

Dr Caerwyn Ash

Digital Communications

Communication has been one of the deepest needs of the human race throughout recorded history. It is essential to forming social unions, to educating the young, and to expressing a myriad of emotions and needs.

The various communication disciplines in engineering have the purpose of providing technological aids to human communication. One could view the smoke signals and drum rolls of primitive societies as being technological aids to communication, but communication technology as we know it today became important with telegraphy, then telephony, then video, then computer communication, such as these notes, digitally scanned and sent using email or dropbox.

Initially these technologies were developed as separate networks and were viewed as having little in common. As these networks grew, however, the fact that all parts of a given network had to work together, coupled with the fact that different components were developed at different times using different design methodologies, caused an increased focus on the underlying principles and architecture understanding required for

Continued system evolution.

This need for basic principles was probably best understood at American Telephone & Telegraph (AT&T) where Bell Laboratories was created as the research and development arm of AT&T. Perhaps the greatest contribution was the creation of information theory by Claude Shannon in 1948. For perhaps the first 25 years of its existence, information theory was regarded as a beautiful theory but not as a central guide to the architecture and design of communication systems. After that time, both device technology and the engineering understanding of the theory were sufficient to enable system development to follow information theoretic principles.

Digital communication systems, by definition, are communication systems that use such a digital sequence as an interface between the source and the channel input.

The idea of converting an analog source output to a binary sequence was quite revolutionary in 1948, and the notion that this should be done before channel processing was even more revolutionary. Today with digital cameras, digital video, mobile phones, iPads, the idea of digitizing any data source is commonplace even among the most technophobic. The notion of a binary interface before channel transmission is almost as commonplace.

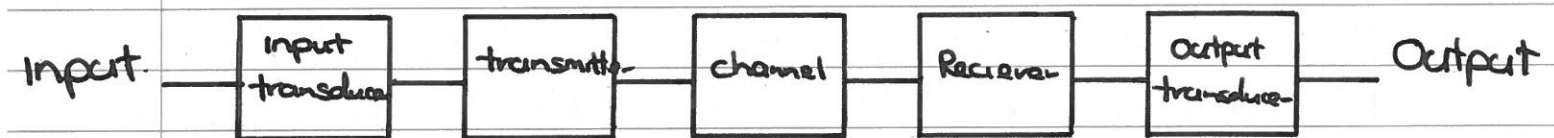
There are a number of reasons why communication systems now usually contain a binary interface between source and channel.

- * Digital hardware has become so cheap, reliable, and miniaturized, that digital interfaces are eminently practical.
- * A standardized binary interface between source and channel simplifies implementation and understanding, since source coding/decoding can be independently of the channel, and similarly channel coding/decoding can be done independently of the source.

- * A standardized binary interface between source and channel simplifies networking, which now reduces to sending binary sequences through the network.
- * One of the most important of Shannon's information theoretic results is that if a source can be transmitted over a channel in any way at all, it can be transmitted using a binary interface between source and char. This is known as the source/channel separation theorem.

Block Diagram of a communications system

Figure 1 shows a commonly used model for a communication system. Although it suggests a system for communication between two remotely located points, this block diagram is also applicable to remote sensing systems such as radar or sonar, in which the system input and output may be located at the same site. Regardless of the particular application and configuration, all information transmission systems invariably involve three major sub-systems, a transmitter, the channel, and a receiver.



Input transducer

The wide variety of possible sources of information results in many different forms for messages. Regardless of their exact form, messages may be categorized as analog or digital. The former may be modeled as function of a continuous-time variable (eg pressure, temperature, speech, music) whereas the latter consist of discrete symbols (eg written text). Almost invariably the message produced by a source must be converted by a transducer to a form suitable for the particular type of communication system employed. For example in electrical communications speech waves are converted by a microphone to voltage variations.

Transmitter

The purpose of the transmitter is to couple the message to the channel. Although it is not uncommon to find the input transducer directly coupled to the transmission medium, as for example in some intercom systems, it is often necessary to modulate a carrier wave with the signal from the input transducer. modulat is the systematic variation of

Some attribute of the carrier, such as amplitude, phase, or frequency, in accordance with a function of the message signal. There are several reasons for using a carrier and modulating it. Important ones are.....

- ① for ease of radiation
- ② to reduce noise and interference
- ③ for channel assignment
- ④ for multiplexing or transmission of several messages over a single channel
- ⑤ to overcome equipment limitations

Channel

The channel can have many different forms, the most familiar is the channel that exists between the transmitting antenna of a commercial radio station and the receiving antenna of a radio. In this channel, the transmitted signal propagates through the atmosphere, or free space, to the receiving antenna. However it is not uncommon to find the transmitter hard-wired to the receiver, as in current cable TV / internet boxes. This channel is vastly different from the radio example. However, all channels have one thing

In common, the signal undergoes degradation from transmitter to receiver. Although this degradation may occur at any point of the communication system block diagram, it is customarily associated with the channel alone. This degradation often results from noise and other undesired signals or interference but also may include other distortion effects as well, such as fading signal levels, multiple transmission paths, and filtering.

