3F4: Data Transmission

Handout 4: Detection of PAM in white Gaussian noise

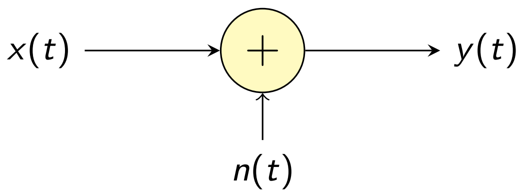
Ramji Venkataramanan

Signal Processing and Communications Lab, CUED

`ramji.v@eng.cam.ac.uk`

Lent Term 2019

Modelling the noise



$n(t)$ is a random signal modelled as a **Gaussian white noise process**:
For each t, $n(t)$ is Gaussian with zero mean and autocorrelation fucntion

$$\mathbb{E}[n(t)n(t+\tau)] = \frac{N_0}{2}\delta(\tau), \qquad \text{for all } t, \tau. \tag{1}$$

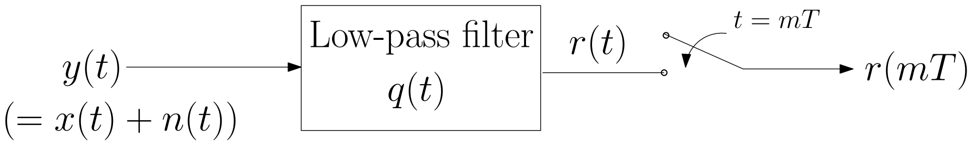
- This implies that the power spectral density (PSD) is:

$$S_n(f) = N_0/2, \qquad -\infty < f < \infty.$$

- This appears like an unrealistic definition, because the power (variance) $\mathbb{E}[n(t)^2] = \int_{-\infty}^{\infty} S_n(f)df$ is infinite.
- But does not pose a problem as the receive filter $q(t)$ is low-pass and rejects all frequency components outside a band, say $[-B, B]$.
- Effectively, what we are saying is $n(t)$ has PSD $S_n(f) = \frac{N_0}{2}$ for $f \in [-B, B]$, and we don't care what $S_n(f)$ is outside this band.
- We take it to be $\frac{N_0}{2}$ for all f, just for mathematical convenience.

Effect of noise on PAM demodulation

Recall that $x(t) = \sum_k X_k p(t - kT)$. The demodulator is :



Therefore,

$$r(t) = \underbrace{x(t) \star q(t)}_{r_s(t)} + \underbrace{n(t) \star q(t)}_{r_n(t)}$$

If we have chosen $q(t)$ so that the overall filter satisfies Nyquist pulse criterion, then recall from Handout 2:

$$r_s(mT) = X_m.$$

Therefore the demodulator output sampled at mT, denoted by Y_m, is

$$\underbrace{r(mT)}_{Y_m} = r_s(mT) + r_n(mT) = X_m + \underbrace{r_n(mT)}_{N_m}. \tag{2}$$

We have

$$N_m = \int_{-\infty}^{\infty} q(u)n(mT - u)du = \int_{-\infty}^{\infty} n(u)q(mT - u)du \tag{3}$$

We now compute the joint distribution of the random variables $\{N_m\}_{m \in \mathbb{Z}}$:

- We assume that the receive filter is the matched filter, i.e., $q(t) = p(-t)$ and the overall filter $g(t) = p(t) \star p(-t)$ satisfies the Nyquist pulse criterion

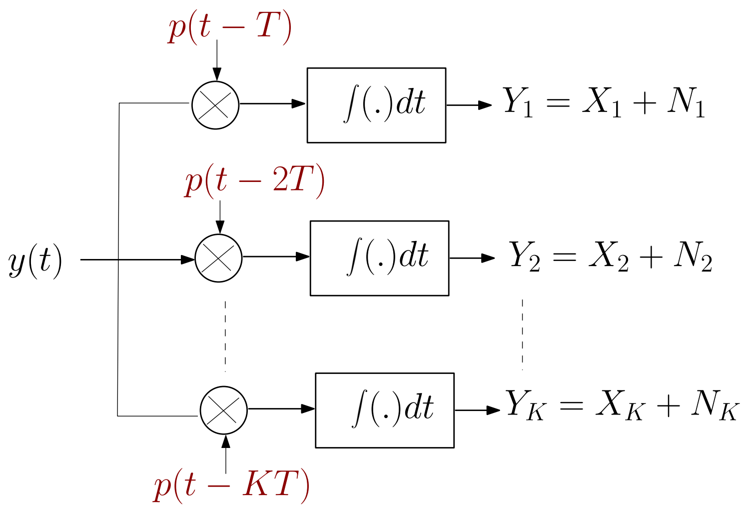
- Recall from the end of Handout 2 that the functions $\{p(t - nT)\}_{n \in \mathbb{Z}}$ are orthonormal, i.e.,

