Digital Transmission through Bandlimited AWGN Channels

Engr. Muhammad Abu Bakar Siddique

July 18, 2020

Abstract

اَلْحَمْدُ لِلّٰهِ رَبِّ الْعٰلَمِيْنَ وَالصَّلٰوةُ وَالسَّلَامُ عَلٰى خَاتَمِ النَّبِيِّيْنَ وَالْمُرْسَلِيْنَ وَعَلٰى آلِهِ الطَّيِّبِيْنَ وَأَصْحَابِهِ الْأَخْيَارِ أَجْمَعِيْنَ وَمَنْ تَبِعَهُمْ بِإِحْسَانٍ إِلٰى يَوْمِ الدِّيْنِ. أَمَّا بَعْدُ

8.1 Pulse Amplitude Modulation in AWGN

- No. of pulse levels in binary modulation is TWO
- No. of pulse levels in PAM is M,
- Binary modulation is special case of PAM with $M = 2$
- No. of bits per pulse is $k = \log_2 M$
- Each pulse has a waveform $s_m(t)$, where $m = \{1, 2, \ldots, M\}$
- All pulses are level shifted version of basic pulse $p(t)$

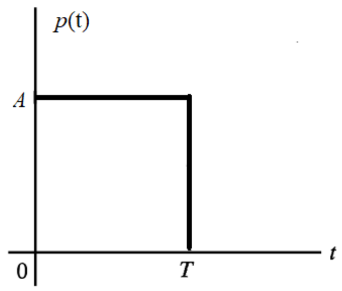

Figure 8.1: The basic pulse $p(t)$

$$p(t) = \begin{cases} A, & 0 \leq t < T \\ 0, & \text{otherwise} \end{cases} . \tag{8.1.1}$$

So the other pulses are level shifted version of $p(t)$

$$s_m(t) = A_m p(t), \quad m = 1, 2, \cdots, M \tag{8.1.2}$$

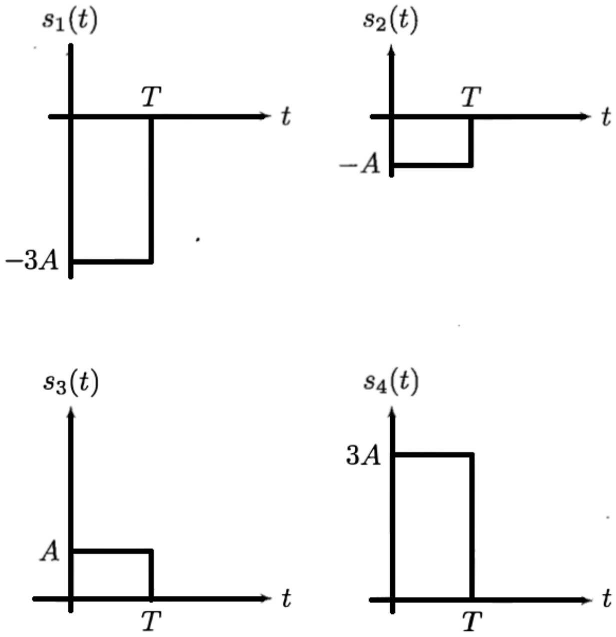

Figure 8.2: The 4-PAM pulses with amplitudes $-3A, -A, A$ and $3A$. Here M=4 is used.

The basic pulse has magnitude A.

The magnitudes of M pulses are

$$[-(M-1), \cdots, -3, -1, 1, 3, \cdots, (M-1)]A \tag{8.1.3}$$

In total there are M levels from $-(M-1)A$ to $(M-1)A$ in increments of 2A.

The most negative pulse $s_1(t)$ has amplitude multiplier $A_1 = -(M-1)$

The highest positive pulse $s_M(t)$ has amplitude multiplier $A_M = (M-1)$

So an arbitrary pulse $s_m(t)$ has amplitude multiplier A_m

$$A_m = (2m - M - 1), \quad m = 1, \cdots, M \tag{8.1.4}$$

To find orthonoraml basis we apply Gram-Schmidt algorithm. It is easier if we start with the pulse having amplitude A, that is the pulse $p(t)$ defined in Equation (8.1.1).

