1B Paper 6: Communications
Handout 4: Digital Baseband Modulation

Ramji Venkataramanan

Signal Processing and Communications Lab
Department of Engineering
ramji.v@eng.cam.ac.uk

Lent Term 2015
Data Transmission

We have seen how analogue sources can be digitised. E.g., An MPEG or QuickTime file is a stream of bits

Now we have to transport those bits across a channel:

(Digitised source)

110 001 100 111

\[ \leftrightarrow \]

\[ \ldots 10110010001101010 \ldots \]

110 000 100 111

Transmitter

Channel

Receiver

input waveform

output waveform
The transmitter (Tx) does two things:

1. **Encoding**: Adding redundancy to the source bits to protect against noise

2. **Modulation**: Transforming the coded bits into waveforms.

The receiver (Rx) does:

- **Demodulation**: noisy output waveform $\rightarrow$ output bits
- **Decoding**: Try to correct errors in the output bits and recover the source bits
Modulation/Demodulation

We’ll first consider the modulation and demodulation blocks assuming that the channel encoder/decoder are fixed, and look at the design of the channel encoder and decoder later.

(Encoded bits)
10110100

Modulator

\[ x(t) \]

Demodulator

\[ y(t) \]

\[ n(t) \]

10100110

We now study a digital baseband modulation technique called *Pulse Amplitude Modulation* (PAM) & analyse its performance over an Additive White Gaussian Noise (AWGN) channel.
The Symbol Constellation

The digital modulation scheme has two basic components.

1. The first is a mapping from bits to real/complex numbers, e.g.

   $0 \rightarrow -A, \quad 1 \rightarrow A$  (binary symbols)

   $00 \rightarrow -3A, \quad 01 \rightarrow -A, \quad 10 \rightarrow A, \quad 11 \rightarrow 3A$  (4-ary symbols)

The set of values the bits are mapped to is called the *constellation*, e.g., the 4-ary constellation above is $\{-3A, A, -A, 3A\}$.

Once we fix a constellation, a sequence of bits can be uniquely mapped to constellation symbols. E.g., with constellation $\{-A, A\}$

$0 1 0 1 1 1 0 0 1 0 \rightarrow -A, \ A, -A, \ A, \ A, -A, -A, -A, \ A$

With constellation $\{-3A, -A, A, 3A\}$, the same sequence of bits is mapped as

$01 01 11 00 10 \rightarrow -A, \ -A, \ 3A, \ -3A, \ A$

In a constellation with $M$ symbols, each symbol represents $\log_2 M$ bits.
The Pulse Shape

The second component of Pulse Amplitude Modulation is a unit-energy baseband waveform denoted \( p(t) \), called the pulse shape. E.g., a sinc pulse or a rect pulse:

\[
p(t) = \frac{1}{\sqrt{T}} \text{sinc} \left( \frac{\pi t}{T} \right) \quad \text{or} \quad p(t) = \begin{cases} 
\frac{1}{\sqrt{T}} & \text{for } t \in (0, T] \\
0 & \text{otherwise}
\end{cases}
\]

\( T \) is called the symbol time of the pulse.

A sequence of constellation symbols \( X_0, X_1, X_2, \ldots \) is used to generate a baseband signal as follows:

\[
x_b(t) = \sum_k X_k p(t - kT)
\]

Thus we have the following important steps to associate bits with a baseband signal \( x_b(t) \):

\[
\ldots 0 \ 1 \ 0 \ 1 \ 1 \ 1 \ 0 \ 0 \ 1 \ 0 \ldots \longrightarrow X_0, X_1, X_2, \ldots \longrightarrow \sum_k X_k p(t - kT)
\]
Rate of Transmission

The modulated baseband signal is $x_b(t) = \sum_k X_k p(t - kT)$. With the rect pulse shape

$$p(t) = \begin{cases} \frac{1}{\sqrt{T}} & \text{for } t \in (0, T] \\ 0 & \text{otherwise} \end{cases}$$

and $X_k \in \{+A, -A\}$, $x_b(t)$ looks like

Every $T$ seconds, a new symbol is introduced by shifting the pulse and modulating its amplitude with the symbol.