Receiver

The receiver's function is to extract the desired message from the received signal at the channel output and to convert it to a form suitable for the output transducer. Although amplification may be one of the first operations performed by the receiver, especially in radio communications, where the received signal may be extremely weak, the main function of the receiver is to demodulate the received signal. Often it is desired that the receiver output be a scaled, possibly delayed version

of the message signal at the modulator input, although in some cases a more general function of the input message is desired. However, as a result of the presence of noise and distortion, this operation is less than ideal.

Output transducer

The output transducer completes the communication system. This device converts the electrical signal at its input into the form desired by the system user. Perhaps the most common output transducer is a loudspeaker. Other examples, such as tape recorders, personal computers, meters, cathode-ray tubes, ect.

Standardized interfaces and layering

Large communication systems such as the Internet have incredible complexity, made up of an enormous variety of equipment made by different manufacturers at different times following different design principles. Such complex networks need to be based on some simple architectural principles in order to be understood, managed, and maintained.

Two such fundamental architectural principles are standardized interfaces and layering.

A standardized interface allows the user or equipment on one side of the interface to ignore all details about the other side of the interface except for certain specified interface characteristics. The idea of layering in communication functions into a string of separate layers as illustrated

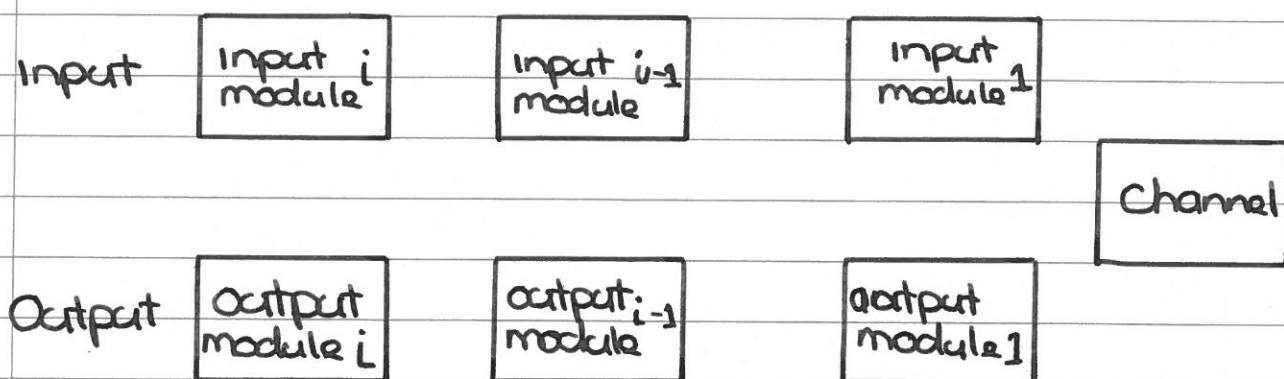


Figure 2: layers & interfaces

Communication Sources

The source might be discrete i.e. it might produce a sequence of discrete symbols, such as letters from the english or chinese alphabet, binary symbols from a computer file etc. Alternatively the source might produce an analog waveform, such as a voice signal from a microphone, the output sensor, a video waveform etc. or a sequence of images such as an x-ray.

Whatever the nature of the source, the output from the source will be modeled as a sample function of a random process. and in fact this was not appreciated before Shannon developed information theory in 1948.

The study of communication before 1948 was based on Fourier analysis, basically one studied the effect of passing sine waves through various kinds of systems. and viewed the source signal as a superposition of sine waves.

Shannon's view however was that if the recipient knows that a sine wave of a given frequency is to be communicated, why not simply regenerate it at the output rather than send it over a long distance. Or if the recipient knows that a sine wave of unknown frequency is to be communicated, why not simply send the frequency rather than the entire waveform.

Transmission Channels

There are many types of transmission channels. We will discuss the characteristics, advantages, and disadvantages of three common types, electromagnetic wave propagation, guided electromagnetic wave, and optical channels. The characteristics of all three maybe explained on the basis of electromagnetic-wave propagation phenomena. However the characteristics and applications of each are different enough to consider seperatly.

electromagnetic-wave propagation channels
The possibility of the propagation of electromagnetic waves was predicted in 1864 by James Clerk Maxwell (1831-1879), a Scottish mathematician who based his theory on the experimental work of Michael Faraday. Heinrich Hertz (1857-1894) a german physicist carried out experiments between 1886-1888 using a rapidly oscillating spark to produce electromagnetic waves, thereby experimentally proving maxwell's predictions. Therefore by the latter part of the nineteenth century

The physical basis for many modern inventions utilizing electromagnetic-wave propagation such as radio, television, radar, mobile phones were already established.

