

MECH0010 Topic Notes

UCL

HD

December 30, 2020

Contents

I Control	11
1 Introduction to Control Systems	12
1.1 Open loop control	12
1.2 Closed loop control	14
1.3 Block diagram representation of control loops	15
1.4 Basic definitions/concepts	16
1.5 Notation	17
1.6 Control system design process	18
1.7 Summary	18
1.8 Tutorial	19
2 Modelling Linear Systems	21
2.1 Linear Term Invariant (LTI) systems	21
2.1.1 Superposition	21
2.1.2 Homogeneity	22
2.1.3 Time invariance	23
2.1.4 Linearisation: Example 1	23
2.2 Dynamic systems - Laplace Transform	24
2.2.1 Dynamic systems as ODEs	24
2.2.2 Laplace Transforms	25
2.2.3 Transfer Functions	27

2.2.4	Transfer Functions of Mechanical Components	28
2.2.5	Combining components	29
3	First and Second Order Systems	33
3.1	First Order Systems	33
3.1.1	Returning to the Time Domain	33
3.1.2	Unit Step Response	36
3.2	Second Order Systems	37
3.2.1	Returning to the Time Domain	39
3.2.2	Unit Step Response	40
3.3	Tutorial	40
4	System Time Response - 1st order	43
4.1	Impulse functions/responses	43
4.1.1	Impulse response of a system: Dirac delta function	43
4.1.2	Impulse response on a first order system	44
4.1.3	Impulse response of a system	44
4.2	Input functions: Impulse - Step - Ramp	46
4.2.1	Step input	46
4.2.2	Ramp input	48
4.2.3	Summary of Input Functions	49
4.3	First Order System Step Response	49
4.4	First Order System Ramp Response	52
4.5	Understanding Poles and Zeros	53
4.5.1	Example	53
4.5.2	s-Plane	54
4.6	Open Loop Motor Speed Control	55

5 System Time Response - 2nd order	57
5.1 Understanding poles and zeros	57
5.1.1 Example	58
5.1.2 s-Plane	58
5.2 Effect of damping ratio on the transfer function of a 2nd order system	59
5.2.1 2nd order transfer function	59
5.2.2 Characteristic equation	59
5.2.3 Overdamped 2nd order system	60
5.2.4 Critically damped 2nd order system	61
5.2.5 Undamped 2nd order system	62
5.3 Applying a step input to 2nd order systems - damping	63
5.3.1 Overdamped step input response of 2nd order systems	63
5.3.2 Critically damped step input response of 2nd order systems . .	64
5.3.3 Underdamped step input response of 2nd order systems	64
5.3.4 Undamped step input response of 2nd order systems	65
5.4 Underdamped step input responses	66
5.4.1 Performance Criteria	66
5.5 Ramp input function responses	68
5.5.1 Applying inputs	68
5.5.2 Undamped response	69
5.5.3 Underdamped response	69
5.5.4 Critically damped response	70
5.5.5 Overdamped response	70
5.5.6 Summary	71
5.6 Transfer function - unity feedback system	71
5.7 Summary	72

6 The s-plane and Steady State Errors	74
6.1 The s-Plane - Poles and Zeroes	74
6.1.1 Example: Servo as a second order system	74
6.1.2 Transfer functions	75
6.1.3 System poles and zeroes	75
6.1.4 Transfer functions in the s-plane	76
6.1.5 Summary	78
6.2 Effects of the damping ratio on poles	79
6.2.1 Transfer functions of second order systems	79
6.2.2 Example: Servomechanism in the s-plane	81
6.3 Steady state performance	84
6.3.1 Final value theorem (FVT)	84
6.3.2 Steady state error	85
6.3.3 Steady state position error	86
6.3.4 Steady state velocity lag	87
6.3.5 Servo steady state error	88
7 PID Controllers	90
7.1 System types	90
7.1.1 System type - calculating error	91
7.1.2 System type for unit feedback control systems - calculating error	91
7.1.3 Type 0 servo - spring return	92
7.2 Proportional control	93
7.2.1 Proportional control - damping ratio trade off	93
7.3 Velocity feedback	94
7.3.1 Doing better - compensation methods	94
7.3.2 Velocity feedback - disadvantages	96

7.3.3	Alternative controllers	97
7.4	PID control	97
7.4.1	Working example: PID control for servo-mechanism	98
7.4.2	Proportional error	98
7.4.3	PID control	99
7.4.4	Derivative error	99
7.4.5	PID control	102
7.4.6	Integral error	102
7.4.7	PID control	105
II	Instrumentation	107
8	General Measurement Systems	108
8.1	Transducers	108
8.2	Signal condition circuits	108
8.3	Amplifiers	110
8.4	Output stage	110
8.5	Feedback-control stage	110
8.6	Calibration	110
8.7	Standards	111
8.8	Units and dimensions	112
9	Strain Gauges	113
9.1	Strain gauges	113
9.1.1	Strain	113
9.1.2	Stress	113
9.1.3	Strain gauges	114
9.1.4	Simple wire strain gauge	115

9.1.5	Gauge factor	115
9.1.6	Foil strain gauge	115
9.1.7	How can we measure strain	116
9.2	Wheatstone bridge	116
9.2.1	Temperature compensation	117
9.2.2	Effect of Poisson's ratio on a dummy gauge	119
9.3	Practical Aspects of Strain Gauge Measurement	120
9.3.1	Strain gauge rosette	120
9.3.2	Bending Strain	120
9.3.3	Tensile or compressive strain in a bending member	121
9.3.4	Bridge balancing	122
9.3.5	Semiconductor strain gauge	123
9.4	Summary	125
10	Measurements of Displacement	126
10.1	Capacitance-Based Displacement Transducers	126
10.1.1	Capacitive Displacement Transducers	126
10.1.2	Measurement of Capacitance	128
10.1.3	Differential Capacitance Displacement Transducers	128
10.1.4	Capacitive Bridge	129
10.2	Inductance-Based Displacement Transducers	131
10.2.1	Linear Variable Differential Transformer (LVDT)	131
10.3	Resistance-Based Displacement Transducers	132
10.4	Measuring Rotational Displacement	134
10.4.1	Optical Incremental Shaft Encoder	134
10.4.2	Optical Absolute Shaft Encoder	136
10.5	Measuring Large Linear Displacement	137

10.5.1	Measurement of Large Linear Displacements	137
10.5.2	Wave Propagation Methods	137
10.6	Other Methods of Measuring Position and/or Displacement	139
10.6.1	Projection Mapping	139
10.6.2	Global Positioning System (GPS)	139
10.7	Datasheet	140
11	Laboratory measurement	141
11.1	Amplifiers	141
11.1.1	Equivalent op-amp circuit and conditions of an ideal op-amp .	143
11.1.2	Non-zero voltage with zero current?	144
11.1.3	Saturation	144
11.1.4	Negative feedback	145
11.1.5	Basic feedback amplifier configurations	145
11.1.6	Basic inverting feedback amplifier	146
11.1.7	Basic non-inverting feedback amplifier	147
11.1.8	Summary	147
11.1.9	Voltage amplification summary	148
11.1.10	Difference amplifier	148
11.1.11	On a bread board	150
11.2	Analogue-Digital (AD) conversion	150
11.2.1	Comparator	150
11.2.2	Example ADC mechanism using comparators	151
11.2.3	Other types of ADC	153
11.2.4	Typical issues to be considered	153
11.3	Interface devices to PC and programming platform	154
11.3.1	Arduino	154

11.3.2 NI products and LabVIEW	155
11.4 Experimental Errors	155
11.4.1 Systematic errors	156
11.4.2 Combining errors in a single quantity	157
11.4.3 Various combinations of quantities	157
11.4.4 General rule for combining errors	158
11.4.5 Example	159
11.4.6 What if we do not know the relationship between error sources and measured variables?	160
11.5 Reference information about least-square curve fitting	160
11.5.1 Least square analysis	160
 12 Measurement of Force and Acceleration	 163
12.1 Measuring Forces	163
12.1.1 Elastic Sensing Elements	164
12.1.2 Strain Gauge Load Cell	165
12.1.3 Force Transducer Using the Force-Balance Principle	167
12.1.4 Piezoelectric Force Transducers	167
12.2 Measuring Acceleration	169
12.2.1 Seismic-Mass Accelerometer	169
12.2.2 Strain Gauge Accelerometer	171
12.2.3 Potentiometric Accelerometer	172
12.2.4 Servo Accelerometer	173
12.2.5 Piezoelectric Accelerometer	174
12.2.6 MEMS Accelerometer	175
 13 Measurement of Velocity, Torque and other Variables	 177
13.1 Measurement of Velocity	177

13.1.1	Velocity transducers with a frequency output	177
13.1.2	Tachogenerator	179
13.2	Measurement of Torque	181
13.2.1	Torque measurement with strain gauges	181
13.3	Measurement of Other Variables	182
13.3.1	Voltage – voltmeter	182
13.3.2	Light – photodiode	183
13.3.3	Light – LDR (light dependent resistor)	185
13.3.4	Other Important Measurements	185
14	Measurement of Temperature	187
14.1	Standard Temperature for Calibration	187
14.2	Measurement Using Changes in Volume/Pressure	188
14.3	Thermoelectric Temperature Sensors (Thermocouples)	188
14.3.1	Thermocouples in practice	189
14.3.2	Thermocouple materials and construction	190
14.3.3	Thermocouple structure	191
14.4	Basic Metallic Resistance Thermometers	191
14.5	Thermistors	193
14.5.1	Measuring circuits for resistive sensors	195
14.6	Pyrometer and Radiation	196
14.6.1	Black body	196
14.6.2	Radiation pyrometers	196
14.7	Temperature Response	197
14.7.1	Transient temperature response	198

Part I

Control

Chapter 1

Introduction to Control Systems

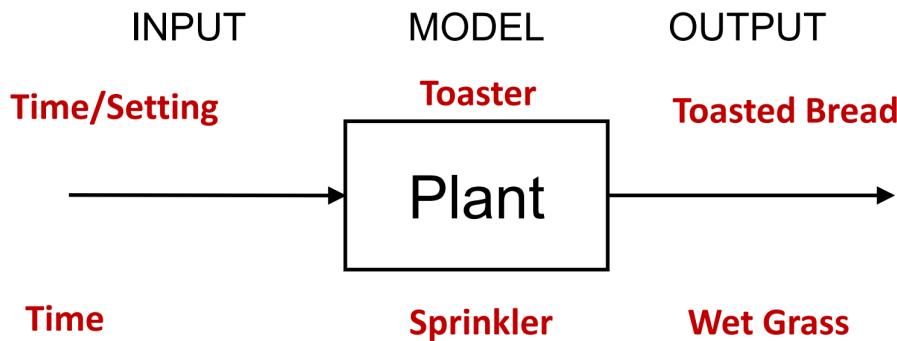
A **control system** is a mechanism which alters the future state of a particular system. **Control theory** is the method of selecting appropriate inputs. Control theory usually concerns dynamic behaviour of physical systems - the goal is to design a controller which leads to the system exhibiting the desired behaviour. Control systems have a large range of applications throughout engineering such as autopilot systems for ships and aircraft, radar tracking, robotics, machinery plant control and machine tools.

Methodology

- **Modelling** - Mathematical model of a system or "transfer function." Comes from detailed analysis of a system and often involves simplifications
- **Prediction** - The model is used to predict the behaviour of the system for a range of parameters and expected excitations
- **Design** - Design a controller which achieves its operating objectives and test it on the model
- **Test** - Here, theory is taken into reality and we compare our model to the physical hardware. Testing of the validity of our assumptions also takes place
- **Iterate** - Improve controller design through updated models

1.1 Open loop control

In an **open loop** control system, the input to the system is not dependant on any previous outputs. The output of the system is not being observed to confirm whether the desired output has been achieved.

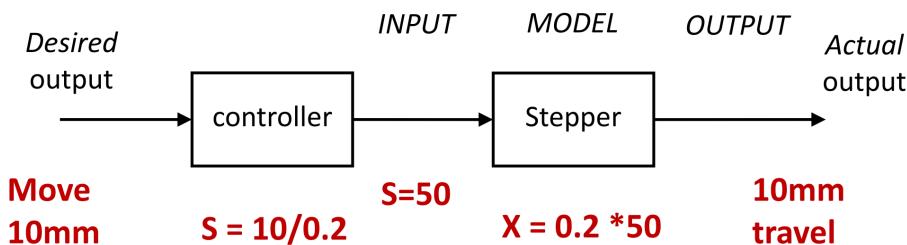


These control systems are simple and low cost. They are used in systems where feedback is not important. For example, a washing machine runs for 90 minutes, regardless of whether the clothes are dry after 60 minutes.

3D Printer example

Consider the stepper motors used to move the print head in x and y with belt and pulley system. Here, open loop control is used because the stepper motors are simple to control and have a relationship between input and output. For example, consider a stepper motor which moves 0.2mm each step (S) for an integer number of steps.

Plant model: $X = 0.2 S$
 Control strategy: $S = X/0.2$



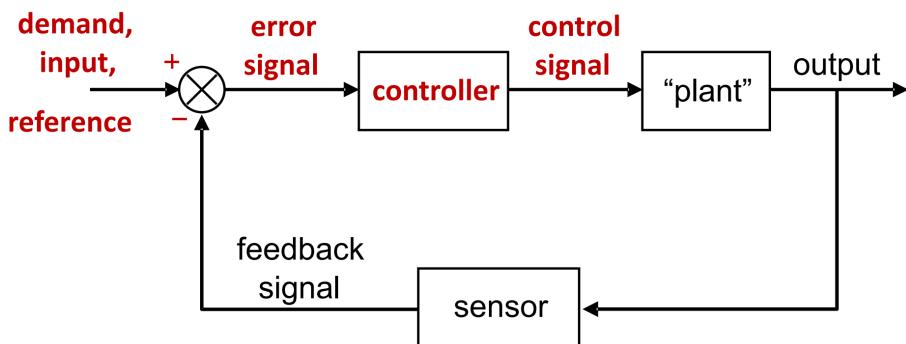
Implementing such a controller could be done with the following pseudocode:

```
Function MovePrintHeadX(Input_mm)
//convert to number of steps needed
Steps_S = Input_mm / 0.2
//send step command to motor
Stepper.step(Steps_S)
```

However, the assumption used to model our system is only valid if the stepper motors is within its specification i.e. friction, load, temperature, power are all nominal. As stated before, the system does not observe if the steps were successful or not.

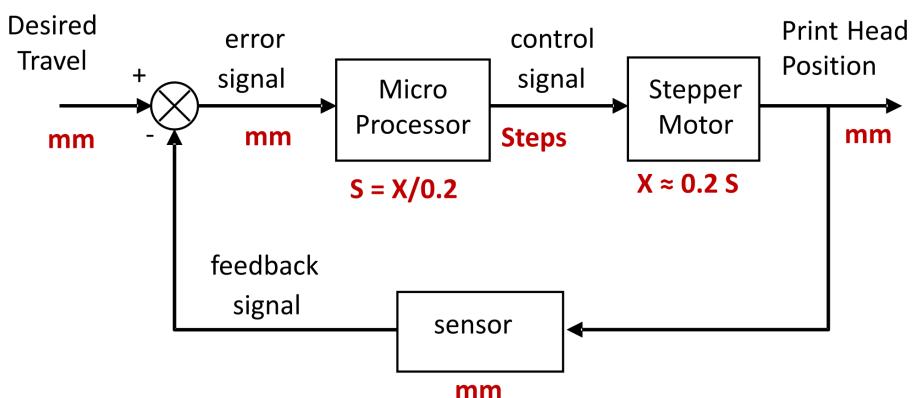
1.2 Closed loop control

By observing the output and comparing it against your desired output, it is possible to update the input to the system. The observed output is called the **feedback signal**.



The difference between the desired output and the feedback signal is known as the **error signal**. This is the signal that the controller uses. Feedback signals allow powerful controllers to be designed.

Taking our 3D printer for example, we can add a position sensor on the x-axis.



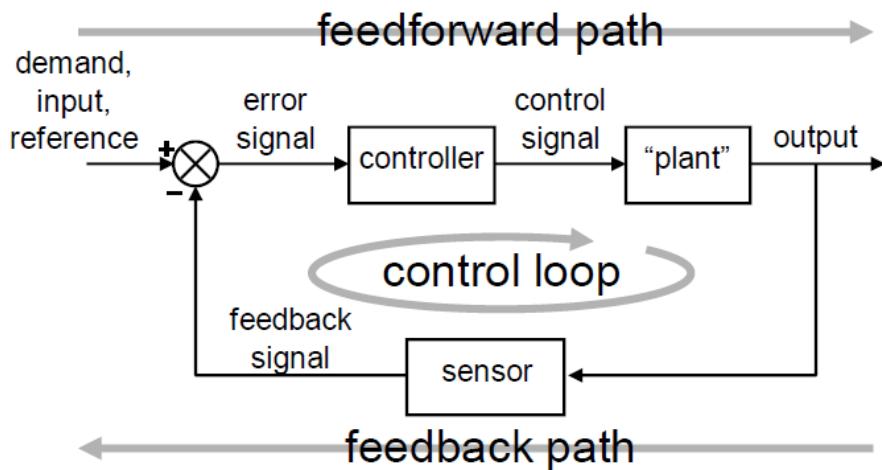
The pseudocode for this system could look like this

```

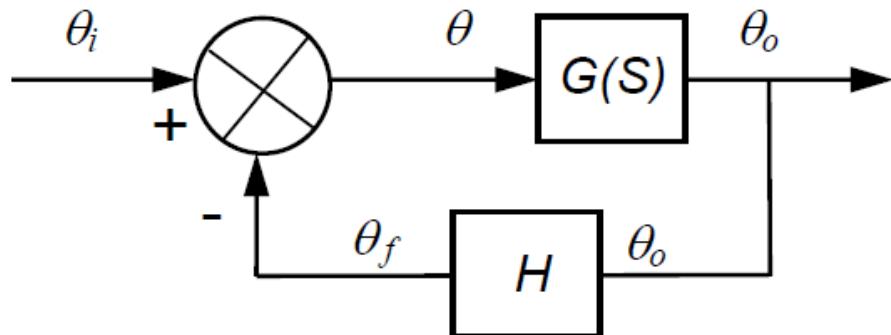
while
    // get measurement of position
    measured_value_mm = Sensor.getValue
    // calc error signal
    error_mm = setpoint_mm - measure_value_mm
    // convert to steps
    Steps_S = error_mm / 0.2
    // send step command to motor
    Stepper.Step(Steps_S)
end
  
```

We must consider whether this extra complexity is required or necessary in our system. Closed loop systems are not a magic bullet; they require careful modelling to predict system behaviour and a considered choice of parameters to prevent the system becoming unstable.

1.3 Block diagram representation of control loops



Control systems are often represented by block diagrams which show the information flow.



Flow path indicates the direction of data flow (from left to right). Values are maintained at a branch.

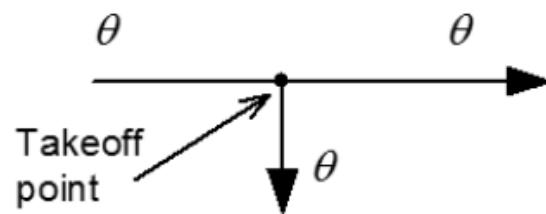


Figure 1.1: Flow path

Function block. The function acts on the input to produce the output. $x_{out} = f(x_{in})$

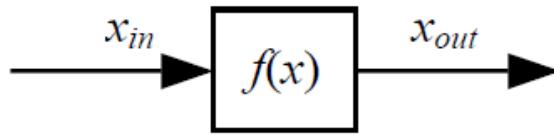


Figure 1.2: Function block

Comparator. The signals θ_1 and θ_2 are compared according to the signs (+ or -) and the result is θ_3 . They **must** have the same units. In this case $+\theta_1 - \theta_2 = \theta_3$.

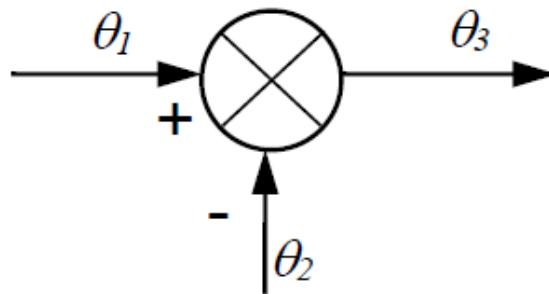
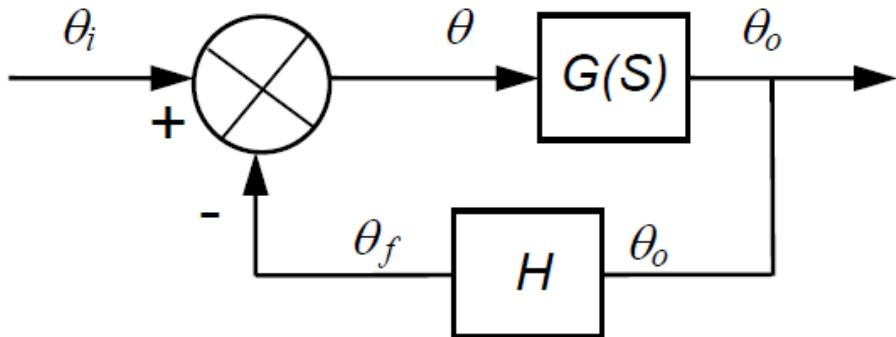


Figure 1.3: Comparator



1.4 Basic definitions/concepts

The following definitions are based on the standards of the IEEE (Institute of Electrical and Electronics Engineers)

- A **system** is an arrangement, set or collection of components connected or related in such a manner as to form an entirety or whole.
- A **control system** is an arrangement of physical components connected or related in such a manner as to command, direct or regulate itself or another

system. The components act together to perform a function not possible with any of the individual parts.

- The **input** is the stimulus or excitation applied to a system usually in order to produce a specified response.
- The **output** is the actual response obtained.
- An **open loop control system** is one in which the input is independent of the output.
- A **closed loop control system** is one in which the input is somehow dependent on the output.
- **Error signal (or actuating signal)** is the difference between the reference input and the feedback, in closed loop control systems it is this signal which is sent to the plant not the reference signal.

1.5 Notation

Large complex systems can be broken down into interconnected smaller ones and reduced further into a number of blocks.

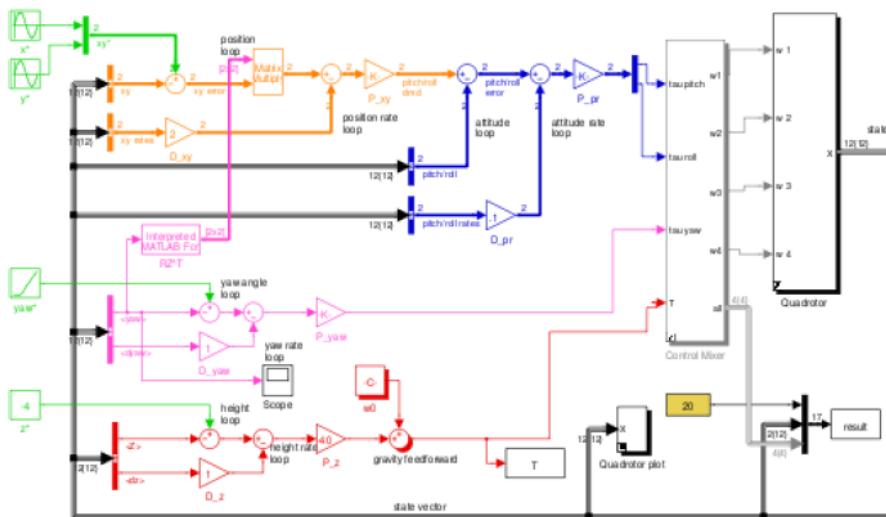
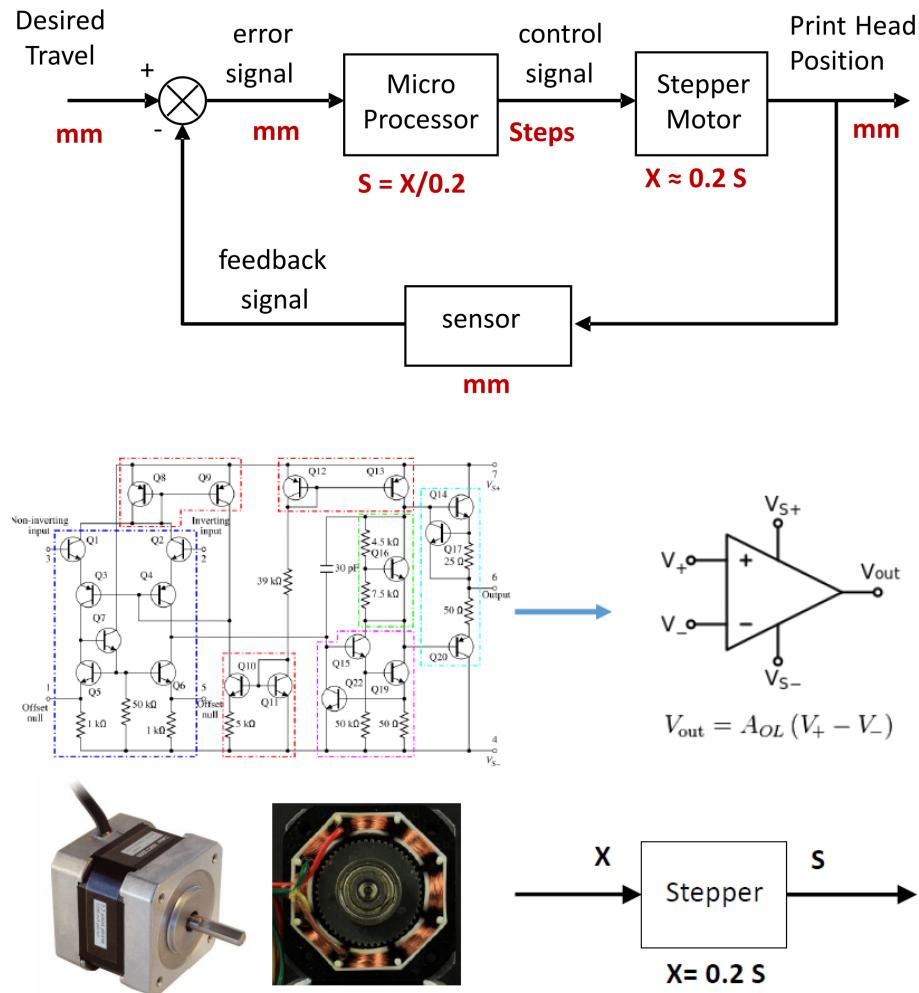


Figure 1.4: A quadcopter with lots of PID controllers running in parallel. Each colour represents a control loop for position, altitude, yaw etc.

Complex systems can be abstracted to a signal block if the behaviour can be adequately modelled. Consider modelling the stepper motor from the 3D printer, in reality a complex device, as a single block. This is analogous to representing a complex op-amp circuit as a single unit.



1.6 Control system design process

1. Establish control goals
2. Identify the variables to control
3. Write the specifications for the variables
4. Establish the system configuration and identify the actuators
5. Obtain a model of the process, the actuator and the sensor
6. Describe a controller and select the parameters to be adjusted
7. Optimise the parameters and analyse the performance

1.7 Summary

- Control systems are interconnected components which are configured to provide a desired response.

- Two broad categories of control systems: open loop (no feedback of output) and closed loop (feedback signal).
- Successful design of the controller requires consideration of the design goals, definition of the specifications, system definition, modelling and subsequent analysis. Controller design is an iterative process.

1.8 Tutorial

3D printer design choice

The positioning of the print head is in open loop. Considering the comparative advantages of open and closed loop control, what would have lead the designers to make this choice?

Open loop control is used where the observation of the output is not necessary/relevant to the control of the system. Considering the accuracies required for 3D printing, which is around $\pm 0.05\text{mm}$, we need to make sure that our print head moves to distances within this tolerance, otherwise our prints will be warped and misshaped. We must also consider that in an open system, small errors may lead to large errors over a certain period of operation, despite the print head moving distances within tolerance. There are also external factors affecting the belt and pulley system of the print head e.g. heat, dust, friction, load. These must all be within specification for the system to work properly.