The transmission rate is $\frac{1}{T}$ symbols/sec or $\frac{\log_2 M}{T}$ bits/second.
Desirable Properties of the Pulse Shape $p(t)$

$p(t)$ is chosen to satisfy the following important objectives:

1. We want $p(t)$ to decay quickly in time, i.e., the effect of symbol $X_k$ should not start much before $t = kT$ or last much beyond $t = (k + 1)T$.

2. We want $p(t)$ to be approximately band-limited. For a fixed sequence of symbols $\{X_k\}$, the spectrum of $x_b(t)$ is

\[
X_b(f) = \mathcal{F} \left[ \sum_k X_k p(t - kT) \right] = P(f) \sum_k X_k e^{-j2\pi fkT} \]

Hence the bandwidth of $x_b(t)$ is the same as that of the pulse $p(t)$.

3. The retrieval of the information sequence from the noisy received waveform $x_b(t) + n(t)$ should be simple and relatively reliable. In the absence of noise, the symbols $\{X_k\}_{k \in \mathbb{Z}}$ should be recovered perfectly at the receiver.
Orthonormality of pulse shifts

Consider the third objective, namely, simple and reliable detection. To achieve this, the pulse is chosen to have the following “orthonormal shifts” property:

\[
\int_{-\infty}^{\infty} p(t - kT) p(t - mT) \, dt = \begin{cases} 
1 & \text{if } k = m \\
0 & \text{if } k \neq m 
\end{cases}
\]  \hspace{1cm} (1)

We’ll see how this property makes signal detection at the Rx simple.

- This property is satisfied by the rect pulse shape

\[
p(t) = \begin{cases} 
\frac{1}{\sqrt{T}} & \text{for } t \in (0, T] \\
0 & \text{otherwise} 
\end{cases}
\]

- The sinc pulse \( p(t) = \frac{1}{\sqrt{T}} \text{sinc} \left( \frac{\pi t}{T} \right) \) also has orthonormal shifts! (You will show this in Examples Paper 9, Q.2)
Time Decay vs. Bandwidth Trade-off

The first two objectives say that we want $p(t)$ to:

1. Decay quickly in time
2. Be approximately band-limited

But . . . faster decay in time $\iff$ larger bandwidth

Consider the pulse $p(t) = \frac{1}{\sqrt{T}} \text{sinc} \left( \frac{\pi t}{T} \right)$

The sinc is perfectly band-limited to $W = \frac{1}{2T}$

But decays slowly in time $|p(t)| \sim \frac{1}{|t|}$
Next consider the rectangular pulse

\[
p(t) = \begin{cases} \frac{1}{\sqrt{T}} & \text{for } t \in (0, T] \\ 0 & \text{otherwise} \end{cases}
\]

This pulse is perfectly time-limited to the symbol interval \([0, T)\).

But . . .

- Decays slowly in freq. \(|P(f)| \sim \frac{1}{|f|}
- \text{Main-lobe bandwidth } = \frac{1}{T}
In practice, the pulse shape is often chosen to have a *raised cosine* spectrum

\[
P(f) = \begin{cases} 
1 & \text{for } |f| < 1 \frac{T}{2} \\
0 & \text{for } |f| > 1 \frac{T}{2} 
\end{cases}
\]

Bandwidth slightly larger than \( \frac{1}{2T} \); decay in time \(|p(t)| \sim \frac{1}{|t|^3}\)

A happy compromise!

- More on raised cosine pulses in 3F4
- For intuition, it often helps to envision \(x_b(t)\) with a rect pulse, though it is never used in practice
PAM Demodulation

Now, assume that we have picked a constellation and a pulse shape satisfying the objectives, and we transmit the baseband waveform

\[ x_b(t) = \sum_k X_k p(t - kT) \]

over a baseband channel \( y(t) = x_b(t) + n(t) \)

How does the receiver recover the information symbols \( \{X_0, X_1, X_2, \ldots\} \) from \( y(t) \)?

