# MECH0089 Control and Robotics

## Digital Control Systems

**LECTURER**

Prof. Helge Wurdemann

Original notes prepared by Dr. P. Fromme, Prof. R. Bucknall, and Dr. D. Western with modifications by Dr. H. Wurdemann

# Literature / Sources

These lecture notes are based on previous notes by Dr. Richard Bucknall and Matt Greenshields.

More extensive and theoretical treatment than in this brief introduction can be found in the following literature:

[1] Digital Control Systems
B.C. Kuo
Saunders College Publishing

[2] Design of Feedback Control Systems
R.T. Stefani, B. Shahian, C.J. Savant Jr., and G.H. Hostetter
Oxford University Press

[3] Feedback Control of Dynamic Systems
G.F. Franklin, J.D. Powell, & Abbas Emami-Naeini
Pearson
N.B. Older versions are titled 'Digital Control of Dynamic Systems' by Franklin & Powell

[4] Real-Time Computer Control: An Introduction
S. Bennett
Prentice Hall International

[5] Control Systems Theory
O.I. Elgerd
McGraw-Hill

[6] The Art of Control Engineering
K. Dutton, S. Thompson, B. Barraclough
Addison-Wesley

[7] Control System Design and Simulation
J. Golten, A. Verwer
McGraw-Hill

# 1. Introduction to Digital Control Systems

## 1.1 Review and Classification of Control Systems

Control systems can be either open loop or closed loop. Open loop control systems are those without feedback, i.e. those which do not depend upon the output of the process or system being controlled for the control action to occur. In other words, the resulting change in output from the process or system is not fed back to the controller.

Examples of open loop control systems are:

- An (old-fashioned) washing machine where the cycle of washing, spinning, etc. depend solely upon the action of a time switch.

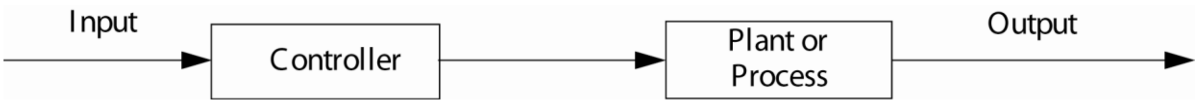


Fig. 1.1 Open Loop Control System

Closed loop control systems rely upon feedback so that the output from the controlled process has a direct effect upon the control action.

Examples of closed loop control systems are:

- Central heating systems with a thermostat: the boiler is controlled (on or off) to keep the air temperature within set limits. A thermometer provides feedback.

- A cistern where water flow is regulated by a float which gradually closes the filling valve as the water level rises. This is known as automatic feedback.

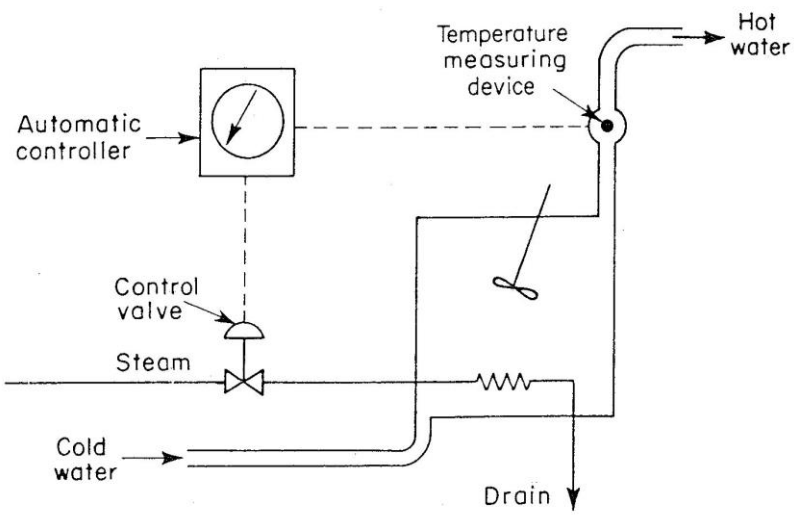


Fig. 1.2 Closed Loop Control System

The types of control systems of interest in this course are automatic (closed loop) control systems. Such control systems are used extensively to control the following parameters in process control applications.

