support) of R.

Example 8.10. Back to Example 8.6.

$$\hat{s}(r) = \begin{cases} s^{(1)}, & r \in \mathcal{D}_1 = (-\infty, \tau) \\ s^{(2)}, & r \in \mathcal{D}_2 = [\tau, \infty) \end{cases}$$

8.11. Gaussian Noise: When the noise N is Gaussian with mean 0 and standard deviation σ_N,

$$f_N(n) = \frac{1}{\sqrt{2\pi}\sigma_N} e^{-\frac{1}{2}\left(\frac{n}{\sigma_N}\right)^2}.$$

Definition 8.12. In general, a **Gaussian (normal) random variable** X with mean m and standard deviation σ is characterized by its probability density function (PDF):

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}.$$

To talk about such X, we usually write $X \sim \mathcal{N}(m, \sigma^2)$. Probability involving X can be evaluated by

$$P[X \in A] = \int_A f_X(x)dx.$$

In particular,

$$P[X \in [a,b]] = \int_a^b f_X(x)dx = F_X(b) - F_X(a)$$

where $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$ is called the cumulative distribution function (CDF) of X.

We usually express probability involving Gaussian random variable via the Q function which is defined by

$$Q(z) = \int_z^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Note that $Q(z)$ is the same as $P[S > z]$ where $S \sim \mathcal{N}(0,1)$; that is $Q(z)$ is the probability of the "tail" of $\mathcal{N}(0,1)$.

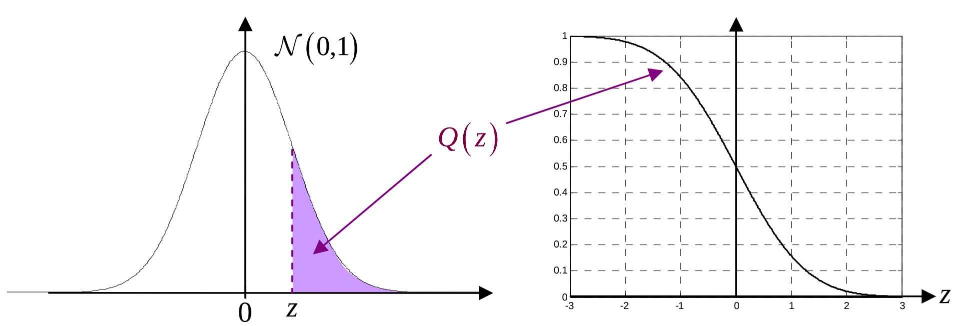


Figure 25: Q-function

It can be shown that

- Q is a decreasing function
- $Q(0) = \frac{1}{2}$
- $Q(-z) = 1 - Q(z)$
    - This is useful for converting the argument of the Q function to positive value.
- For $X \sim \mathcal{N}(m, \sigma^2)$,

$$P[X > c] = Q\left(\frac{c-m}{\sigma}\right).$$

8.13. Three important noise probabilities for $N \sim \mathcal{N}(0, \sigma_N^2)$:

$P[N > c] = Q\left(\frac{c}{\sigma_N}\right)$, $P[N < c] = 1 - Q\left(\frac{c}{\sigma_N}\right)$ $P[a < N < b] = Q\left(\frac{a}{\sigma_N}\right) - Q\left(\frac{b}{\sigma_N}\right)$

Note that all strict inequalities above can also be replaced by the ones that also include equalities because the noise is a continuous random variable and hence including one particular noise value does not change probability.