Stepper motors are simple electronic components and can be manufactured to have a relationship between input and output. Hence, the need to observe whether the print head has moved a certain distance may be inconsequential. In applications where an extremely accurate print is required, a closed loop system would allow for a self calibrating system, which could allow the printer to work in different operational environments without the need to recalibrate or change components. However, this adds complexity to the controller and requires additional components to be added to the controller, increasing cost.

How might the designers overcome the problems mentioned without closed loop control?

Maintenance of the components in the system is important to make sure that everything is in the correct working order. When everything is kept clean and within the specification that the system is designed to work in, the controller will be able to work optimally. For example, calibrating the "0" position of the print head after cleaning the belts and pulleys from dust.

The designers may also implement a calibration operation into the controller before each phase. This may include implementing a "0" point into the hardware of the

rails that the print head is attached to. The controller would move the print head to this position simply by moving the maximum distance left and down. Using this as a reference for each print would eliminate systematic distance errors.

Washing machine

**Think of the control loops inside are they open or closed loop? Why?
What could the sensors be?**

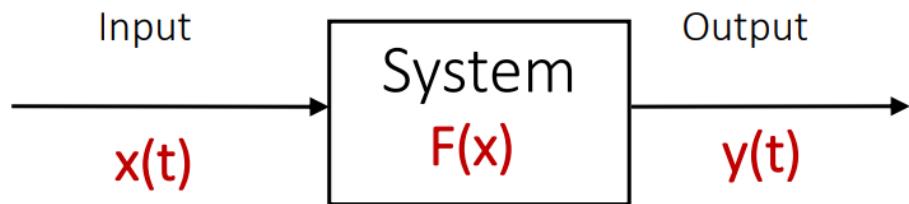
Let us consider the possible subsystems inside a washing machine.

- Timer - controls how long the washing machine operates. Open loop system. We only need a simple timing circuit with a minimum accuracy of around $\pm 1\text{min}$. Even if the washing machine runs for 85 minutes instead of 90 (some degree of inaccuracy), this is insignificant.
- Motor - this is to spin the tub. Open loop system. Motors are simple electronic components with a relationship between input and output. An error signal provides no useful information for the operation of the motor.
- Temperature control - controls how hot or cold the water in the washing machine is. Open loop system. Most washing machines simply have a hot water connection for hot water and cold water. A solenoid valve can control the proportion of hot water to cold water that enters the tub. For washing machines with a heating element, the amount of power put through the coil will determine the final temperature of the water (relationship between input and output is known). Adding a thermistor to measure the output water temperature may be possible but ultimately unnecessary as the tolerance for the temperature can be quite high ($\pm 5^\circ\text{C}$)
- Water valve control - controls how much water enters and leaves the drum. Closed loop control. The level of water in the drum needs to only be measured when the drum is full of water. Knowing the level of water at any given moment is unnecessary, thus using a simple pressure switch, will be sufficient in letting the controller know that the tub has reached capacity and close the input valve. After a set amount of time, the drain valve may open and since the amount of water in the tub when it is full is known, the drain time can simply be calculated as a function of its volume. This means that we do not need to check whether the water has drained from the tub.

Chapter 2

Modelling Linear Systems

Previously, we have considered systems in the general sense, with a generic function block approach.



Now we will investigate how we model physical systems and obtain the function block $F(x)$ for electrical and mechanical components.

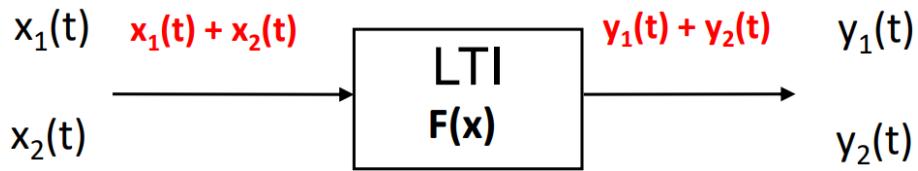
2.1 Linear Term Invariant (LTI) systems

The focus of this course and indeed much of control theory itself focuses on modelling physical systems as **linear** and **time invariant**. LTI systems have three key properties:

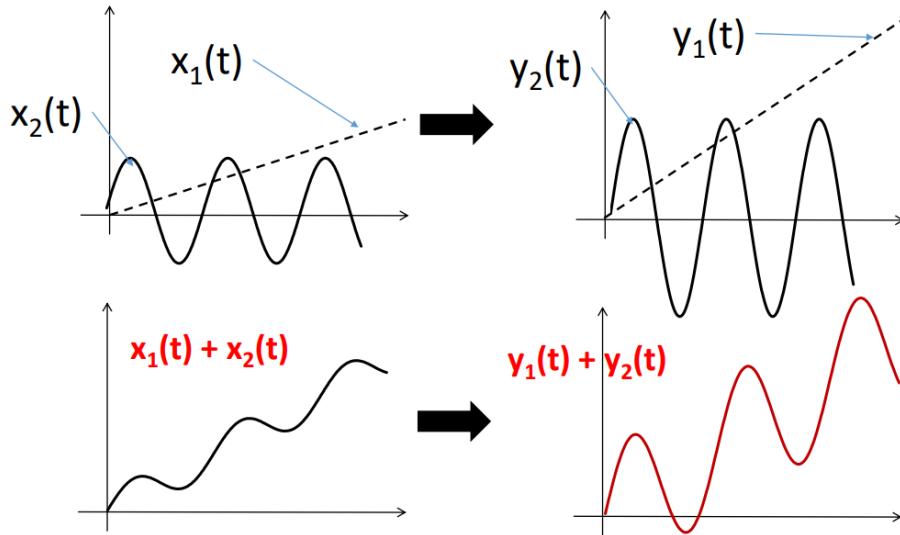
- Obey principle of superposition
- Homogeneity
- Time invariance

2.1.1 Superposition

If input $x_1(t)$ produces output $y_1(t)$ and input $x_2(t)$ produces $y_2(t)$, then input $x_1(t) + x_2(t)$ produces output $y_1(t) + y_2(t)$

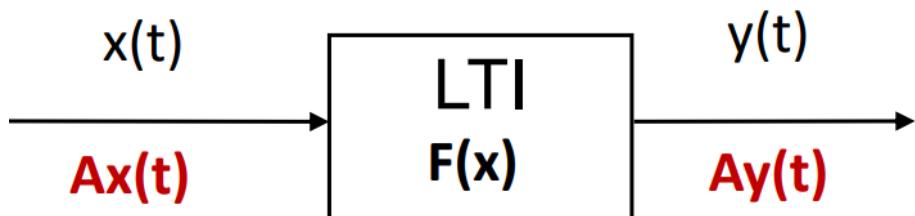


Say for a system which doubles the input $F(x) = 2x$

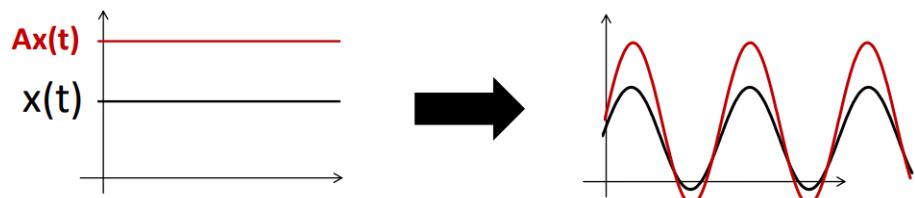


2.1.2 Homogeneity

If the input to the system $x(t)$ is scaled by a magnitude scale factor A , then the output $y(t)$ is also scaled by the same factor.

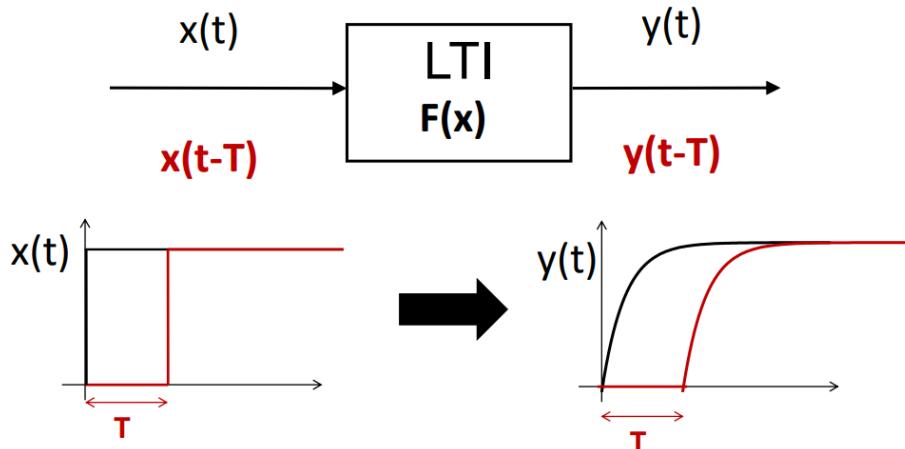


For example, consider a system which generates a sine wave at a given amplitude, with a set frequency:



2.1.3 Time invariance

If input is applied at time $t = 0$ or T s from now, the output is identical with the exception of a delay of T s.



Are these models suitable for physical systems?

These three requirements, whilst simple, are so stringent that **almost no physical LTI system truly exists**. Consider a car engine - the performance deteriorates over time, to stretch it further, would you expect a system to give the same output after a time delay T of 10 years? Even simple systems such as a resistor in an electrical circuit have non-linearities - a scaling factor A could be chosen for $x(t)$ which would mean too much current flows and the resistor melts.

Most practical systems are not linear, but often we can assume they behave linearly **under certain conditions/assumptions**. Linear systems are **much** easier to solve! There are **analytic** solutions with standard tools used to solve the equations. Whereas for non-linear problems it is often necessary to solve them numerically.

2.1.4 Linearisation: Example 1

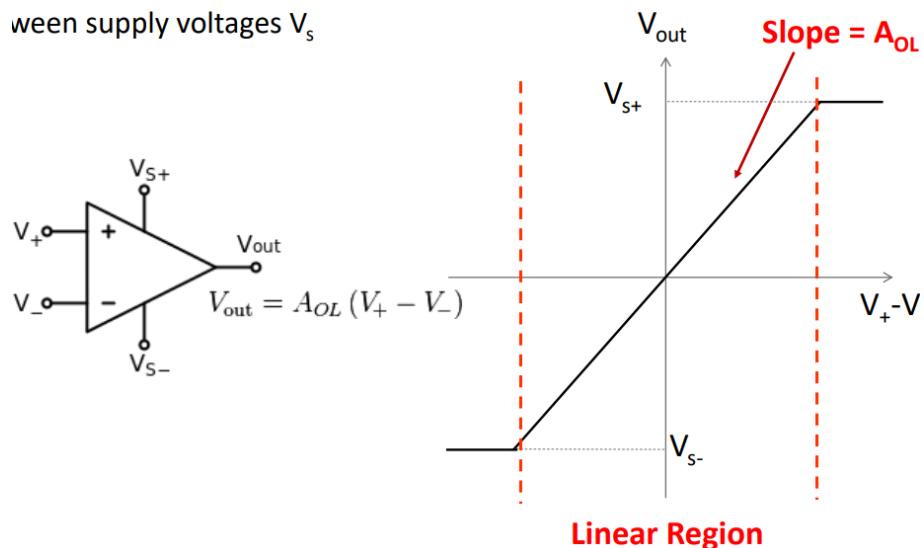
For a simple system such as a spring, across all possible compressions or extensions the response is non-linear:

$$F = -kX \quad (2.1)$$

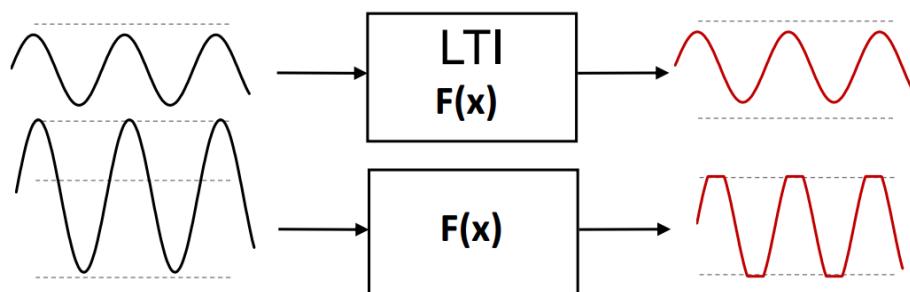
Hooke's law is only a linear approximation of the true response. However, if we choose the operating range of the spring correctly, the response is within the linear region. This approximation is **valid**.

Linearisation: Example 2

Similarly an op-amp has a **linear region** where the output signals are between supply voltages V_S



Keep signals within these regions and the linear assumption holds:

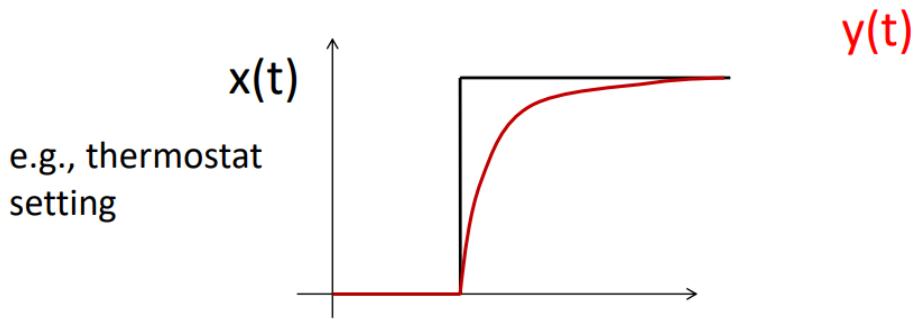


Add LTI example

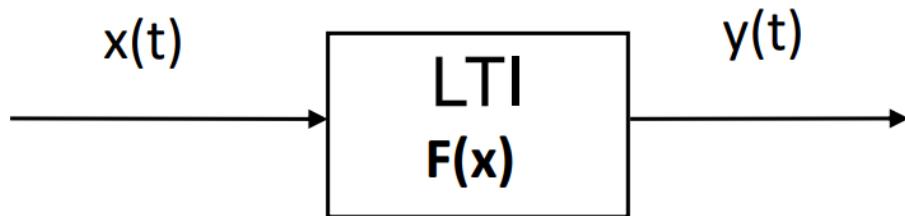
2.2 Dynamic systems - Laplace Transform

2.2.1 Dynamic systems as ODEs

Ideal systems would respond **instantaneously** to inputs, however real world systems require some time to adjust to changes and are thus known as **dynamic systems** as the output changes over time.



As we are interested in describing something that **changes** with time, it is useful to express the function block of the system $F(t)$ as an ordinary differential equation (ODE).



$$\frac{d^n y}{dt^n} + a_{n-1} \frac{d^{n-1} y}{dt^{n-1}} + \dots + a_2 \frac{d^2 y}{dt^2} + a_1 \frac{dy}{dt} + a_0 = bx \quad (2.2)$$

- x is input function or forcing function
- y is output
- n is **order** of the ODE
- $a_0\dots$ are coefficients. These **completely characterise the system**

2.2.2 Laplace Transforms

Because of our **linear assumptions** we can use Laplace transforms to simplify solving the ODEs. The Laplace transforms of a signal (function) x is the function $X = \mathcal{L}(x)$ defined by

$$X(s) = \int_{0^-}^{\infty} x(t)e^{-st} dt \quad (2.3)$$

For those $s \in \mathbf{C}$ for which the integral makes sense.

- X is a complex-valued function of complex numbers
- s is called the (complex) **frequency variable** with units s^{-1} , t is called the **time variable** (in sec); st is unitless

- $s = \sigma + j\omega$

As we shall see:

- Differential operators are replaced with algebraic variables
- Algebraic equations are much easier to manipulate & solve
- Standard forms exist for many physical systems

Laplace Transforms: Example 1

Let's find Laplace transform $x(t) = e^t$:

$$X(e^t) = \int_{0^-}^{\infty} e^t e^{-st} dt \quad (2.4)$$

$$X(e^t) = \int_{0^-}^{\infty} e^{(1-s)t} dt \quad (2.5)$$

$$Xe^t = \frac{1}{1-s} e^{(1-s)t} \Big|_{0^-}^{\infty} \quad (2.6)$$

$$X(e^t) = \frac{1}{1-s} \times 0 - \frac{1}{1-s} \times 1 = \frac{1}{s-1} \quad (2.7)$$

Laplace Transforms: Example 2

Constant or **unit step** $x(t) = 1$ (for $t \geq 0$)

$$X(s) = \int_{0^-}^{\infty} 1 e^{-st} dt \quad (2.8)$$

$$X(s) = \int = -\frac{1}{s} e^{-st} \Big|_{0^-}^{\infty} \quad (2.9)$$

$$X(s) = -\frac{1}{s} \times 0 - \left(-\frac{1}{s}\right) \times 1 = \frac{1}{s} \quad (2.10)$$

Laplace Transforms: Example 3

Sinusoid: first express $x(t) = \cos(\omega t)$ as:

$$x(t) = \frac{1}{2} e^{i\omega t} + \frac{1}{2} e^{-i\omega t} \text{ (Euler's formula)} \quad (2.11)$$

$$X(s) = \int_{0^-}^{\infty} e^{-st} \left(\frac{1}{2} e^{i\omega t} + \frac{1}{2} e^{-i\omega t} \right) dt \quad (2.12)$$

$$X(s) = \frac{1}{2} \int_{0^-}^{\infty} e^{(-s+i\omega)t} dt + \frac{1}{2} \int_{0^-}^{\infty} e^{(-s-i\omega)t} dt \quad (2.13)$$

$$X(s) = \frac{1}{2} \frac{1}{s - i\omega} + \frac{1}{2} \frac{1}{s + i\omega} = \frac{s}{s^2 + \omega^2} \quad (2.14)$$

You can look up these transforms in a table. The Laplace variable, s , can be considered to represent the differential operator (very useful for control engineering):

$$s \equiv \frac{d}{dt} \quad (2.15)$$

$$\frac{1}{s} \equiv \int_{0^-}^{\infty} dt \quad (2.16)$$

Laplace transform of time derivative $\frac{dx}{dt}$:

$$L \left\{ \frac{dx}{dt} \right\} = \int_{0^-}^{\infty} \frac{dx}{dt} e^{-st} dt \quad (2.17)$$

Integrating by parts

$$L \left\{ \frac{dx}{dt} \right\} = s \int_{0^-}^{\infty} x(t) e^{-st} dt + [x(t)e^{-st}]_{0^-}^{\infty} \quad (2.18)$$

The initial condition $x(0^-)$ is often zero in practice

$$L \left\{ \frac{dx}{dt} \right\} = sX(s) - x(0^-) = sX(s) \quad (2.19)$$

We can substitute this result to solve higher order derivatives:

$$L \left\{ \frac{d^2 x}{dt^2} \right\} = sL \left\{ \frac{dx}{dt} \right\} - \frac{d^2 x}{dt^2}(0^-) \quad (2.20)$$

$$L \left\{ \frac{d^2 x}{dt^2} \right\} = s^2 X(s) - sx(0^-) - \frac{d^2 x}{dt^2}(0^-) = s^2 X(s) \quad (2.21)$$

So more generally, with all initial conditions set to zero:

$$L \left\{ \frac{d^n x}{dt^n} \right\} = s^n X(s) \quad (2.22)$$

2.2.3 Transfer Functions

After we have taken the Laplace transform of the differential equation of a system, it's useful to rearrange to give the system **output** as the product of the system **input** and the system **transfer function**.

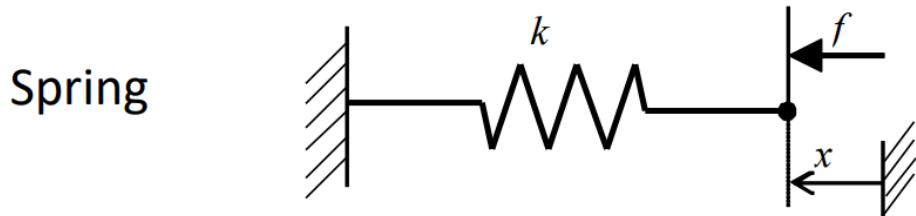


The transfer function of a linear system is defined as the **ratio** of the Laplace transform of the output variable to the Laplace transform of the input variable, with all the initial conditions assumed to be zero.

2.2.4 Transfer Functions of Mechanical Components

Spring

Convention is input force, output displacement. Balance **forces**



Time domain equation (Hooke's Law):

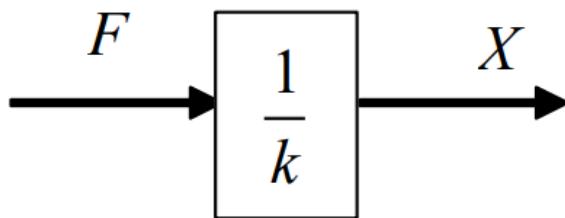
$$f(x) = kx \quad (2.23)$$

k is stiffness in N m^{-1} . Laplace domain equation:

$$F(s) = kX(s) \quad (2.24)$$

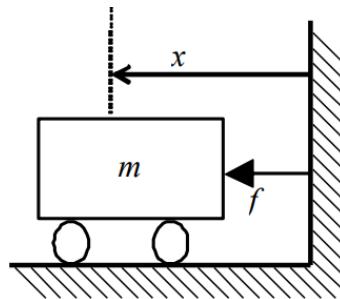
Transfer function:

$$G(s) = \frac{X(s)}{F(s)} = \frac{1}{k} \quad (2.25)$$



Mass

Inertial load - mass



Time domain equation from Newton's 2nd law of motion

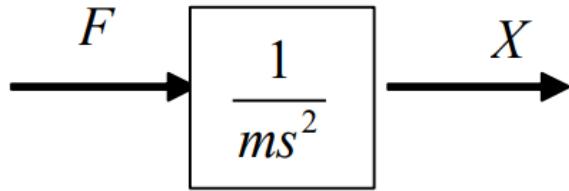
$$f(t) = m \frac{d^2 x(t)}{dt^2} \quad (2.26)$$

Laplace domain equation

$$F(s) = ms^2 X(s) \quad (2.27)$$

Transfer function

$$\frac{X(s)}{F(s)} = \frac{1}{ms^2} \quad (2.28)$$



Damper

Below is a dashpot - a viscous damper. They resist motion through friction. The damping coefficient is in terms of c with units $\text{N m}^{-1} \text{s}^{-1}$. Time domain equation:

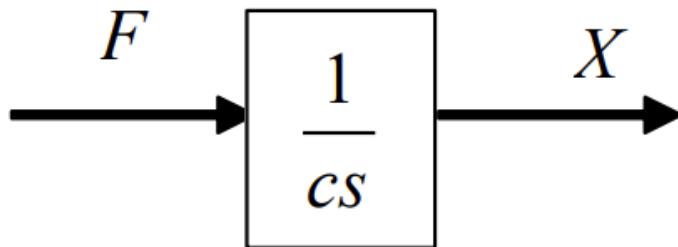
$$f(x) = x \frac{dx}{dt} = csx \quad (2.29)$$

Laplace domain equation:

$$F(s) = csX(s) \quad (2.30)$$

Transfer function

$$\frac{X(s)}{F(s)} = \frac{1}{cs} \quad (2.31)$$



2.2.5 Combining components

Springs

Springs in parallel

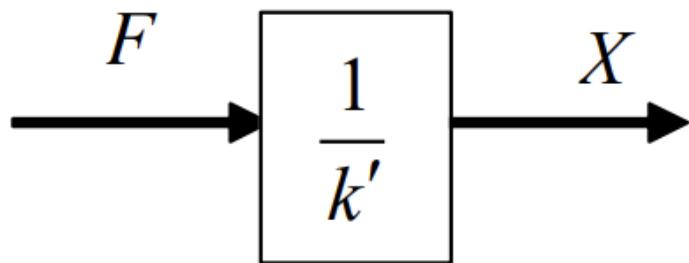
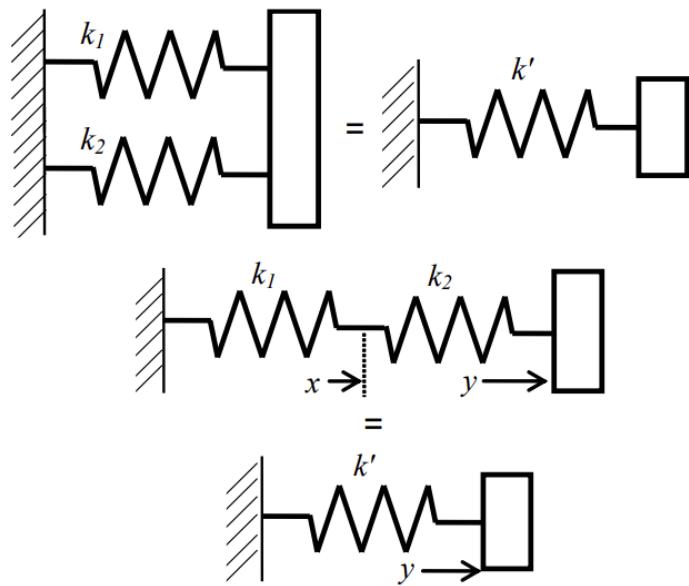
$$k' = k_1 + k_2 \quad (2.32)$$

$$G(s) = \frac{X(s)}{F(s)} = \frac{1}{k'} \quad (2.33)$$

Springs in series

$$\frac{1}{k'} = \frac{1}{k_1} + \frac{1}{k_2} \quad (2.34)$$

$$G(s) = \frac{X(s)}{F(s)} = \frac{1}{k'} \quad (2.35)$$



Dashpots behave in a similar way to springs. Parallel:

$$c' = c_1 + c_2 \quad (2.36)$$

and in series:

$$\frac{1}{c'} = \frac{1}{c_1} + \frac{1}{c_2} \quad (2.37)$$

Resistor

Convention is input voltage, output current. Balance voltages. Ohms law is \$V = IR\$. Time domain:

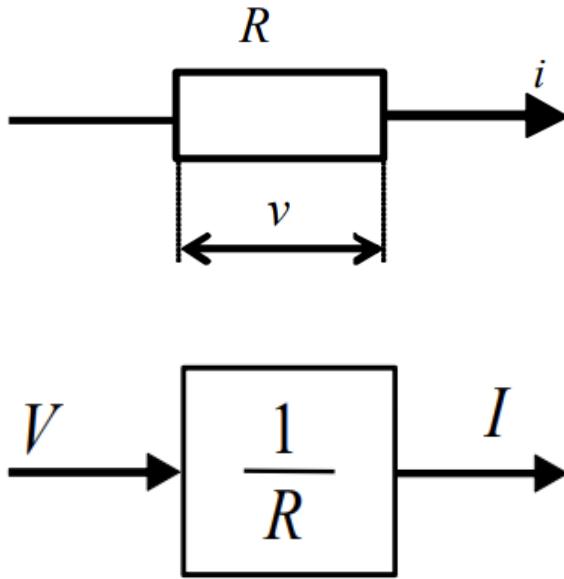
$$v(t) = i(t)R \quad (2.38)$$

Laplace domain:

$$V(s) = I(s)R \quad (2.39)$$

Transfer function:

$$G(s) = \frac{I(s)}{V(s)} = \frac{1}{R} \quad (2.40)$$



Capacitor

Either definition of current/voltage relationship gives same result. Time domain:

$$i(t) = C \frac{dv}{dt} \quad (2.41)$$

$$v(t) = \frac{1}{C} \int i(t) dt \quad (2.42)$$

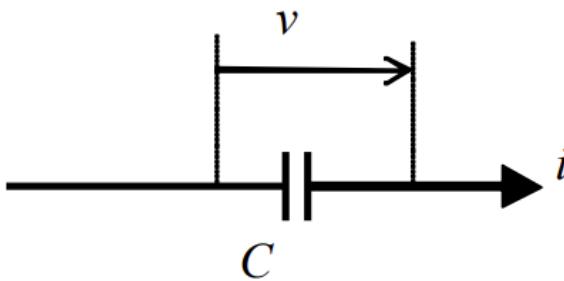
Laplace domain:

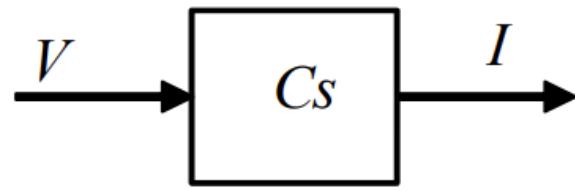
$$I(s) = CsV(s) \quad (2.43)$$

$$V(s) = \frac{1}{C} \frac{1}{s} I(s) \quad (2.44)$$

Transfer function:

$$G(s) = \frac{I(s)}{V(s)} = Cs \quad (2.45)$$





Inductor

An inductor resists changes of current by generating a voltage in opposition via magnetic induction. From Faraday's Law:

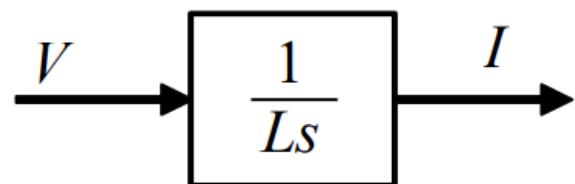
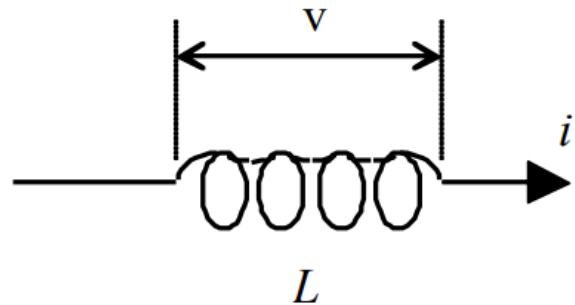
$$v(t) = L \frac{di}{dt} \quad (2.46)$$

Laplace domain

$$V(s) = LsI(s) \quad (2.47)$$

Transfer function:

$$G(s) = \frac{I(s)}{V(s)} = \frac{1}{Ls} \quad (2.48)$$



Chapter 3

First and Second Order Systems

19/10/2020

3.1 First Order Systems

All first order systems i.e. those with only $\frac{dx}{dt}$ take the following “standard” forms:

$$\frac{X(s)}{Y(s)} = \frac{\alpha}{1 + Ts} = \frac{\gamma}{1 + \tau s} \quad (3.1)$$

Where:

- τ, γ is the **gain**
- T, τ is the **time constant**

This function is commonly known as an exponential time delay, or lag.