$$\int_{-\infty}^{\infty} p(u)p(u - nT)du = \begin{cases} 1, & n = 0 \\ 0, & n = \pm 1, \pm 2, \ldots \end{cases}$$

Signal space interpretation of PAM detection

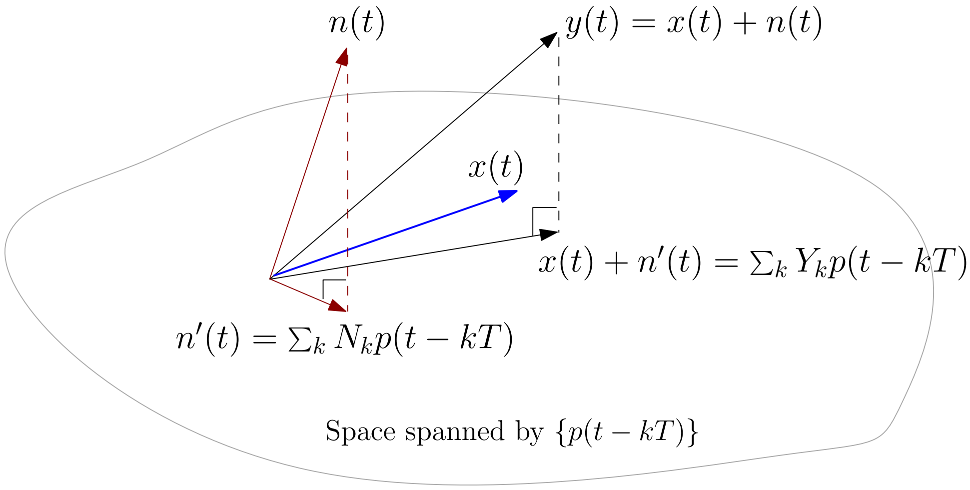
- The PAM signal $x(t) = \sum_k X_k p(t - kT)$ lies in the space spanned by the orthonormal set $\{p(t - kT)\}_{k \in \mathbb{Z}}$.
- X_k is the projection coefficient of $x(t)$ along the basis function $p(t - kT)$.
- Also, $N_k = \int_{-\infty}^{\infty} n(u) q(kT - u) du = \int_{-\infty}^{\infty} n(u) p(u - kT) du$.
- Thus, N_k is the projection coefficient of the noise waveform $n(t)$ along $p(t - kT)$.

Thus the matched filter receiver (p.3) is equivalent to the following receiver, which computes inner products with each of the basis functions:



Transmitted PAM signal: $x(t) = \sum_k X_k p(t - kT)$

Geometric interpretation of demodulator:



- The signal lies in the space spanned by the orthonormal basis $\{p(t - kT)\}_{k \in \mathbb{Z}}$
- The demodulator is only affected by $n'(t)$, the component of the $n(t)$ that lies in the signal space.
- The receiver projects $y(t) = x(t) + n(t)$ onto the signal space. The component of $y(t)$ in this space is $x(t) + n'(t)$
- For any $m \in \mathbb{Z}$, we can extract Y_m from $x(t) + n'(t)$ by taking inner product with $p(t - mT)$. (This is what the demodulator does.)

Projection coefficients of noise

We will use the following general result about the coefficients of projection of white noise along *any* orthonormal set.

Distribution of projection coefficients of white noise

Let $\{\phi_m(t)\}_{m\in\mathbb{Z}}$ be any orthonormal set of functions, and $n(t)$ be a white noise process with autocovariance defined in Eq. (1). For $m \in \mathbb{Z}$, let

$$N_m = \int_{-\infty}^{\infty} n(t)\phi_m(t)dt$$

Then $\{N_m\}_{m\in\mathbb{Z}}$ are i.i.d. Gaussian with zero mean and variance $\frac{N_0}{2}$

Proof: For each m, N_m is a linear combination of *jointly* Gaussian rvs $\{n(t),\ t\in\mathbb{R}\}$. Hence the rvs $\{N_m\}_{m\in\mathbb{Z}}$ are jointly Gaussian.