- Pressure
- Flow control
- Temperature
- Energy
- Speed

## Automatic Control

Automatic control systems are used extensively in process control systems and are broadly divided into two different camps, analogue and digital control systems. Analogue control systems were extensively used up until the 1990's and were pneumatic, electronic or mechanical in nature. Digital control systems have superseded analogue control systems in most areas and are entirely electronic in nature. Digital control systems have come about because of the development of the digital computer and in particular the microprocessor. Today almost all new control systems in whatever application are digital by nature. A study of digital control systems is in itself exhaustive and such depth of study is probably best suited to a digital electronic engineer. However the main principles of digital control are important to all engineering students because of their increasing and widely accepted use in all areas of life.

## Comparison of Digital and Analogue Control Systems

The advantages of digital control systems over analogue control systems are:

- smaller and lighter
- flexible, one design suited to multitude of systems
- versatile through use of software which can be modified easily
- cheap as processor and computing costs fall
- reliable
- can carry out recording and statistical analysis in addition to control
- communication in multi-processor digital systems allows the design of large systems
- can be 'embedded and forgotten'
- sophisticated control algorithms can be written

The disadvantages are:

- limitations in resolution (analogue is infinite)
- finite digital word length can increase instability
- digital systems are advancing rapidly and older systems get forgotten
- corrupted or poor software validation can result in catastrophic failure
- 'black box' requires specialists to understand what is inside

Digital control systems of on-going processes are often referred to as real time systems. Real time systems are used in many sectors including engineering, finance and medicine. Studied during this course will be **clock based systems** in which the digital control system obtains data and emits signal stimuli to the process at fixed time intervals known as the sampling rate. This may be only a few milli-seconds in the case of the control surfaces a fly-by-wire aircraft or seconds in the case of pH sampling in a chemical process plant.

## Multi-Level Control Systems

Digital control systems allow the design of large systems with different levels due to the possibility of communication in multi-processor digital systems and can additionally carry out recording and statistical analysis.

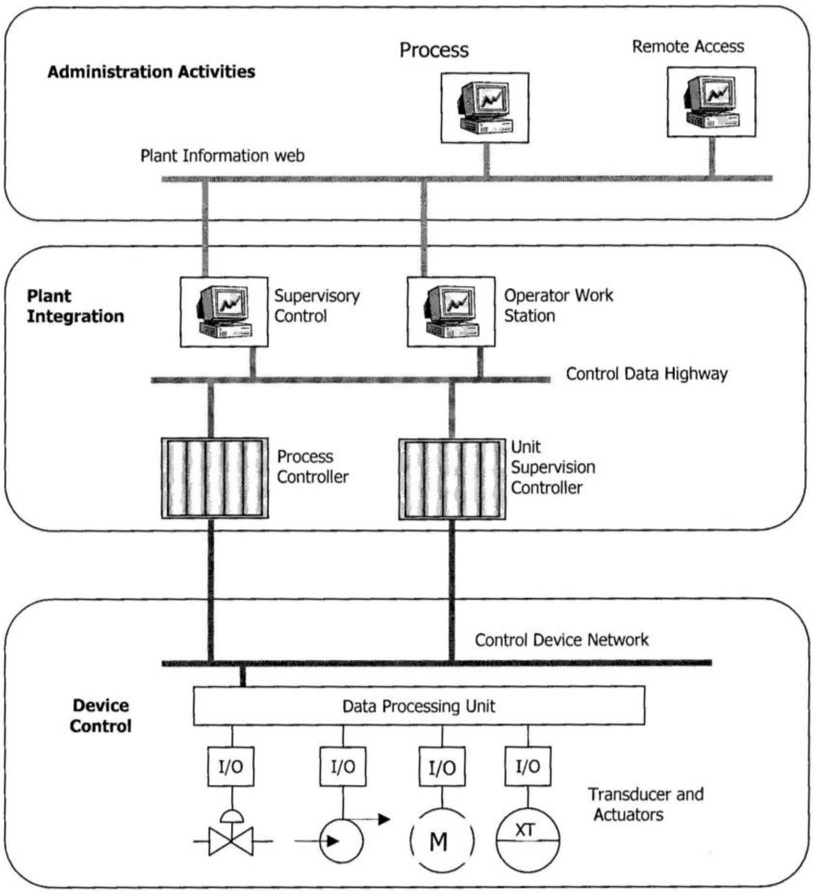


Fig. 1.3 Breakdown of Control System into Different Levels

## Example: A simplified digital autopilot control system

The control diagram below shows a simplified version of an digital autopilot. The main objective of the loop is to ensure that any changes made by the pilot commands are achieved safely and within the capability of the aircraft and once they have been made are maintained until a new attitude command is given. The position and rate information are obtained by transducers and the error signals calculated before being converted to the digital domain where the main control processes undertakes to calculate the most appropriate response of the aircraft control surfaces (airframe dynamics).