So the first orthonormal basis is

$$\psi(t) = \frac{p(t)}{\mathscr{E}_p} \tag{8.1.5}$$

Now we calculate energy of $p(t)$

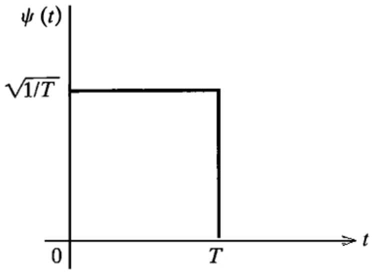

Figure 8.3: The basis functions obtaind from normalizing $p(t)$ by its energy.

$$
\begin{aligned}
\mathscr{E}_p &= \int_0^\infty (p(t))^2 dt \\
&= \int_0^T A^2 dt \\
&= A^2 |t|_0^T \\
&= A^2 T
\end{aligned}
\tag{8.1.6}
$$

So the basis function is

$$
\psi(t) = \frac{p(t)}{A\sqrt{T}} \tag{8.1.7}
$$

The good news is that it is the only basis function and all other signals lie parallel to this basis. It can be proved in few lines

There is just one orthonormal basis

Let us take an arbitrary waveform $s_m(t)$ and othonormalize it to $\psi(t)$. First calculate the correlation s_m between $s_m(t)$ and $\psi(t)$ as

$$
\begin{aligned}
s_m &= \int_{-\infty}^\infty s_m(t)\psi(t) dt \\
&= \int_{-\infty}^\infty (A_m p(t))(\frac{p(t)}{A\sqrt{T}}) \\
&= \frac{A_m}{A\sqrt{T}} \int_{-\infty}^\infty (p(t))^2 dt \\
&= \frac{A_m}{A\sqrt{T}} \int_0^T (A)^2 dt \\
&= \frac{A^2 A_m}{A\sqrt{T}} T \\
&= A A_m \sqrt{T}
\end{aligned}
\tag{8.1.8}
$$

The orthogonal part $d_m(t)$ of $s_m(t)$ is calculated by subtracting from $s_m(t)$ its projection along $\psi(t)$.

$$\begin{aligned}
d(t) &= s_m(t) - s_m\psi(t) \\
&= s_m(t) - AA_m\sqrt{T}\frac{p(t)}{A\sqrt{T}} \\
&= s_m(t) - A_m p(t) \\
&= s_m(t) - s_m(t) \\
&= 0
\end{aligned} \tag{8.1.9}$$

Therefore no orthogonal basis other than $\psi(t)$ exists.

Energy in each signal

Now that we have found the correlation s_m between $s_m(t)$ and $\psi(t)$, the energy of $s_m(t)$ pulse can be calculated as

$$s_m(t) = s_m\psi(t) \tag{8.1.10}$$

$$\begin{aligned}
\mathscr{E}_m &= \int_{-\infty}^{\infty}(s_m(t))^2 dt \\
&= \int_{-\infty}^{\infty}(s_m\psi(t))^2 dt \\
&= s_m^2\int_{-\infty}^{\infty}(\psi(t))^2 dt \\
&= s_m^2 \\
&= (AA_m\sqrt{T})^2 \\
&= A^2A_m^2T
\end{aligned} \tag{8.1.11}$$

The energy calculation integral of any basis function, such as $\int_{-\infty}^{\infty}(\psi(t))^2dt$, is always easy as its value is always ONE. We ourselves make it one as part of Gram-Schmidt process.

8.1.1 Average Energy

We assume that all M symbols arrive with equal probability. Therefore probability of each symbol is $1/M$ The average energy considering all waveforms can be calculated as follows:

$$
\begin{aligned}
\mathscr{E}_{avg} &= \frac{1}{M}(\mathscr{E}_1 + \cdots + \mathscr{E}_m + \cdots + \mathscr{E}_M) \\
&= \frac{1}{M}\sum_{m=1}^{M}\mathscr{E}_m \\
&= \frac{1}{M}\sum_{m=1}^{M}(AA_m\sqrt{T})^2
\end{aligned} \tag{8.1.12}
$$

Where we used the value of $\mathscr{E}_m$ from Equation (8.1.11)