Mean: For each m:

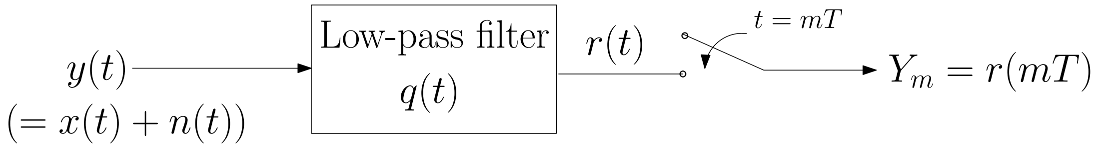
$$\mathbb{E}[N_m] = \mathbb{E}\Big[\int_{-\infty}^{\infty} n(t)\phi_m(t)dt\Big] = \int_{-\infty}^{\infty} \mathbb{E}[n(t)]\phi_m(t)dt = 0$$

Covariance: For each pair of integers m, ℓ:

$$\begin{aligned}
\mathbb{E}[N_m N_\ell] &= \mathbb{E}\Big[\int_{-\infty}^{\infty} n(t)\phi_m(t)dt \int_{-\infty}^{\infty} n(s)\phi_\ell(s)dt\Big] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{E}[n(t)n(s)]\ \phi_m(t)\phi_\ell(s)dt\ ds \\
&\overset{(a)}{=} \int_{-\infty}^{\infty}\Big[\int_{-\infty}^{\infty} \frac{N_0}{2}\delta(t-s)\ \phi_m(t)dt\Big]\ \phi_\ell(s)ds \\
&\overset{(b)}{=} \frac{N_0}{2}\int_{-\infty}^{\infty} \phi_m(s)\phi_\ell(s)ds \overset{(c)}{=} \begin{cases} \frac{N_0}{2}, & m=\ell \\ 0, & m\neq\ell \end{cases}
\end{aligned}$$

Here (a) follows from the autocovariance function of the white noise process, Eq. (1). Step (b) follows from the sifting property of the δ-function, and (c) follows from the orthonormality of $\{\phi_n\}_{n\in\mathbb{Z}}$. □

In the PAM setting, we apply the result with $\{\phi_m\} = \{p(t-mT)\}$ to see that the random variables $\{N_m\}_{m\in\mathbb{Z}}$ in (3) are i.i.d. $\sim \mathcal{N}(0, \frac{N_0}{2})$.



We have shown that the demodulator output sampled at times $\{mT\}$ is:

$$Y_m = X_m + N_m, \qquad m = 0, 1 \ldots$$

where the noise random-variables $\{N_m\}$ are i.i.d. Gaussian $\sim \mathcal{N}(0, \frac{N_0}{2})$.

The next step is *detection*: to determine the transmitted constellation symbol X_m from the noisy observation Y_m. But first note that:

- We have converted the continuous-time problem detection problem with $y(t) = x(t) + n(t)$ into a discrete-time one with $Y_m = X_m + N_m$ for integers m.
- The key reason this is possible is that the signal lies in the space of the orthonormal set of functions $\{p(t - mT)\}_{m \in \mathbb{Z}}$.
- Therefore, we can work with $\{Y_m\}$, the coefficients of projection of $y(t)$ along these orthonormal functions.
- Only the component of the noise in the direction of these orthonormal functions affects the detection of the symbols $\{X_m\}$.
- The symbol time T is determined by the bandwidth allocated for transmission.

Optimal Detection

$$Y_m = X_m + N_m, \qquad m = 0, 1, \ldots$$

How to optimally detect the transmitted symbol X_m from the demodulator output Y_m?

Optimality will be defined in terms of *probability of detection error*, i.e., if the detected symbol is $\hat{X}_m$, then we want to minimize $P(\hat{X}_m \neq X_m)$.