8.14. For additive noise channel where $R = S + N$, $S \perp\!\!\!\perp N$, and $N \sim \mathcal{N}(0, \sigma_N^2)$

$$\hat{s}_{MAP}(r) = \underset{s \in \{s^{(1)}, s^{(2)}, \ldots, s^{(M)}\}}{\arg\max}\; p_S(s) \underbrace{f_N(r-s)}_{\cancel{\frac{1}{\sqrt{2\pi}\,\sigma_N}} e^{-\frac{1}{2}\left(\frac{r-s}{\sigma_N}\right)^2}}$$

$$= \arg\max_s \left(\ln p_S(s)\right) - \frac{1}{2}\left(\frac{r-s}{\sigma_N}\right)^2 = \arg\max_s \left(2\sigma_N^2 \ln p_S(s) - (r-s)^2\right)$$

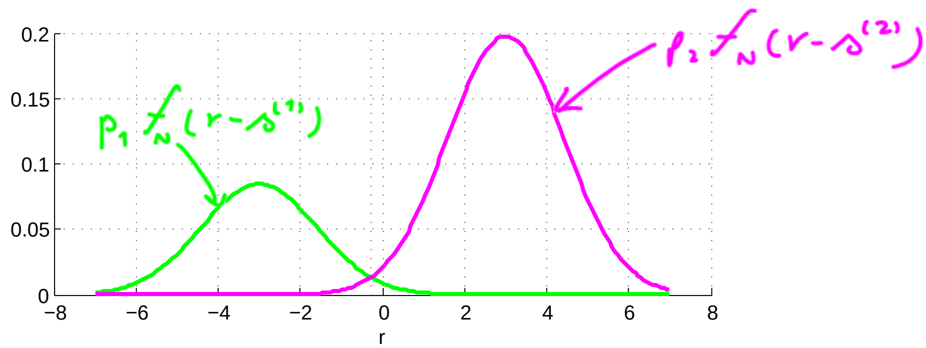
$$\hat{s}_{\mathrm{MAP}}(r) = \arg\max_{s \in S} \left(2\sigma_N^2 \ln p_S(s) - (r-s)^2\right)$$
$$= \arg\max_{s \in S} \left(\sigma_N^2 \ln p_S(s) - \frac{E_s}{2} + s \cdot r\right),$$

and

$$\hat{s}_{\mathrm{ML}}(r) = \arg\min_{s \in S} (r-s)^2 = \arg\min_{s \in S} d(r, s).$$

minimum distance detector

Example 8.15. In a binary antipodal signaling scheme, the message S is randomly selected from the alphabet set $\mathcal{S} = \{-3, 3\}$ with $P[S = -3] = 0.3$ and $P[S = 3] = 0.7$. The message is corrupted by an independent additive noise $N \sim \mathcal{N}(0, 2)$. Find the MAP detector $\hat{s}_{\text{MAP}}(r)$.



For a given r, $\hat{s}_{MAP}(r) = s^{(2)}$

$$p_2 f_N(r-s^{(2)}) > p_1 f_N(r-s^{(1)})$$

$$p_2 \cancel{\frac{1}{\sqrt{2\pi}\sigma_N}} e^{-\frac{1}{2}\left(\frac{r-s^{(2)}}{\sigma_N}\right)^2} > p_1 \cancel{\frac{1}{\sqrt{2\pi}\sigma_N}} e^{-\frac{1}{2}\left(\frac{r-s^{(1)}}{\sigma_N}\right)^2}$$

$A^2 - B^2 = (A+B)(A-B)$

$$e^{-\frac{1}{2\sigma_N^2}\left((r-s^{(2)})^2-(r-s^{(1)})^2\right)} > \frac{p_1}{p_2}$$

$$-\frac{1}{2\sigma_N^2}\left((2r-s^{(1)}-s^{(2)})(s^{(1)}-s^{(2)})\right) > \ln\frac{p_1}{p_2}$$

so, $\hat{s}_{MAP}(r) = \begin{cases} s^{(1)} & \text{if } r < \tau_{MAP} \\ s^{(2)} & \text{if } r \geq \tau_{MAP} \end{cases}$

$$r > \underbrace{\frac{\sigma_N^2 \ln\frac{p_1}{p_2}}{s^{(2)}-s^{(1)}} + \frac{s^{(1)}+s^{(2)}}{2}}_{\tau_{MAP}} = -0.2824$$