Now using the value of A_m from Equation (8.1.4), we get

$$
\begin{aligned}
\mathscr{E}_{avg} &= \frac{1}{M}\sum_{m=1}^{M}(AA_m\sqrt{T})^2 \\
&= \frac{1}{M}\sum_{m=1}^{M}(AA_m\sqrt{T})^2 \\
&= \frac{A^2T}{M}\sum_{m=1}^{M}(A_m)^2 \\
&= \frac{A^2T}{M}\sum_{m=1}^{M}(2m - M - 1)^2 \\
&= \frac{A^2T}{M}\sum_{m=1}^{M}(2m - (M + 1))^2 \\
&= \frac{A^2T}{M}\sum_{m=1}^{M}(4m^2 + (M + 1)^2 - 4m(M + 1)) \\
&= \frac{A^2T}{M}\left(4\sum_{m=1}^{M}m^2 + (M + 1)^2\sum_{m=1}^{M}1 - 4(M + 1)\sum_{m=1}^{M}m\right)
\end{aligned} \tag{8.1.13}
$$

Simplify using the following summation formulae in above equation

$$\sum_{m=1}^{M}1 = M \tag{8.1.14}$$

$$\sum_{m=1}^{M}m = \frac{M(M + 1)}{2} \tag{8.1.15}$$

$$\sum_{m=1}^{M}m^2 = \frac{M(M + 1)(2M + 1)}{6} \tag{8.1.16}$$

$$\begin{aligned}\mathscr{E}_{avg} &= \frac{A^2T}{M}\left(\frac{4M(M+1)(2M+1)}{6} + (M+1)^2M - 4(M+1)\frac{M(M+1)}{2}\right)\\ &= \frac{A^2TM(M+1)}{6M}[4(2M+1)+6(M+1)-12(M+1)]\\ &= \frac{A^2T(M+1)}{6}[8M+6M-12M+4+6-12]\\ &= \frac{A^2T(M+1)}{6}[2M-2]\\ &= \frac{2A^2T(M+1)(M-1)}{6}\\ &= \frac{A^2T(M+1)(M-1)}{3}\\ &= \frac{A^2T(M^2-1)}{3}\end{aligned} \tag{8.1.17}$$

So $\frac{A^2T(M^2-1)}{3}$ is the energy carried by one pulse on average. As we noted earlier, with M different signals we can transmit $k = \log_2 M$ bits per signal. From average energy we calculate the bit energy as

$$\begin{aligned}\mathscr{E}_b &= \mathscr{E}_{avg}/k\\ &= \frac{A^2T(M^2-1)}{3k}\\ &= \frac{A^2T(M^2-1)}{3\log_2 M}\end{aligned} \tag{8.1.18}$$

8.1.2 Distance between two waveforms

The distance between two waveforms $s_1(t)$ and $s_2(t)$ is defined as

$$(\text{Distance})^2 = \int_{-\infty}^{\infty}(s_1(t)-s_2(t))^2dt \tag{8.1.19}$$

We define a variable d which is half the distance.

$$(2d)^2 = \int_{-\infty}^{\infty}(s_1(t)-s_2(t))^2dt \tag{8.1.20}$$

For simplicity we calculate the distance between two waveforms which are on either side of zero, i.e. $p(t)$ and $-p(t)$

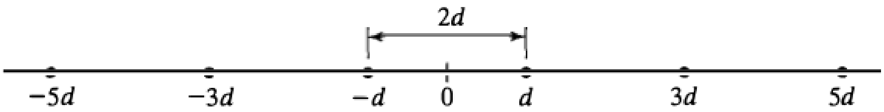

Figure 8.4: The pulse levels are shown by dots on basis axis. Distance between any two pulses is d.

$$\begin{aligned} 4d^2 &= \int_{-\infty}^{\infty} [p(t) - (-p(t))]^2 dt \\ &= \int_0^T (A - (-A))^2 dt \\ &= \int_0^T (2A)^2 dt \\ &= 4A^2 T \end{aligned} \tag{8.1.21}$$