3.1.1 Returning to the Time Domain

To get the response of the system in the time domain we need to convert the transfer function from the Laplace domain. Tables of common inverse Laplace functions have already been created; all we need to do arrange our transfer function into one of these standard forms.

Laplace transforms – Table			
$f(t) = L^{-1}\{F(s)\}$	$F(s)$	$f(t) = L^{-1}\{F(s)\}$	$F(s)$
$a \quad t \geq 0$	$\frac{a}{s} \quad s > 0$	$\sin \omega t$	$\frac{\omega}{s^2 + \omega^2}$
$at \quad t \geq 0$	$\frac{a}{s^2}$	$\cos \omega t$	$\frac{s}{s^2 + \omega^2}$
e^{-at}	$\frac{1}{s + a}$	$\sin(\omega t + \theta)$	$\frac{s \sin \theta + \omega \cos \theta}{s^2 + \omega^2}$
te^{-at}	$\frac{1}{(s + a)^2}$	$\cos(\omega t + \theta)$	$\frac{s \cos \theta - \omega \sin \theta}{s^2 + \omega^2}$
$\frac{1}{2}t^2e^{-at}$	$\frac{1}{(s + a)^3}$	$t \sin \omega t$	$\frac{2\omega s}{(s^2 + \omega^2)^2}$
$\frac{1}{(n-1)!}t^{n-1}e^{-at}$	$\frac{1}{(s + a)^n}$	$t \cos \omega t$	$\frac{s^2 - \omega^2}{(s^2 + \omega^2)^2}$
e^{at}	$\frac{1}{s - a} \quad s > a$	$\sinh \omega t$	$\frac{\omega}{s^2 - \omega^2} \quad s > \omega $
te^{at}	$\frac{1}{(s - a)^2}$	$\cosh \omega t$	$\frac{s}{s^2 - \omega^2} \quad s > \omega $
$\frac{1}{b-a}(e^{-at} - e^{-bt})$	$\frac{1}{(s + a)(s + b)}$	$e^{-at} \sin \omega t$	$\frac{\omega}{(s + a)^2 + \omega^2}$
$\frac{1}{a^2}[1 - e^{-at}(1 + at)]$	$\frac{1}{s(s + a)^2}$	$e^{-at} \cos \omega t$	$\frac{s + a}{(s + a)^2 + \omega^2}$
t^n	$\frac{n!}{s^{n+1}} \quad n = 1, 2, 3..$	$e^{at} \sin \omega t$	$\frac{\omega}{(s - a)^2 + \omega^2}$
$t^n e^{at}$	$\frac{n!}{(s - a)^{n+1}} \quad s > a$	$e^{at} \cos \omega t$	$\frac{s - a}{(s - a)^2 + \omega^2}$
$t^n e^{-at}$	$\frac{n!}{(s + a)^{n+1}} \quad s > a$	$1 - e^{-at}$	$\frac{a}{s(s + a)}$
\sqrt{t}	$\frac{\sqrt{\pi}}{2s^{3/2}}$	$\frac{1}{a^2}(at - 1 + e^{-at})$	$\frac{1}{s^2(s + a)}$
$\frac{1}{\sqrt{t}}$	$\sqrt{\frac{\pi}{s}} \quad s > 0$	$f(t - t_1)$	$e^{-t_1 s} F(s)$
$g(t) \cdot p(t)$	$G(s) \cdot P(s)$	$f_1(t) \pm f_2(t)$	$F_1(s) \pm F_2(s)$
$\int f(t)dt$	$\frac{F(s)}{s} + \frac{f^{-1}(0)}{s}$	$\delta(t)$ unit impulse	1 all s
$\frac{df}{dt}$	$sF(s) - f(0)$	$\frac{d^2f}{df^2}$	$s^2 F(s) - sf(0) - f'(0)$
$\frac{d^n f}{dt^n}$	$s^n F(s) - s^{n-1}f(0) - s^{n-2}f'(0) - s^{n-3}f''(0) - \dots - f^{n-1}(0)$		

Figure 3.1: Laplace Transformation Table

Example: Parallel Spring & Damper (Laplace Domain found previously)

$$\frac{X(s)}{Y(s)} = \frac{\alpha}{1 + Ts} = \frac{\gamma}{1 + \tau s} \quad (3.2)$$

Looking at the table, we can find the relevant inverse function; in this case (3rd from the top):

$$a = \frac{1}{\tau} \quad (3.3)$$

Hence:

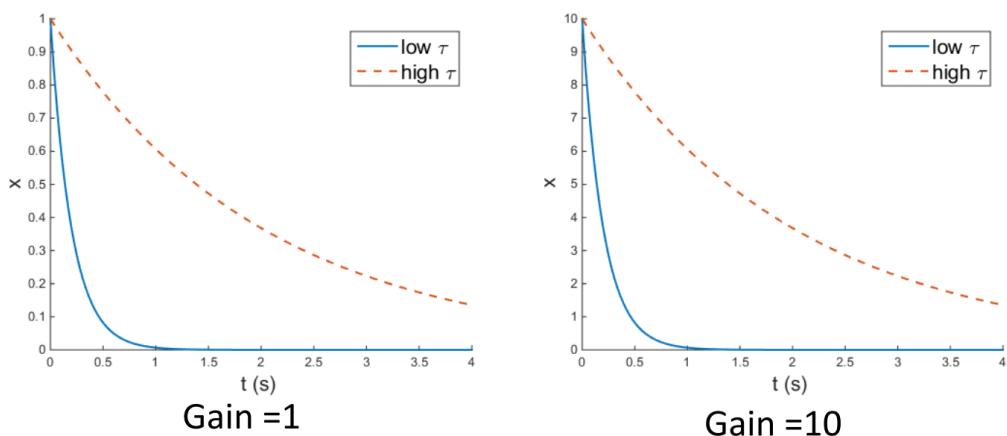
$$\frac{X(s)}{Y(s)} = \frac{\frac{\gamma}{\tau}}{\frac{1}{\tau} + s} = \frac{\gamma}{\tau} \frac{1}{\frac{1}{\tau} + s} \quad (3.4)$$

$$x(t) = L^{-1} \left\{ \frac{\gamma}{\tau} \frac{1}{\frac{1}{\tau} + s} \right\} \quad (3.5)$$

$$= \frac{\gamma}{\tau} L^{-1} \left\{ \frac{1}{\frac{1}{\tau} + s} \right\} \quad (3.6)$$

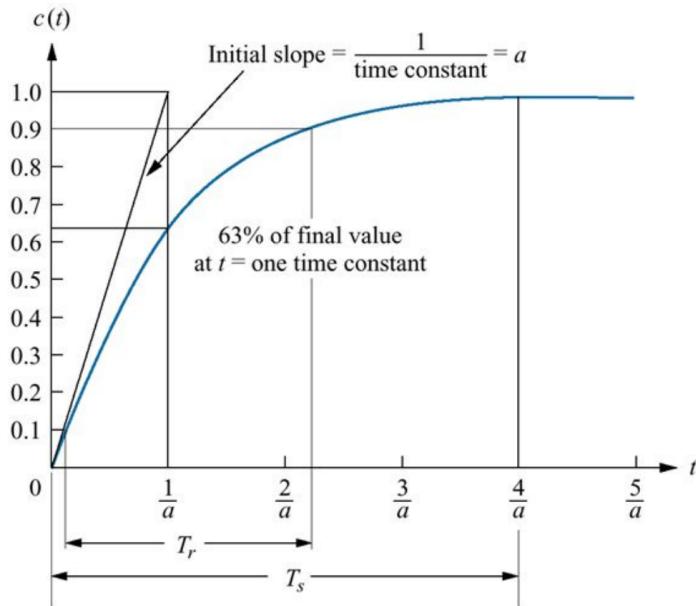
$$= \frac{\gamma}{\tau} e^{-\frac{t}{\tau}} \quad (3.7)$$

This is an exponential decay – the only variable needed to define the system is the time constant. The gain merely scales the response.



Having a negative time constant makes no physical sense, so the exception of $T = 0$, the response will always be an exponential decay.

3.1.2 Unit Step Response



If a step input is given to a first order system, this will result in an exponential increase. This might be something like changing the room temperature. When talking about unit step input, the amplitude of the input will vary from 0 to 1. Some important definitions are:

- Initial slope = $\frac{1}{\text{time constant}} = a$
- T_r = Raising time = how quick the exponential function achieves 90% of the final value
- τ = Time constant = how quick the exponential function achieves 63% of the final value
- T_s = Settling time = how quick the exponential function becomes constant

Remember, we have modelled mechanical and electrical systems, and arrived at the same equations. This could be biological, financial, meteorological... This is good, because you only need to understand one system to understand hundreds. However, this means the models can appear abstract (or possibly arbitrary), so it is important to keep in mind the physical systems we are trying to model.

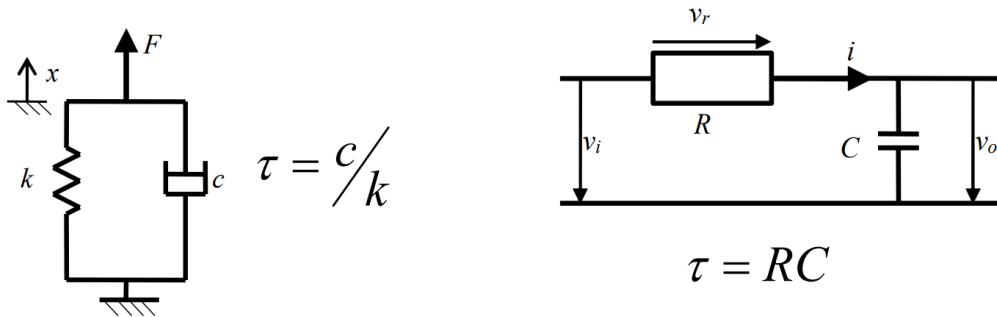


Figure 3.2: Left: Rate of decay determined by ratio of spring return force to viscous friction — Right: Rate of decay determined by total impedance of R and C

3.2 Second Order Systems

The standard form for second order systems is shown below:

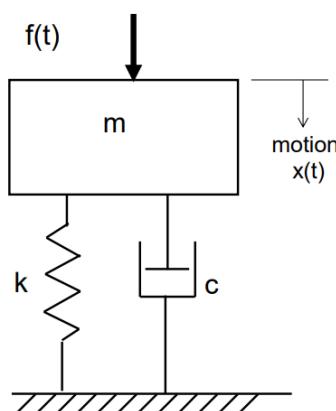
$$G(s) = \gamma \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad (3.8)$$

Where:

- γ is the **gain**
- ω_n is the **natural frequency**
- ζ is the **damping ratio**

This function is known as a damped oscillator, in that it produces harmonic sinusoidal oscillations which decay over time. This type of system appears everywhere in physics, as well as in engineering. Even to the extent that some higher order systems are simplified to become second order, just because it is so well understood.

Example – Mass Spring Damper



Consider a simple mechanical system Mass/Spring/Damper (MSD). As before, we wish to relate the input force to the output displacement i.e. Transfer function desired:

$$G(s) = \frac{X(s)}{F(s)} \quad (3.9)$$

Balancing forces as function of time:

$$f(t) = f_S(t) + f_D(t) + f_I(t) \quad (3.10)$$

Where:

- $f_S(t) = kx(t)$

- $f_D(t) = c \frac{dx(t)}{dt}$

- $f_I(t) = m \frac{d^2 x(t)}{dt^2}$

So time domain force equilibrium is:

$$f(t) = kx(t) + c \frac{dx(t)}{dt} + m \frac{d^2 x(t)}{dt^2} \quad (3.11)$$

Converting to Laplace domain:

$$F(s) = kX(s) + csX(s) + ms^2X(s) \quad (3.12)$$

$$F(s) = X(s)(k + cs + ms^2) \quad (3.13)$$

$$G(s) = \frac{X(s)}{F(s)} = \frac{1}{ms^2 + cs + k} \quad (3.14)$$

To obtain the Laplace domain in the "standard" form, the coefficient of the highest order of s on the denominator should be 1:

$$G(s) = \frac{\frac{k}{m}}{s^2 + s\frac{c}{m} + \frac{k}{m}} \frac{1}{k} \quad (3.15)$$

$$= \gamma \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad (3.16)$$

Where:

- $\omega_n = \sqrt{\frac{k}{m}}$

- $\zeta = \frac{c}{2\sqrt{km}}$

- $\gamma = \frac{1}{k}$

3.2.1 Returning to the Time Domain

As before, we use the inverse Laplace transform to get the time domain response. To reiterate, the benefit of standard forms is that the transforms are given in the tables:

$$G(s) = \gamma \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad (3.17)$$

$$x(t) = L^{-1} \left\{ \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \right\} \quad (3.18)$$

$$\frac{x(t)}{x(0)} = \gamma \frac{\omega_n}{\sqrt{1 - \zeta^2}} e^{-\zeta\omega_n t} \sin(\omega_n \sqrt{1 - \zeta^2} \cdot t) \quad (3.19)$$

γ is not shown as it is unaffected by the Laplace transform. The equation above looks complicated, but investigating each term yields:

$$\gamma \frac{\omega_n}{\sqrt{1 - \zeta^2}} \quad (3.20)$$

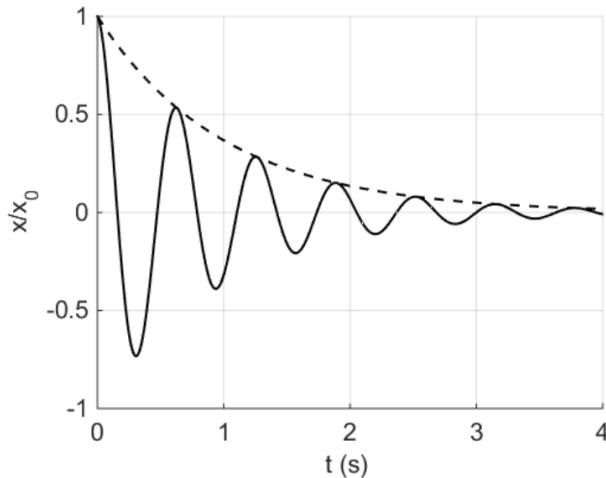
$\gamma \omega_n \zeta$ are all constants, so the whole term is just a number.

$$e^{-\zeta\omega_n t} \quad (3.21)$$

It is an exponential function, which depending on whether $-\zeta\omega_n$ is positive or negative, increases or decays.

$$\sin(\omega_n \sqrt{1 - \zeta^2} \cdot t) \quad (3.22)$$

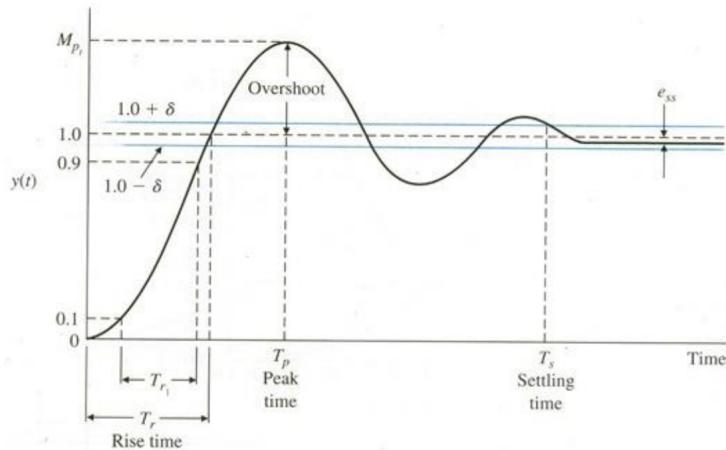
$\omega_n \zeta$ are just constants, so assuming $\zeta < 1$, this is just a sine wave.



This is an exponentially decaying sinusoidal oscillation, with:

- Frequency: $\omega_n \sqrt{1 - \zeta^2} \cdot t$
- Decay: $e^{-\zeta\omega_n t}$
- Gain: $\gamma \frac{\omega_n}{\sqrt{1 - \zeta^2}}$

3.2.2 Unit Step Response

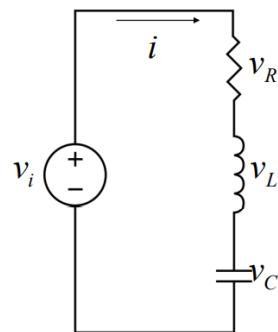


Some properties that should be known, but will be investigated later are:

- Slope at the time step input is made ($t = 0$) = 0
- T_p = Peak Time
- T_s = Settling Time
- T_r = Rise Time
- Overshoot

3.3 Tutorial

LRC Filter



A more complex filtering circuit is an LRC filter, also known as a harmonic oscillator circuit, which allows for a narrower range of frequencies to be amplified or attenuated. These circuits are used in analogue radios with a variable capacitor or

inductor connected to the dial to select the frequency. To describe the relationship between the input voltage V_i , and the voltage across the capacitor V_C :

$$G(s) = \frac{V_C(s)}{V_i(s)} \quad (3.23)$$

Balancing voltages using Kirchoff's Law:

$$v_i = v_R + v_L + v_C \quad (3.24)$$

Where:

- $v_r(t) = i(t)R$
- $v_c(t) = \frac{1}{C} \int i(t) dt$
- $v_L(t) = L \frac{di(t)}{dt}$

So time domain voltage is:

$$v_i(t) = i(t)R + \frac{1}{C} \int i(t) dt + L \frac{di(t)}{dt} \quad (3.25)$$

Converting to Laplace domain:

$$V_i(s) = I(s)R + \frac{1}{Cs} I(s) + LsI(s) \quad (3.26)$$

$$V_i(s) = I(s) \left(R + \frac{1}{Cs} + Ls \right) \quad (3.27)$$

Getting the laplace domain and rearranging the $v_c(t)$ term yields:

$$v_c(t) = \frac{1}{C} \int i(t) dt \quad (3.28)$$

$$V_c(s) = \frac{1}{Cs} I(s) \quad (3.29)$$

$$I(s) = V_C(s)Cs \quad (3.30)$$

Inputting equation (3.30) into equation (3.27):

$$V_i(s) = CsV_c(s) \left(R + \frac{1}{Cs} + Ls \right) \quad (3.31)$$

$$G(s) = \frac{V_C(s)}{V_i(s)} = \frac{1}{LCs^2 + CRs + 1} \quad (3.32)$$

To obtain the Laplace domain in the "standard form":

$$G(s) = \frac{\frac{1}{LC}}{s^2 + s\frac{R}{L} + \frac{1}{LC}} \quad (3.33)$$

$$= \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad (3.34)$$

Where:

- $\omega_n = \sqrt{\frac{1}{LC}}$

- $\zeta = \frac{R}{2} \sqrt{\frac{C}{L}}$

- $\gamma = 1$

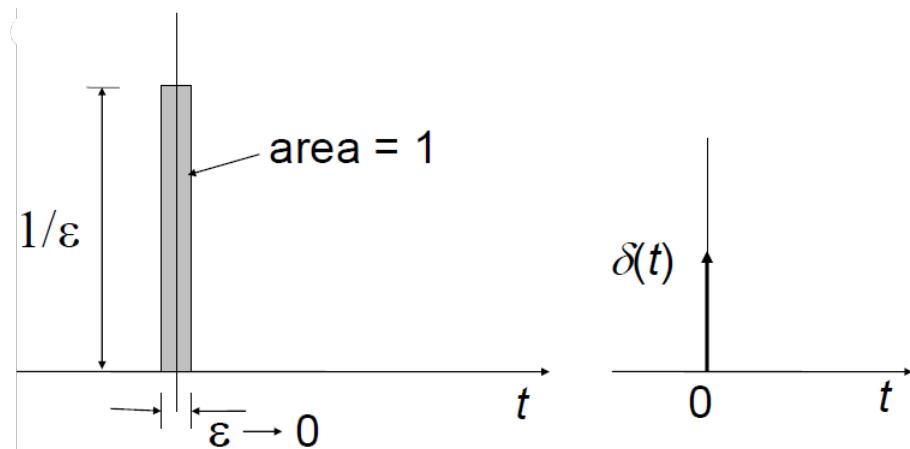
Chapter 4

System Time Response - 1st order

4.1 Impulse functions/responses

4.1.1 Impulse response of a system: Dirac delta function

A useful tool in analysing the transient response of a system is the impulse signal, a unit (amplitude = 1) pulse infinitesimally small, with area = 1. Formally this is known as the **Dirac delta function (impulse function)**.



The Dirac delta function is a non-physical, singularity function with the following definition:

$$\delta(t) = \begin{cases} 0 & \text{for } t \neq 0 \\ \text{undefined} & \text{for } t = 0 \end{cases} \quad (4.1)$$

but with the requirement that

$$\int_{-\infty}^{\infty} \delta(t) dt = 1 \quad (4.2)$$

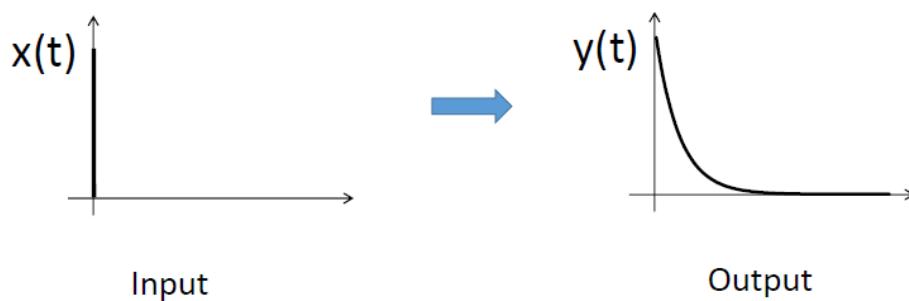
so taking the Laplace transform of this is also just 1

$$\mathcal{L}(\delta(t)) = \int_{0^-}^{\infty} \delta(t)e^{-st} dt = 1 \quad (4.3)$$

Thus, the impulse response of the system is equal to the transfer function and from this it can be shown that **any** arbitrary signal can be described as a summation of impulse responses.

4.1.2 Impulse response on a first order system

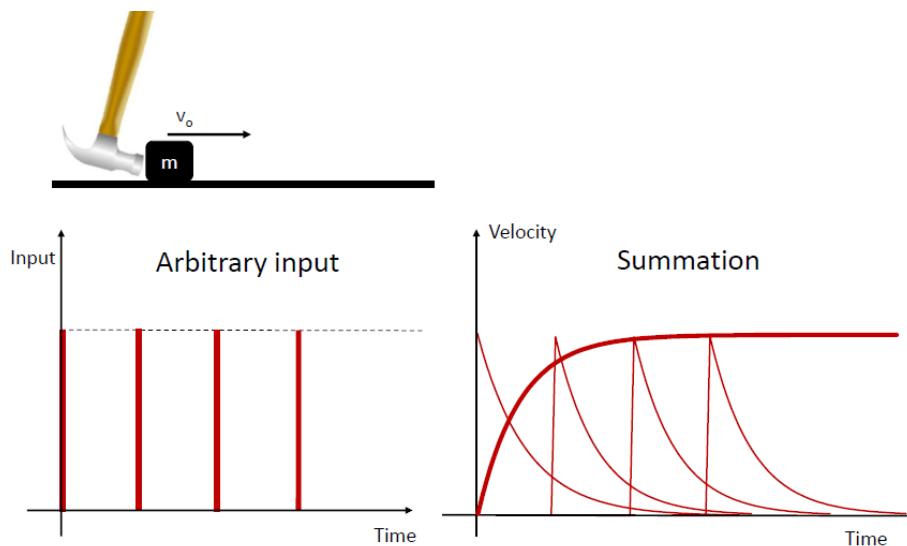
Take a first order response for example, the transfer function and thus the impulse response looks like this:



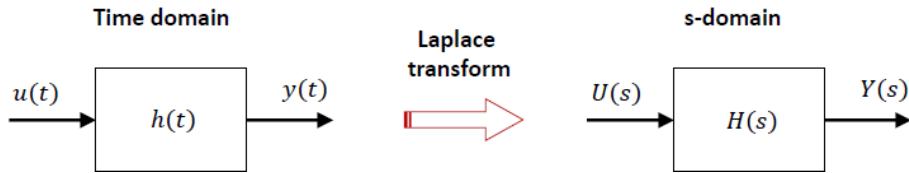
Due to our LTI assumptions:

- Scaling the input scales the output
- Superposition of inputs equals superposition of outputs
- Time invariance

4.1.3 Impulse response of a system



Time vs Frequency Domain



- u is the impulse function to the system
- h is called the impulse response of the system
- H is called the transfer function (TF) of the system

$$y(t) = \int_0^{\infty} h(\tau)u(t - \tau) d\tau = \int_0^{\infty} h(\tau - t)u(\tau) d\tau \quad (4.4)$$

with $0 \leq \tau \leq t$

$$y(t) = h(t) \cdot u(t) \quad (4.5)$$

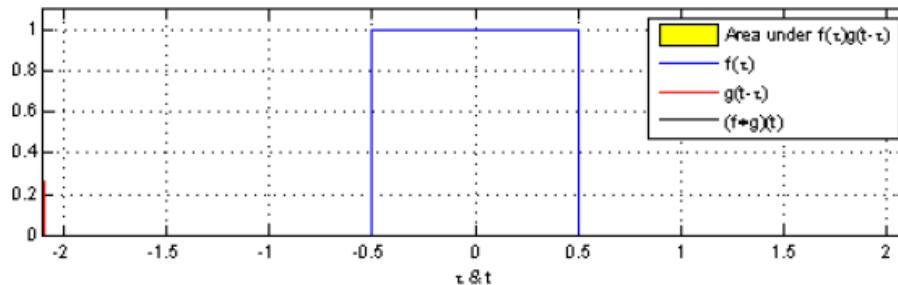
This is called convolution.

$$Y(s) = H(s) \cdot U(s) \quad (4.6)$$

This is called multiplication.