- This process is called *demodulation*

- We will see that the orthonormal shift property of \( p(t) \) leads to a simple and elegant demodulator
Matched Filter Demodulator

Let us first understand the operation assuming no noise, i.e.,

\[ y(t) = x_b(t) = \sum_k X_k p(t - kT) \]

\[ y(t) \rightarrow \text{Filter} \]
\[ h(t) = p(-t) \]
\[ y(t) \rightarrow r(t) \]
\[ t = mT \]
\[ r(mT) \]

\[ y(t) \] is passed through a filter with impulse response \( h(t) = p(-t) \)

This is called a **matched filter**. The filter output is

\[ r(t) = y(t) \ast h(t) = x_b(t) \ast h(t) \quad (\text{assuming no noise}) \]

\[ = \int_{-\infty}^{\infty} x_b(\tau) h(t - \tau) d\tau \]

\[ = \int_{-\infty}^{\infty} x_b(\tau) p(\tau - t) d\tau \]
The Output of the Matched Filter

\[ y(t) \xrightarrow{\text{Filter}} h(t) = p(-t) \xrightarrow{\text{Filter}} y(t) \]

\[ r(t) = \int_{-\infty}^{\infty} x_b(\tau)p(\tau - t)\,d\tau = \sum_k X_k \int_{-\infty}^{\infty} p(\tau - kT)p(\tau - t)\,d\tau \]

By sampling the filter output at time \( t = mT \), \( m = 0, 1, 2, \ldots \), you get

\[ r(mT) = \sum_k X_k \int_{-\infty}^{\infty} p(\tau - kT)p(\tau - mT)\,d\tau = X_m \]

because of the orthonormal shifts property of \( p(t) \)

\[ \int_{-\infty}^{\infty} p(\tau - kT)p(\tau - mT)\,d\tau = \begin{cases} 1 & \text{if } k = m \\ 0 & \text{if } k \neq m \end{cases} \]

The orthonormal shifts property is crucial for this demodulator to work!
Demodulation with Noisy $y(t)$

Now consider the noisy case. The receiver gets $y(t) = x(t) + n(t)$

\[ y(t) \xrightarrow{\text{Filter}} r(t) \xrightarrow{t = mT} r(mT) \]

The matched filter output is

\[
  r(t) = y(t) * h(t) = x_b(t) * h(t) + n(t) * h(t)
  = \sum_k X_k \int_{-\infty}^{\infty} p(\tau - kT)p(\tau - t) d\tau + \int_{-\infty}^{\infty} n(\tau)p(\tau - t) d\tau
\]

Sampling at $t = mT$, $m = 0, 1, 2, \ldots$, we now get

\[
  r(mT) = X_m + N_m
\]

where $N_m$ is noise part of the filter output at time $mT$:

\[
  N_m = \int_{-\infty}^{\infty} n(\tau)p(\tau - mT) d\tau
\]
Properties of the Noise

Let us denote \( r(mT) \), the sampled output at time \( mT \), by \( Y_m \).

\[ Y_m = X_m + N_m, \quad m = 0, 1, 2, \ldots \]

Note that this is a *discrete-time channel*. We have converted the continuous-time problem into a discrete-time one of detecting the symbols \( X_m \) from the noisy outputs \( Y_m \).

To do this, we first need to understand the properties of the noise \( N_m \). Recall that

\[ N_m = \int_{-\infty}^{\infty} n(\tau)p(\tau - mT)d\tau \]

\( N_m \) is a random variable whose distribution depends on the statistics of the random process \( n(t) \).

You will learn about random processes and their characterisation in 3F1 & 3F4, but this is outside the scope of this course. For now, we will directly specify the distribution of \( N_m \) and analyse the detection problem.
\[ Y_m = X_m + N_m, \quad m = 0, 1, 2, \ldots \]

Modelling \( n(t) \) as a Gaussian process leads to the following important characterisation of \( N_m \):

- For each \( m \), \( N_m \) is a Gaussian random variable with zero mean, and variance \( \sigma^2 \) that can be estimated empirically.
- Further \( N_1, N_2, \ldots \) are independent.
- Thus the sequence of random variables \( \{N_m\}, m = 0, 1, \ldots \) are independent and identically distributed as \( \mathcal{N}(0, \sigma^2) \).