From here we get the value of d and d squared as follows

$$d^2 = A^2 T \tag{8.1.22}$$

Or

$$d = A\sqrt{T} \tag{8.1.23}$$

Now we can write $\mathscr{E}_{avg}$ in Equation (8.1.17) in terms of d as

$$\mathscr{E}_{avg} = \frac{d^2(M^2-1)}{3} \tag{8.1.24}$$

Similarly, we can write $\mathscr{E}_b$ in Equation (8.1.18) terms of d as

$$\mathscr{E}_b = \frac{d^2(M^2-1)}{3\log_2 M} \tag{8.1.25}$$

We can write d^2 in terms of $\mathscr{E}_{avg}$ and $\mathscr{E}_b$

$$d^2 = \frac{3\mathscr{E}_{avg}}{(M^2-1)} \tag{8.1.26}$$

$$d^2 = \frac{3\mathscr{E}_b \log_2 M}{(M^2-1)} \tag{8.1.27}$$

8.2 Probability of Error in PAM

The PAM modulated signal is received by the correlation detector

The received signal is $r(t)$

This is the sum of symbol waveform and the channel Additive White Gaussian Noise (AWGN)

$$r(t) = s_m(t) + n(t) \tag{8.2.1}$$

Or if the waveform is represented in the form of its basis function $\psi(t)$

$$r(t) = s_m\psi(t) + n(t) \tag{8.2.2}$$

The AWGN noise is modeled by a Gaussian random variable.

A Gaussian random variable is described completely by two parameters, its mean, μ and its variance σ^2

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} \tag{8.2.3}$$

AWGN is the Gaussian R.V. with zero mean, $\mu_n = 0$ and variance $\sigma_x^2 = N_0/2$

$$\begin{aligned} f(n) &= \frac{1}{\sqrt{2\pi N_0/2}} e^{-\frac{(n-0)^2}{2N_0/2}} \\ &= \frac{1}{\sqrt{\pi N_0}} e^{-\frac{n^2}{N_0}} \end{aligned} \tag{8.2.4}$$

$N_0/2$ is the power spectral density of the noise measured in Watts/Hz

The AWGN noise power is uniformally distributed our complete frequency band, hence it got the name white. White light contains all the wavelengths (frequencies) of light.

The signal power can be computed in three ways

From its time description as

$$\mathscr{P}_x = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} |x(t)|^2 dt \tag{8.2.5}$$

From its frequency domain representation as

$$\mathscr{P}_x = \lim_{T\to\infty} \frac{1}{T} \int_{-\infty}^{\infty} |X(f)|^2 df \tag{8.2.6}$$

where $X(f)$ is the Fourier transform of $x(t)$

From its statistical properties as

$$\mathscr{P}_x = \mathbf{E}[X^2] = \int_{-\infty}^{\infty} x^2 f(x) dx \tag{8.2.7}$$

where x is the random variable and $f(x)$ is its probability density function.

Homework: Compute the signal power of a pulse train having peak-peak amplitude A and period T from its time domain, frequency domain and probabilistic representations for following cases

1) When the pulses amplitudes are 0 and A

2) When the pulses amplitude are $-A/2$ and $A/2$

Greater the noise power N_0, greater the spread of the noise values about zero.

Why the Gaussian noise?

There are two reasons

1) There is large portion of thermal noise in communication systems. Heated electronics components produce white noise, both at transmitter and receiver side. It is modelled as noise added by the channel. Thermal noise has Gaussian distribution.

2) Due to central moment theorem, if multiple random processes add up, their cumulative probability density function is Gaussian distribution.

The correlation type detector applies the following operations on received signal $r(t)$

$$r(t) = s_m\psi(t) + n(t) \tag{8.2.8}$$

This received signal is then multiplied with the basis function to get $x(t)$

$$\begin{aligned} x(t) &= (s_m\psi(t) + n(t))\psi(t) \\ &= s_m\psi^2(t) + n(t)\psi(t) \end{aligned} \tag{8.2.9}$$

Next this signal is passed through an integrator to get $y(t)$

$$\begin{aligned} y(t) &= \int_0^t x(t)dt \\ &= \int_0^t (s_m\psi(t) + n(t))\psi(t)dt \\ &= \int_0^t s_m\psi^2(t)dt + \int_0^t n(t)\psi(t)dt \end{aligned} \tag{8.2.10}$$