Convolution example

Essentially, the steps for convolving two signals are to first reflect the signal g , then offset the reflected signal. Then calculate the area under the graph for every offset, by sliding $-g$. The convolution at each time point is equal to the area under the intersection of functions. For two pulses, the result is a triangle wave:



However, the calculations to obtain this result in the time domain are complicated, but are only multiplication in the Laplace domain.

Getting the Time Response

The procedure to describe the time response for LTI systems is thus:

- Express the input, $u(t)$, in Laplace notation, $U(s)$
- Use this to find output $Y(s)$, usually by multiplying $U(s)$ by the transfer function $Y(s) = U(s)G(s)$
- Use inverse Laplace transforms (from tables) to express $Y(s)$ as a function of time, $y(t)$

4.2 Input functions: Impulse - Step - Ramp

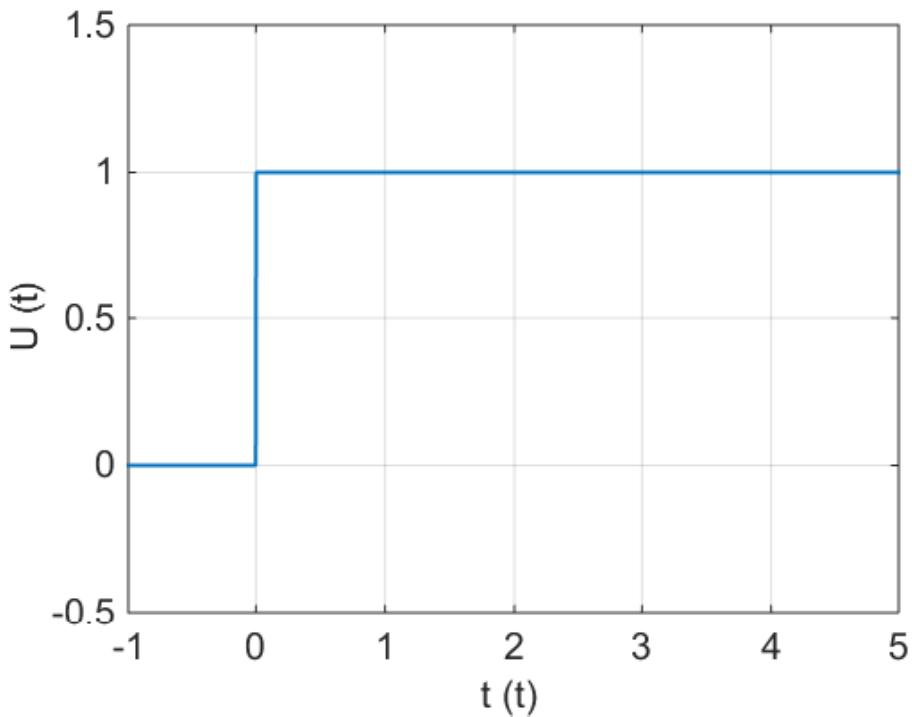
Now we will consider some standard inputs and look at the response of first and second order systems:

- Impulse
 - The Laplace transform is 1, so the response to an impulse is by the definition the transfer function
- Step
- Ramp

There are many others, particularly sinusoidal inputs or other discontinuous inputs, which are important in control loops, but we will focus on the two classic examples.

4.2.1 Step input

A step input is a discontinuous function, which is zero for all negative values of t and 1 for all positive values.



Laplace transform

$$x(t) = U(t) \quad (4.7)$$

$$\mathcal{L}\{U(t)\} = \int_{0^-}^{\infty} e^{-st} dt = \left[-\frac{1}{s} e^{-st} \right]_{0^-}^{\infty} \rightarrow \frac{1}{s} \quad (4.8)$$

(4.9)

Or for a gain of A

$$x(t) = AU(t) \quad (4.10)$$

$$\mathcal{L}\{AU(t)\} = \frac{A}{s} \quad (4.11)$$

Applications

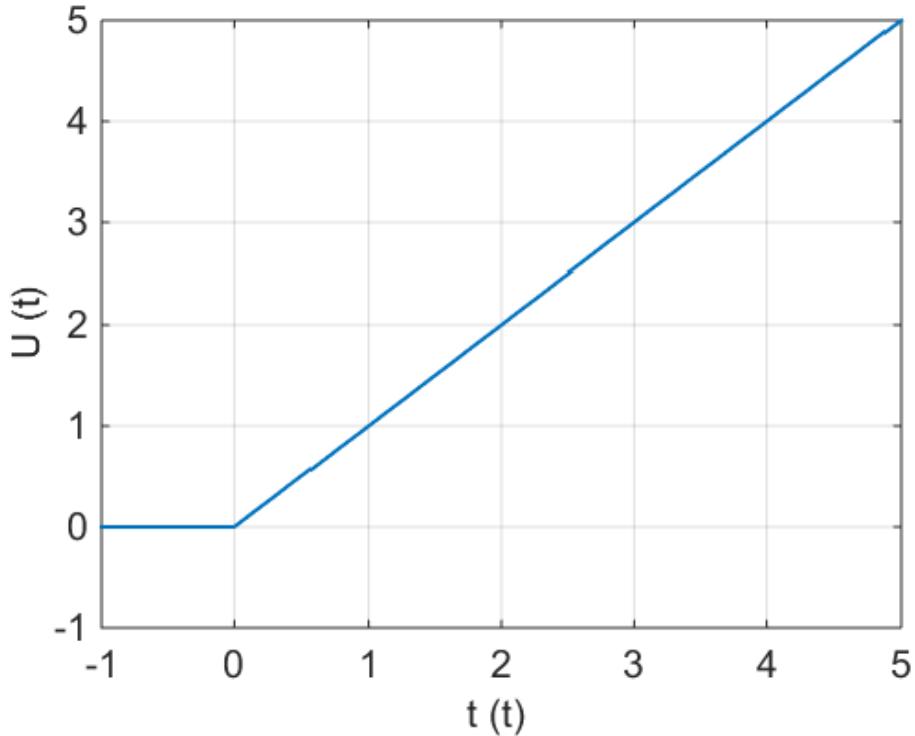
The step response is extremely useful in control theory for describing the behaviour of the system. In part because it incorporates the "transient" behaviour - from the sudden change from zero to one, as well as the "steady state" behaviour as the system settles down to a single value. It also replicates many real world control applications such as:

- Position control - move to a $X = 10\text{mm}$ position and stay
- Speed control - go to 33 RPM
- Temperature - heat element on 3D printer to $230\text{ }^\circ\text{C}$

Also, unlike the Dirac impulse - it is physically realisable.

4.2.2 Ramp input

A ramp input has a value of t for all t values above zero and zero elsewhere, often is scaled by a gain A .



Laplace transform

$$x(t) = at \quad (4.12)$$

$$\mathcal{L}\{at\} \int_{0^-}^{\infty} ate^{-st} dt = -a \left[\frac{t}{s} e^{-st} \right]_0^{\infty} + a \int_0^{\infty} \frac{1}{s} e^{-st} dt \quad (4.13)$$

$$= a \left[-\frac{1}{s^2} e^{-st} \right] \rightarrow \frac{a}{s^2} \quad (4.14)$$

Applications

Ramp inputs are useful in understanding the steady state behaviour of a system i.e. when t goes to infinity. Practical examples of control applications using ramp inputs are

- Servo motors - shaft **position** rather than speed

- Ovens for PCB manufacturing etc. - strict linear **profile** of temperature required as opposed to "get to this temperature quickly"
- CNC milling machine, move in X direction and constant rate

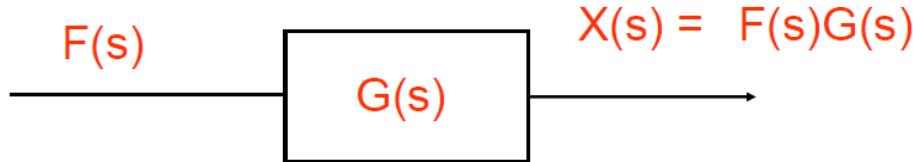
4.2.3 Summary of Input Functions

$$\text{Impulse } F(s) = A \quad (4.15)$$

$$\text{Step } F(s) = \frac{A}{s} \quad (4.16)$$

$$\text{Ramp } F(s) = \frac{A}{s^2} \quad (4.17)$$

For a **unit** response, $A = 1$. We can apply these inputs to the LTI system by multiplying the transfer function by the input, both in terms of s .



4.3 First Order System Step Response

Electromagnets are used in motors both rotary and linear, as well as in power transfer and also magnetic levitation in trains. First obtain the transfer function.

$$\frac{I(s)}{V(s)} \quad (4.18)$$

Balancing voltage gives:

$$v(t) = v_R(t) + v_I(t) \quad (4.19)$$

where

$$v_R(t) = i(t)R \quad (4.20)$$

$$v_I(t) = L \frac{di(t)}{dt} \quad (4.21)$$

Substitution gives:

$$v(t) = Ri(t) + L \frac{di(t)}{dt} \quad (4.22)$$

In the Laplace domain

$$V(s) = RI(s) + LsI(s) \quad (4.23)$$

$$\frac{I(s)}{V(s)} = \frac{1}{(Ls + R)} \quad (4.24)$$

For a **step** input of 1 V (unit step)

$$V(s) = \frac{1}{s} \quad (4.25)$$

$$I(s) = V(s) \frac{1}{(Ls + R)} \quad (4.26)$$

$$= \frac{1}{s(Ls + R)} \quad (4.27)$$

Sadly there is no direct equivalent in the tables for this transform. The first takes is to split the expression up using partial fractions, into expressions that are give in the tables.

$$I(s) = \frac{1}{s(Ls + R)} = \frac{\frac{1}{L}}{s(s + \frac{R}{L})} \text{ (remove coeff for s)} \quad (4.28)$$

$$= \frac{k_1}{s} + \frac{k_2}{(s + \frac{R}{L})} \text{ partial fraction expansion} \quad (4.29)$$

where

$$k_1(s + \frac{R}{L}) + k_2s = \frac{1}{L} \quad (4.30)$$

Looking at $s = 0$

$$k_1 \frac{R}{L} = \frac{1}{L} \rightarrow k_1 = \frac{1}{R} \quad (4.31)$$

Looking at $s = -\frac{R}{L}$

$$-k_2 \frac{R}{L} = \frac{1}{L} \rightarrow k_2 = -\frac{1}{R} \quad (4.32)$$

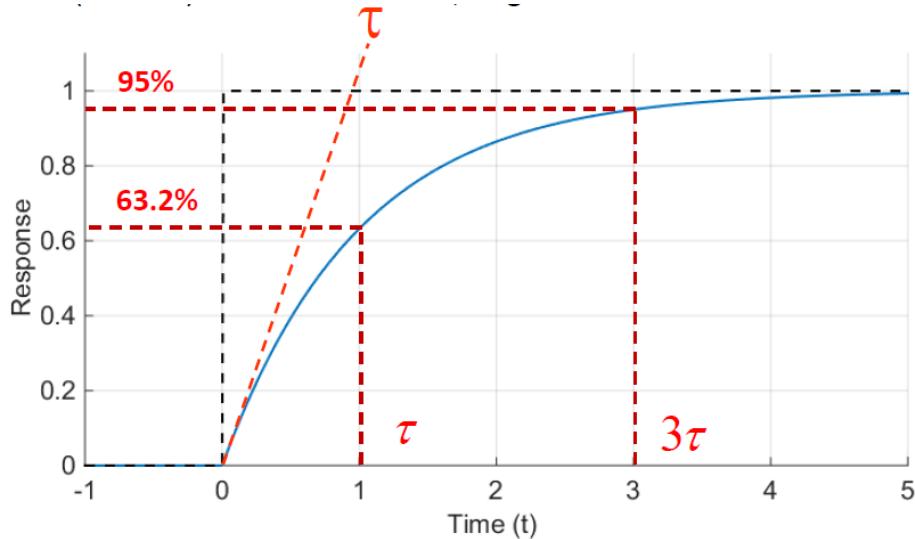
Which yields

$$I(s) = \frac{1}{R} \left[\frac{1}{s} - \frac{1}{(s + \frac{R}{L})} \right] \quad (4.33)$$

Looking at the Laplace tables, we can now use entries for $\frac{1}{s}$ and $\frac{1}{(s+a)}$, considering $t > 0 \therefore u(t) = 1$.

$$I(t) = \frac{1}{R} \left[1 - e^{-\frac{R}{L}t} \right] \quad (4.34)$$

Which has the familiar form of a first order exponential rise, with a steady state gain of $\frac{1}{R}$. Time constant $\tau = \frac{L}{R}$. For now assume that L and $R = 1$, so gain and time constant are 1.



Thus for a known time constant, it is possible to calculate the response at any point after the step. A common parameter of a first order system is the **settling time** which is the time taken to reach 95% of the desired value, or 3τ . The inverse process is also used in system identification, because the response is so simple - we can measure the step response and calculate the time constant of the system. If we do convolution in the time domain you get the same too.

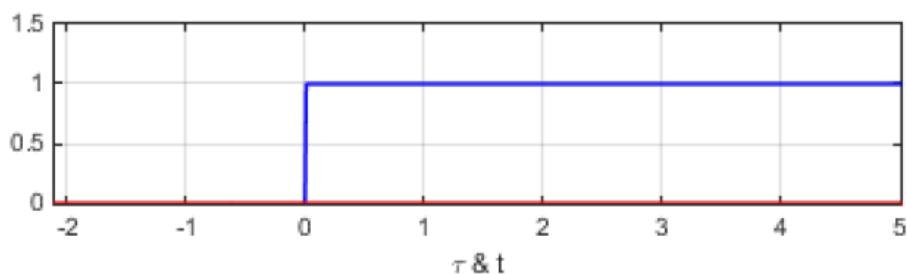
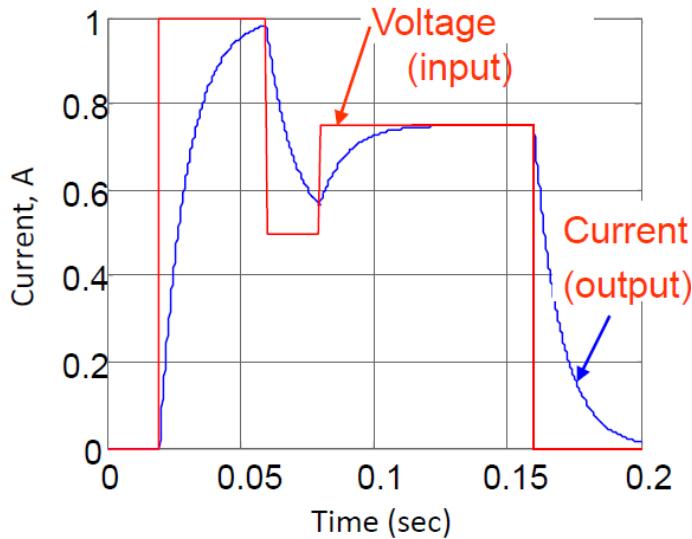


Figure 4.1: Area under $f(\tau)g(1 - \tau)$, blue line $f(\tau)$, red line $g(t - \tau)$, black line $(f + g)t$.

These "step responses" can be repeated for different values of V and for different starting values of $I(t)$. The response will always be behind the input for all time constant > 0 , and thus these systems are referred to as **first order lags**.



4.4 First Order System Ramp Response

For a **ramp** input of gain 1 (unit ramp)

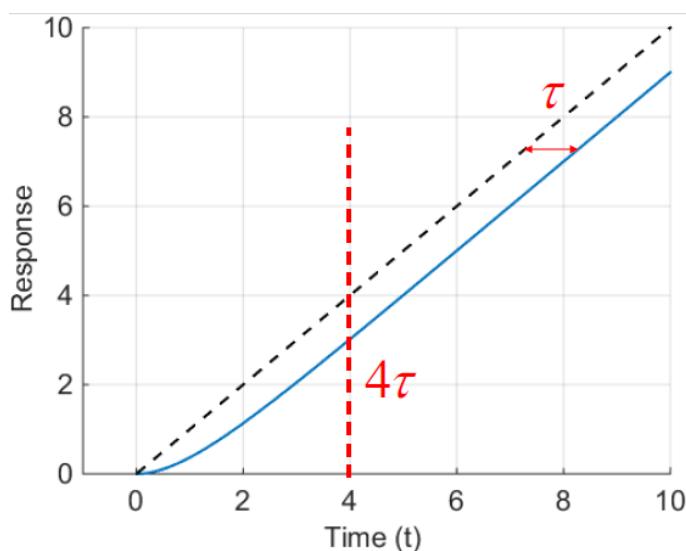
$$\frac{I(s)}{V(s)} = \frac{1}{(Ls + R)} \quad (4.35)$$

$$V(s) = \frac{1}{s^2} \quad (4.36)$$

$$I(s) = \frac{1}{s^2(Ls + R)} = \frac{\frac{1}{L}}{s^2(s + \frac{R}{L})} \quad (4.37)$$

Again, this is not in the Laplace tables, so rearrange use partial fractions, giving us the general solution:

$$I(t) = t - \tau \left(1 - e^{-\frac{t}{\tau}} \right) \quad (4.38)$$



After 4 time constants, we can say that the system is in the **steady state** where the difference between the input and output is no longer changing. The steady state lag is then equal to τ

4.5 Understanding Poles and Zeros

As defined, the transfer function is a rational function in the complex variable $s = \sigma + jw$, that is

$$G(s) = \frac{b_m s^m + b_{m-1} s^{m-1} + \dots + b_1 s + b_0}{a_n s^n + a_{n-1} s^{n-1} + \dots + a_1 s + a_0} \quad (4.39)$$

It is often convenient to factor the polynomials in the numerator and denominator, and to write the transfer function terms of those factors:

$$G(s) = \frac{N(s)}{D(s)} = K \frac{(s - z_1)(s - z_2)\dots(s - z_{m-1})(s - z_m)}{(s - p_1)(s - p_2)\dots(s - p_{n-1})(s - p_n)} \quad (4.40)$$

Where the numerator and denominator polynomials, $N(s)$ and $D(s)$, have real coefficients defined by the system's differential equation and $K = \frac{b_m}{a_n}$. Poles and zeros are found by

$$N(s) = 0 \text{ and } D(s) = 0 \quad (4.41)$$

All of the coefficients of polynomials $N(s)$ and $D(s)$ are real, therefore the poles and zeros must be either purely real, or appear in complex conjugate pairs.

The poles and zeros are properties of the transfer function and therefore of the differential equation describing the input-output system dynamics. Together with the gain constant, they completely characterize the differential equation, and provide a complete description of the system.

4.5.1 Example

A linear system is described by the differential equation:

$$\frac{d^2 y}{dt^2} + 2 \frac{dy}{dt} + 5y = 3 \frac{du}{dt} + 12u \quad (4.42)$$

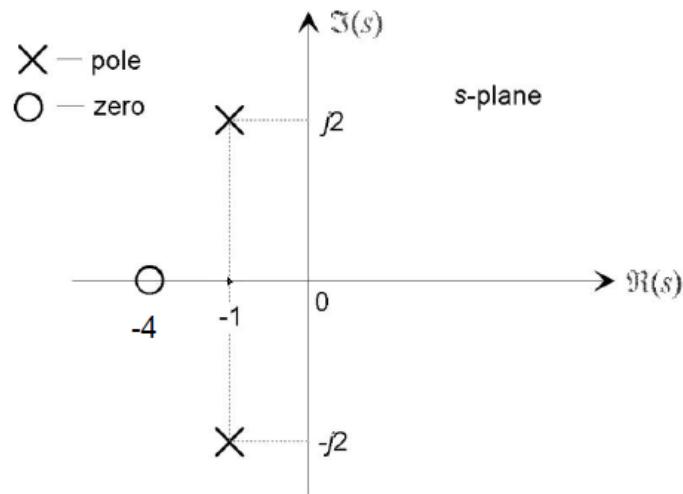
$$\mathcal{L} \left\{ \frac{d^2 y}{dt^2} + 2 \frac{dy}{dt} + 5y \right\} = s^2 Y(s) + 2sY(s) + 5Y(s) \quad (4.43)$$

$$\mathcal{L} \left\{ 3 \frac{du}{dt} + 12u \right\} = 3sU(s) + 12U(s) \quad (4.44)$$

$$N(s) = 3s + 12 = 0 \rightarrow z_1 = -4 \quad (4.45)$$

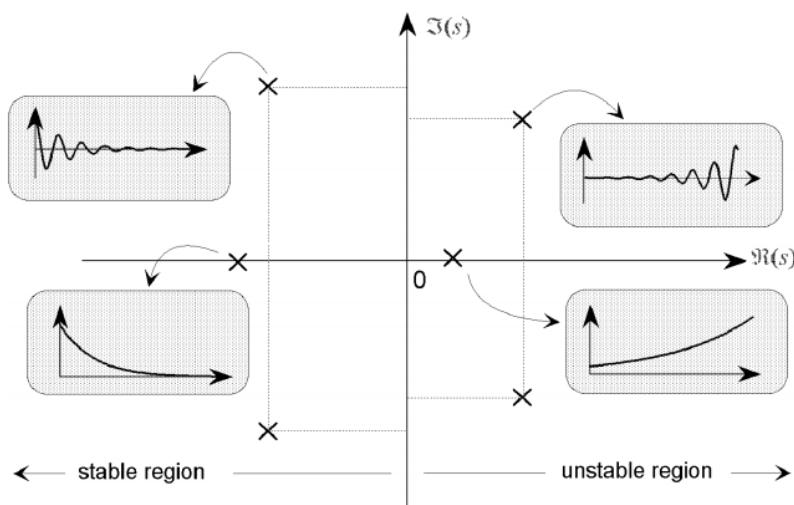
$$D(s) = s^2 + 2s + 5 \rightarrow p_{1/2} = \frac{-2 \pm \sqrt{2^2 - 4 \times 5}}{2} = -1 \pm j2 \quad (4.46)$$

$$G(s) = \frac{N(s)}{D(s)} = 3 \frac{(s - (-4))}{(s - (-1 + j2))(s - (-1 - j2))} \quad (4.47)$$

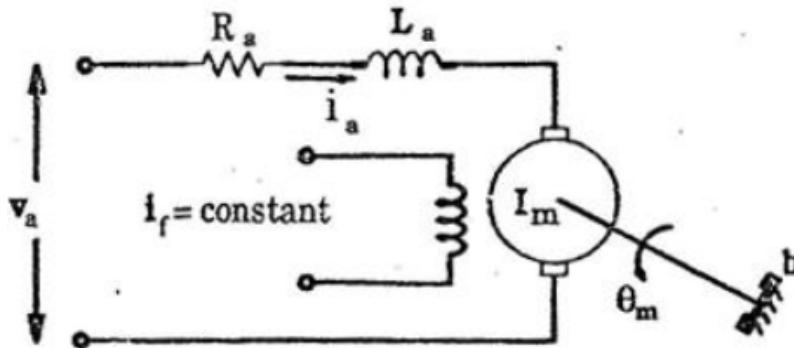


4.5.2 s-Plane

We can see how the time domain representation changes as we move around the s-plane. The imaginary axis corresponds to the sinusoidal component, and the real axis corresponds to the exponential.



4.6 Open Loop Motor Speed Control



For a DC motor connected to an inertial load, there is a combination of electrical and mechanical system equations determining the transfer function for angular position with respect to voltage.

$$\frac{\Theta_m(s)}{v_a(s)} \quad (4.48)$$

First consider the voltage balance:

$$v_a = R_a i_a + L_a \frac{di_a}{dt} + e_m \quad (4.49)$$

Where e_m is the back e.m.f of the motor. We can also write the above in Laplace:

$$V_a(s) = R_a I_a(s) + L_a s I_a(s) + E_m(s) \quad (4.50)$$

Next consider the torque balance in motor:

$$T(t) = I_m \frac{d^2 \theta_m}{dt^2} + b \frac{d\theta_m}{dt} \quad (4.51)$$

Which gives the following in Laplace:

$$T(s) = I_m s^2 \Theta_m + b s \Theta_m(s) \quad (4.52)$$

The electrical and mechanical sides are connected by motor constant which describe the characteristics of the motor (and given by manufacturer).

$$e_m = K_e \frac{d\theta_m}{dt} \quad (4.53)$$

$$T = K_t i_a \quad (4.54)$$

Or in the Laplace domain:

$$E_m(s) = K_e s \Theta_m(s) \quad (4.55)$$

$$T(s) = K_t I_a(s) \quad (4.56)$$

Combining this with our previous equations:

$$V_a(s) = R_a I_a(s) + L_a s I_a(s) + E_m(s) \quad (4.57)$$

$$T(s) = I_m s^2 \Theta_m(s) + b s \Theta_m(s) \quad (4.58)$$

We arrive at the equation of voltage with respect to angular position:

$$V_a(s) = \frac{R_a b}{K_t} [s(\tau_a s + 1)(\tau_m s + 1)] \Theta_m(s) + K_e s \Theta_m(s) \quad (4.59)$$

”Armature” time constant related to electrical side:

$$\tau_a = \frac{L_a}{R_a} \quad (4.60)$$

Motor time constant related to mechanical side:

$$\tau_m = \frac{I_m}{b} \quad (4.61)$$

The transfer function becomes:

$$\frac{\Theta_m(s)}{V_a(s)} = \frac{\frac{K_t}{R_a b}}{s \left[\tau_m \tau_a s^2 + (\tau_m + \tau_a)s + \left(\frac{K_e K_t}{R_a b} + 1 \right) \right]} \quad (4.62)$$

Which looks very complicated, but we can make some simplifying assumptions, which quickly make things simple again. The electrical time constant is normally very small when compared to the mechanical one, as manufacturers try to keep the resistance and inductance low, so we can neglect τ_a . The result, then simplifies to the following equation:

$$\frac{\Theta_m(s)}{V_a(s)} = \frac{K}{s(\tau s + 1)} \quad (4.63)$$

$$K = \frac{K_a K_t}{(b R_a + K_t K_b)} \quad (4.64)$$

Where K is the remaining leftover constants. This looks **much** simpler now, but we can take a step further, rather than consider the output position, lets look at the transfer function with respect to speed ω

$$\frac{\omega_m(s)}{V_a(s)} = \frac{d \left(\frac{\Theta_m(s)}{V_a(s)} \right)}{dt} = s \left(\frac{K}{s(\tau s + 1)} \right) \quad (4.65)$$

$$\frac{\omega_m(s)}{V_a(s)} = \frac{K}{(\tau s + 1)} \quad (4.66)$$

This is now just a first order system.

Chapter 5

System Time Response - 2nd order

5.1 Understanding poles and zeros

As defined, the transfer function is a rational function in the complex variable $s = \sigma + j\omega$, that is:

$$G(s) = \frac{b_m s^m + b_{m-1} s^{m-1} + \dots + b_1 s + b_0}{a_m s^m + a_{m-1} s^{m-1} + \dots + a_1 s + a_0} \quad (5.1)$$

It is often convenient to factor the polynomials in the numerator and denominator, and to write the transfer function in terms of those factors:

$$G(s) = \frac{N(s)}{D(s)} = K \frac{(s - z_1)(s - z_2)\dots(s - z_{m-1})(s - z_m)}{(s - p_1)(s - p_2)\dots(s - p_{n-1})(s - p_n)} \quad (5.2)$$

where the numerator and denominator polynomials, $N(s)$ and $D(s)$, have real coefficients defined by the system's differential equation and $K = \frac{b_m}{a_n}$. Poles and zeros are found by

$$N(s) = 0 \text{ and } D(s) = 0 \quad (5.3)$$

All of the coefficients of polynomials $N(s)$ and $D(s)$ are real, therefore the poles and zeros must be either purely real, or appear in complex conjugate pairs.