Given $Y = y$, the optimal detection rule that minimizes the probability of detection error is the **'Maximum a posteriori probability rule' (MAP)** rule:

$$\begin{aligned} \hat{X} &= \arg\max_{c \in \mathcal{C}} P(X = c \mid Y = y) \\ &= \arg\max_{c \in \mathcal{C}} P(X = c)\, f(y \mid X = c), \end{aligned} \tag{4}$$

where $\mathcal{C}$ is the constellation (set of PAM symbols) and $f(y \mid X = c)$ is the conditional density of Y given $X = c$.

Proof of optimality of MAP: Let $X \in \mathcal{C}$ denote the true transmitted symbol, $\hat{X}$ be the MAP decoded symbol, and let $\tilde{X} = g(Y)$ be the decoded symbol using any other decoding rule g.
The probability of *correct detection* for $\tilde{X} = g(Y)$ is:

$$\begin{aligned} P(X = \tilde{X}) &= \int_{y \in \mathbb{R}} P(X = \tilde{X}(y) \mid Y = y)\, f(y) dy \\ &\leq \int_{y \in \mathbb{R}} \max_{c \in \mathcal{C}} P(X = c \mid Y = y)\, f(y) dy \\ &\stackrel{(a)}{=} \int_{y \in \mathbb{R}} P(X = \hat{X} \mid Y = y)\, f(y) dy \\ &= P(X = \hat{X}) \end{aligned}$$

In the above, $f(y) = \sum_{c \in \mathcal{C}} P(X = c) f(Y = y | X = c)$, and the equality (a) follows from the definition of the MAP rule.
Therefore, the MAP rule $\hat{X}$ maximises the probability of correct detection, or equivalently, minimizes the probability of detection error.
Finally, to obtain the second representation of the MAP rule (4), we write

$$P(X = c \mid Y = y) = \frac{P(X = c) f(Y = y \mid X = c)}{f(y)},$$

and note that the denominator $f(y)$ is the same for all $c \in \mathcal{C}$. □

If all the symbols in the constellation are equally likely, i.e., $P(X = c)$ is the same for all symbols $c \in \mathcal{C}$), then the MAP rule in (4) becomes

$$\hat{X} = \arg\max_{c \in \mathcal{C}} f(Y = y \mid X = c).$$

This is called the **maximum-likelihood** (ML) decoding rule.

Since $Y = X + N$, with $N \sim \mathcal{N}(0, \frac{N_0}{2})$,

$$\begin{aligned} f(Y = y \mid X = c) &= f(X + N = y \mid X = c) \\ &= f(c + N = y \mid X = c) \\ &\stackrel{(a)}{=} f(N = y - c) = \frac{1}{\sqrt{\pi N_0}} e^{-(y-c)^2/N_0}, \end{aligned}$$

where (a) holds because noise N is independent of transmitted symbol X. Therefore the ML rule is

$$\hat{X} = \arg\max_{c \in \mathcal{C}} e^{-(y-c)^2/N_0} = \arg\min_{c \in \mathcal{C}} (y - c)^2$$

If all the constellation symbols are equally likely, the optimum detector simply chooses the symbol *closest* to the output.
(Also called "nearest-neighbour" or "minimum-distance" decoding)

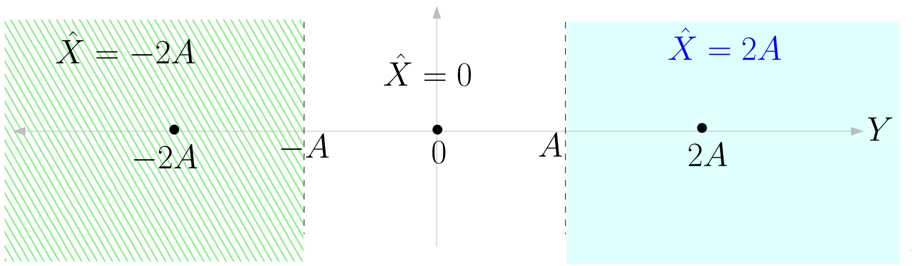
MAP Decoding Example: 3-ary PAM

We will now analyse PAM performance a three-symbol constellation $\{-2A, 0, +2A\}$, under the assumption that the symbols are equally likely.

You have already seen a similar analysis in 1B Comms. In the examples paper, you will explore MAP decoding for PAM symbols with different probabilities.