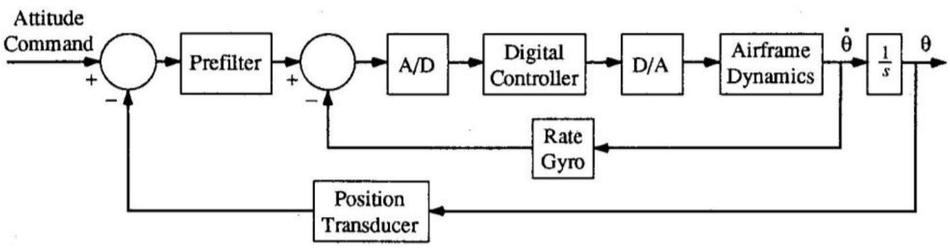


Fig. 1.4 Basic Fly by wire system for modern civilian aircraft control, from [1].

## 1.2. Analogue and Digital Control Systems

Any controller can be considered as an engine for processing information. The input to a controller may include a demanded value for the controlled system output, the actual controlled system output and other information including, for example, information about uncontrollable external influences on and the internal state of the controlled system.

## The Basic Analogue Control Loop

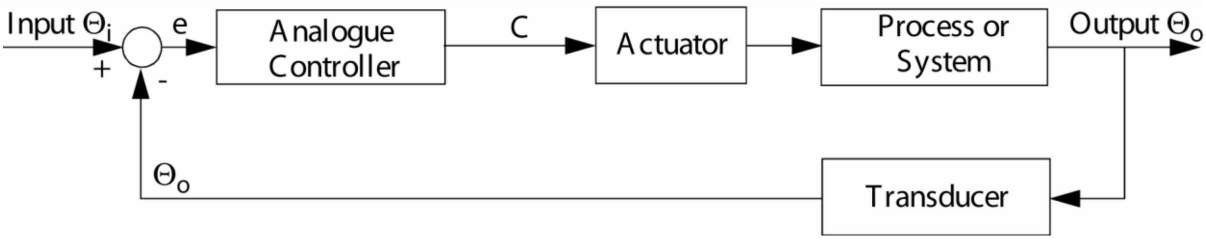


Fig. 1.7 Typical single-loop analogue control system

In an analogue control system information often takes the form of voltages or currents from transducers measuring the physical quantities of the controlled system. Similarly, an output from an analogue control system is a voltage or current sent to an actuator to produce a physical input to the controlled system. These voltages are considered to vary continuously with time and have an infinite resolution, though this may not actually be the case in practice due to sensor and actuator non-linearity.

A basic control loop can be considered by examining a typical analogue control system. A simple analogue control system will consist of the following key components:

**Set point (Input):**
This is the desired output required from the process and is set manually by the operator or automatically.

**Summing Junction:**
This computes the difference between the set point and the measured feedback signal and is known as the **error signal**.

**Controller:**
The controller acts upon the error signal to change the actuator so to give the desired output. In analogue control systems appropriate signals are achieved by using carefully designed analogue electronic circuits, pneumatic controls or mechanical links.

**Actuator:**
This is used to change the input to the process (e.g. the opening and closing of a pneumatic control valve).

**Process:**
This is the system that is being controlled. The actuator, the process, or some combination of the two can all be referred to as **plants**.