The basic physical principle involved is the coupling of electromagnetic energy into a propagation medium, which can be free space or the atmosphere to an antenna. Many different propagation modes are possible, depending on the physical configuration of the antenna and the characteristics of the propagation medium. The simplest case is propagation from a point source in a medium that is infinite in extent. The propagating wave fronts (surfaces of constant phase) in this case would be concentric spheres. Such a model might be used for the propagation of electromagnetic energy from a distant spacecraft to earth. Another idealized model which approximates the propagation of radio waves from a radio antenna is that of a conducting line perpendicular to an infinite conducting plane.

Digital Transmission

Timing is the operation that significantly distinguishes a digital communication system from an analog system. Before a receiver can even begin to decide which of the various symbols it is receiving, it must establish symbol timing. That is, it must synchronize a local clock with the timing of the received symbols.

What do we gain with digital systems over analog systems?

An analog system can be implemented with relatively few components, whereas a digital system requires significantly more hardware.

1. Stability

Digital systems are inherently time invariant. Key system parameters are imbedded in algorithms that change only if reprogrammed, making for greater accuracy in signal reproduction. With analog hardware, the signal and its parameters are subject to change with component aging, external temperatures, and other environmental factors.

2. flexibility

Once digital hardware is in place, we have a great deal of flexibility in changing the system. This enables us to employ a multitude of signal processing algorithms efficiently

- Ⓐ improve signal fidelity
- Ⓑ do error correction/detection for data accuracy
- Ⓒ perform encryption for privacy and security
- Ⓓ employ compression algorithms to remove redundancies
- Ⓔ allow for multiplexing of various types of signals such as voice, picture, video, text and so on,
- Ⓕ an algorithm can be easily and remotely modified.

3. Reliable reproduction

An analog message traveling through a channel becomes degraded by distortion and noise. While we can employ amplifiers (repeaters) to boost the signal, they amplify the noise as well as the signal and only increase distortion. Thus distortion becomes cumulative. Making a photocopy of another photocopy is an example of this phenomenon. As we will see with digital communication signal reproduction is extremely reliable whether we employ regenerative repeaters for a long haul digital channel.

The task of a digital communication system is to transfer a digital message from the source to the destination. Finite transmission bandwidth sets an upper limit to the symbol rate, and noise causes errors to appear in the output message. Thus signaling rate and error probability play role in digital communication similar to those of bandwidth and signal-to-noise ratio in analog communications.

Principles of Digital Transmission

A digital communication system is made up of several components

Source

The input to a digital system is in the form of a sequence of digits. The input could be the output from such sources as a data set, a computer, a digitized voice signal (PCM or DM), a digital fax, text message, television or telemetry.

Multiplexer

Generally speaking, the capacity of a practical channel transmitting data is much larger than the data rate of individual sources. To utilize this capacity effectively, we combine sources through a digital multiplexer using the process of interleaving. Thus a channel is time shared by several messages simultaneously.

Line Coder

The output of a multiplexer is coded into electrical pulses or waveforms for the purpose of transmission over the channel. This process is called "line coding" or "transmission coding"

Regenerative Repeater

Regenerative repeaters are used at regularly spaced intervals along a digital transmission line to detect the incoming digital signal and regenerate new clean pulses for further transmission along the line. This process periodically eliminates, and thereby combats, the accumulation of noise and signal distortion along the transmission path.

line Coding

Digital data can be transmitted by various transmission or line codes, such as on-off, NRZ, Unipolar RZ, NRZ-mark, manchester coding. A line code should have the following properties:

1. Transmission bandwidth

It should be as small as possible.

2. Power efficiency

for a given bandwidth and a specified detection error probability, the transmitted power should be as small as possible.

3. Error detection and correction capability

It should be possible to detect, and preferably correct, detection errors. In a bipolar case, for example, a single error will cause bipolar violation and can easily detected.

4. Favourable power spectral density

It is desirable to have zero PSD at $\omega = 0$ (dc), because ac coupling and transformers are used at the repeaters. Significant power in low frequency components cause dc wander in the pulse stream when ac coupling is used.

5. Adequate timing content

It should be possible to extract timing or clock information from the signal

6. Transparency

It should be possible to transmit a signal correctly regardless of the pattern of 1's and 0's. If the data is coded that for every possible sequence of data the coded signal is received faithfully, the code is transparent.

line coding consists of representing the digital signal to be transported, by an amplitude and time discrete signal, that optimally tuned for the specific properties of the physical channel (and of the receiving equipment). The common types of line coding are unipolar, polar, bipolar, and manchester-coding.

much compound jitter that it would either lose or gain an entire bit time, and then be out of sync with the transmitter. This is because a long stream of ones or zeros would not "change state" at all, but instead would ~~suffer~~ stay at a single value. Manchester codes say that every single bit time will have a transition in the middle of the bit time. Because there are more transitions however manchester codes also require the highest bandwidth of all the line codes.

Non-Return to zero codes (NRZ)

NRZ are complicatedly named especially considering that the NRZ code does return to a zero value. An NRZ code is just a simple square wave, assigning one value to binary one, and another value to binary zero. NRZ codes are more bandwidth efficient than bipolar ones.