**Detection**

- At the Rx, how do we detect the information symbol \( X_m \) from \( Y_m \) for \( m = 0, 1, \ldots \)?
- Remember that each \( X_m \) belongs to the symbol constellation.
Detection for Binary PAM

Let’s start with a simple binary constellation, then generalise. Consider a constellation where each $X_m \in \{-A, A\}$. This is called binary PAM or BPSK (‘Binary Phase Shift Keying’)

$$Y = X + N$$

The detection problem is now:

*Given $Y$, how to decide whether $X = A$ or $X = -A$?*

Observe that:

- $Y = A + N$ if $X = A$ and $Y = -A + N$ if $X = -A$
- $N$ is distributed as $\mathcal{N}(0, \sigma^2)$
- Therefore the pdf $f(Y|X = A)$ is Gaussian with mean $A$ and variance $\sigma^2$
- Similarly the pdf $f(Y|X = -A)$ is Gaussian with mean $-A$ and variance $\sigma^2$

Note: Adding a constant to a random variable just shifts the mean, does not change the shape of the distribution
Let $\hat{X}$ denote the decoded symbol. When the symbols $A$ and $-A$ are \textit{a priori} equally likely, the optimal detection rule is:

$$\hat{X} = A \text{ if } f(Y | X = A) \geq f(Y | X = -A)$$

$$\hat{X} = -A \text{ if } f(Y | X = -A) > f(Y | X = A)$$

\textit{“Choose the symbol from which $Y$ is most likely to have occurred”}

- This decoder is called the \textbf{maximum-likelihood decoder}
- This decoder is intuitive and seems sensible, and is in fact, the optimal detection rule when all the constellation symbols are equally likely (we will not prove this here)
- It is then a special case of the Maximum a Posteriori (MAP) detection rule, which minimises the probability of detection error (module 4F5)
\[ f(Y|X = -A) \quad f(Y|X = A) \]

The detection rule can be compactly written as

\[ \hat{X} = \arg \max_{x \in \{A, -A\}} f(Y|X = x) \]

Thus the detection rule is just: \( \hat{X} = A \) if \( Y \geq 0 \), \( \hat{X} = -A \) if \( Y < 0 \)

“Choose the constellation symbol closest to the output \( Y \)”
Decision Regions

The detection rule partitions the space of $Y$ (the real line) into decision regions.

For binary PAM, we just derived the following decision regions:

\[ \hat{X} = -A \quad \hat{X} = A \]

Q: When does the detector make an error?

A: When $X = A$ and $Y < 0$, or When $X = -A$ and $Y > 0$

We will calculate the probability of error shortly, but let’s first find the detection rule for general PAM constellations.
Detection for General PAM Constellations

The detection rule can easily be extended to a general constellation $C$

- E.g., $C$ may be the 3-ary constellation $\{-2A, 0, 2A\}$ or a 4-ary constellation $\{-3A, -A, A, 3A\}$

- The maximum-likelihood principle is the same: "Choose the constellation symbol from which $y$ is most likely to have occurred"

\[
\hat{X} = \arg \max_{x \in C} f(Y|X = x) \\
= \arg \max_{x \in C} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y-x)^2}{2\sigma^2}} = \arg \min_{x \in C} (Y - x)^2
\]

Thus, the detection rule for any PAM constellation boils down to:

"Choose the constellation symbol closest to the output $Y"
Example: 3-ary PAM

\[ Y = X + N, \quad N \sim \mathcal{N}(0, \sigma^2) \]

What is the optimal detection rule and the associated decision regions if \( X \) belongs to the 3-ary constellation \( \{-2A, 0, 2A\} \)?