Next the signal $y(t)$ is sampled at end of the symbol time, T

$$\begin{aligned} y = y(T) &= s_m\int_0^T \psi^2(t)dt + \int_0^T n(t)\psi(t)dt \\ &= s_m + n \end{aligned} \tag{8.2.11}$$

where we used the face that energy of basis is one. n is the correlation of n(t) along the basis function. It is a sample of the AWGN Gaussian random variable and has mean $\mu_n = 0$ and $\sigma_n = N_0/2$

$$n = \int_0^T n(t)\psi(t)dt \tag{8.2.12}$$

Now y is a new random variable. s_m is a constant which is inner product of $s_m(t)$ with basis function $\psi(t)$.

$$\begin{aligned}\mu_y &= \mathbf{E}[y] \\ &= \mathbf{E}[s_m + n] \\ &= \mathbf{E}[s_m] + \mathbf{E}[n] \\ &= s_m + \mathbf{E}[n] \\ &= s_m + \mu_n\end{aligned} \tag{8.2.13}$$

where we use the fact that expected value of a constant value is that constant value.

Since $\mu_n = 0$

$$\mu_y = s_m \tag{8.2.14}$$

The variance of y is

$$\begin{aligned}\sigma_y^2 &= \mathbf{E}[(y-\mu_y)^2] \\ &= \mathbf{E}[(s_m + n - s_m)^2] \\ &= \mathbf{E}[n^2] \\ &= \mathbf{E}[n^2] - 0 \\ &= \mathbf{E}[n^2] - (\mathbf{E}[n])^2 \\ &= \sigma_n^2 \\ &= N_0/2\end{aligned} \tag{8.2.15}$$

where we used the definition of variance as $Var(y) = \mathbf{E}[(y-\mu_y)^2]$.

The sampled y is a Gaussian random variable haveing mean s_m and variance $\sigma_n^2 = N_0/2$. The probability density function of y is a bell curve centred at s_m.

$$f(y) = \frac{1}{\sqrt{\pi N_0}} e^{-\frac{(y-s_m)^2}{N_0}} \tag{8.2.16}$$

The function of the detector is to map received y to the symbol used at receiver s_m.

Though y has high probability to be near its mean value of s_m but there is non zero probability that its value is far off from s_m.

If the y is within d distance of the mean value, then it would be correctly detected as s_m. We calculated d earlier in Equation (8.1.23.)

If y is below its mean value then receiver will make a mistake and consider it s_{m-1}. Also if y is above its mean value more than d, then receiver will make a mistake and consider it s_{m+1}

For the smallest pulse $s_1(t)$ the error is half of others. Error is only made when y is greater than $s_1 + d$. If y is less than $s_1 - d$ no mistake is made as every value below s_1 is correctly detected as s_1.

Similiar is the case for the highest pulse as its error occurs only when y is less than $s_m - d$ and no error is made if $y > (s_m + d)$

Therefore the probability of error for the two pulses at extremes is half of the error for all other pulses

The error is calculated by integrating the received random variable in the undesired region.

The probability of error when s_1 is transmitted is given by

$$\begin{aligned} P_{e1} &= P(y > d) \\ &= \int_d^\infty f(y)dy \\ &= \int_d^\infty \frac{1}{\sqrt{\pi N_0}} e^{-\frac{(y-s_m)^2}{N_0}} dy \end{aligned} \tag{8.2.17}$$

So the error can be calculated by evaluating the integral in Equation (8.2.25)). Unfortunately, however, does not have any closed form. We have only tables available for standard normal random variable probability cumulative density functions (CDF). Som eof the popular CDF functions are given below.

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-t^2/2}\, dt$$

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{0}^{x} e^{-t^2}\, dt$$

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_{x}^{\infty} e^{-t^2}\, dt$$

Communication systems books usually use the error-function and complementary-error-function tables. For statistics the $\Phi(x)$ is usually used. The other popluar choice is $Q(x)$ function which is also the one used by our text book. So we make error calculations using Q-function.