- The observed symbol is $Y = X + N$, where $N \sim \mathcal{N}(0, N_0/2)$.
- With equally likely symbols, the optimal (ML) decoding rule is

$$\hat{X} = \begin{cases} -2A & \text{if } Y < -A \\ 0 & \text{if } -A \leq Y < A \\ 2A & \text{if } Y > A \end{cases}$$



Q: How would the decision boundaries change if 0 had higher probability than the other two points?

Probability of decoding error

An error occurs when:

1. $X = -2A$ and $Y \geq -A$, or
2. $X = 0$ and $|Y| > A$, or
3. $X = 2A$ and $Y \leq A$

The probability of detection error is

$$\begin{aligned} P_e &= P(\hat{X} \neq X) \\ &= P(X = -2A)P(Y \geq -A | X = -2A) + P(X = 0)P(|Y| > A | X = 0) \\ &\quad + P(X = 2A)P(Y \leq A | X = 2A) \\ &= \frac{1}{3}\left[P(Y \geq -2A | X = -2A) + P(|Y| > A | X = 0) + P(Y \leq A | X = 2A)\right] \end{aligned} \tag{5}$$

(The last inequality uses the fact that the symbols are equally likely.)

Let us compute each of the probabilities on the RHS.

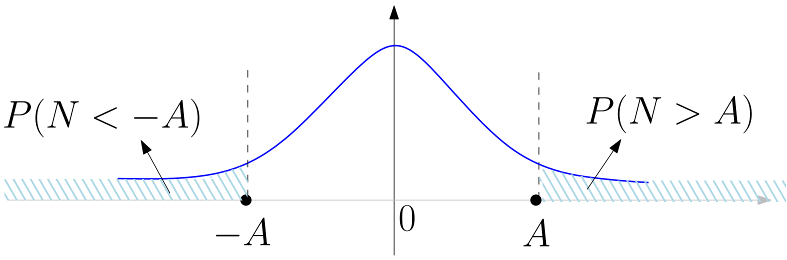
$$P(Y \geq -A | X = -2A) = P(-2A + N > -A \mid X = -2A)$$
$$= P(N \geq A \mid X = -2A) \overset{(a)}{=} P(N \geq A) \qquad (6)$$

(a) is true because the noise random variable N is **independent** of the transmitted symbol X. Similarly,

$$P(Y \leq A | X = 2A) = P(2A + N \leq A \mid X = 2A)$$
$$= P(N \leq -A \mid X = -2A) = P(N \leq -A), \qquad (7)$$

and

$$P(|Y| > A | X = 0) = P(|0 + N| > A \mid X = 0)$$
$$P(|N| > A) = P(N < -A) + P(N > A), \qquad (8)$$



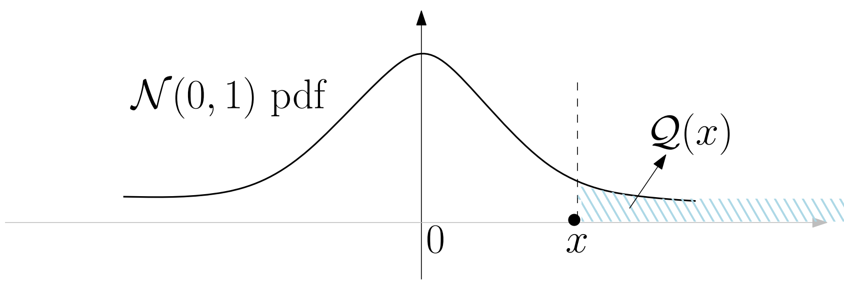
As N is Gaussian $\sim \mathcal{N}(0, N_0/2)$, the symmetry of the pdf implies $P(N < -A) = P(N > A)$.