**Transducer (sensor):**
Used to measure the actual output from the process.

# The Basic Digital Control Loop

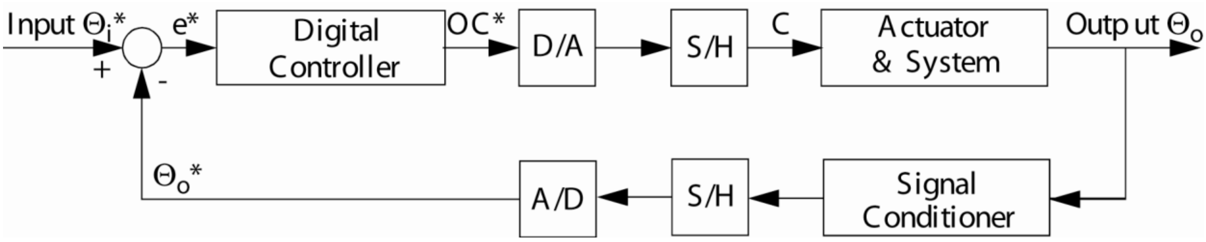


Fig. 1.8 Typical single-loop digital control system for an analogue system

In a digital control system information is quantized in level, according to the resolution of the Analogue to Digital Converter (A/D), and available only at the sampling instants. The control system is referred to as 'quantized discrete time system'. The system being controlled is still usually a continuous time system, as it is usually a real physical process.

In order to understand the difference between the information used in a digital control system and the analogue, continuous time information present in the controlled system the processes of 'Data Sampling' and 'Data Reconstruction' must be examined. These are the processes relating the digital controller input and output and the controlled system.

The basic digital control loop is similar to the analogue control loop except additional components are included in their design these are:

**Digital computing system (Controller):**

Undertakes to do all the processing work including storing the desired signal, comparing the desired signal with the feedback signal and producing an appropriate output to drive a servo-mechanism that in turn drives the actuator.

**Analogue to digital (A/D) converter:**

Undertakes to convert the measured analogue signal from the transducer and convert it to a digital word that the processor can understand. The digital word is a sampled value of the input analogue signal.

**Digital to analogue (D/A) converter:**

Undertakes to convert the digital output word from the microprocessor and convert it to an analogue signal for the servo-mechanism and actuator.

**Sample and Hold (S/H):**

The analogue signal arriving at the digital controller will be sampled at a given rate, known as the sampling rate or sampling frequency. To ensure that the analogue signal is captured, it is sampled and then held across the sampling period.

**Conditioner:**

The signal is processed prior to conversion to ensure that it is within defined bounds such as amplitude, frequency and phase.

In the next chapter, we will take a closer look at the components of a digital control system.

# 1.3. Components of a Control System

### 1.3.1 Instrumentation: Transducers and Actuators

In any control system, parameters must be measured. Parameter measurement is achieved using transducers designed specifically to measure pressure, temperature, energy etc. This can be done using electrical, electronic and mechanical means. Transducers measure the parameter of interest and convert it to another form of energy, e.g., pressure to mechanical torque in a gauge, temperature to a voltage, acceleration to current etc. In modern transducers the output is usually analogue voltage or digital. There are many different transducer types to measure a range of different parameters. For all transducers used in control systems, knowledge of the static and dynamic performance is important.

For static conditions, output signals from the transducer are compared to the theoretical or 'ideal' output to establish the transducer's sensitivity, accuracy and precision. The dynamic performance gives information on steady-state response, dynamic response and response to cyclic variation. Both static and dynamic conditions are important. There is usually a requirement for the converted energy source to be 'conditioned'. Such conditioning may include amplification to strengthen the signal ($\mu$V to V) and the use of filtering to remove unwanted noise prior to transmission, as signals may be measured some distance from the controller.

### 1.3.2 Sample and Hold (S/H)

The analogue signal arriving at the digital controller will be sampled at a given rate, known as the sampling rate or sampling frequency. To ensure that the analogue signal is captured, it is sampled and then held across the sampling period regardless of what happens to the analogue input signal. At the next period the analogue signal is sampled again and then held until the next period begins.