Unipolar

unipolar is simply a square wave with $+AV$ being a binary one, and $0V$ being a binary zero. unipolar is convenient as computer circuits use unipolar internally and it requires little effort to expand this system outside the computer. unipolar has a dc term, but a relatively narrow bandwidth.

Bipolar

Bipolar operates using a bipolar voltage supply rail. Binary ones being a negative voltage and binary zeros being a positive voltage.

Manchester

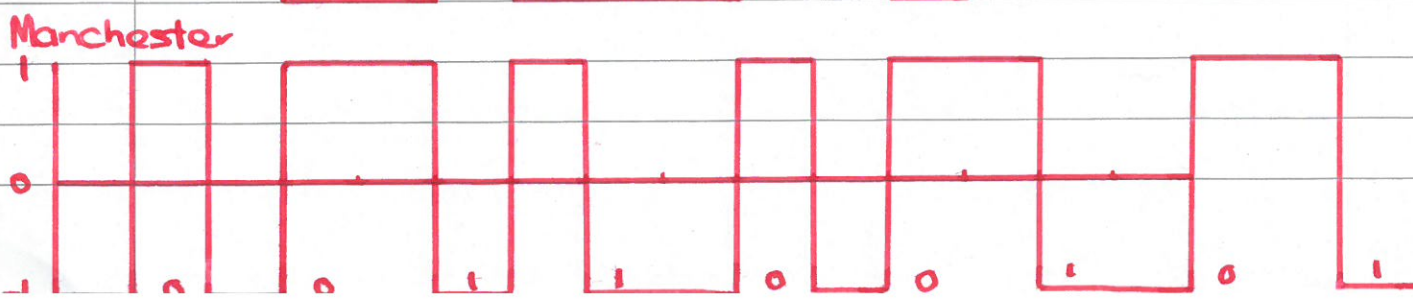
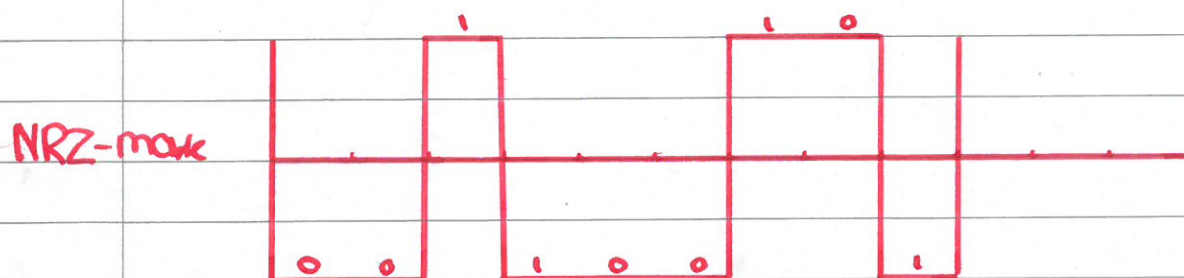
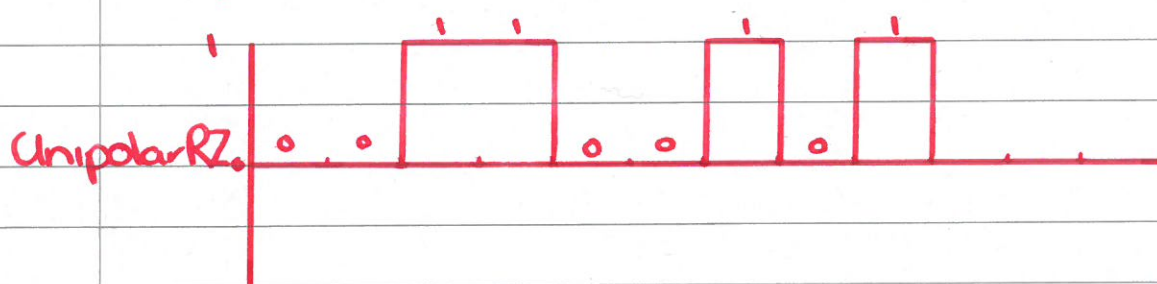
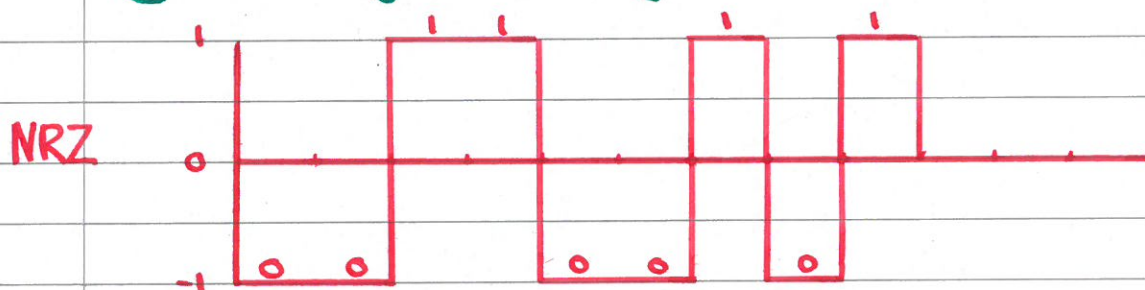
Manchester codes were an invention that allows for timing information to be sent along with the data. In an NRZ code, if there is a long stream of ones or zeros, the receiver could conceivably suffer so

Worked Example

Given the binary data stream 001100101, draw the encoded waveform for each of the following line codes.