Therefore, the probability of error in (5) becomes ...

$$P_e = \frac{1}{3}\left[4P(N \geq A)\right]$$
$$= \frac{4}{3} P\left(\frac{N}{\sqrt{N_0/2}} \geq \frac{A}{\sqrt{N_0/2}}\right) = \frac{4}{3}\mathcal{Q}\left(\frac{A}{\sqrt{N_0/2}}\right)$$

We have normalised N by its standard deviation to express the probability in terms of the $\mathcal{Q}$ function:

$$\mathcal{Q}(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$



- $\mathcal{Q}(x)$ is the probability that a **standard Gaussian** $\mathcal{N}(0,1)$ random variable takes value greater than x
- Also note that $\mathcal{Q}(x) = 1 - \Phi(x)$, where $\Phi(.)$ is the cdf of a $\mathcal{N}(0,1)$ random variable

P_e in terms of signal-to-noise ratio $\frac{E_b}{N_0}$

The *average energy per symbol* E_s of the constellation is

$$E_s = \frac{1}{3}[(2A)^2 + 0^2 + (-2A)^2] = \frac{8}{3}A^2$$

- For an M-point constellation with equally likely symbols, the average energy per symbol $E_s = E_b \log_2 M$
- Here $M = 3$, hence the *average energy per bit* can be computed using

$$E_s = \frac{8}{3}A^2 = E_b \log_2 3.$$

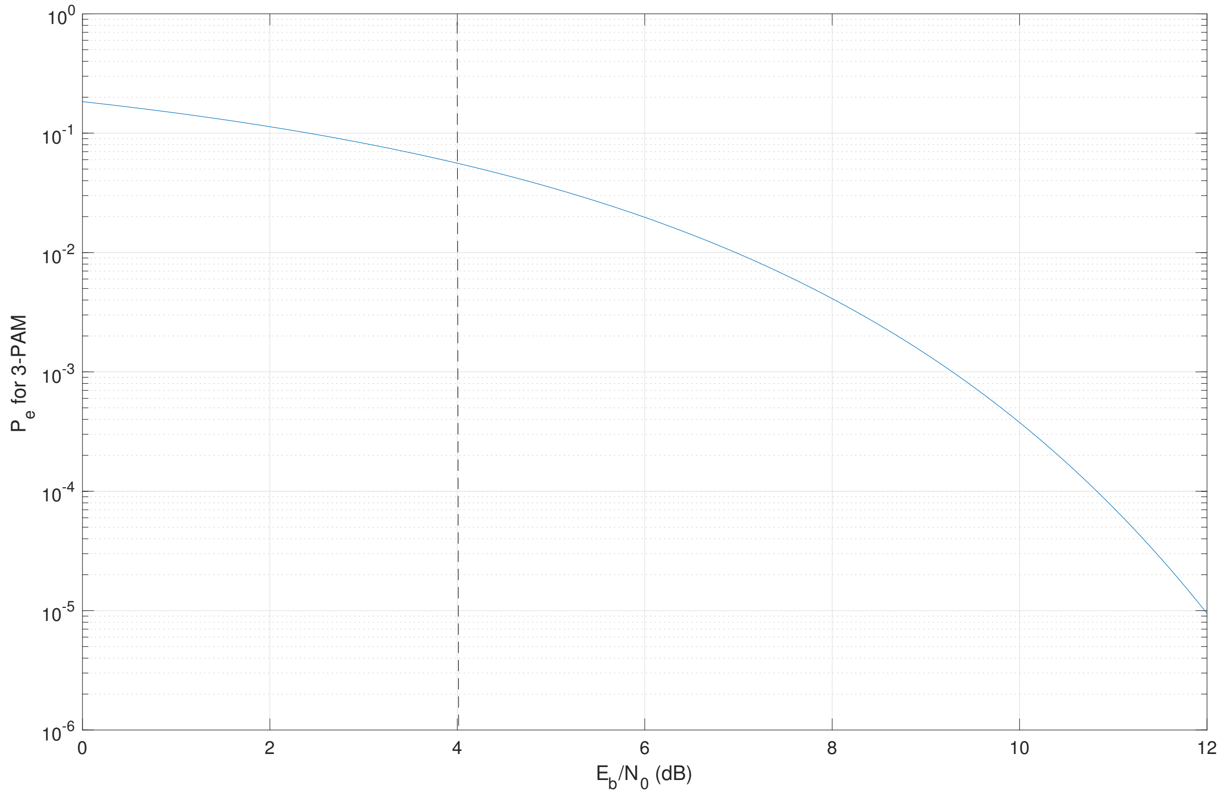
We can therefore write

$$P_e \leq \frac{4}{3}\mathcal{Q}\left(\frac{A}{\sqrt{N_0/2}}\right) = \frac{4}{3}\mathcal{Q}\left(\sqrt{\frac{3\log_2 3}{4}\,\frac{E_b}{N_0}}\right)$$

$\frac{E_b}{N_0}$ is a key signal-to-noise parameter of a transmission scheme.