8.16. The error probability of a detector can be found via its success probability

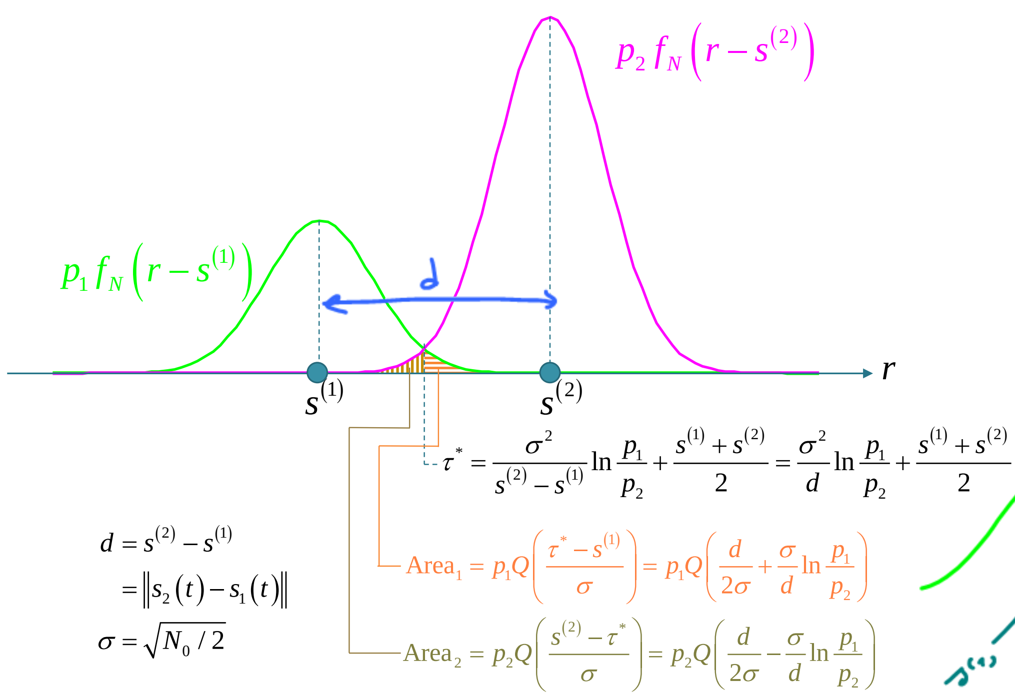
$$\begin{aligned}
P(\mathcal{C}) &= \sum_{i=1}^{M} P\left(\mathcal{C}|S = s^{(i)}\right) P\left[S = s^{(i)}\right] = \sum_{i=1}^{M} P\left[R \in D_i \middle| S = s^{(i)}\right] p_i \\
&= \sum_{i=1}^{M} p_i P\left[S + N \in D_i \middle| S = s^{(i)}\right] = \sum_{i=1}^{M} p_i P\left[N + s^{(i)} \in D_i\right] \\
&= \sum_{i=1}^{M} p_i \int\limits_{D_i} f_N\left(r - s^{(i)}\right) dr = \sum_{i=1}^{M} \int\limits_{D_i} p_i f_N\left(r - s^{(i)}\right) dr.
\end{aligned}$$

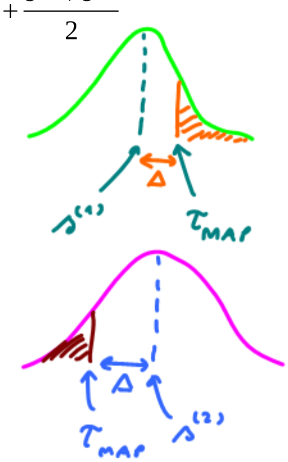
This gives

$$
\begin{aligned}
P(\mathcal{E}) &= 1 - P(\mathcal{C}) \\
&= \sum_{i=1}^{M} p_i \int_{D_i^c} f_N\left(r - s^{(i)}\right) dr = \sum_{i=1}^{M} \int_{D_i^c} p_i f_N\left(r - s^{(i)}\right) dr.
\end{aligned}
$$

Although, at first, the above expressions may look complicated, it is similar to what we used when we did in Section 3.