The $Q(x)$ tables the probablity $P(X > x)$ of a zero mean $\mu_x = 0$ and unit variance $\sigma_x^2 = 1$ Gaussian random variable to have value greater than x.

$$\begin{aligned} Q(x) &= P(X > x) \\ &= \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt \end{aligned} \tag{8.2.18}$$

With small change if variables we can also see the following definition of $Q(x)$

$$
\begin{aligned}
Q(x) &= P(X < -x) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt
\end{aligned}
\tag{8.2.19}
$$

Few points to note for Q(x) are

$$
\begin{aligned}
Q(0) &= 0.5 \\
Q(-x) &= 1 - Q(x) \\
Q(\infty) &= 0 \\
Q(-\infty) &= 1
\end{aligned}
$$

If we have to evaluate $P(y > d)$ for a normal random variable with mean μ and variance σ^2

$$
\begin{aligned}
P(y > d) &= \int_{d}^{\infty} f(y) dy \\
&= \int_{d}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2} dy
\end{aligned}
$$

Let us make variable changes in above equation

$$
\begin{aligned}
\frac{x^2}{2} &= \frac{(y-\mu)^2}{2\sigma^2} \\
\implies x^2 &= \frac{(y-\mu)^2}{\sigma^2} \\
\implies x &= \frac{(y-\mu)}{\sigma}
\end{aligned}
\tag{8.2.20}
$$

Differentiating Equation (8.2.20)

$$
\begin{aligned}
dx &= \frac{1}{\sigma} dy \\
\implies dy &= \sigma dx
\end{aligned}
\tag{8.2.21}
$$

Also the limits will be changed

$$
\begin{aligned}
y = d &\implies x = \frac{d-\mu}{\sigma} \\
y = \infty &\implies x = \frac{\infty - \mu}{\sigma} = \infty
\end{aligned}
\tag{8.2.22}
$$

$$
\begin{aligned}
P(y > d) &= \int_{d}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2} dy \\
&= \int_{\frac{d-\mu}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{x^2/2} \sigma dx \\
&= \int_{\frac{d-\mu}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{x^2/2} dx \\
&= Q(\frac{d-\mu}{\sigma})
\end{aligned} \tag{8.2.23}
$$

Hence probability of error can be calculated for any Gaussian random variable using the $Q(x)$ function.

Now we are able to complete Equation (8.2.25) to calculate $P(y > d)$. We have earlier calculated the mean value of y, μ_y in Equation (8.2.13) and its variance σ_y in Equation (8.2.15). Substituting these values in Equation (8.2.23) we calculate the error probability in terms of Q(x).

$$
\begin{aligned}
P_{e1} &= P(y > d) \\
&= Q(\frac{d-0}{\sqrt{N_0/2}}) \\
&= Q(\sqrt{\frac{2d^2}{N_0}})
\end{aligned} \tag{8.2.24}
$$

Now using the value of d^2 calculated in Equation (8.1.22) in above equation we get

$$
\begin{aligned}
P_{e1} &= P(y > d) \\
&= Q\left(\frac{d-0}{\sqrt{N_0/2}}\right) \\
&= Q\left(\sqrt{\frac{2d^2}{N_0}}\right)
\end{aligned} \tag{8.2.25}
$$

P_{e1} is the error in detection when $s_1(t)$ was transmitted. By symmetry P_{eM} when $s_M(t)$ is transmitted is also P_{e1}. For each of the other $(M-2)$ pulses, the probability of error is twice that, i.e. $2P_{e1}$. If each symbol is equiprobable, probability of occurance of one symbol is $1/M$.

$$
P(s_1(t)) = P(s_2(t)) = \cdots = P(s_M(t)) = 1/M \tag{8.2.26}
$$

So the average probability of error per symbol is

$$\begin{aligned} P_e &= P(s_1)P_{e1} + P(s_2)(2P_{e1}) + \cdots + P(s_{M-1})(2P_{e1}) + P(s_M)P_{e1} \\ &= \frac{P_{e1} + 2P_{e1} + \cdots + 2P_{e1} + P_{e1})}{M} \\ &= \frac{2(M-1)}{M}P_{e1} \end{aligned} \tag{8.2.27}$$