- ① NRZ
- ② Unipolar RZ
- ③ NRZ-mark
- ④ Manchester coding

on your diagrams indicate the direction of positive time and assume a binary value of one is represented by a positive signal and binary zero is represented by a negative signal (8 marks)



Modulation

Pulse code modulation (PCM)

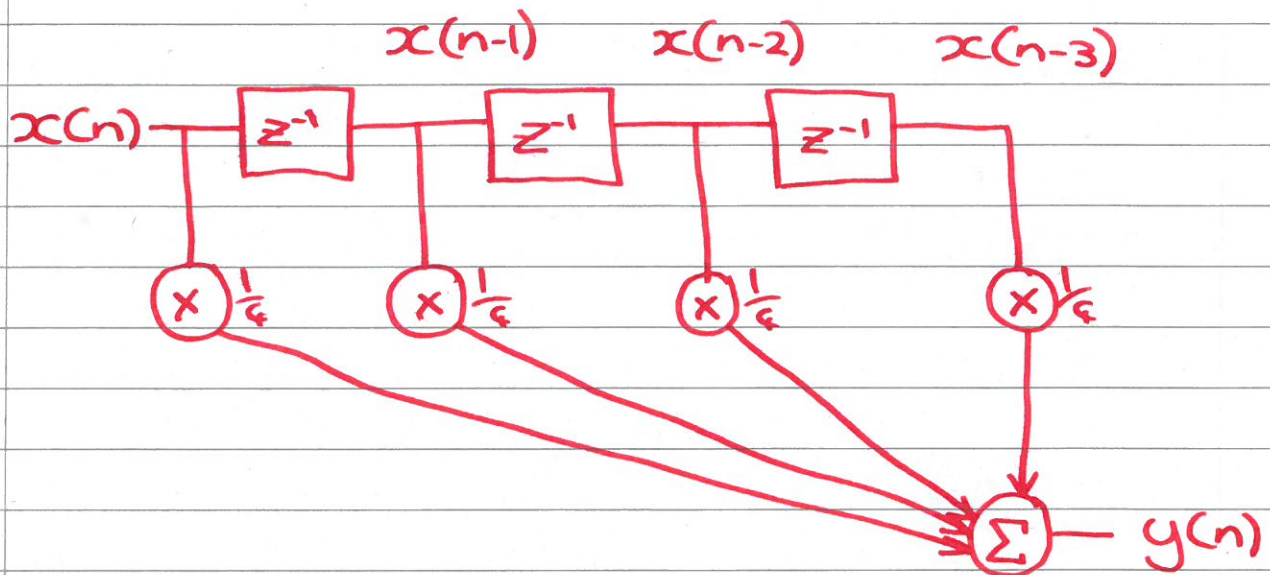
Pulse code modulation is a method used to digitally represent sampled analog samples. It is the standard form of digital audio in computers, CD's, digital audio applications. In a PCM stream, the amplitude of the analog signal is sampled at uniform intervals and each sample is quantized to the nearest value within a range of digital steps.

Delta modulation

Delta modulation is an analog to digital and digital to analog conversion technique used for transmission of voice information where quality is not of primary importance. Delta modulation is the simplest form of differential pulse code modulation, where difference between successive samples is encoded into n-bit data streams. To achieve a high signal to noise ratio, delta modulation must use oversampling techniques, that is the analog signal is sampled at a rate several times higher than the Nyquist's rate.

Worked Example

Provide a sketch illustrating how a 4th order moving average filter can be realised? (2 marks)



Determine the output of a 4th order moving average filter for the following input values