Example 8.17. Back to Example 8.15.





$$
\begin{aligned}
P(\mathcal{E}) &= p_1 Q\left(\frac{\tau^* - s^{(1)}}{\sigma}\right) + p_2 Q\left(\frac{s^{(2)} - \tau^*}{\sigma}\right) \\
&= p_1 Q\left(\frac{d}{2\sigma} + \frac{\sigma}{d}\ln\frac{p_1}{p_2}\right) + p_2 Q\left(\frac{d}{2\sigma} - \frac{\sigma}{d}\ln\frac{p_1}{p_2}\right)
\end{aligned}
$$

We can see from the last expression that the error probability of the optimal (MAP) detector depends on $s^{(1)}$ and $s^{(2)}$ only through their distance d.

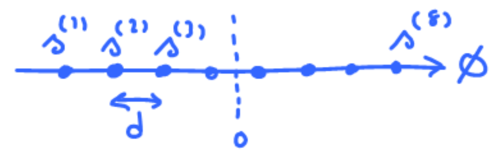
1 d

Definition 8.18. In **"standard"** multi-level **PAM**, we required that spacing between all adjacent signals to be the same. Furthermore, all M signals 2 are equally likely. To minimize the average energy, we also require that the 3 constellation is "centered" around zero.

Suppose the distance between adjacent signals is d, then the M signals are represented in the constellation by

$s^{(1)} = \frac{d}{2}(2-1-M) = -\frac{d}{2}(M-1)$

$$
s^{(j)} = \frac{d}{2}(2j - 1 - M).
$$



$s^{(M)} = \frac{d}{2}(2M-1-M) = \frac{d}{2}(M-1)$

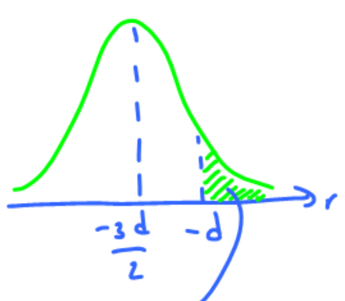
When M is even,

$S = \left\{ \pm\frac{d}{2}, \pm\frac{3d}{2}, \ldots, \pm(M-1)\frac{d}{2} \right\}$