P_e is often plotted as a function of $\frac{E_b}{N_0}$.

P_e vs $\frac{E_b}{N_0}$ for 3-ary PAM



E.g., if we want to guarantee $P_e \leq 10^{-4}$, then need $\frac{E_b}{N_0}$ at least 10.4 dB

To get a sense of how P_e decays with E_b/N_0, we can use the bound

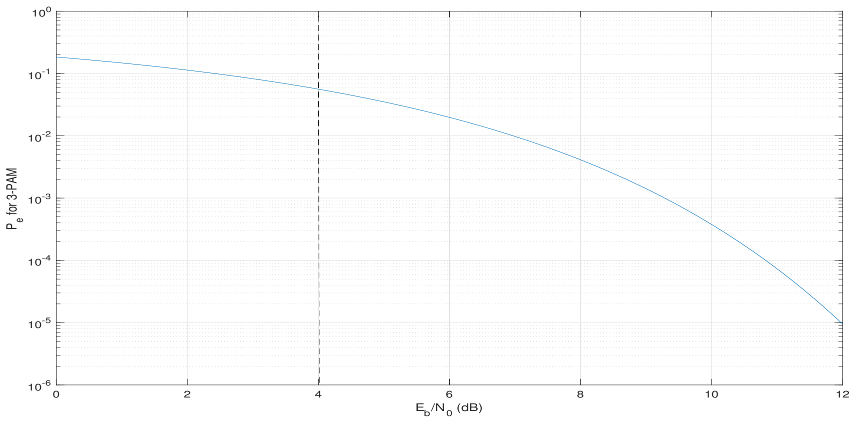
$$\mathcal{Q}(x) \leq \frac{1}{2} e^{-x^2/2}, \qquad \text{for } x \geq 0.$$

This bound is a pretty good approximation for large x.

Using this approximation, for 3-PAM we obtain:

$$P_e \leq \frac{4}{3} \mathcal{Q}\left(\sqrt{\frac{3 \log_2 3}{4} \frac{E_b}{N_0}}\right) \leq \frac{2}{3} e^{-\frac{3 \log_2 3}{8} \frac{E_b}{N_0}}$$

- As we increase $\frac{E_b}{N_0}$, the error probability drops **exponentially** in $\frac{E_b}{N_0}$.
- We will later see how *coding* can give a steeper decay of probability of error with E_b/N_0



Aside (not examinable):
The vertical dashed line is the 'Shannon limit' for rate $\log_2 3$ bits/transmission over a additive white Gaussian noise channel:

- This is the minimum snr required by *any* coding scheme with the same rate of transmission as 3-PAM ($\log_2 3$ bits/ transmission).
- If you have taken 3F7, you have seen that to achieve low probability of error at rates close to capacity, one can use random coding, joint typicality decoding etc.
- Clearly 3-PAM is not close to capacity-achieving. E.g., snr required to achieve $P_e \leq 10^{-4}$ is more than 5 dB from Shannon limit.
- Adding a good LDPC outer code can give steep drop in error prob. curve $\Rightarrow$ reliable decoding at snrs closer to Shannon limit

Summary

- Noise waveform $n(t)$ modelled as white Gaussian noise with constant PSD $N_0/2$
- Matched filter demodulator projects received signal $y(t) = x(t) + n(t)$ along the basis functions $\{p(t - kT)\}_{k \in \mathbb{Z}}$
- Next step is *detection* of $\{X_k\}$ from $\{Y_k = X_k + N_k\}$, where $\{N_k\}$ are iid $\sim \mathcal{N}(0, \frac{N_0}{2})$
- Optimal detection rule is the MAP rule: find the most likely symbol given the observation. For a constellation $\mathcal{C}$, MAP rule is

$$\hat{X} = \arg\max_{c \in \mathcal{C}} P(X = c)\, f(Y = y \mid X = c).$$

- If all constellation symbols are equally likely, then MAP rule reduces to maximising $f(Y = y \mid X = c)$, i.e., maximum likelihood
- Probability of detection error P_e can be bounded using $\mathcal{Q}$-functions, and can be expressed in terms of snr E_b/N_0

You can now do all the questions in Examples Paper 1.