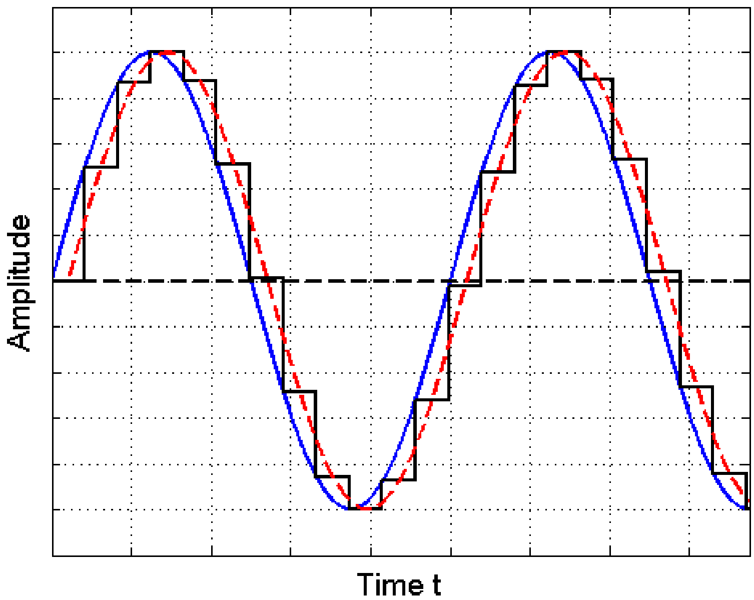


Fig. 1.9 Time lag introduced by a zero order hold: original, sampled-and-held (A-D converted), and then converted back to analogue.

The simplistic operation above is referred to as a **zero order hold** (ZOH) extrapolator, described mathematically as

$f_k(t) = f(kT)$, where ***f(t)*** is approximated for ***kT*** $\leq$ ***t < (k+1)T***.

An ***N-th*** order hold is a device that creates a continuous signal by fitting an order ***N*** polynomial through the past ***N+1*** discrete data values and extrapolating over the next sample period. Zero order holds or “Boxcar generators” are the almost universally applied type of hold because they are simple devices and they typically perform adequately thanks to the ease of achieving high sample rates in modern systems. A zero order hold simply takes the last value of the discrete data and holds the continuous output constant at that level until the next discrete datum.

However, even the ZOH introduces a slight time lag to the sampled signal. This lag depends on the sample period and may cause instabilities in the control system. Hence it needs to be accounted for in an analysis of the system. This is especially true in low-power applications, since slow sample rates are desirable to reduce power consumption.

The choice of the sampling period T relative to the maximum frequency content of the time signal is of utmost importance. The minimal permissive sampling frequency of twice the maximum system frequency is discussed in further detail in chapter 1.4 about the sampling theorem. For practical applications we usually aim to sample at about 10 to 20 times the signal frequency to achieve good control characteristics. In order to replace an analogue continuous controller with a digital controller without significant adaptations the sampling frequency should be about 20 times the maximum system frequency.

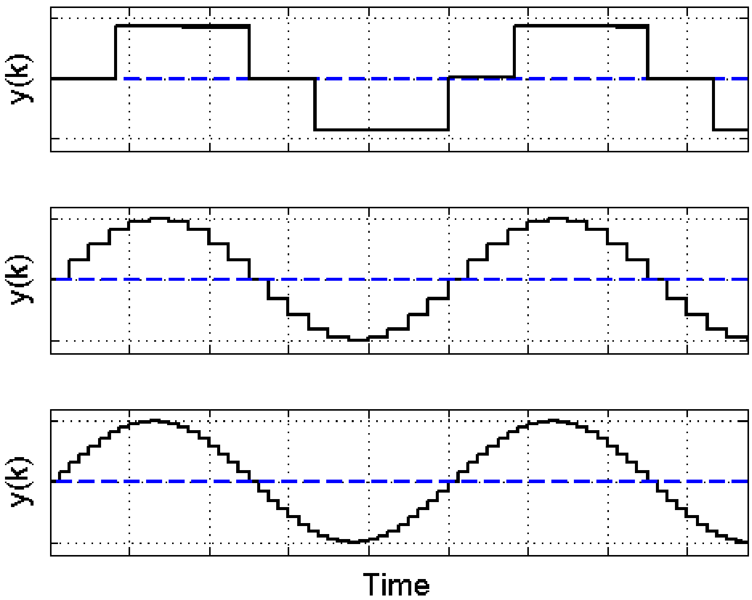


Fig. 1.10 Sampled time signals at 6, 20, and 40 times the signal frequency.

### 1.3.3 Analogue to Digital Converter (A/D)

From the point of view of discretisation of the input signal in time a fast A/D behaves very similarly to an ideal sampler. The effect of the capture period is usually negligible and the data may be transferred to the digital controller with a minimal delay. The effect of the amplitude quantization of the signal must also be considered, however, and this can have an effect on the stability of the total system. The fineness of amplitude quantization of an A/D depends on the number of significant binary digits in the conversion.

Systems using between 8 and 14 bits are common. A 24-bit device would be considered a ‘high resolution’ system. The effect of amplitude quantization is equivalent to putting a staircase non-linearity before a sampler. The maximum quantization error for an n-bit conversion is $2^{-n}$ of the maximum representable voltage (full scale).