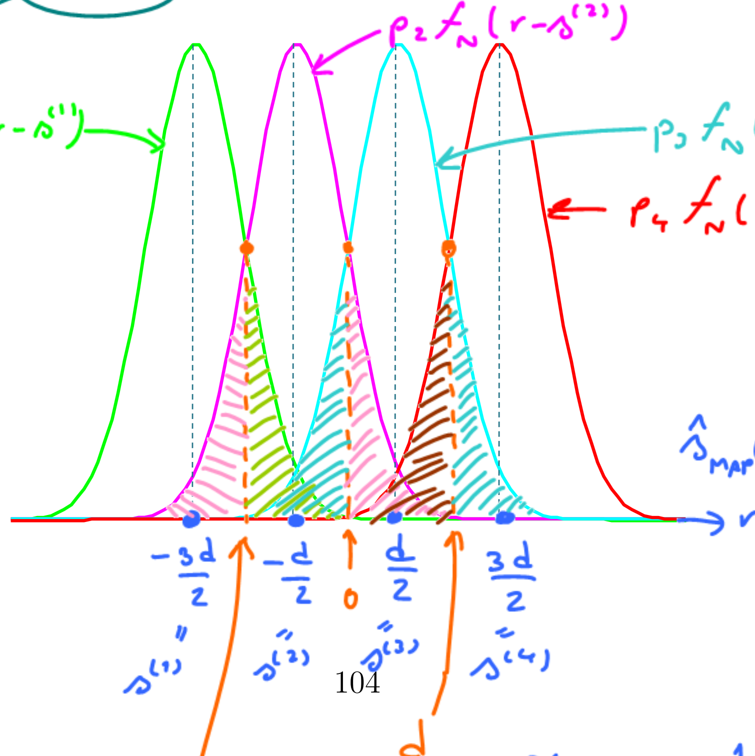
$M=4$

Example 8.19. Standard Quaternary PAM under additive Gaussian noise

$p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$



area $= \frac{1}{4} Q\left(\frac{d/2}{\sigma_N}\right)$



$$
\hat{s}_{MAP}(r) = \begin{cases} s^{(1)}, & r < -d, \\ s^{(2)}, & -d \le r < 0, \\ s^{(3)}, & 0 \le r < d, \\ s^{(4)}, & r \ge d, \end{cases}
$$

$P(\mathcal{E}) = 6 \times \frac{1}{4} Q\left(\frac{d/2}{\sigma_N}\right) = \frac{3}{2} Q\left(\frac{d}{2\sigma_N}\right)$

Average symbol energy: $E_s = \sum_{j=1}^{M} \underbrace{p_j}_{1/4} E_j = \frac{1}{\cancel{4}\,2}\left(\frac{9d^2}{4}+\frac{d^2}{4}+\cancel{\frac{d^2}{4}+\frac{9d^2}{4}}\right) = \frac{5}{4}d^2$

Average bit energy: $E_b = \frac{E_s}{\log_2 M} = \frac{5/4\, d^2}{2} = \frac{5}{8}d^2 \Rightarrow d = 2\sqrt{\frac{2}{5}E_b}$

$$P(\mathcal{E}) = \frac{3}{2}Q\left(\frac{1}{\cancel{2}\sigma_N}\cancel{2}\sqrt{\frac{2}{5}E_b}\right) = \frac{3}{2}Q\left(\sqrt{\frac{2}{5}\frac{E_b}{\sigma_N^2}}\right)$$

$$= \frac{3}{2}Q\left(\sqrt{\frac{4}{5}\frac{E_b}{N_0}}\right)$$

$\sigma_N^2 = \frac{N_0}{2}$

8.20. Correlation detector: Recall, from (50), that for additive noise channel where $R = S + N$, $S \perp\!\!\!\perp N$, and $N \sim \mathcal{N}(0, \sigma_N^2)$,

$$\hat{s}_{\mathrm{MAP}}(r) = \arg\max_{s \in S} \left(\underbrace{\sigma_N^2 \ln p_S(s) - \frac{E_s}{2}}_{\eta} + r \cdot s \right).$$

Equivalently,

$$\hat{w}_{\mathrm{MAP}}(r) = \arg\max_{j \in \{1,2,\ldots,M\}} \left(\sigma_N^2 \ln p_j - \frac{E_j}{2} + r \cdot s^{(j)} \right).$$

Now,

$$r \cdot s^{(j)} = \langle r(t), s_j(t) \rangle = \int_{-\infty}^{\infty} r(t) s_j(t) dt.$$

8.21. Matched filter implementation of the optimal detector: In practice, to calculate the correlation (inner-product) $\langle r(t), s(t) \rangle$ above, we use filtering. Recall that when a signal $r(t)$ passes through a filter whose impulse response is $h(t)$, the output of the filter is given by

$$\{r * h\}(t) = \int_{-\infty}^{\infty} r(\tau) h(t - \tau)\, d\tau$$

Let's try $h(t) = s^*(T - t)$ for some constant T. Then,

$$h(t - \tau) = s^*(T - (t - \tau)) = s^*(T - t + \tau).$$

Therefore,

$$\{r * h\}(t) = \int_{-\infty}^{\infty} r(\tau) s^*(T - t + \tau)\, d\tau.$$

In particular,

$$\{r * h\}(T) = \int_{-\infty}^{\infty} r(\tau) s^*(\tau)\, d\tau = \langle r(t), s(t) \rangle .$$

Conclusions: Implementation of optimal (MAP) detector can be done by matched filters.