The poles and zeros are properties of the transfer function, and therefore of the differential equation describing the input-output system dynamics. Together with the gain constant, the completely characterise the differential equation, and provide a complete description of the system.

5.1.1 Example

A linear system is described by the differential equation:

$$\frac{d^2 y}{dt^2} + 2 \frac{dy}{dt} + 5y = 3 \frac{du}{dt} + 12u \quad (5.4)$$

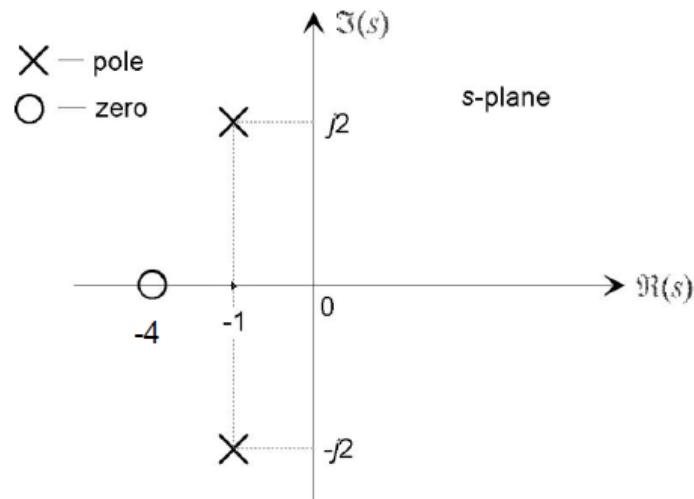
$$\mathcal{L} \left\{ \frac{d^2 y}{dt^2} + 2 \frac{dy}{dt} + 5y \right\} = s^2 Y(s) + 2sY(s) + 5Y(s) \quad (5.5)$$

$$\mathcal{L} \left\{ 3 \frac{du}{dt} + 12u \right\} = 3sU(s) + 12U(s) \quad (5.6)$$

$$N(s) = 3s + 12 = 0 \rightarrow z_1 = -4 \quad (5.7)$$

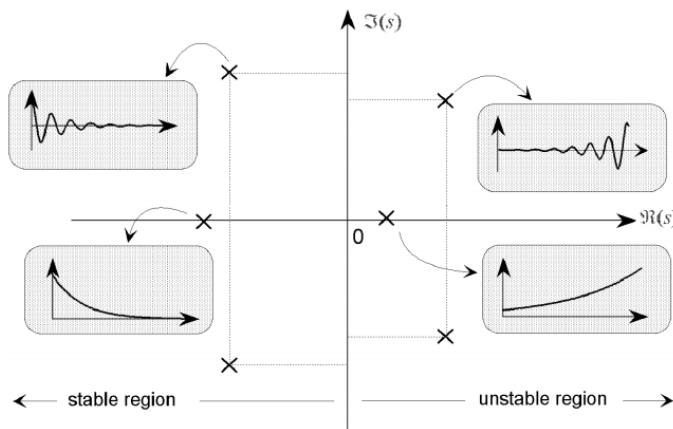
$$D(s) = s^2 + 5s + 5 \rightarrow p_{\frac{1}{2}} = \frac{-2 \pm \sqrt{2^2 - 4 \times 5}}{2} = -1 \pm j2 \quad (5.8)$$

$$G(s) = \frac{N(s)}{D(s)} = 3 \frac{s - (-4)}{(s - (-1 + j2))(s - (-1 - j2))} \quad (5.9)$$



5.1.2 s-Plane

We can see how the time domain representation changes as we move around the s-plane. The imaginary axis corresponds to the sinusoidal component, and the real axis corresponds to the exponential.



5.2 Effect of damping ratio on the transfer function of a 2nd order system

5.2.1 2nd order transfer function

The standard form of second order systems is shown below:

$$G(s) = \gamma \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad (5.10)$$

where, γ is gain, ω_n is the natural frequency and ζ is the damping ratio. Luckily the equation is 'standard form' and the time domain result is given directly in Laplace tables:

$$\gamma \frac{\omega_n}{\sqrt{1 - \zeta^2}} e^{-\zeta\omega_n t} \sin(\omega_n \sqrt{1 - \zeta^2} t) \quad (5.11)$$

But this is only for $\zeta < 1$. Given the damping ratio for a mass spring damper

$$\zeta = \frac{c}{2\sqrt{mk}}, \quad \omega_n = \sqrt{\frac{k}{m}} \quad (5.12)$$

It is clear through the choice of m , k and c you could create a system with $\zeta \geq 1$.

5.2.2 Characteristic equation

$$G(s) = \gamma \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad (5.13)$$

The denominator determines the behaviour of the transfer function, as the numerator is constant. This is known as the **characteristic equation**. It makes life

much simpler if we first try to factor the denominator into first order terms like $(s + a)(s + b)$. So we require the roots of the equation:

$$s^2 + 2\zeta\omega_n s + \omega_n^2 \quad (5.14)$$

Which can be determined from the quadratic equation:

$$s = -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1} \quad (5.15)$$

Thus it is clear that the roots of the characteristic equations and thus the behaviour of the transfer function is determined by the damping ratio.

- Overdamped: $\zeta > 1$, $s = -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1}$
- Critically damped: $\zeta = 1$, $s = -\omega_n$
- Underdamped: $0 < \zeta < 1$, $s = -\zeta\omega_n \pm j\omega_n\sqrt{1 - \zeta^2}$
- Undamped: $\zeta = 0$, $s = \pm j\omega_n$

Normally, only the underdamped case is given directly in the Laplace tables, so we need a bit of manipulation to get the others.

5.2.3 Overdamped 2nd order system

The characteristic equation has two real and negative roots

$$\zeta > 1 \quad (5.16)$$

$$s = -\zeta\omega_n + \omega_n\sqrt{\zeta^2 - 1} = -\alpha \quad (5.17)$$

$$s = -\zeta\omega_n - \omega_n\sqrt{\zeta^2 - 1} = -\beta \quad (5.18)$$

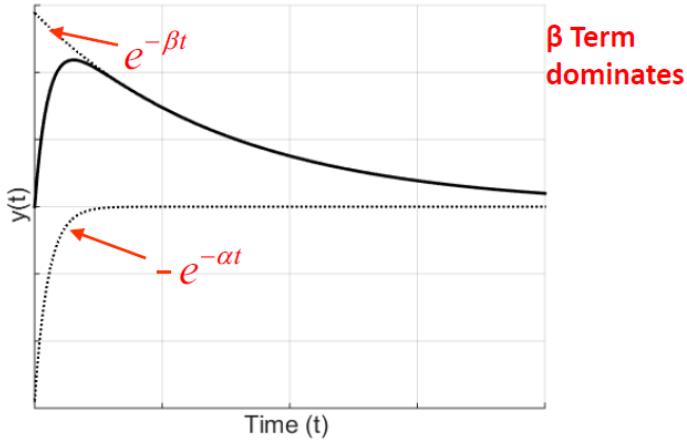
$$\frac{\omega^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} = \frac{\alpha\beta}{(s + \alpha)(s + \beta)} \quad (5.19)$$

Note: $\alpha\beta = \omega_n^2$. Using partial fractions we can split this into two separate first order systems:

$$\left(\frac{\alpha\beta}{\beta - \alpha} \right) \frac{1}{S + \alpha} + \left(\frac{\alpha\beta}{\alpha - \beta} \right) \frac{1}{S + \beta} \quad (5.20)$$

which gives the following in the time domain:

$$x(t) = \alpha\beta \left(\frac{e^{-\alpha t}}{\beta - \alpha} + \frac{e^{-\beta t}}{\alpha - \beta} \right) \quad (5.21)$$



There is no sinusoidal term anymore, the response is only a combination of exponential decays, a more complex but similar response to a first order system.

5.2.4 Critically damped 2nd order system

Special case with a single repeated root:

$$\zeta = 1 \quad (5.22)$$

$$s = -\omega_n \quad (5.23)$$

$$\frac{\omega^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} = \frac{\omega_n^2}{(s + \omega_n)^2} \quad (5.24)$$

To get the time response, we first take the numerator outside:

$$L^{-1} \left[\frac{\omega_n^2}{(s + \omega_n)^2} \right] = \omega_n^2 L^{-1} \left[\frac{1}{(s + \omega_n)^2} \right] \quad (5.25)$$

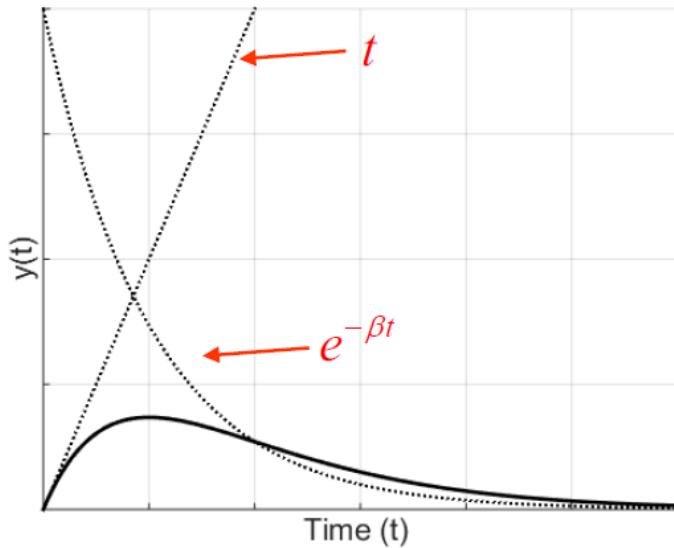
We can then use the inverse Laplace rule:

$$\text{if } L^{-1} [F(s)] = f(t) \text{ then } L^{-1} [F(s-a)] = e^{at} f(t) \quad (5.26)$$

$$\omega_n^2 L^{-1} \left[\frac{1}{(s + \omega_n^2)} \right] = \omega_n^2 e^{-\omega_n t} L^{-1} \left[\frac{1}{s^2} \right] \quad (5.27)$$

$\frac{1}{s^2}$ is the standard form of a ramp, or more generally from the Laplace tables:

$$L^{-1} \left[\frac{n!}{s^{n+1}} \right] = t^n \quad (5.28)$$



The result is an exponential decay, multiplied by a ramp $= \omega_n^2 e^{-\omega_n t} t$.

5.2.5 Undamped 2nd order system

Another special case with two imaginary roots:

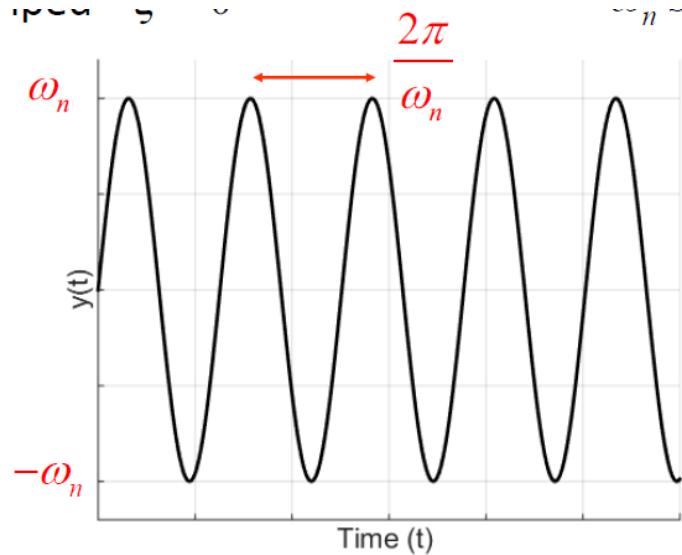
$$\zeta = 0 \quad (5.29)$$

$$s = \pm j\omega_n \quad (5.30)$$

$$\frac{\omega^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} = \frac{\omega_n^2}{s^2 + \omega_n^2} \quad (5.31)$$

Which, using the standard form for an oscillator is:

$$\omega_n L^{-1} \left[\frac{\omega_n}{s^2 + \omega_n^2} \right] = \omega_n \sin(\omega_n t) \quad (5.32)$$



So the response is now just a sinusoid, with no exponential components.

5.3 Applying a step input to 2nd order systems - damping

Previously we saw for a step of A the output can be calculated from

$$Y(s) = \frac{A}{s} G(s) \quad (5.33)$$

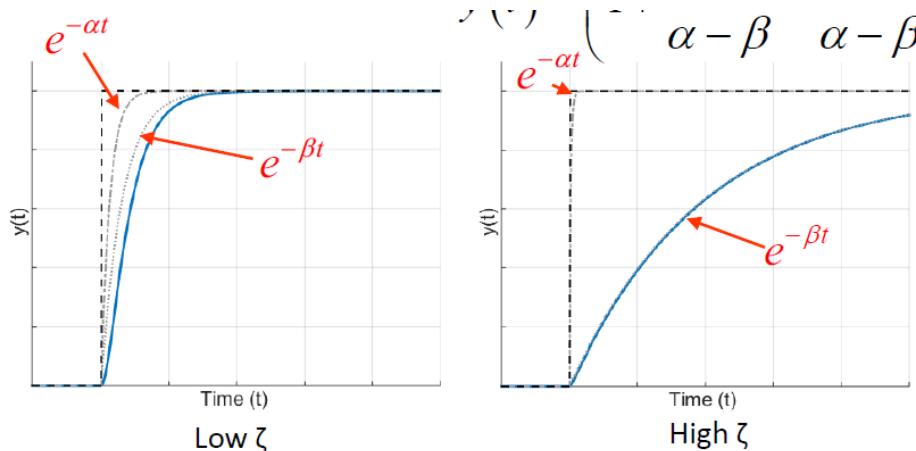
$$Y(s) = \frac{1}{s} \cdot \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s\omega_n^2} \quad (5.34)$$

All of the following can be found through partial fraction decomposition and the standard Laplace tables.

5.3.1 Overdamped step input response of 2nd order systems

$$\zeta > 1 \quad (5.35)$$

$$y(t) = \left(1 = \frac{\beta e^{-\alpha t}}{\alpha - \beta} - \frac{\alpha e^{-\beta t}}{\alpha - \beta} \right) \quad (5.36)$$

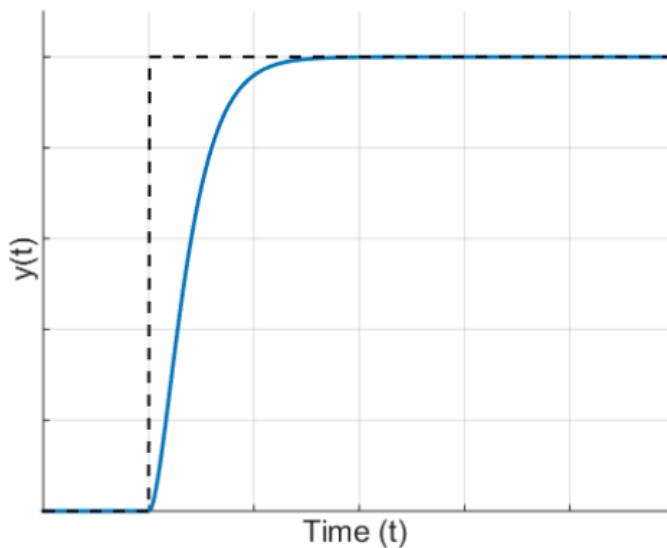


The slowest rise from the β exponential dominates, as ζ increases the α term becomes negligible, and system becomes first order.

5.3.2 Critically damped step input response of 2nd order systems

$$\zeta = 1 \quad (5.37)$$

$$y(t) = (1 - e^{-\omega_n t} - \omega_n e^{-\omega_n t} t) \quad (5.38)$$

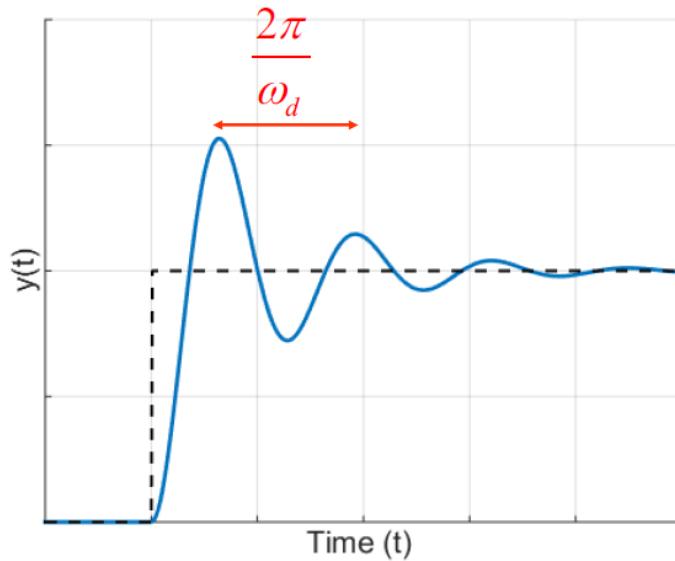


Fastest possible rise without oscillating.

5.3.3 Underdamped step input response of 2nd order systems

$$0 < \zeta < 1 \quad (5.39)$$

$$y(t) = 1 - \frac{1}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_n t} \sin \left(\omega_n \sqrt{1-\zeta^2} t + \arccos \zeta \right) \quad (5.40)$$



It is sometimes written in terms of the damped natural frequency:

$$\omega_d = \omega_n \sqrt{1 - \zeta^2} \quad (5.41)$$

$$y(t) = 1 - \frac{1}{\sqrt{1 - \zeta^2}} e^{-\zeta \omega_n t} \sin(\omega_d t + \arccos \zeta) \quad (5.42)$$

5.3.4 Undamped step input response of 2nd order systems

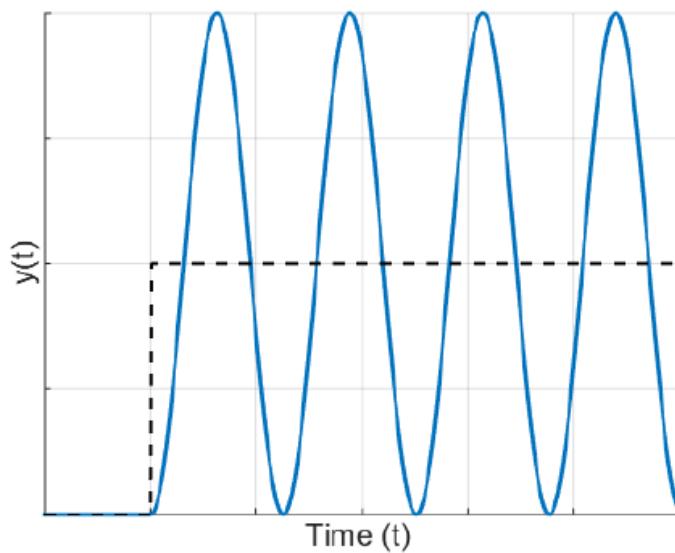
$$\zeta = 0 \quad (5.43)$$

$$y(t) = (1 - \cos(\omega_n t)) \quad (5.44)$$

As damping ratio reaches zero, the oscillations no longer decay

$$\omega_d = \omega_n \sqrt{1 - 0} = \omega_n \quad (5.45)$$

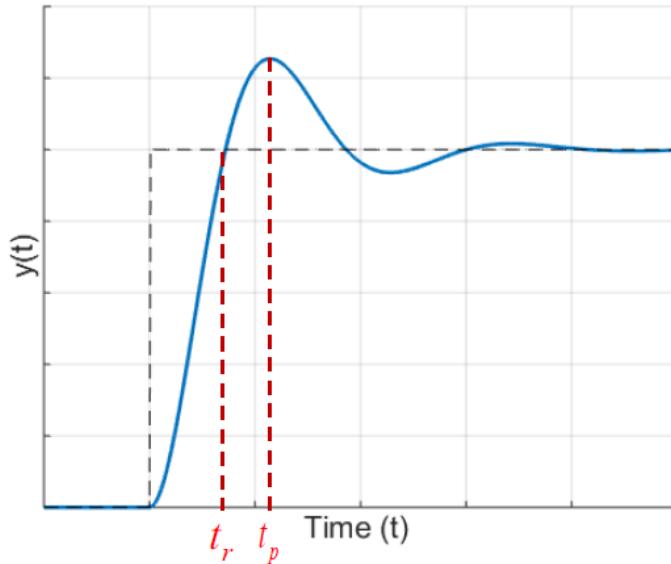
So oscillations are at natural frequency with pi phase shift.



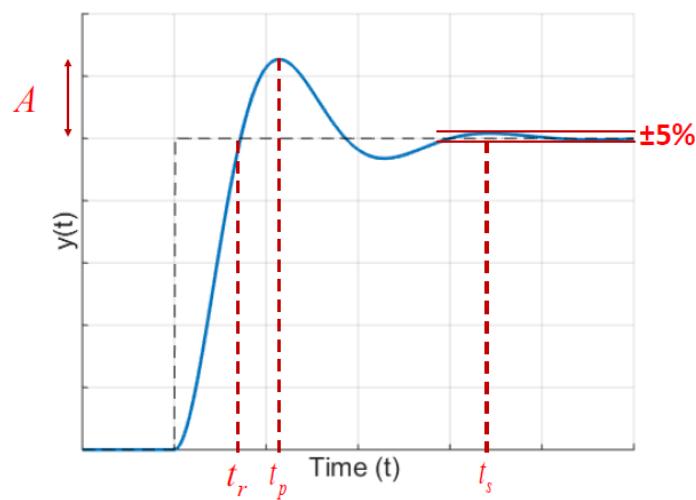
5.4 Underdamped step input responses

5.4.1 Performance Criteria

Rise time t_r is the time to reach steady state value (for first time). Peak time t_p is the time to initial overshoot.



Peak overshoot - value of overshoot above steady state: $y(t_p) - y_{ss} = A$. Settling time t_s . Time for response to reach and maintain a certain ratio of the steady state value. Commonly $\pm 5\%$ or $\pm 2\%$



These criteria can be read from the graph, but having equations for them allows for a system parameters ζ and ω_n to be estimated. It is also useful for quantifying the effect of ζ on the response. Rise time can be found by setting the response y to 1

and then rearranging:

$$t_r = \frac{1}{\omega_d} \left(\pi - \arctan \left(\frac{\sqrt{1-\zeta^2}}{\zeta} \right) \right) \omega_d = \omega_n \sqrt{1-\zeta^2} \quad (5.46)$$

So a higher damping ratio gives a slower response, but lower damping ratios give a more oscillatory response. Also, that for **overdamped** systems where $\zeta > 1$, the rise time is infinite. The peak time occurs at half a period of oscillation, or can be found from finding first minima by setting derivative to zero:

$$t_p = \frac{\pi}{\omega_n \sqrt{1-\zeta^2}} = \frac{\pi}{\omega_d} \quad (5.47)$$

Peak overshoot

$$1 + A = 1 - \frac{1}{\sqrt{1-\zeta^2}} e^{\frac{-\zeta\pi}{\sqrt{1-\zeta^2}}} \sin(\pi + \arccos \zeta) \quad (5.48)$$

$$A = \frac{\sqrt{1-\zeta^2}}{\sqrt{1-\zeta^2}} e^{\frac{-\zeta\pi}{\sqrt{1-\zeta^2}}} = e^{\frac{-\zeta\pi}{\sqrt{1-\zeta^2}}} \quad (5.49)$$

Often, this is given as a percentage:

$$A = 100 e^{\frac{-\zeta\pi}{\sqrt{1-\zeta^2}}} \quad (5.50)$$

This can be rearranged to find the required damping ratio for a required overshoot:

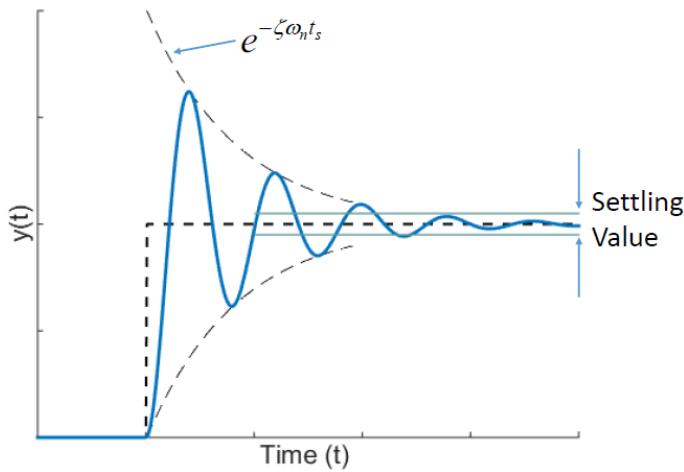
$$\zeta = \sqrt{\frac{\left(\ln \frac{A}{100}\right)^2}{\pi^2 + \left(\ln \frac{A}{100}\right)^2}} \quad (5.51)$$

$$\frac{A}{100} \rightarrow 0 \quad (5.52)$$

$$\zeta \rightarrow 1 \quad (5.53)$$

Settling time

The envelope of the oscillation is described by the decaying exponential term, so to find the settling time, we can just consider this term alone.



For example, the common target is 5%:

$$e^{-\zeta\omega_n t_s} = 0.05 \quad (5.54)$$

$$-\zeta\omega_n t_s = \ln 0.05 = -3 \quad (5.55)$$

$$t_s(5\%) = \frac{3}{\zeta\omega_n} \quad (5.56)$$

Or similarly:

$$t_s(2\%) = \frac{4}{\zeta\omega_n} \quad (5.57)$$

So settling time increases with a **decreasing** damping ratio, for underdamped systems only.

5.5 Ramp input function responses

5.5.1 Applying inputs

Previously, we saw for a step of A the output can be calculated from:

$$Y(s) = \frac{A}{s^2} G(s) \quad (5.58)$$

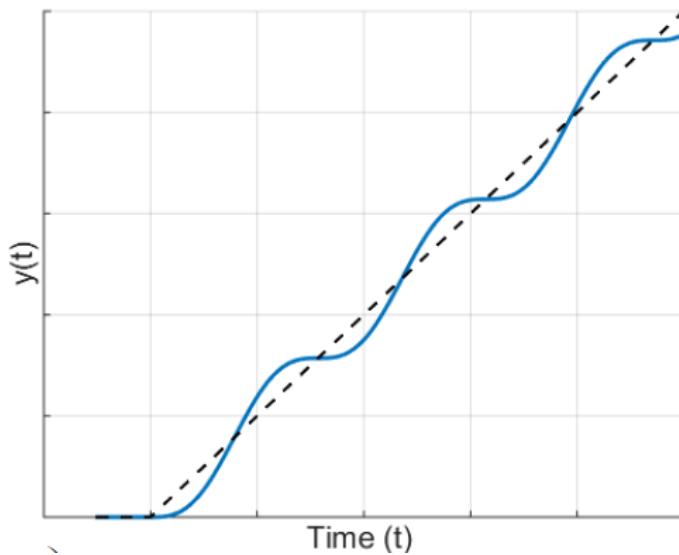
$$Y(s) = \frac{1}{s^2} \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad (5.59)$$

All the following can be found through partial fraction decomposition and the standard Laplace tables (and loads more manipulation!). All these expressions are pretty complicated, but the important thing is to understand how the response changes with damping ratio, which follows a similar trend to the step response.

5.5.2 Undamped response

$\zeta = 0$: undamped - sustained oscillations around ramp input $y = t$.