The non-linearity introduces a quantization noise into the system and makes limit cycles more likely. This can be the limiting factor on the controller gain. The quantization noise for a 12-bit conversion, for example, gives a signal-to-noise ratio (SNR) of

$$\text{SNR} = 20 \log_{10} 2^{12} = 72.2 \text{ dB}.$$

Higher resolution A/Ds are now much cheaper and faster, making it easier to avoid this problem. However, in low-memory or low-power applications, using a lower resolution is desirable, so it is important to understand the implications for controller performance.

Example: A self-driving car uses a distance sensor with an 8-bit digital controller to sense how far away obstacles are while parking. The sensor has a working range of 10 cm to 5 metres. What is this best resolution this system can achieve?

$$\frac{5m - 0.1m}{2^8} = \frac{4.9m}{256} = 1.9cm$$

An analogue to digital converter is a micro-electronic circuit that takes a single analogue input and converts it to a single digital word under the control of the microprocessor. There are several different ways to achieve the conversion with the four common methods being successive approximation, integration, counter and parallel types. Whatever the process of converting a signal into a digital word, it takes a finite time which ever method is used known as the conversion time. This conversion time will contribute to delays of the type shown in Fig 1.9, affecting the performance of a control system.

If several input values have to be read into the system, quite often a component called a multiplexer is used to allow the conversion of several analogue signals with a single A/D converter.

The number system used in digital systems is binary.

1. To convert a binary number to decimal then:

$$11001 = \frac{2^4\ 2^3\ 2^2\ 2^1\ 2^0}{1\ \ 1\ \ 0\ \ 0\ \ 1} = 1x2^4 + 1x2^3 + 1x2^0 = 25$$

2. To convert decimal number to binary then:

$47 = 47 \div 2^5 = 1$
remainder $15 = 15 \div 2^4 = 0$
remainder $15 = 15 \div 2^3 = 1$
remainder $7 = 7 \div 2^2 = 1$
remainder $3 = 3 \div 2^1 = 1$
remainder 1

Answer is therefore 101111.

Because of the long lengths of the binary numbers that can be involved, human beings (and computer engineers in particular) prefer to use hexadecimal (base 16), which are identified by the prefixed by #.

### 1.3.4 Digital to Analogue Converter (D/A)

The Digital to Analogue Converter (D/A) is the component used as a zero order hold type data reconstructor in nearly all digital control systems. Like the A/D it is not perfect, suffering from a small time delay and having amplitude quantized output. The effect of the quantization is here equivalent to a staircase non-linearity placed after a perfect zero order hold and again the describing function method may be used for analysis.

Data reconstruction is the process of regenerating a continuous time signal from a discrete time signal, usually to drive an actuator at the input to the controlled system. In theory perfect data reconstruction could be achieved by taking the impulse series output from the digital controller and passing it through an analogue filter with uniform gain ***T*** below the Nyquist frequency, zero gain above it and no phase lag or lead.

Unfortunately it is not possible to achieve the impulse input to the filter, nor the perfect reconstructor filter itself, as it requires knowledge of future values of the impulse series. In practice data reconstruction is achieved imperfectly by the use of 'Holds', such as the zero-order hold described previously.

The zero order hold effectively acts as an integrator of the impulse series, holding each integrated value for a time ***T***, until it is discarded (subtracted from its output). Its Laplace transfer function may be found by considering its effect on the discrete impulse series, which will be used when evaluating a discrete transform model of a combined zero order hold and controlled system.

### 1.3.5 Digital Processor

The basic functional blocks of any microprocessor are: the control unit, the arithmetic logic unit (ALU), the registers, the memory, a clock, and the input and output circuits. The control unit is the part of the microprocessor that determines what is going to happen next. To do this it refers to the 'instructions' written in binary called machine code which is stored in the 'program memory'.

The 'program' may have originally been written in a user-friendly 'high level language' such as C++, Fortran or Pascal. A programme written in a high level language is converted to form a set of machine-code binary instructions using a 'compiler or assembler'

The microprocessor needs to remember where in the programme it currently is and it may also need to remember other information such as data is being stored. To do this it uses 'pointers' and these 'pointers' are stored in 'registers'. Registers can also be used for other things such as storing binary numbers prior to any manipulation by the ALU and of course the results of the manipulation.