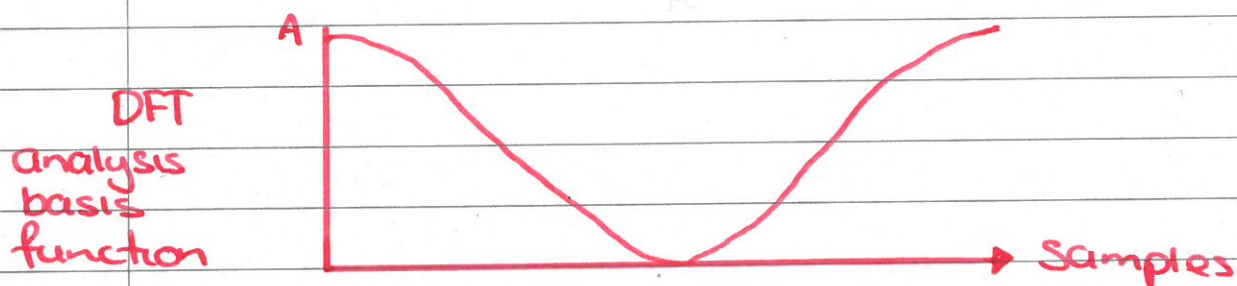
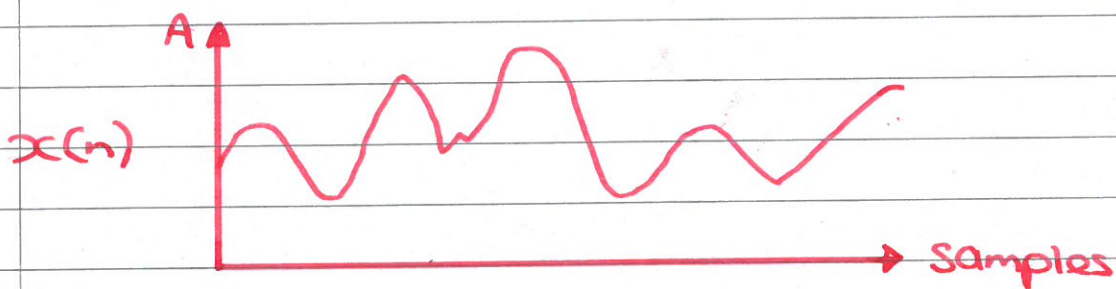
n	0	1	2	3	4	5	6	7	8
x	22	24	21	19	26	19	20	21	22

n	x	y
0	22	5.5
1	24	$5.5 + 6 = 11.5$
2	21	$5.5 + 6 + 5.25 = 16.75$
3	19	$5.5 + 6 + 5.25 + 4.75 = 21.5$
4	26	$6 + 5.25 + 4.75 + 6.5 = 22.5$
5	19	$5.25 + 4.75 + 6.5 + 4.75 = 21.25$
6	20	$4.75 + 6.5 + 4.75 + 5 = 21$
7	21	$6.5 + 4.75 + 5 + 5.25 = 21.5$
8	22	$4.75 + 5 + 5.25 + 5.5 = 20.5$
9	0	$5 + 5.25 + 5.5 = 15.75$
10	0	$5.25 + 5.5 = 10.75$
11	0	$5.5 = 5.5$

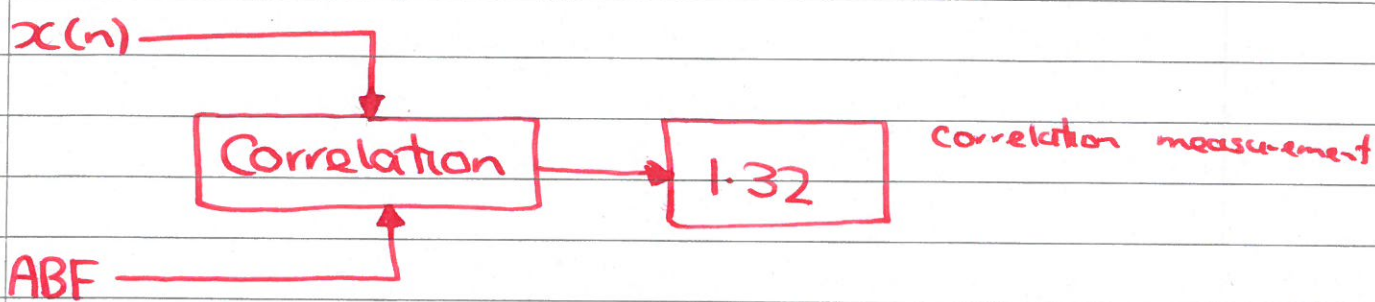
The Discrete Fourier Transform (DFT)

The DFT analyses a time domain signal $x(n)$ to determine the signal's frequency content $X(k)$.

The DFT operates by comparing $x(n)$ against signals known as sinusoidal basis functions.



The comparison is achieved using a mathematical process called correlation. Correlation can be thought of as a measure of similarity between signals or as a measure of how strongly present one signal is in another.