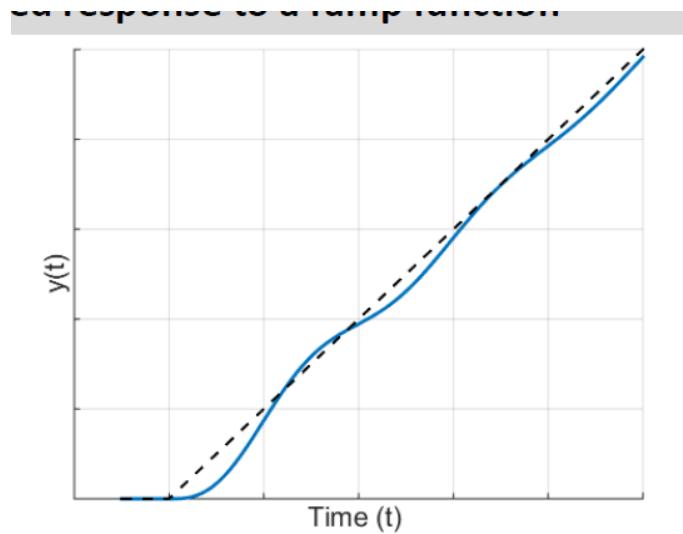
$$y(t) = \left(t - \frac{\sin(\omega_n t)}{\omega_n} \right) \quad (5.60)$$



5.5.3 Underdamped response

$\zeta < 1$: underdamped - exponential decaying oscillatory response around delayed ramp function.

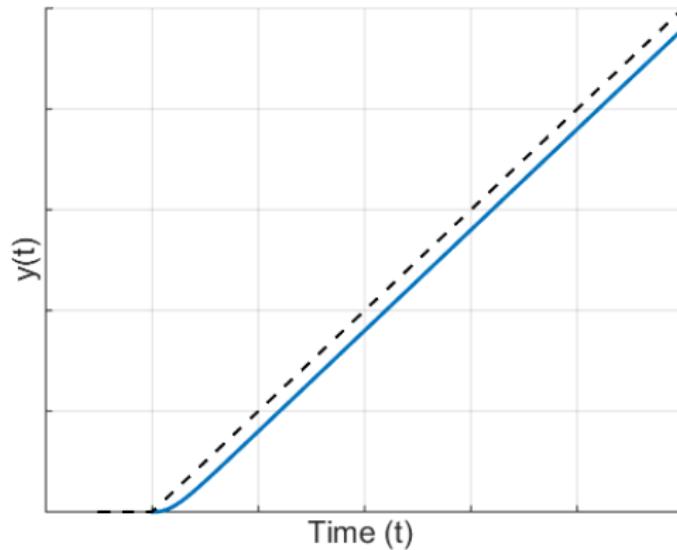
$$y(t) = \frac{1}{\omega_n} \left(\omega_n t - 2\zeta + e^{-\zeta\omega_n t} \left(2\zeta \cos \left(\omega_n t \sqrt{1 - \zeta^2} \right) - \frac{1}{\sqrt{1 - \zeta^2}} (1 - 2\zeta^2) \sin \left(\omega_n t \sqrt{1 - \zeta^2} \right) \right) \right) \quad (5.61)$$



5.5.4 Critically damped response

$\zeta = 1$: Critically damped - fastest response with no overshoot, still lagging input.

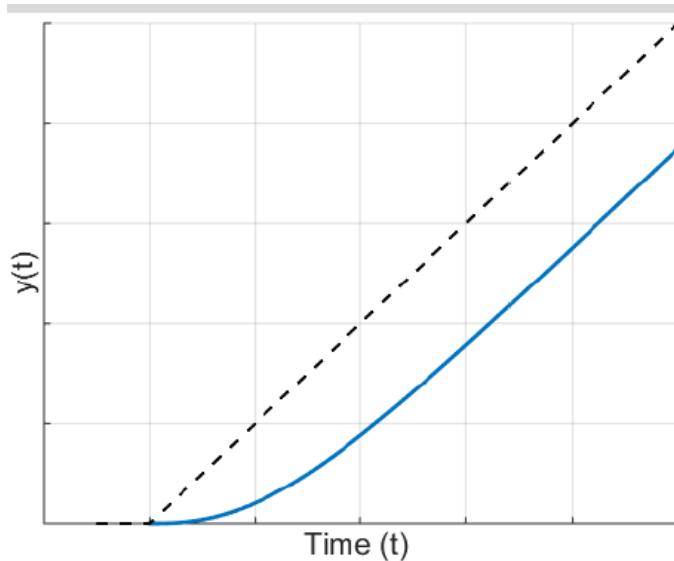
$$y(t) = \frac{1}{\omega_n} (\omega_n t - 2 + (2 + \omega_n t) e^{-\omega_n t}) \quad (5.62)$$



5.5.5 Overdamped response

$\zeta > 1$: overdamped - cannot overshoot, lag increases.

$$y(t) = \left(t - \frac{2\zeta}{\omega_n} + \frac{\beta e^{-\alpha t}}{\alpha - \beta} - \frac{\alpha e^{-\beta t}}{\alpha - \beta} \right) \quad (5.63)$$



5.5.6 Summary

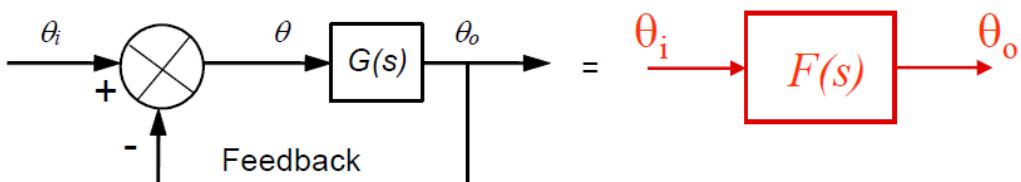
Much more complicated response! but follows the same pattern as with the step response with changing damping ratio. The most important thing to notice here, is that the system does not match desired value - there is still a lag:

$$y(t) = t - \frac{2\zeta}{\omega_n} \quad (5.64)$$

So a low damping ratio is needed to closely match the input value, but this comes at the expense of overshoot and settling time.

5.6 Transfer function - unity feedback system

Recall the block diagram for a closed loop feedback system, in this case with 'unity' feedback - i.e. the error signal is passed directly to the summation junction:



$G(s)$ is referred to as the **forward loop transfer function**, i.e. the combined transfer function of all the amplifiers controller, plant, systems in the feedforward path. In the case of unity feedback, it is also the **open loop transfer function**. We can find the overall transfer function $F(s)$ or **closed loop transfer function**

through a bit of manipulation:

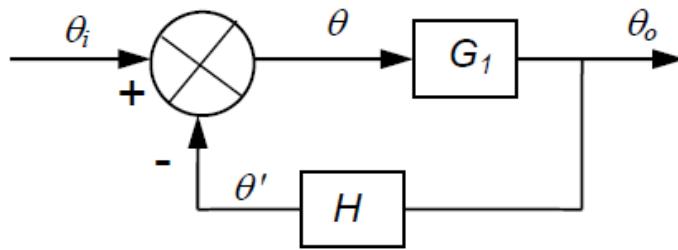
$$\theta = \theta_i - \theta_0 \quad (5.65)$$

$$G(s)\theta = \theta_0 \quad (5.66)$$

$$\theta_0 = G(s)(\theta_i - \theta_0) \quad (5.67)$$

$$\frac{\theta_0}{\theta_i} = F(s) = \frac{G(s)}{1 + G(s)} \quad (5.68)$$

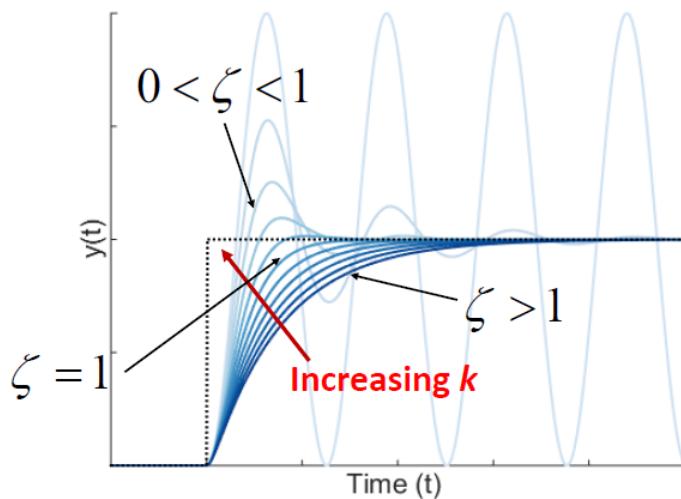
Or more generally for a non unity gain system:



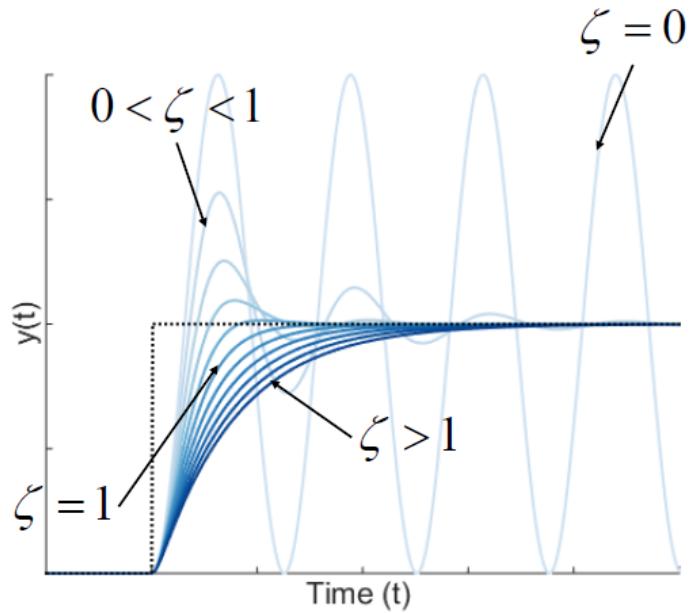
$$F_1 = \frac{\theta_0}{\theta_i} = \frac{G_1}{1 + G_1 H} \quad (5.69)$$

Lets take an example to show how this transfer function can relate to the systems we have studied so far.

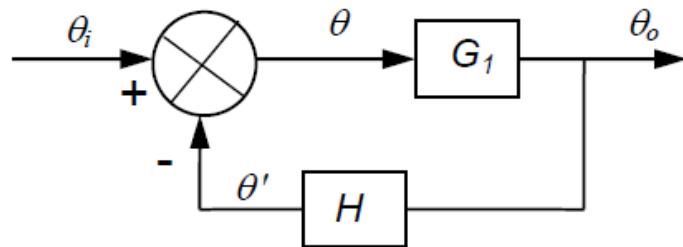
5.7 Summary



The overdamped ($\zeta > 1$) and critically damped ($\zeta = 1$) responses are seldom desirable for control systems due to their long settling time. $\zeta = 0.7$ gives a fast response without excessive overshoot or oscillations.



The response of second order systems is much more complicated than that of a first order system, with considerably different responses arising from changes in damping ratio.



Negative feedback systems have a transfer function of the form:

$$F_1 = \frac{\theta_0}{\theta_i} = \frac{G_1}{1 + G_1 H} \quad (5.70)$$

Which for a remote position servo is equal to:

$$F(s) = \frac{\Theta_0(s)}{\Theta_i(s)} = \frac{\frac{k}{I}}{s^2 + \frac{f}{I}s + \frac{k}{I}} \quad (5.71)$$

Where $\omega_n = \sqrt{\frac{k}{I}}$ and $\zeta = \frac{f}{2\sqrt{kI}}$. The gain determines the damping ratio, and this the characteristics of the response of the system.

Chapter 6

The s-plane and Steady State Errors

6.1 The s-Plane - Poles and Zeroes

6.1.1 Example: Servo as a second order system

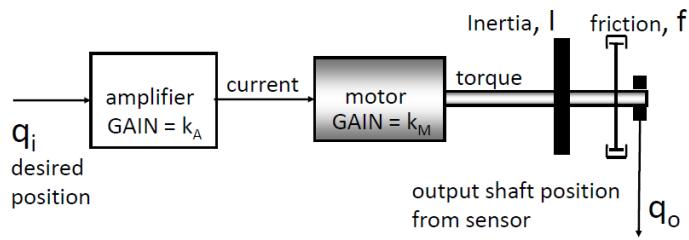


Figure 6.1:

We saw last week that a DC motor servo with position control has a transfer function of a second order system.

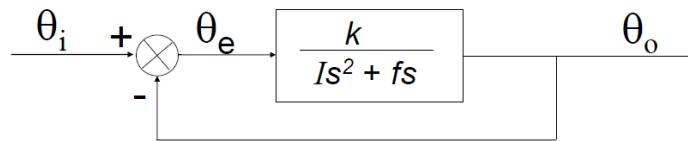


Figure 6.2:

$$F(s) = \frac{G(s)}{1 + G(s)} = \frac{\frac{K}{I}}{s^2 + \frac{f}{I}s + \frac{K}{I}} \quad (6.1)$$

Where, $\omega_n = \sqrt{\frac{k}{I}}$ (with increasing k , ω_n increases) and $\zeta = \frac{f}{2\sqrt{kI}}$ (with increasing k , ζ decreases).

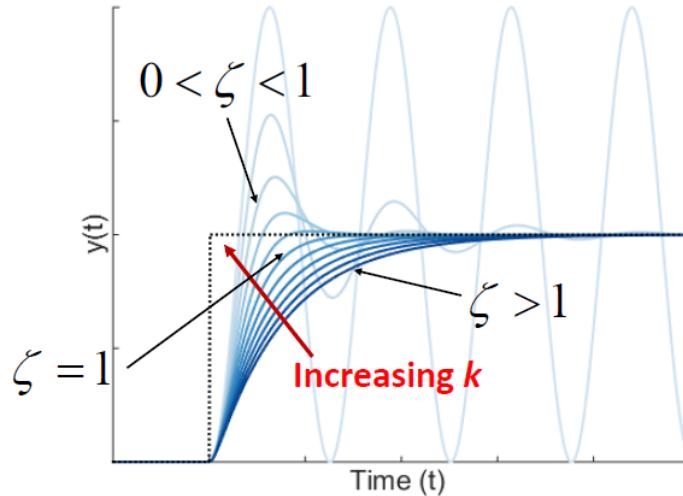


Figure 6.3:

The overdamped ($\zeta > 1$) and critically damped ($\zeta = 1$) responses are seldom desirable for control systems due to their long settling time. $\zeta = 0.7$ gives a fast response without excessive overshoot or oscillations.

6.1.2 Transfer functions

As we have seen, transfer functions allow for important characteristics of the system response to be determined, without having to solve complete differential equations. Generally, transfer functions can be derived from the differential equations thusly:

$$a_n \frac{d^n y}{dt^n} + a_{n-1} \frac{d^{n-1} y}{dt^{n-1}} + \dots + a_2 \frac{d^2 y}{dt^2} + a_1 \frac{dy}{dt} + a_0 y = b_0 x + b_1 \frac{dx}{dt} + b_2 \frac{d^2 x}{dt^2} + \dots + b_{m-1} \frac{d^{m-1} x}{dt^{m-1}} + b_m \frac{d^m x}{dt^m} \quad (6.2)$$

(Solving equation (6.2) is hard!). Where a are the output coefficients and b are the input coefficients, combined these **entirely** characterise the system. The Laplace domain transfer function is then given by:

$$G(s) = \frac{b_m s^m + b_{m-1} s^{m-1} + \dots + b_1 s + b_0}{a_n s^n + a_{n-1} s^{n-1} + \dots + a_1 s + a_0} \quad (6.3)$$

6.1.3 System poles and zeroes

As defined, the transfer function is a rational function in the complex variable $s = \sigma + j\omega$, that is:

$$G(s) = \frac{b_m s^m + b_{m-1} s^{m-1} + \dots + b_1 s + b_0}{a_n s^n + a_{n-1} s^{n-1} + \dots + a_1 s + a_0} \quad (6.4)$$

It is often convenient to factor the polynomials in the numerator and denominator; and to write the transfer function in terms of those factors:

$$G(s) = \frac{N(s)}{D(s)} = K \frac{(s - z_1)(s - z_2)\dots(s - z_{m-1})(s - z_m)}{(s - p_1)(s - p_2)\dots(s - p_{n-1})(s - p_n)} \quad (6.5)$$

where the numerator and denominator polynomials, $N(s)$ and $D(s)$, have real coefficients defined by the system's differential equation. Poles and zeroes are found by:

$$N(s) = 0 \text{ and } D(s) = 0 \quad (6.6)$$

All of the coefficients of polynomials $N(s)$ and $D(s)$ are real, therefore the poles and zeroes must be either purely real, or appear in complex conjugate pairs.

The poles and zeros are properties of the transfer function, and therefore of the differential equation describing the input-output system dynamics. Together with the gain constant, they completely characterise the differential equation and provide a complete description of the system.

6.1.4 Transfer functions in the s-plane

A system is characterised by its poles and zeroes as they allow the reconstruction of the input/output differential equation. It is possible to get a sense of the system dynamics from plotting the poles and zeroes graphically on the s-plane. A pole is commonly represented by a cross (\times) and a zero by a circle (\circ). Recall that, as s is a complex number, we can plot the value on a plane, with a real and imaginary axis.

$$s = \sigma + j\omega \quad (6.7)$$

By plotting the locations of the poles and zeroes on this plane, we can obtain a considerable amount of information about the response of the system without having to take the inverse Laplace transform. Each pole corresponds to a component of the time domain response, so from the plot it is possible to determine:

- What components exist
- Their relative important (and possible simplifications)
- How they change with gain

Real poles - exponentials

First lets look at the real component only:

$$s = \sigma \quad (6.8)$$

$$e^s \rightarrow e^{-\sigma} \quad (6.9)$$

Corresponding to $Ce^{-\sigma t}$

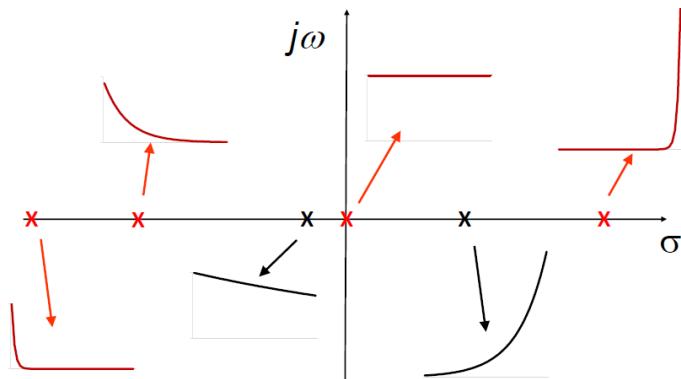


Figure 6.4:

Poles located on the real axis have an exponential component only, with the pole location determining the rate of decay. Poles close to origin decay slowly, where poles far away decay rapidly. Poles located in the right hand side exponentially **increase**, whereas poles on the left hand side **decrease**. A pole on the origin is a flat 'DC' response.

Imaginary poles - sinusoids

First lets looks at the imaginary component only:

$$s = \pm j\omega \quad (6.10)$$

$$e^s \rightarrow e^{\pm j\omega} \quad (6.11)$$

Which, from Euler's formula gives a sinusoidal response:

$$e^{jx} = \cos x + j \sin x \quad (6.12)$$

Poles located on the imaginary axis are in conjugate pairs.

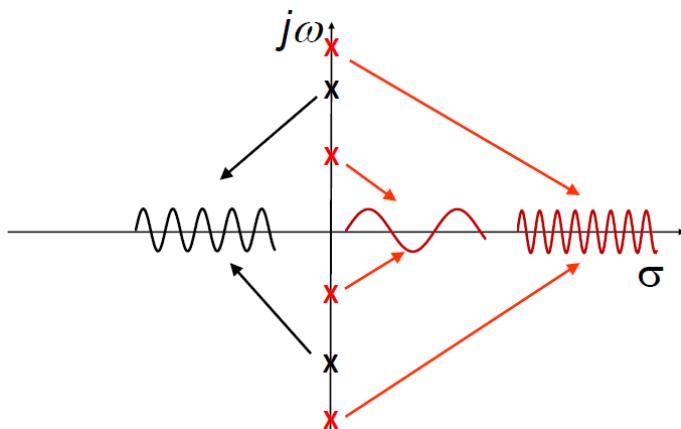


Figure 6.5:

The imaginary poles generate an oscillatory component with a constant amplitude. Poles closer to the origin have a low frequency, which increases the further poles are from the origin.

Complex poles

Commonly, the poles are a combination of the two components.

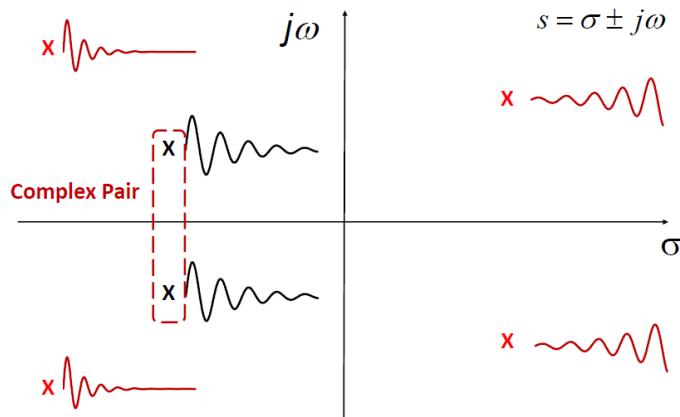


Figure 6.6:

A complex conjugate pair of complex poles generates an exponentially decaying sinusoid in the form:

$$Ae^{-\sigma t} \sin(\omega t + \phi) \quad (6.13)$$

A and ϕ are determined by the initial conditions, with ω specifying oscillations and σ the rate of decay. Poles located in the left hand side decay to zero, whereas poles in the right hand side increase to infinity, thus making the system **unstable**.

6.1.5 Summary

The impulse response of each pole in the transfer function depends upon its location in the s-plane:

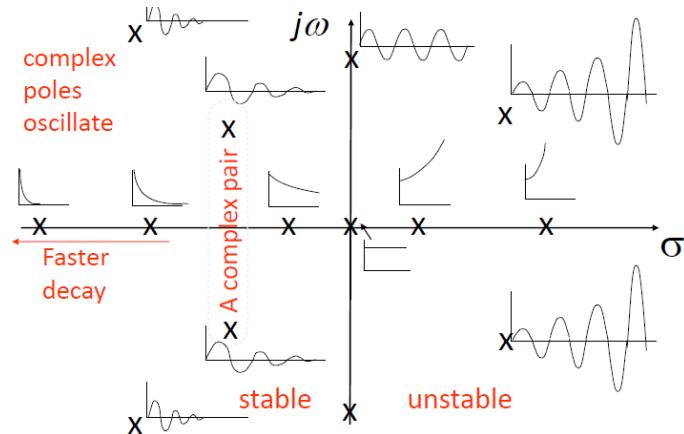


Figure 6.7:

6.2 Effects of the damping ratio on poles

6.2.1 Transfer functions of second order systems

As we have seen, the roots of the transfer function of a second order system depends upon the damping coefficient:

$$G(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n + \omega_n^2} \quad (6.14)$$

We can plot the poles of the characteristic equation as with any other transfer function:

$$s = -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1} \quad (6.15)$$

Overdamped

Both roots are real, and thus are on the real axis:

$$s_1 = -\zeta\omega_n + \omega_n\sqrt{\zeta^2 - 1} \quad (6.16)$$

$$s_2 = -\zeta\omega_n - \omega_n\sqrt{\zeta^2 - 1} \quad (6.17)$$

No oscillations in the step response.

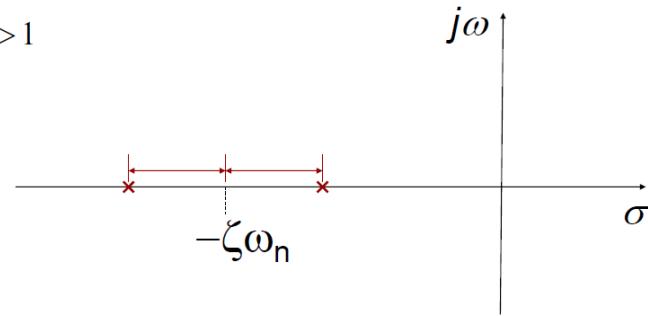


Figure 6.8:

The **right** most pole with smallest value of σ dominates the response (this was the β term from Lecture 5). Further increasing damping ratio, widens the gap between the two poles and shifts them both further away from the origin. Conversely, decreasing the damping ratio brings the two poles closer together, until we reach the special case where $\zeta = 1$.

Critically damped

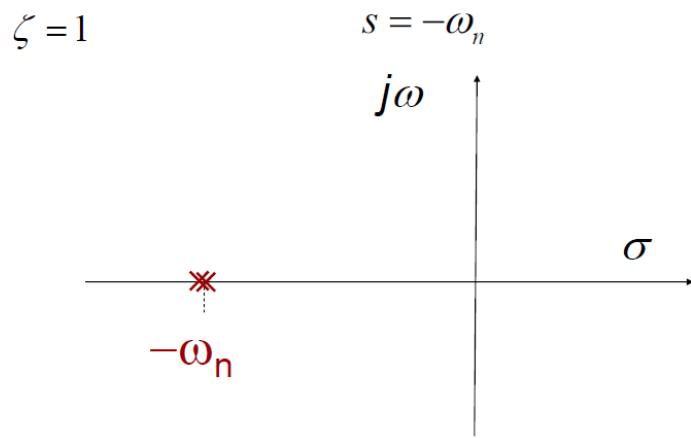


Figure 6.9:

The roots of the equations coincide, and the system is a product of two equal first order lags.

Underdamped

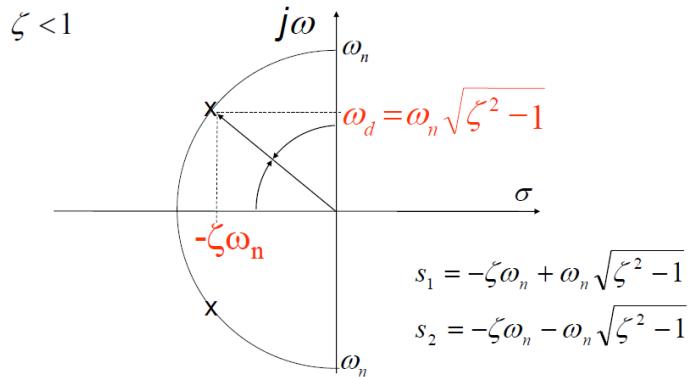


Figure 6.10:

Both roots complex, and form conjugate pair. System is a product of exponential function and oscillatory component.