The ALU carries out all mathematical manipulations such as adding, subtracting and logic operations such as NOT and NOR. This is the heart of the microprocessor.

The other 'bits' needed to make a microprocessor work are the memory, clock and input and output devices to enable data to by passed to and from the outside world. The clock is used to synchronise operations making sure everything is timed correctly so the 'programme instructions' can run smoothly.

Memory is complex in that it is used to store the 'programme instructions', important information regarding start-up and shut down, and data. The data may be binary representation of parameters such as temperature, pressure, etc., or it may be just mean values of these parameters as calculated by the ALU.

## 1.4. Sampling Theorem

Data sampling is the process of converting a continuous function of time, ***f(t)***, into a discrete function of time, ***f*(kT)***, where ***T*** is the sampling interval and ***k = 0 .. n*** is the number of the sample starting at ***t = 0***. The sampling process can be visualized by considering the sampler as a switch which is closed at each sampling instant for a very short time, ***p***. The limit of the resulting signal as ***p*** tends to zero is an ideal sampled signal.

Sometimes in practice a multiplexer switch or sample and hold is used to sample the input signal to an A/D, so this model of the sampling process is quite realistic. In practice the switch cannot close for an infinitely short time and ideal sampling is not achievable. The actual sampling period is called the 'capture time' of the A/D and is usually so short that an ideal sampling model may be used.

An analytical representation of the sampled signal may be found by considering the sampling process as 'Impulse Modulation'. The continuous time signal ***f(t)*** is considered to by modulated or multiplied by a train of Dirac delta functions or unit impulses. This model may be conveniently used to find a Fourier or Laplace representation of ***f*(kT)***, though, as these are continuous time transforms of discrete time signals, they are not best suited to discrete time analysis. Discrete time transforms, like the z transform, are more useful, as we will see in later chapters.

**Sampled spectra and frequency aliasing**

The sampled signal ***f*(t)*** only has values at the sampling instants ***t=kT***, but is intended to be a representation of a continuous time signal with an infinite number of values between the sampling instants. There is obviously information lost in the sampling process. One form of information loss is the 'aliasing' of higher frequencies in the continuous time signal, ***f(t)***, so that they become indistinguishable from lower frequencies in the sampled signal. This can be illustrated by considering the effect of sampling on the different frequency components of ***f(t)***.

The continuous Fourier Transform of ***f*(t)*** could be written as

$$F(\omega) = \int_{-\infty}^{\infty} f*(t)e^{-j\omega t}dt \text{ .}$$

If ***f*(t)*** includes a large oscillation at frequency $\omega_0$, we will of course get a large value for $F(\omega_0)$. However, we will also get an identical value for at any frequency $\omega_n$ satisfying

$$\omega_n = \omega_1 + \frac{2\pi n}{T}$$

because for all values of ***t = kT*** (the only non-zero contributions to our Fourier Transform) the exponential factor has the same value, i.e.

$$e^{-j\omega_0 kT} = e^{-j\omega_n kT}$$

for all k and all n. In practice, this means that when we observe a large value at a particular frequency in our Fourier Transform of a sampled signal, we don’t know whether the original oscillation actually occurred at that frequency of at some other frequency differing by a multiple of $2\pi/T$. So to avoid confusion, we must restrict the bandwidth of our signal **before sampling**.

If frequencies are present in ***f(t)*** which do vary by more than ***2π/T*** then they are not reconstructable from the sampled signal ***f*(kT)***. They are said to have been 'aliased' because they appear with a frequency that is different from their true frequency.
This leads to **Shannon's Sampling Theorem** (or the sampling theorem):

**'The sampling frequency applied to a signal must be at least twice as great as the highest frequency in the input signal to prevent aliasing frequencies'.**

This is mathematically expressed as:

$$f_{sampling} = \frac{1}{T} \geq 2 f_{\max}$$

The frequency $f_{Ny} = \frac{1}{2T}$ is called the Nyquist frequency of the sampling process. So we could also write the above criterion as $f_{Ny} \geq f_{\max}$.

Aliasing must be considered when sampling signals from real processes, as they are rarely frequency limited. Higher frequency components of the signal may be removed with an analogue 'anti-aliasing filter' before the sampler, as long as they do not contain information of interest about the system.