The “nearest symbol to \( Y \)” decoding rule yields

\[ \hat{X} = \begin{cases} 
-2A & \text{if } Y < -A \\
0 & \text{if } -A \leq Y < A \\
2A & \text{if } Y > A 
\end{cases} \]
Probability of Detection Error

\[ Y = X + N \]

Consider binary PAM with \( X \in \{A - A\} \). The decision regions are:

\[ \hat{X} = -A \quad \hat{X} = A \]

The detector makes an error when \( X = A \) and \( Y < 0 \), or when \( X = -A \) and \( Y > 0 \)

The probability of detection error is

\[ P_e = P(\hat{X} \neq X) \]

\[ = P(X = -A)P(\hat{X} = A \mid X = -A) + P(X = A)P(\hat{X} = -A \mid X = A) \]

\[ = \frac{1}{2}P(\hat{X} = A \mid X = -A) + \frac{1}{2}P(\hat{X} = -A \mid X = A) \]

(The symbols are equally likely \( \Rightarrow P(X = A) = P(X = -A) = \frac{1}{2} \))
Let us first examine $P(\hat{X} = A \mid X = -A)$

$$P(\hat{X} = A \mid X = -A) = P(Y > 0 \mid X = -A)$$
$$= P(-A + N > 0 \mid X = -A)$$
$$= P(N > A \mid X = -A) \overset{(a)}{=} P(N > A)$$

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

$$P(\hat{X} = -A \mid X = A) = P(Y < 0 \mid X = A)$$
$$= P(A + N < 0 \mid X = A)$$
$$= P(N < -A \mid X = A) = P(N < -A)$$
The probability of detection error is therefore

\[ P_e = \frac{1}{2} P(\hat{X} = A \mid X = -A) + \frac{1}{2} P(\hat{X} = -A \mid X = A) \]
\[ = \frac{1}{2} P(N > A) + \frac{1}{2} P(N < -A) \]
\[ \overset{(b)}{=} P(N > A) \overset{(c)}{=} P\left(\frac{N}{\sigma} > \frac{A}{\sigma}\right) \]

\( (b) \) holds due to the symmetry of the Gaussian pdf \( \mathcal{N}(0, \sigma^2) \):

\[ P(N < -A) \quad P(N > A) \]

\[ -A \quad 0 \quad A \]

\[ \overset{(c)}{=} \]

In \( (c) \), we have expressed the probability in terms of a standard Gaussian random variable with distribution \( \mathcal{N}(0, 1) \)

Recall from 1B Paper 7 (Probability) that if \( N \) is distributed as \( \mathcal{N}(0, \sigma^2) \) then \( \frac{N}{\sigma} \) is distributed as \( \mathcal{N}(0, 1) \)
The $Q$-function

The error probability is usually expressed in terms of the $Q$-function, which is defined as:

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

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

The probability of detection error is therefore

$$P_e = P(N > A) = P\left(\frac{N}{\sigma} > \frac{A}{\sigma}\right) = Q\left(\frac{A}{\sigma}\right) = Q\left(\sqrt{\frac{E_s}{\sigma^2}}\right)$$

where $E_s$ is the average energy per symbol of the constellation:

$$E_s = \frac{1}{2}(A^2 + (-A)^2) = A^2$$

- For a binary constellation, each symbol corresponds to 1 bit. ⇒ the average energy per bit $E_b$ is also equal to $A^2$ in this case
- For an $M$-point constellation, the average energy per symbol $E_s = E_b \log_2 M$
- $\frac{E_b}{\sigma^2}$ is called the signal-to-noise ratio (snr) of the transmission scheme
- $P_e$ can be plotted as a function of the snr $\frac{E_b}{\sigma^2}$ …
$P_e$ vs snr for binary PAM

To get $P_e$ of $10^{-3}$, we need $snr \frac{E_b}{\sigma^2} \approx 9$ dB
Error Probability vs Transmit Power

The probability of error for binary PAM decays rapidly as $\text{snr} \uparrow$:

- $Q(x) \approx e^{-x^2/2}$ for large $x > 0$ $\Rightarrow$ $P_e \approx e^{-\text{snr}/2}$

Can we set the $\text{snr} \frac{E_b}{\sigma^2}$ to be as high as we want, by increasing $E_b$? (i.e., by increasing $A$ since $E_b = E_s = A^2$ for binary PAM)

- The problem is that transmitted power also increases!
- Intuition: 1 symbol with transmitted every $T$ seconds with average energy $E_s$ $\Rightarrow$ transmit power is $E_s / T$
- Thus as you increase the snr, you battery drains faster!