Substituting the value of P_{e1} from Equation (8.2.25)

$$P_e = \frac{2(M-1)}{M}Q\left(\sqrt{\frac{2d^2}{N_0}}\right) \tag{8.2.28}$$

Using Equation (8.1.26) we obtain error probability as

$$P_e = \frac{2(M-1)}{M}Q\left(\sqrt{\frac{6\mathscr{E}_{avg}}{N_0(M^2-1)}}\right) \tag{8.2.29}$$

The symbol error probability in terms of bit energy $\mathscr{E}_b$ derived in Equation (8.1.27) is

$$P_e = \frac{2(M-1)}{M}Q\left(\sqrt{\frac{6\mathscr{E}_b \log_2 M}{N_0(M^2-1)}}\right) \tag{8.2.30}$$

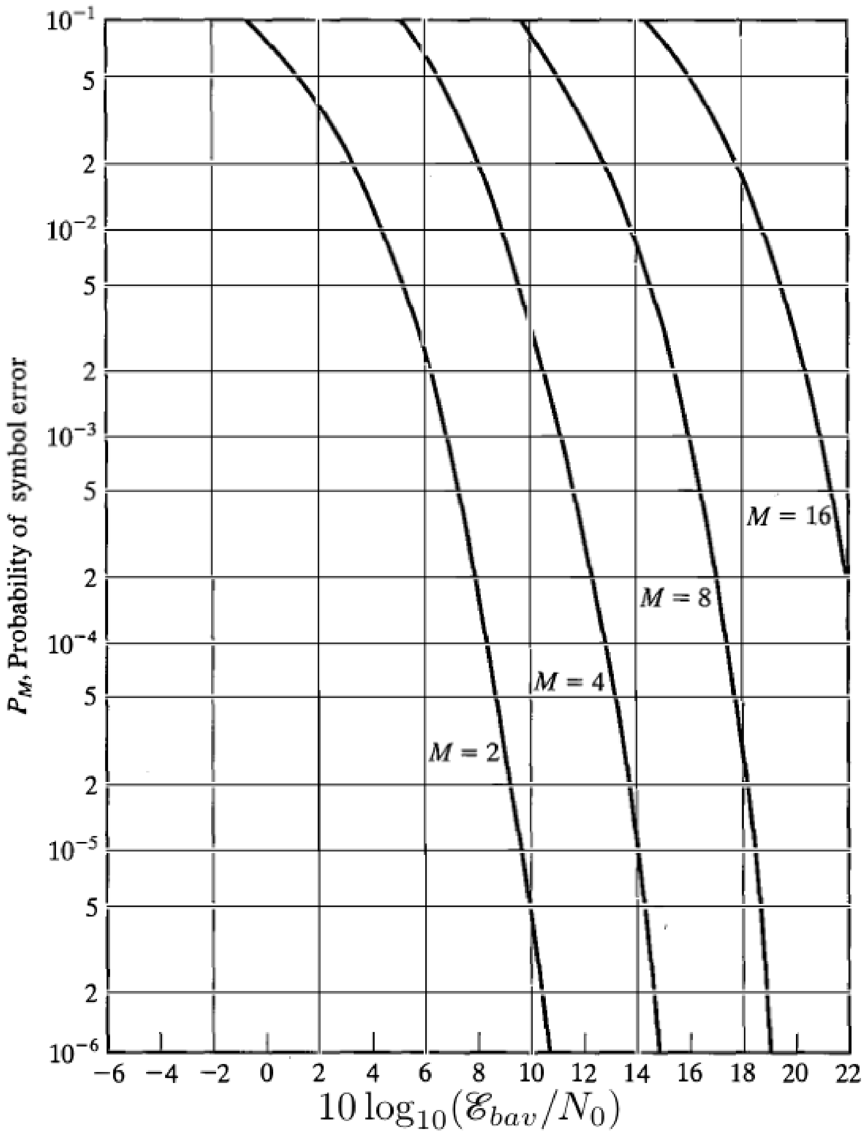

Figure 8.5: The probability of a symbol error as a function of $10\log_{10}(\mathscr{E}_{bav}/N_0)$ in dB for $M = 2, 4, 8, 16$.