6.2.2 Example: Servomechanism in the s-plane

$$F(s) = \frac{\frac{k}{I}}{s^2 + \frac{f}{I}s + \frac{k}{I}} \quad (6.18)$$

$$\zeta = \frac{f}{2\sqrt{kI}} \quad (6.19)$$

$$\omega_n = \sqrt{\frac{k}{I}} \quad (6.20)$$

Considering the second order model for a servo, the roots of the characteristic equation in terms of the parameters I , f and k are:

$$s = -\frac{f}{2I} \pm \frac{1}{2I}\sqrt{f^2 - 4kI} \quad (6.21)$$

Thus if $f^2 < 4kI$ roots are complex and there is an oscillatory response. If $f^2 > 4kI$ roots are real and there is an exponential lag response. Lets consider how the response changes with the parameters, first the gain k . Notice that for the complex root case, the real component is unaffected by k , so an increased gain increases the imaginary component, and thus the frequency of oscillation:

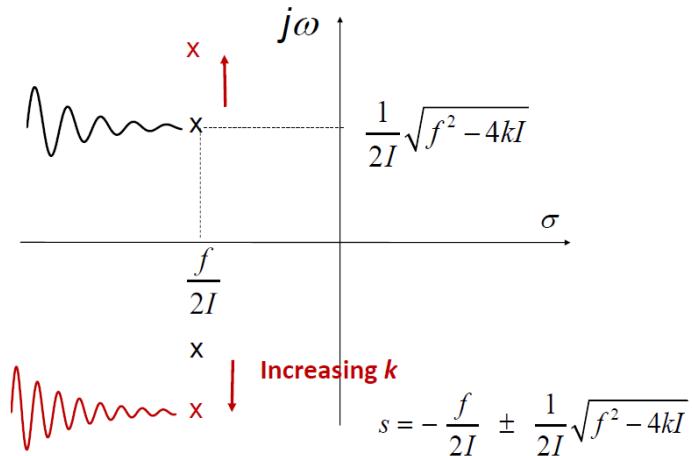


Figure 6.11:

The behaviour is also apparent from the fact that the gain term k affects both ω_n and ζ in an inversely related manner. The real part of the pole is given by $\omega_n\zeta$, so the increase in ω_n is cancelled out by decrease in ζ .

$$s = \zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1} \quad (6.22)$$

Changing friction

Changing the friction coefficient f has a more complicated change on the pole location as it is located both in the real and imaginary components of the complex poles, and does not change ω_n .

$$s = -\frac{f}{2I} \pm \frac{1}{2I}\sqrt{f^2 - fkI} \quad (6.23)$$

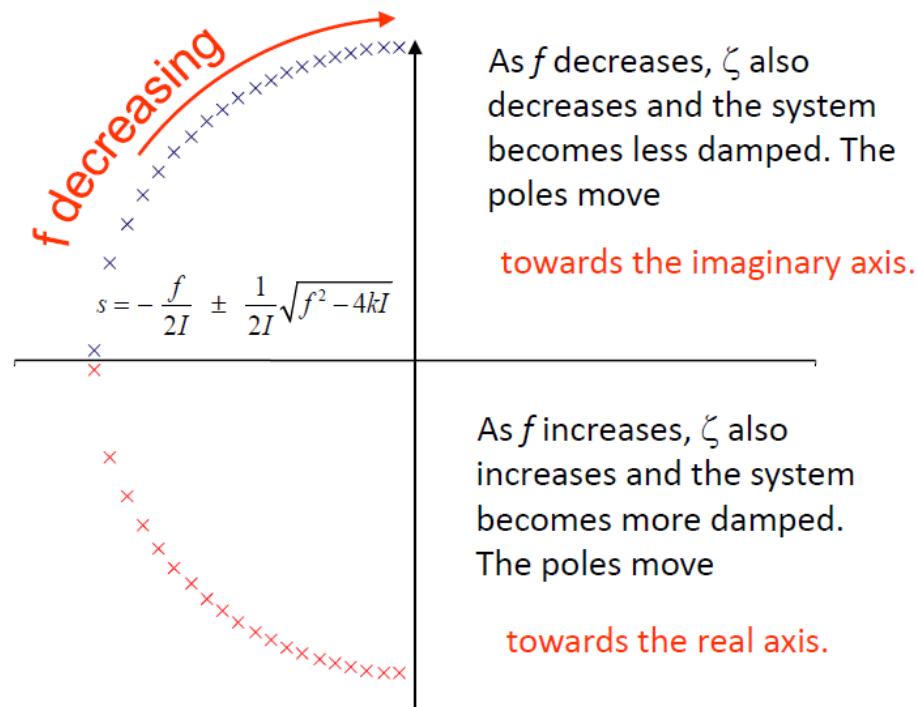


Figure 6.12:

Most servos are designed with a low friction coefficient in mind, but this could increase with wear overtime for example, and thus the behaviour of the system would change.

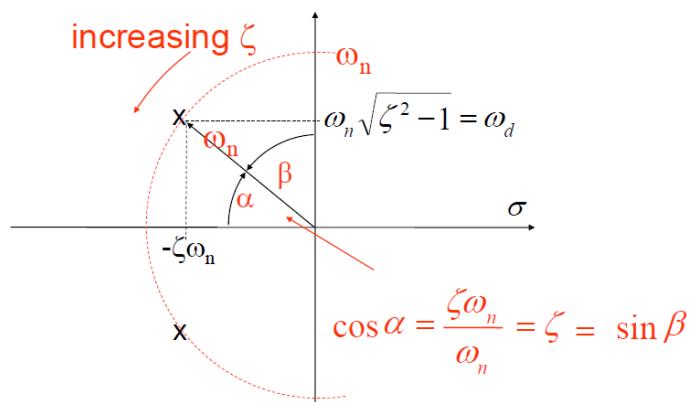


Figure 6.13:

Changing inertia

Unlike the friction term, the inertia affects both the real and imaginary components of the complex poles. In this case, increasing the mass decreases the damping and the natural frequency. $\zeta = \frac{f}{2kI}$ (damping decreases), $\omega_n = \sqrt{\frac{k}{I}}$ (natural frequency decreases).

$$s = -\frac{f}{2I} \pm \frac{1}{2I} \sqrt{f^2 - 4kI} \quad (6.24)$$

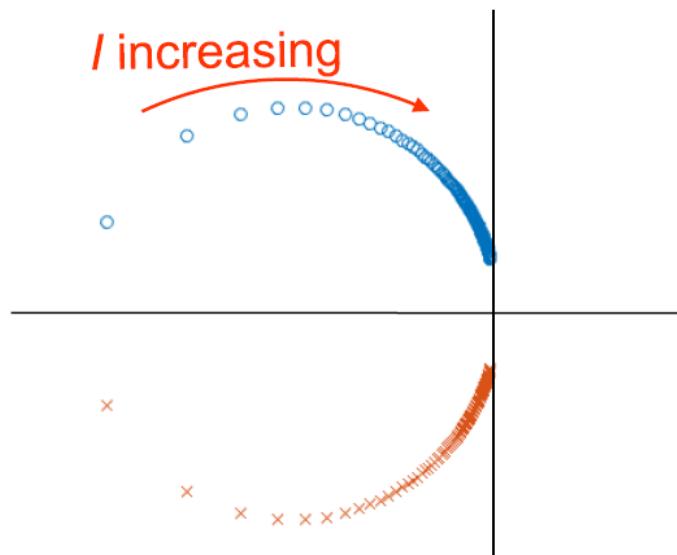


Figure 6.14:

6.3 Steady state performance

Previously we have looked at performance criteria for the transient response of a system: settling time, peak time, overshoot etc. In addition to this, it is important to know how accurately the control system tracks the demand once it has settled down, i.e. it is in the steady state. For example in the case of a ramp or step input:

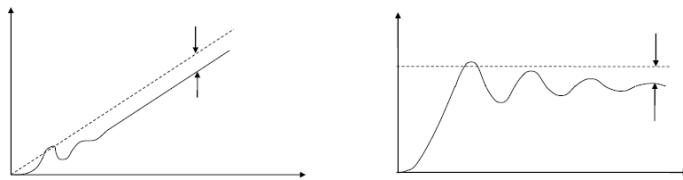


Figure 6.15:

Steady state does not imply systems is not in motion! - e.g. can be moving at constant velocity, or oscillating.

6.3.1 Final value theorem (FVT)

For this we can use the final value theorem, which takes advantage of some handy properties of the Laplace transform.

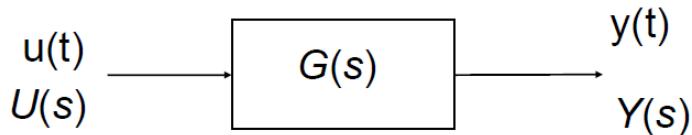


Figure 6.16:

For a constant input, the output will (for some systems) settle down to a steady value that is a multiple of the input. The final value theorem gives the final value reached as:

$$\lim_{t \rightarrow \infty} y(t) = \lim_{s \rightarrow 0} sY(s) \quad (6.25)$$

Note the extra s term in $\lim_{s \rightarrow 0} sY(s)$. So, the final value is found by setting s to zero in the Laplace representation of the output and multiplying by s . There are two checks performed in control theory which confirm valid results for the Final Value Theorem:

1. $Y(s)$ should have no poles in the right half the complex plane.
2. $Y(s)$ should have no poles on the imaginary axis, except at most one pole at $s = 0$.

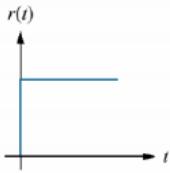
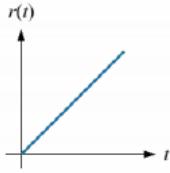
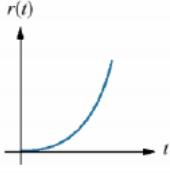
Waveform	Name	Physical interpretation	Time function	Laplace transform
	Step	Constant position	1	$\frac{1}{s}$
	Ramp	Constant velocity	t	$\frac{1}{s^2}$
	Parabola	Constant acceleration	$\frac{1}{2}t^2$	$\frac{1}{s^3}$

Figure 6.17:

6.3.2 Steady state error

We are normally more interested in the final value of the error rather than the output, as the goal of the controller is to drive the error as close to zero. Further

for some inputs such as ramp or sinusoidal input, the final output is not really meaningful.

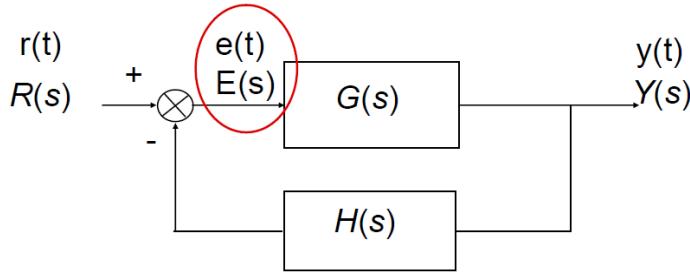


Figure 6.18:

So, we first need to write the error signal with respect to the input:

$$E(s) = R(s) - H(s)Y(s) \quad (6.26)$$

$$Y(s) = G(s)E(s) \quad (6.27)$$

So:

$$E = R - HGE \rightarrow E + HGE = R \quad (6.28)$$

$$E = \frac{R}{1 + GH} \quad (6.29)$$

Therefore, the final value of the error signal $e(t)$ would be:

$$\lim_{t \rightarrow \infty} e(t) = \lim_{s \rightarrow 0} sE(s) \quad (6.30)$$

with

$$E(s) = \frac{R(s)}{1 + G(s)H(s)} \quad (6.31)$$

so

$$\lim_{s \rightarrow 0} sE(s) = \lim_{s \rightarrow 0} \frac{sR(s)}{1 + G(s)H(s)} \quad (6.32)$$

6.3.3 Steady state position error

Using the final value theorem, it is possible to find the steady state error of the control system and thus judge the appropriateness/success of the controller without having to calculate the full dynamic response at all. Remembering back to how long that took with only a second order system, you can imagine how useful this is!

Step input demand

$$R(s) = \frac{A}{s} \quad (6.33)$$

$$E(s) = \frac{R(s)}{1 + G(s)H(s)} \quad (6.34)$$

Thus the final error is:

$$\lim_{t \rightarrow 0} e(t) = \lim_{s \rightarrow 0} \frac{sR(s)}{1 + G(s)H(s)} = \lim_{s \rightarrow 0} \frac{s \frac{a}{s}}{1 + G(s)H(s)} = \lim_{s \rightarrow 0} \frac{a}{1 + G(s)H(s)} = \frac{a}{1 + k_p} \quad (6.35)$$

where $k_p = \lim_{s \rightarrow 0} G(s)H(s)$ is the position error constant.

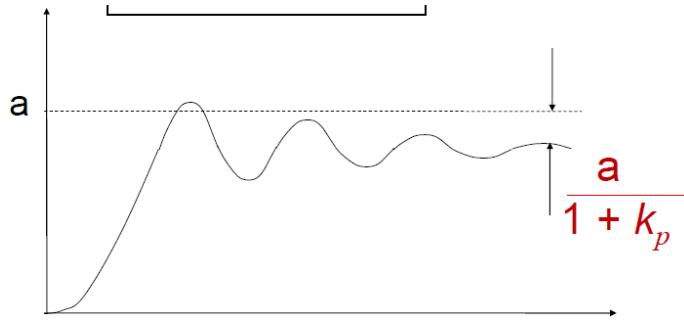


Figure 6.19:

6.3.4 Steady state velocity lag

Ramp input demand

$$R(s) = \frac{A}{s^2} \quad (6.36)$$

$$E(s) = \frac{R(s)}{1 + G(s)H(s)} \quad (6.37)$$

Thus the final error is:

$$\lim_{t \rightarrow \infty} e(t) = \lim_{s \rightarrow 0} \frac{sR(s)}{1 + G(s)H(s)} = \lim_{s \rightarrow 0} \frac{s \frac{a}{s^2}}{1 + G(s)H(s)} = \lim_{s \rightarrow 0} \frac{\frac{a}{s}}{1 + G(s)H(s)} \quad (6.38)$$

$$= \lim_{s \rightarrow 0} \frac{a}{s + sG(s)H(s)} = \frac{a}{k_v} \quad (6.39)$$

where $k_v = \lim_{s \rightarrow 0} sG(s)H(s)$ is the velocity error constant.

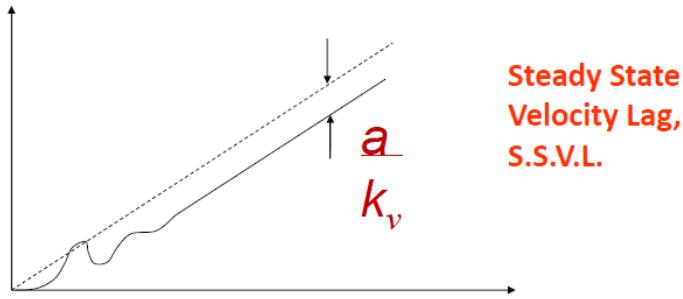


Figure 6.20:

Once again, let's return to the servo example to see how this system would perform given these inputs.

6.3.5 Servo steady state error

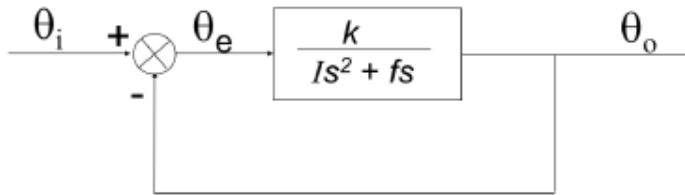


Figure 6.21:

Step input

For a unity gain the steady state error is given by $\frac{a}{1+k_p}$.

$$k_p = \lim_{s \rightarrow 0} G(s)H(s) \quad (6.40)$$

$$k_p = \lim_{s \rightarrow 0} G(s) = \frac{k}{0 \cdot (I \cdot 0 + f)} = \infty \quad (6.41)$$

$$\lim_{t \rightarrow \infty} \theta_e(t) = \frac{a}{1 + k_p} = 0 \quad (6.42)$$

So the position error is zero for all stable gains for a step input, i.e. the servo matches the demand completely.

Ramp input

Velocity error constant:

$$k_v = \lim_{s \rightarrow 0} sG(s) = \frac{k}{Is + f} = \frac{k}{f} \quad (6.43)$$

Steady state velocity error:

$$\lim_{t \rightarrow \infty} \theta(t) = \frac{a}{k_v} = \frac{af}{k} \quad (6.44)$$

This is the lag we saw in Lecture 5 when calculating the ramp response directly.

Chapter 7

PID Controllers

7.1 System types

Specifically, the position error constant k_p and velocity error constant k_v depend upon **the number of poles at the origin** in the open loop system. These correspond to roots in the denominator at $s = 0$, and represent a pure integration. Let's look more generally at the transfer function of our system and introduce the concept of system **type**. By factoring out any s terms from the denominator, we can write the transfer function in the following form:

$$G(s)H(s) = \frac{(s - z_1)(s - z_2)(s - z_3)\dots}{s^p(s - \sigma_1)(s - \sigma_2)(s - \alpha_k + j\omega_k)(s - \alpha_k - j\omega_k)\dots} \quad (7.1)$$

So, if there are p poles at the origin, the system is said to be a 'type p' system.

System type IS NOT THE ORDER OF THE SYSTEM!

For example, the servo motor can be expressed as:

$$\frac{\Theta_0(s)}{\Theta_i(s)} = \frac{k}{s(Is + f)} \quad (7.2)$$

which is a type 1 system. The electromagnet from lecture 2 can be expressed as:

$$\frac{I(s)}{V(s)} = \frac{1}{Ls + R} \quad (7.3)$$

which is a type 0 system. The mass spring damper system can be expressed as:

$$\frac{X(s)}{F(s)} = \frac{1}{ms^2 + cs + k} \quad (7.4)$$

which is a type 0 system.

The system type quickly tells us the form of the SSE for our inputs without manipulation of block diagrams or converting to time domain. This gives an indication of the design of the controller.

7.1.1 System type - calculating error

Essentially, for a zero error we want k_p and k_v to be infinite, or at least be a constant for a finite error, depending on our requirements.

$$SSE = \frac{a}{1 + k_p}, \quad SSVL = \frac{a}{k_v} \quad (7.5)$$

$$k_p = \lim_{s \rightarrow 0} G(s)H(s), \quad k_v = \lim_{s \rightarrow 0} sG(s)H(s) \quad (7.6)$$

7.1.2 System type for unit feedback control systems - calculating error

The **position error constant** k_p for a *type p system* is given by:

- $p > 0, k_p = \lim_{s \rightarrow 0} G(s) = \infty$ - no steady-state position error
- $p = 0, k_p$ is finite - finite position error

The **velocity error constant** for a *type p system* is given by:

- $p > 1, k_v = \lim_{s \rightarrow 0} sG(s) = \infty$ - no velocity error
- $p = 1, k_v$ is finite - steady state velocity lag
- $p = 0, k_v = 0$ infinite lag (completely fails to track)

No. Integrators in denominator = system TYPE	Input type		
	Step $r(t) = a$ $R(s) = \frac{a}{s}$	Ramp $r(t) = at$ $R(s) = \frac{a}{s^2}$	Acceleration $r(t) = \frac{at^2}{2}$ $R(s) = \frac{a}{s^3}$
0	$e_{ss} = \frac{a}{1+k_p}$	$e_{ss} = \infty$	$e_{ss} = \infty$
1	$e_{ss} = 0$	$e_{ss} = \frac{a}{k_v}$	$e_{ss} = \infty$
2	$e_{ss} = 0$	$e_{ss} = 0$	$e_{ss} = \frac{a}{k_a}$

Table 7.1:

where ∞ means the system never settles. The more poles there are at the origin of the open-loop system, the better steady-state tracking performance. However, pure integrations have a highly destabilising effect on the control system!

7.1.3 Type 0 servo - spring return

Previously we have looked at servos with inertia and damping only, without and compliance/springs. This gives a type-1 transfer function. However, when a spring is added, the system becomes type-0.

$$\frac{\Theta_0(s)}{\Theta_i(s)} = \frac{k}{s(Is + f)} \quad (7.7)$$

The question arises, why would you add a spring? Some systems are not bidirectional e.g. a single acting piston in a hydraulic system. A spring is needed to return the piston to the correct position in the cycle. A spring is often added to servo actuators to return to a given position when system is off. HVAC systems use this for fire safety. The transfer function of this servo now becomes:

$$\sum T = I \frac{d^2\theta_0}{dt^2} \quad (7.8)$$

$$I \frac{d^2\theta_0}{dt^2} = T_s + T_f + T_m \quad (7.9)$$

where T_s and T_f oppose the motion of the motor. In time domain:

$$I \frac{d^2\theta_0}{dt^2} = -k_s\theta_0 - f \frac{d\theta_0}{dt} + k_m\theta_i \quad (7.10)$$

In Laplace:

$$Is^2\Theta_0(s) = -k\Theta_0(s) - fs\Theta_0(s) + k_m\Theta_i(s) \quad (7.11)$$

$$Is^2\Theta_0(s) + k_s\Theta_0(s) + fs\Theta_0(s) = k_m\Theta_i(s) \quad (7.12)$$

$$\Theta_0(s)(Is^2 + k_s + fs) = k_m\Theta_i(s) \quad (7.13)$$

Transfer function:

$$\frac{\Theta_0(s)}{\Theta_i(s)} = \frac{k_m}{Is^2 + fs + k_s} = G(s) \quad (7.14)$$

So the closed loop transfer function is now:

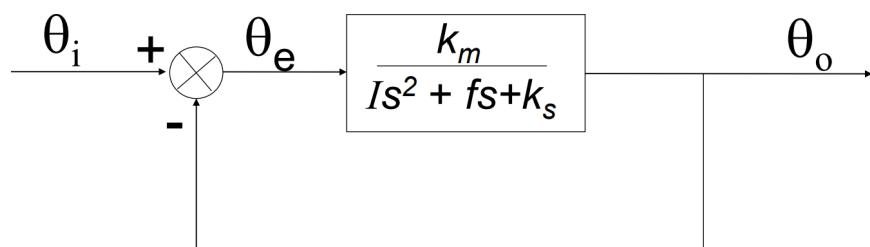


Figure 7.1:

$$F(s) = \frac{G(s)}{1 + G(s)} = \frac{k_m}{Is^2 + fs + k_s + k_m} \quad (7.15)$$

$$k_p = \lim_{s \rightarrow 0} G(s)H(s) = \frac{k_m}{Is^2 + fs + k_s} = \frac{k_m}{k_s} \quad (7.16)$$

$$k_v = \lim_{s \rightarrow 0} sG(s)H(s) = \frac{sk_m}{Is^2 + fs + k_s} = \frac{0}{k_s} \quad (7.17)$$

$$SSE = \frac{a}{1 + k_p} = \frac{ak_s}{k_s + k_m}, \quad SSVL = \frac{a}{k_v} = \infty \quad (7.18)$$

So now the servo will now only travel a fraction of the required distance for a step input, and will completely fail to track a ramp input, never settling to a steady state. We need an improved controller for these systems!

7.2 Proportional control

7.2.1 Proportional control - damping ratio trade off

We have seen the effect of damping ratio for our simple servo control system so far. We have two conflicting requirements:

- For transient response we want something around 0.6-0.8 to reduce overshoot and oscillations. $\zeta = \frac{f}{2\sqrt{kI}}$
- For steady state response, we want the damping ratio as low as possible at the expense of oscillations. $SSVL = \frac{af}{k} = \frac{2\zeta a}{\omega_n}$

For our type 1 servo, this trade off occurs if we want good ramp tracking, something we could possibly neglect in certain applications. However, the type 0 servo SSVL renders it close to unusable in most control applications (with the current controller). Commonly the challenge is to increase ζ , but with our current system, we do not have many options for adjustment.

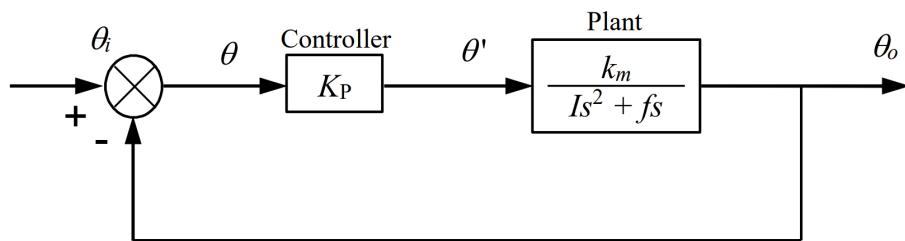


Figure 7.2:

We can either increase our friction f but this increases wear on the components, reducing the life of the system. So that leaves the controller gain k_p to adjust, we can reduce it at the expense of steady state performance. So we need a better controller!

7.3 Velocity feedback

7.3.1 Doing better - compensation methods

What the controller we have seen so far is known as a proportional controller, the control signal is directly proportional to the error signal, scaled by k_p . Compensation methods have been designed to improve **both** transient and steady state performance. One of the earliest and simplest compensation methods is to consider not just the position but the velocity too.

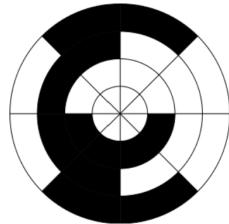


Figure 7.3:

If we take the position from the shaft encoder, and also record the change over time, we can measure the speed. The block diagram of this is just s - as this is the differential operator, we also include another gain here k_{vel} so we can choose the relative contributions of the two feedbacks. So we have the following block diagram.

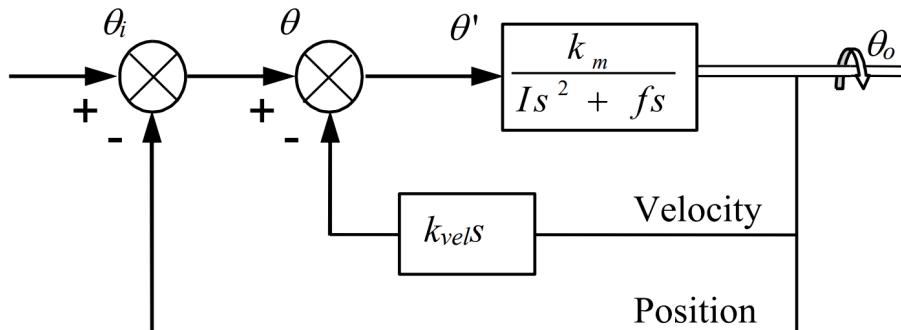


Figure 7.4:

Inner comparator:

$$\theta' = \theta - k_{vel}s\theta_0 \quad (7.19)$$

Outer comparator:

$$\theta = \theta_i - \theta_0 \quad (7.20)$$

If we calculate the new closed loop transfer function - work outwards!

$$\frac{\theta_0}{\theta} = \frac{G}{1 + GH} = F(s) \quad (7.21)$$

$$F(s) = \frac{\frac{k_m}{Is^2 + fs}}{1 + \frac{k_m}{Is^2 + fs} k_{vel}s} = \frac{k_m}{Is^2 + fs + k_m k_{vel}s} \quad (7.22)$$

$$= \frac{k_m}{Is^2 + (f + k_m k_{vel})s} \quad (7.23)$$

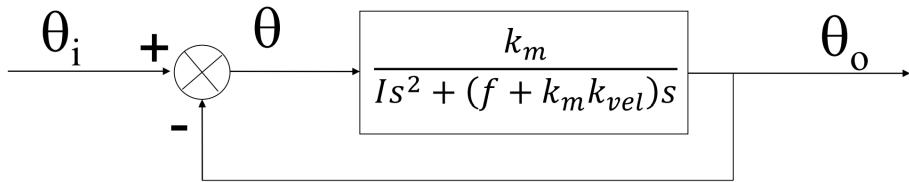


Figure 7.5:

Inner loop transfer function:

$$F(s) = \frac{k_m}{Is^2 + (f + k_m k_{vel})s} \quad (7.24)$$

Total closed loop transfer function:

$$F_T(s) = \frac{F}{1 + F} \quad (7.25)$$

$$F_T(s) = \frac{k_m}{Is^2 + (f + k_m k_{vel})s + k_m} \quad (7.26)$$

$$F_T(s) = \frac{\frac{k_m}{I}}{s^2 + \frac{(f+k_m k_{vel})}{I}s + \frac{k_m}{I}} \quad (7.27)$$

$$\omega_n = \sqrt{\frac{k_m}{I}}, \quad \zeta = \frac{f + k_m k_{vel}}{2\sqrt{kI}} \quad (7.28)$$

So by adding the extra loop, we are able to keep the natural frequency the same, *but increase the damping ratio*. This means we can greatly improve the transient response compared to proportional control, where we can only scale the motor gain k_m through adjusting the parameter k_p . This is evident when considering the position of the poles of the system on the s-plane. Adjusting the motor gain using proportional control (k_p) can only move the poles closer to the real axis. But using velocity feedback (k_{vel}), it is possible to move the poles away from the imaginary axis, thus reducing the decay time.

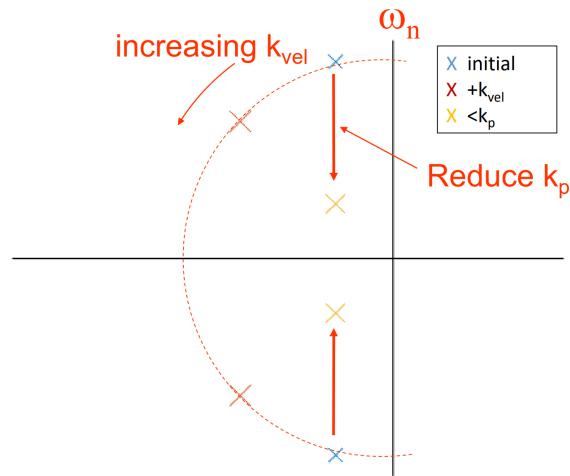


Figure 7.6:

In this example, the initial zeta was $\zeta = 0.3$, in order to achieve the practical estimate of $\zeta = 0.7$ with proportional control, we have to increase both the rise time and settling time. However, with velocity feedback we can achieve a vastly improved transient response.