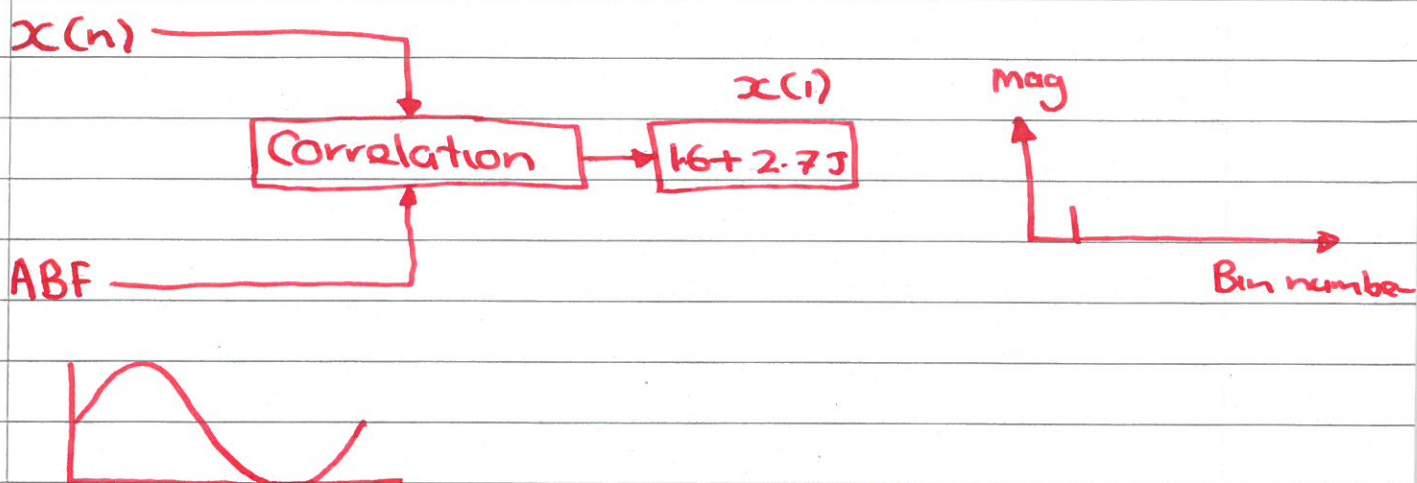
Cosine basis functions with cycle over n steps

The analysis basis functions are the same length as $x(n)$. They are also limited to having an integer number of cycles over the duration of the signal being analysed.

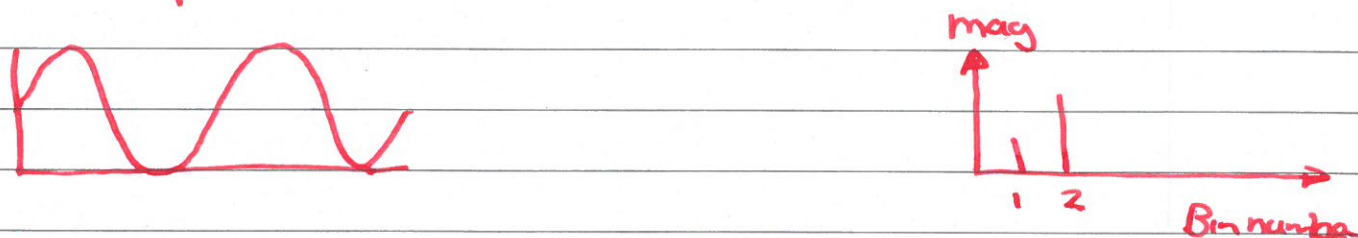
Basis functions are sinusoidal and are both sine waves and cosines.

The result of the comparison with sine wave sinusoidal basis functions are stored as imaginary complex numbers.

The Correlation result is stored in DFT bin 1, which is mathematically denoted as $x(1)$. The magnitude of the correlation result is used to provide the magnitude spectrum of the signal.



Repeat for ABE with 2 cycles over n samples ect



the bin number corresponds to number of cycles associated with ABE.

Bin zero

The signal is compared against a zero frequency basis function that is a signal which is a sequence of ones (or a signal which all amplitude values are ones). In this case the correlation measure is dependant on the average amplitude of the signal $x(n)$. If the average amplitude is zero the correlation measure is zero. If the average of the signal is non zero, the magnitude of the zeroth bin will increase as the deviation from zero average amplitude increases.

The Discrete Fourier Transform equation

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\omega_k n} \quad \text{where } \omega_k = \frac{2\pi k}{N}$$

n = sample number

k = bin number

N = number of samples

complex exponential

Breaks down to cosine waveforms by Euler's theorem

$$e^x = \cos x + j \sin x$$

$$e^{-x} = \cos x - j \sin x$$

$$X(k) = \underbrace{\sum_{n=0}^{N-1} x(n) \cdot \cos(\omega_k n)}_{\text{Real}} - j \underbrace{\sum_{n=0}^{N-1} x(n) \cdot \sin(\omega_k n)}_{\text{Imaginary}}$$

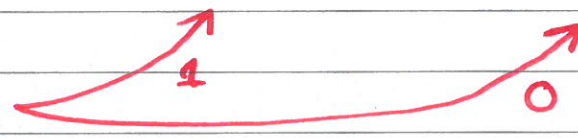
Correlation

Real

Imaginary

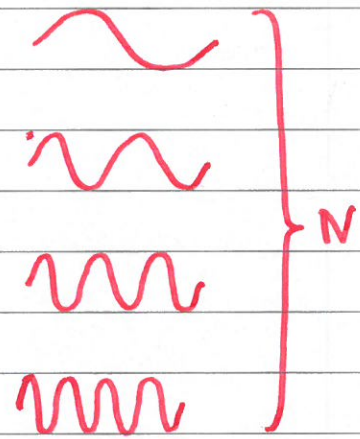
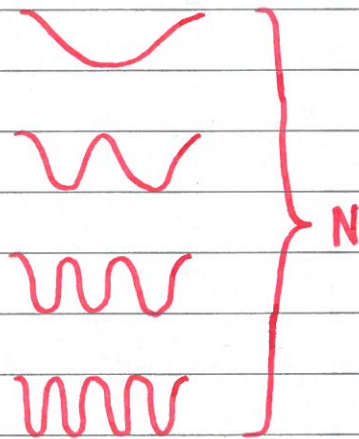
Correlation is the measure of the presence of one signal in another.