We will now prove that for any PAM constellation the power of the baseband PAM signal $x_b(t)$ is $\frac{E_s}{T}$, where $E_s$ is the average symbol energy of the constellation. Recall

$$x_b(t) = \sum_k X_k p(t - kT)$$

Let’s square and integrate ...
Power of PAM signal

Consider the PAM signal from time $-nT$ to $nT$, carrying symbols $X_{-n}, X_{-n+1}, \ldots, X_n$ drawn randomly from the constellation

$$P_{PAM} = \lim_{n \to \infty} \frac{1}{2nT} \mathbb{E} \int_{-nT}^{nT} \left( \sum_{k=-n}^{n} X_k p(t - kT) \right)^2 dt \quad (2)$$

where the expectation is needed because the symbols $\{X_k\}$ are random symbols drawn uniformly from the PAM constellation.

We write

$$\left( \sum_{k=-n}^{n} X_k p(t - kT) \right)^2 = \sum_{k=-n}^{n} X_k^2 p^2(t - kT) + \sum_{k=-n}^{n} \sum_{j \neq k} X_k X_j p(t - kT)p(t - jT)$$

Plug back into (2), integrate each of these sums separately ...
First term \( = \lim_{n \to \infty} \frac{1}{2nT} \sum_{k=-n}^{n} \mathbb{E}[X_k^2] \int_{-nT}^{nT} p^2(t - kT) \, dt \)

\[
(a) \lim_{n \to \infty} \frac{E_s}{2nT} \sum_{k=-n}^{n} \int_{-nT}^{nT} p^2(t - kT) \, dt \\
(b) \frac{E_s}{T} \lim_{n \to \infty} \frac{2n + 1}{2n} = \frac{E_s}{T}
\]

(a) holds because \( \mathbb{E}[X_k^2] = E_s \), the average symbol energy. (b) holds because the pulse shape \( p(t) \) has unit energy.

The second term is

\[
\lim_{n \to \infty} \frac{1}{nT} \sum_{k=-n}^{n} \sum_{j \neq k} \mathbb{E}[X_kX_j] \int_{-nT}^{nT} p(t - kT)p(t - jT) \, dt = 0
\]

because: a) \( \mathbb{E}[X_kX_j] = \mathbb{E}[X_k]\mathbb{E}[X_j] = 0 \) as the symbols \( X_j, X_k \) are independent and the symbol constellation is symmetric around 0; b) further the orthogonal shifts property of \( p(t) \) implies that the integral also \( \to 0 \).

Hence \( P_{PAM} = E_s / T \).
PAM is a way to map a sequence of information bits to a continuous-time baseband waveform:

1. Pick a constellation, map the information bits to symbols $X_1, X_2, \ldots$ in the constellation.
2. These symbols then modulate the amplitude of a pulse shape $p(t)$ to generate the baseband waveform $x_b(t)$.

$$x_b(t) = \sum_k X_k p(t - kT)$$
Desirable properties of the pulse shape $p(t)$:
- $p(t)$ should decay quickly in time; its bandwidth $W$ shouldn’t be too large
- *Orthonormal* shifts property for simple and reliable decoding

At the receiver, first **demodulate** then **detect**:
- The demodulator is a *matched filter* with IR $h(t) = p(-t)$
- Matched filter output is *sampled* at times $\ldots, 0, T, 2T, \ldots$
  At time $mT$, the output is

$$Y_m = X_m + N_m$$

$N_m$ is Gaussian noise with zero mean and variance $\sigma^2$ that can be empirically estimated
- Detection rule: $\hat{X}_m$ = the constellation symbol closest to $Y_m$

Probability of detection error can be calculated:
- Decays exponentially with snr $E_s/\sigma^2$
- $E_s$ is average energy/symbol of the constellation; power of PAM waveform $x_b(t)$ is $E_s/T$