A useful rule is to sample the signal at **10 to 20 times** the maximum frequency of interest and then design an anti-aliasing filter to remove higher frequency components **before sampling** (i.e. an analog filter, implemented in hardware). Various criteria are used to specify the anti-aliasing filter and it is suggested that it should be designed to reduce the aliased signal to one hundredth the power of the non-aliased signal. This can be problematic where signal to noise ratio of the signal is very low.

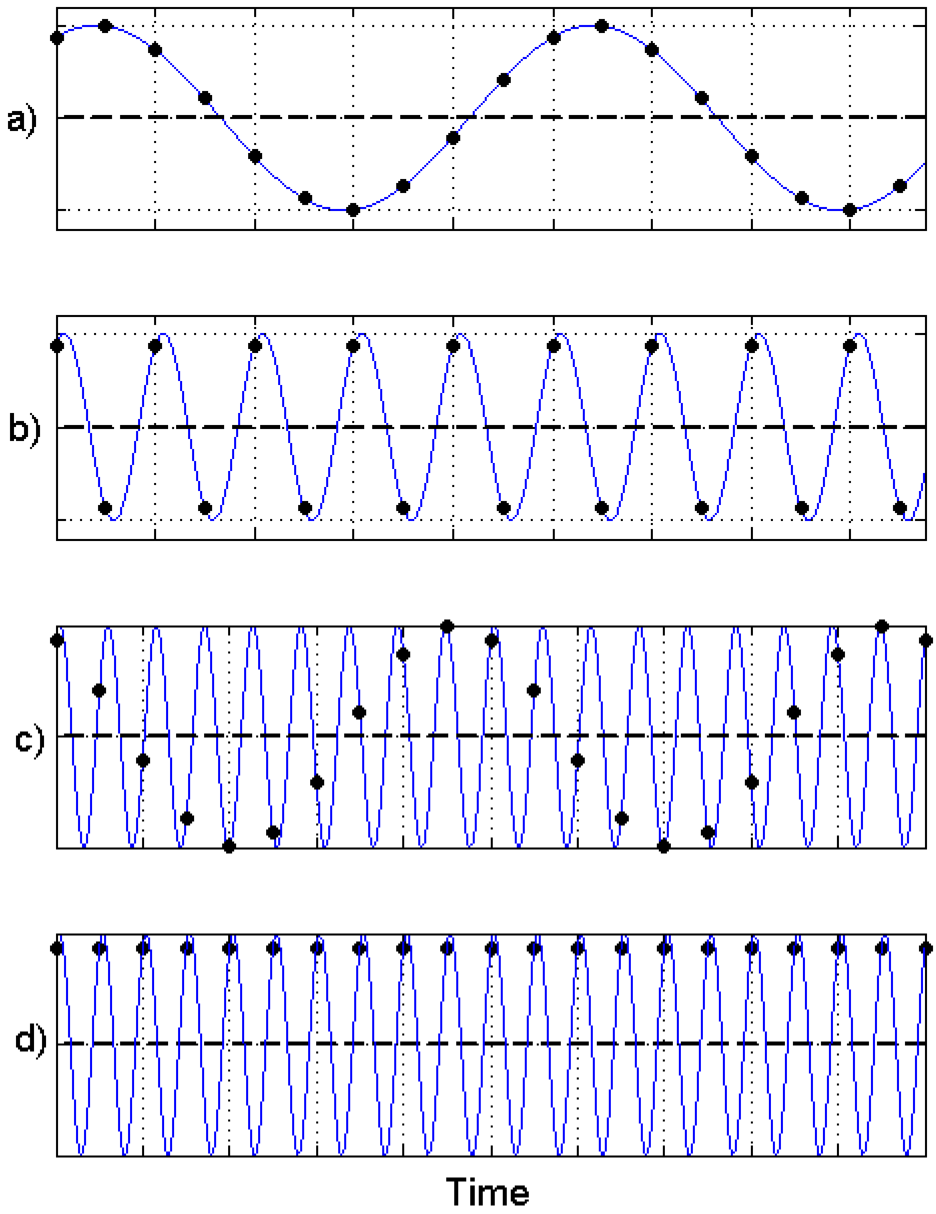


Fig. 1.11 Sinusoidal time signal, sampled at varying sampling frequencies:
a) $f = 0.1\ f_s$, b) $f = 0.5\ f_s$, c) $f = 0.9\ f_s$, d) $f = f_s$.