$$X(k) = \sum_{n=0}^{N-1} x(n) \cos\left(\frac{2\pi k n}{N}\right) - j \sum_{n=0}^{N-1} x(n) \sin\left(\frac{2\pi k n}{N}\right)$$

When $k=0$ 

$$\therefore X(0) = \sum_{n=0}^{N-1} x(n) \times 1$$

$$X(k) = \sum_{n=0}^{N-1} x(n) \cos\left(\frac{2\pi k n}{N}\right) - j \sum_{n=0}^{N-1} x(n) \sin\left(\frac{2\pi k n}{N}\right)$$



Worked Example

Using the discrete Fourier transform (DFT) determine the $k=0, 1, 2, 3, 4$ bin frequency absolute magnitude and phase values for the following sampled data set?

n	x
0	1
1	0.9
2	0
3	-1
4	-0.9
5	0

using
$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j \frac{2\pi k n}{N}}$$

$k=0$

$$X(0) = \sum_{n=0}^{N-1} x(n) e^0$$

$$X(0) = \sum_{n=0}^{N-1} x(n)$$

$$X(0) = \sum_{n=0}^5 x(n)$$

$$X(0) = 1 + 0.9 + 0 + (-1) + (-0.9) + 0$$

$$X(0) = 0$$

$$K=1 \quad X(1) = \sum_{n=0}^{N-1} x(n) e^{-j \frac{2\pi}{6} n}$$

$$\text{real} \quad X(1) = \sum_{n=0}^5 x(n) \cos\left(\frac{2\pi}{6} n\right) \quad \frac{6\pi}{6}$$

$$X(1) = \cos(0) + 0.9 \cos\left(\frac{2\pi}{6}\right) + 0 - \cos(\pi) - 0.9 \cos\left(\frac{8\pi}{6}\right) + 0$$

$$X(1) = 1 + 0.45 + 1 + 0.45$$

$$X(1) = 2.9$$

$$\text{imag} \quad X(1) = - \sum_{n=0}^5 x(n) \sin\left(\frac{2\pi}{6} n\right)$$

$$X(1) = 0.9 \sin\left(\frac{2\pi}{6}\right) - \sin(\pi) - 0.9 \sin\left(\frac{8\pi}{6}\right)$$

$$X(1) = 1.5588$$

$$X(1) = 2.9 - 1.5588j$$

$$k=2 \quad X(2) = \sum_{n=0}^5 x(n) e^{-\frac{j4\pi n}{6}}$$

$$\text{real} \quad X(2) = \sum_{n=0}^5 \cos\left(\frac{4\pi n}{6}\right)$$

$$X(2) = \cos(0) + 0.9\cos\left(\frac{4\pi}{6}\right) - \cos(2\pi) - 0.9\cos\left(\frac{16\pi}{6}\right)$$

$$X(2) = 0$$

$$\text{imag} \quad X(2) = -\sum_{n=0}^5 \sin\left(\frac{4\pi n}{6}\right)$$

$$X(2) = 0.9\sin\left(\frac{4\pi}{6}\right) - \sin(2\pi) - 0.9\sin\left(\frac{16\pi}{6}\right)$$

$$X(2) = 0$$

$x(k)$ is periodic therefore

$$X(0) = 0$$

$$X(1) = 2.9 - 1.5588j$$

$$X(2) = 0$$

$$X(3) = 2.9 - 1.5588j$$

$$X(4) = 0$$

Worked example

If the data was acquired with a sample time equal to $100\mu\text{s}$ determine the resolution of the transformed data in the frequency domain?

using $T_s = 100\mu\text{s}$ and $f_s = \frac{1}{T_s}$

$$f_s = 1$$

$$100 \times 10^{-6}$$

$$f_s = 10 \times 10^3$$

$$f_s = 10\text{ kHz}$$

using $f_{\text{res}} = \frac{f_s}{N}$

$$f_{\text{res}} = \frac{10 \times 10^3}{6}$$

$$f_{\text{res}} = 1666.6\dot{6}$$

$$f_{\text{res}} = 1.66\dot{6}\text{ kHz}$$

What is the difference between the FFT and the DFT function?

The FFT is just an efficient implementation of the DFT. The output of both is the same. However the FFT requires that the signal being analysed being of length $N = 2^m$ samples where m is a non-zero positive integer.