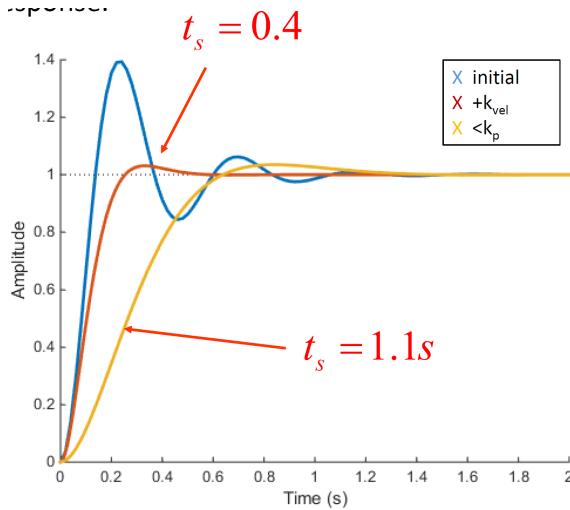


Figure 7.7:

7.3.2 Velocity feedback - disadvantages

This improved transient response comes at the expense of the steady state error, the expression for the SSVL becomes:

$$k_v = \lim_{s \rightarrow 0} sF(s) = \frac{s k_m}{I s^2 + (f + k_m k_{vel}) s} = \frac{k_m}{f + k_m k_{vel}} \quad (7.29)$$

$$SSVL = a \frac{k_m}{f + k_m k_{vel}} \quad (7.30)$$

Compare this SSVL to the one we had previously $\left(\frac{af}{k}\right)$. Using this feedback system allows for the designer to trade off the transient and steady state response dependent upon the requirements of the system. However, using the velocity in this manner is very sensitive to errors or noise from the sensor, which give large instantaneous changes in the error signal. To combat this, a low pass filter is normally employed in the feedback loop before the derivative term.

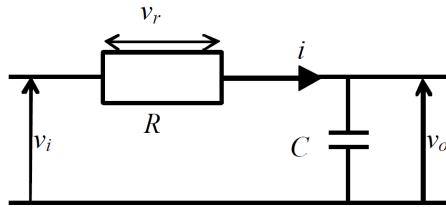


Figure 7.8:

$$= \frac{1}{RCs + 1} \text{ see lecture 2} \quad (7.31)$$

This makes the response more complex and slightly reduces the improvements in transient response obtained.

7.3.3 Alternative controllers

The velocity feedback is rather specific to servomechanisms with a shaft encoder. It is not always possible to have a sensor or a measurement from which both the variable and the derivative can be acquired simultaneously.

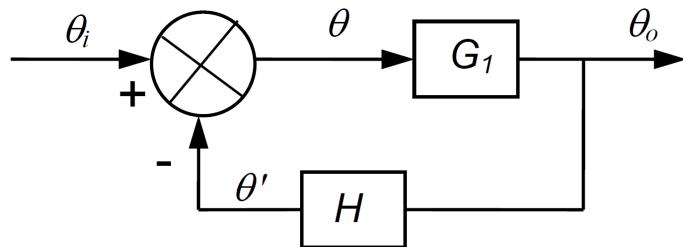


Figure 7.9:

Instead, we can consider what happens to the *error* across time - something necessarily present in all closed loop systems!

7.4 PID control

PID = Proportional, Integral and Derivative controllers: These controllers consist of three terms, all applied to the error signal. They are *incredibly* common in

engineering, accounting for nearly 90% of all use cases.

7.4.1 Working example: PID control for servo-mechanism

These can be modelled as shown below, with a DC motor rotating an inertial load, with a friction or damping resistance.

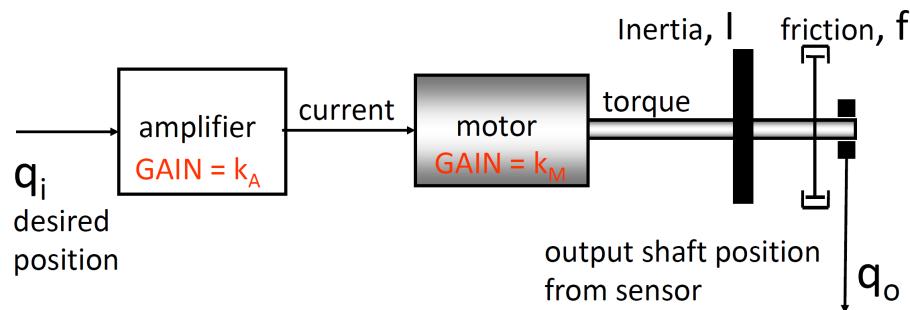


Figure 7.10:

7.4.2 Proportional error

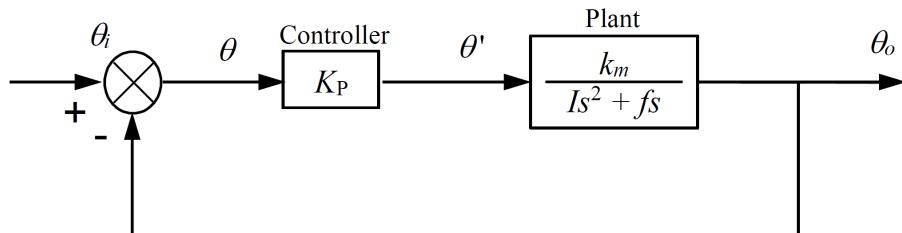


Figure 7.11:

$$F(s) = \frac{G(s)}{1 + G(S)} = \frac{k_p k_m}{Is^2 + fs + k_p k_m} \quad (7.32)$$

We have already seen this in our previous analysis of second order systems, and from changing k_m from last week. To summarise: low values of k_p give stable but slow responses, and high SSVL. High values reduce SSVL but response overshoots considerably. With proportional control the error *and thus the control signal* does not reach zero until the motor is at the desired position.

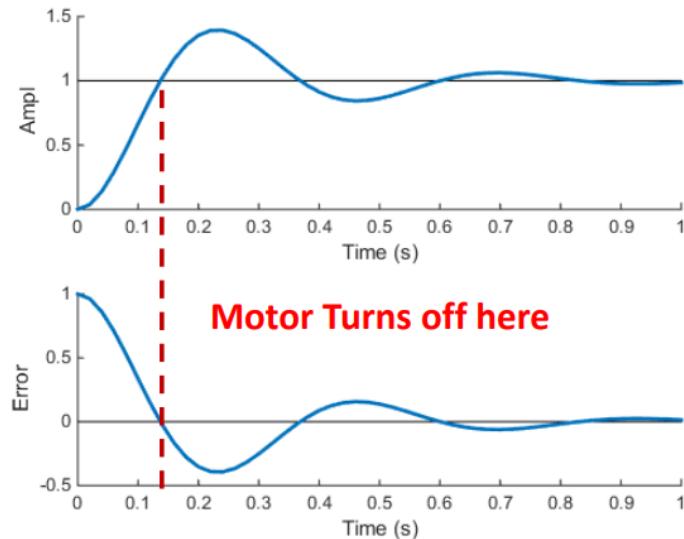


Figure 7.12:

However, due to the inertia for the system, the motor continues to move even without a control signal, resulting in an overshoot.

7.4.3 PID control

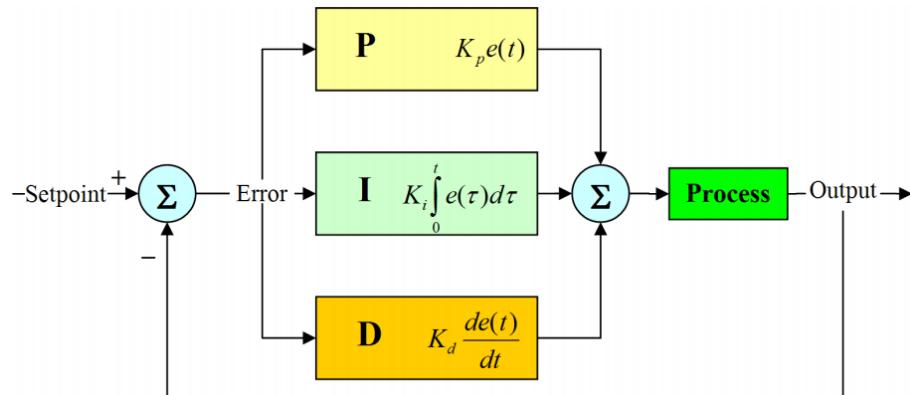


Figure 7.13:

7.4.4 Derivative error

We can replicate this by considering the **derivative** of the error.

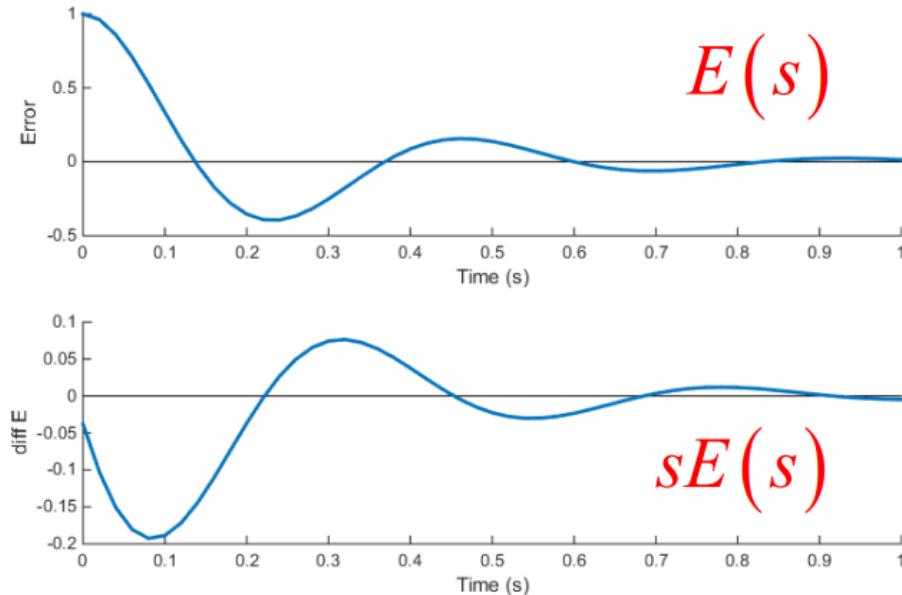


Figure 7.14:

This is negative as the servo approaches the target, so offers a way of restraining the servos forward motion. Lets consider a derivative controller with unity proportional gain.

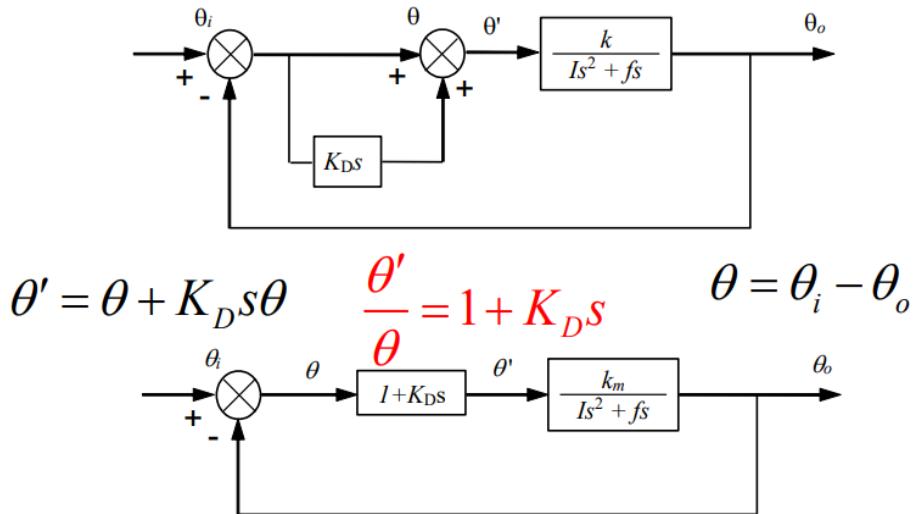


Figure 7.15:

The open loop transfer function now becomes type 1:

$$G' = \frac{(1 + K_D s) k_m}{I s^2 + f s} \quad (7.33)$$

Giving a *closed loop transfer function* of:

$$F(s) = \frac{G'(s)}{1 + G'(s)} = \frac{(1 + K_D s) k_m}{I s^2 + f s + (1 + K_D s) k_m} \quad (7.34)$$

$$F(s) = \frac{k_m K_d s + k_m}{I s^2 + (f + k_m K_D) s + k_m} \quad (7.35)$$

System is still second order but now there is a zero in the numerator. The effect of this is subtle compared to the change in the poles. More importantly, lets consider the SSVL of this system:

$$k_v = \lim_{s \rightarrow 0} s G'(s) = \frac{s (1 + K_D s) k_m}{s (I s + f)} = \frac{k_m}{f} \quad (7.36)$$

So, unlike velocity feedback, the SSVL is unchanged by derivative error. So we can improve transient response without compromising steady state error. The effect on the step response is similar to that of the minor loop velocity feedback.

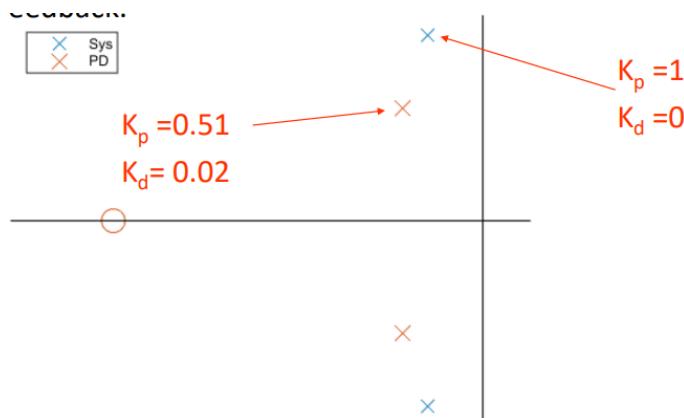


Figure 7.16:

The poles of the system become less oscillatory and decay quicker. This combined with a reduced proportional gain, gives an improved transient response. This is clear when looking at the step response, and the relative contributions of the two error terms.

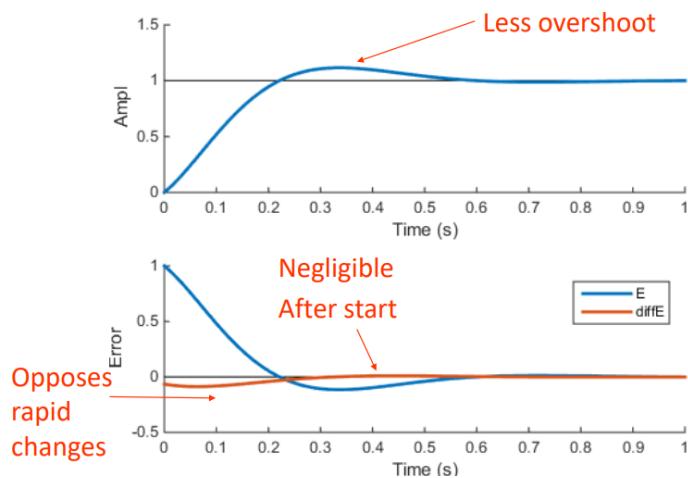


Figure 7.17:

At the start, the derivative term is significant *and in the opposite direction to* the proportional error term, but becomes negligible as the system settles.

7.4.5 PID control

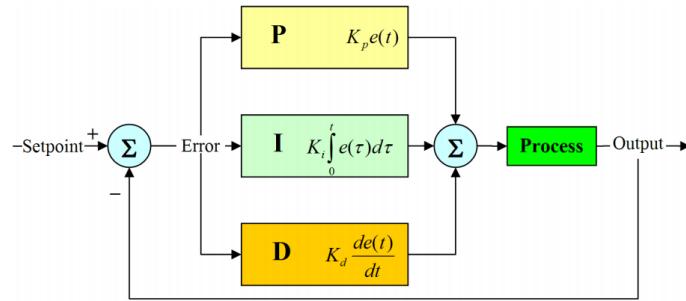


Figure 7.18:

7.4.6 Integral error

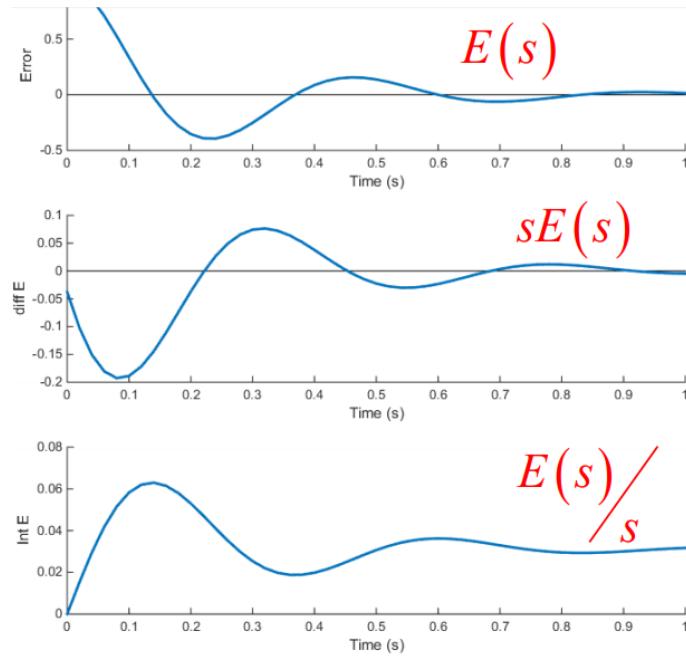


Figure 7.19:

Further, if we consider the **integral** or total error over time. This gradually increases over time, and can be used to magnify the control signal for small errors, and improve SSE. The integral of the error is used for correcting steady state errors, as it adds a pole at zero in the transfer function. This increases type.

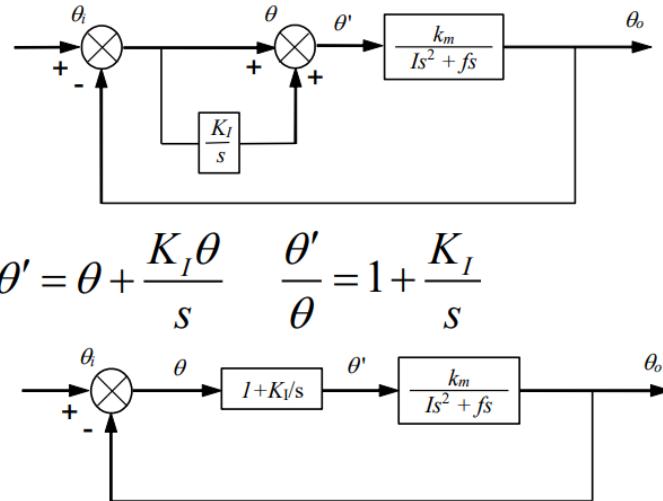


Figure 7.20:

$$G' = \frac{\left(1 + \frac{K_I}{s}\right) k_m}{Is^2 + fs} = \frac{\frac{1}{s}(s + K_I) k_m}{s(Is + f)} \quad (7.37)$$

$$G' = \frac{k_m s + k_m K_I}{s^2 (Is + f)} \quad (7.38)$$

With a closed loop transfer function of:

$$F(s) = \frac{G'(s)}{1 + G'(s)} = \frac{k_m s + k_m K_I}{s^2 (Is + f) + k_m s + k_m K_I} \quad (7.39)$$

$$F(s) = \frac{K - ms + k_m K_I}{Is^3 + fs^2 + k_m s + k_m K_I} \quad (7.40)$$

Eq.7.40 is a third order transfer function. Increasing the system type improves the steady state response:

$$k_v = \lim_{s \rightarrow 0} sG'(s) = \frac{s(k_m s + k_m K_I)}{s^2 (Is + f)} \quad (7.41)$$

$$k_v = \frac{k_m K_I}{0} = \infty \quad (7.42)$$

$$\text{SSVL} = \frac{a}{k_v} = 0 \quad (7.43)$$

We now have no velocity lag! The effect of an extra pole close to the origin - which would make our system *very slow* - is largely cancelled out by the nearby zero.

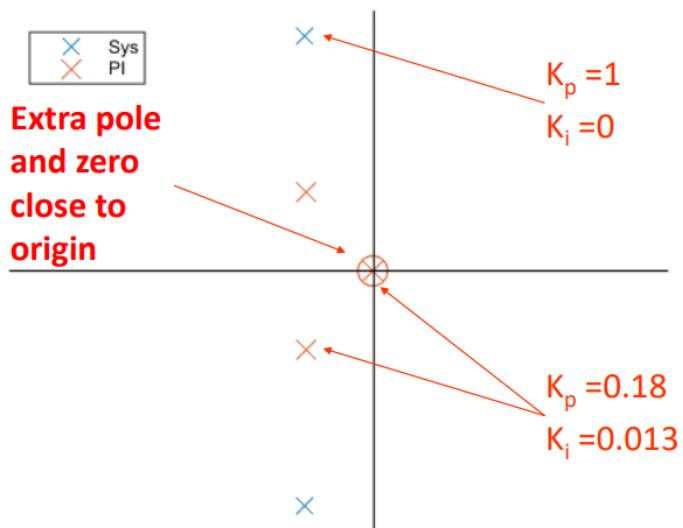


Figure 7.21:

This means the system is broadly similar to a proportional controller, with the exception of improved steady state performance.

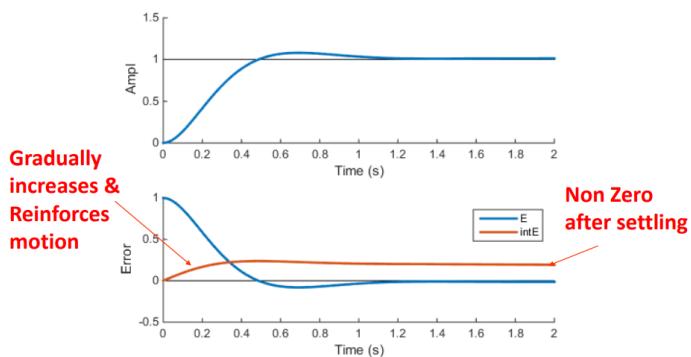


Figure 7.22:

We can see that the integral term has enabled proper tracking of a ramp input.

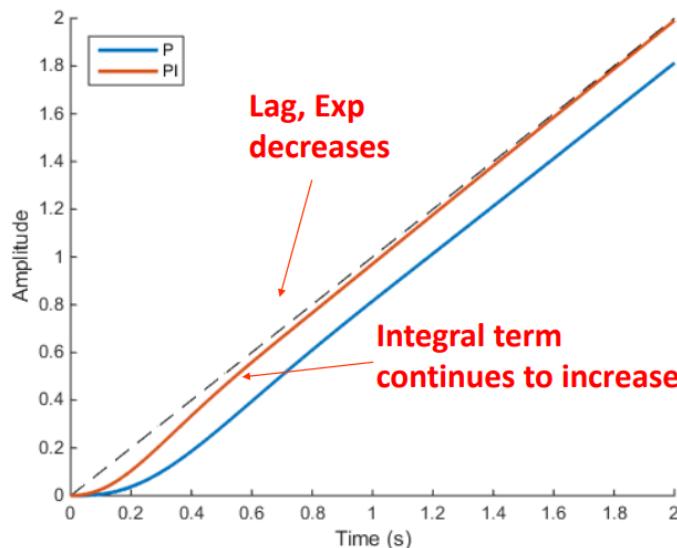


Figure 7.23:

This is useful for overcoming steady state errors arising from nonlinearities not included in the model which prevent proper tracking in reality.

7.4.7 PID control

Putting all three of these controllers together gives the complete PID.

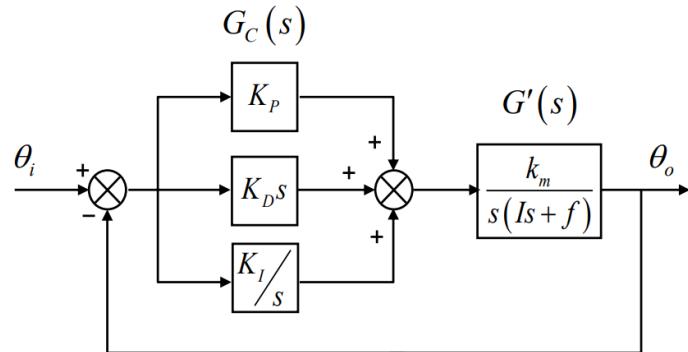


Figure 7.24:

$$G_c(s) = K_P + K_Ds + \frac{K_I}{s} = \frac{K_Ds^2 + K_Ps + K_I}{s} \quad (7.44)$$

Open loop gain:

$$G'(s) = \frac{k_m (K_Ds^2 + K_Ps + K_I)}{s^2 (Is + f)} \quad (7.45)$$

The closed loop transfer function is then:

$$F(s) = \frac{k_m (K_D s^2 + K_P s + K_I)}{s^2 (I s + f) + k_m (K_D s^2 + K_P s + K_I)} \quad (7.46)$$

$$F(s) = \frac{k_m K_D s^2 + k_m K_P s + k_m K_I}{I s^3 + f s^2 + k_m K_D s^2 + k_m K_P s + k_m K_I} \quad (7.47)$$

$$F(s) = \frac{k_m K_D s^2 + k_m K_P s + k_m K_I}{I s^3 + (f + k_m K_D) s^2 + k_m K_P s + k_m K_I} \quad (7.48)$$

Thus, by choosing the appropriate values of K_P , K_D and K_I , it is possible to design a controller with improved transient response **and** decreased/no steady state error (in theory). There is no unique solution for the settings of the three gains, and so the values are dependent upon the specific system and the application.

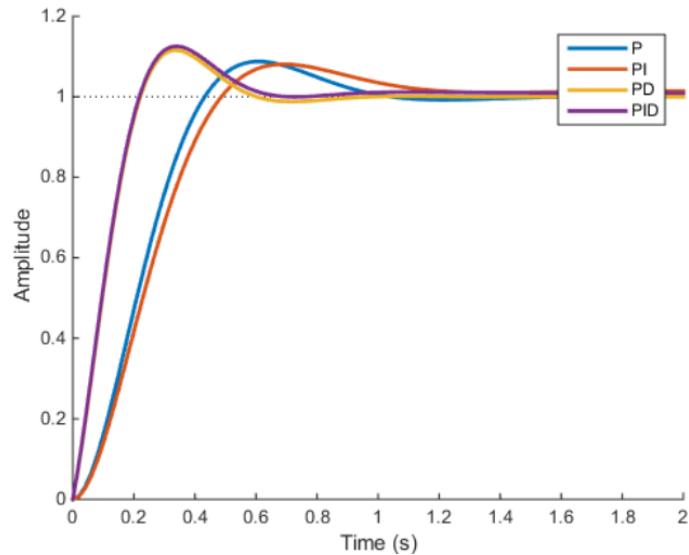


Figure 7.25:

Selecting these parameters is known as **tuning**, and it was (and still is in machine learning circles) a very active area of research. These controllers are used everywhere, and can even be applied with acceptable results with little or no knowledge of the system being controlled!

Part II

Instrumentation

Chapter 8

General Measurement Systems

A measurement system has to be devised such that the relationship between the real value of a variable and the value actually measured is unambiguously known. For example, when placing a weight on a scale, we must know the relationship between the **true** weight and the **measured** weight. The measurement system must allow:

- Easy interpretation of the measured data
- Provide high degree of confidence

8.1 Transducers

A transducer converts the sensed variable into a detectable signal form. Sometimes the device changes a mechanical quantity into a change in an electrical quantity. For example, a strain gauge converts a change in strain in the specimen to a change in electrical resistance in the gauge. Another example is a thermometer which converts thermal expansion of the liquid (due to a rise in temperature) into a mechanical translation.

8.2 Signal condition circuits

This takes the transducer signal and can convert, compensate or manipulate it into a more usable electrical quantity. This may include filters, compensators, modulators, demodulators, integrators or differentiators. For example, a Wheatstone bridge used with the strain gauge converts the change in the electrical resistance of the gauge to a change in voltage.

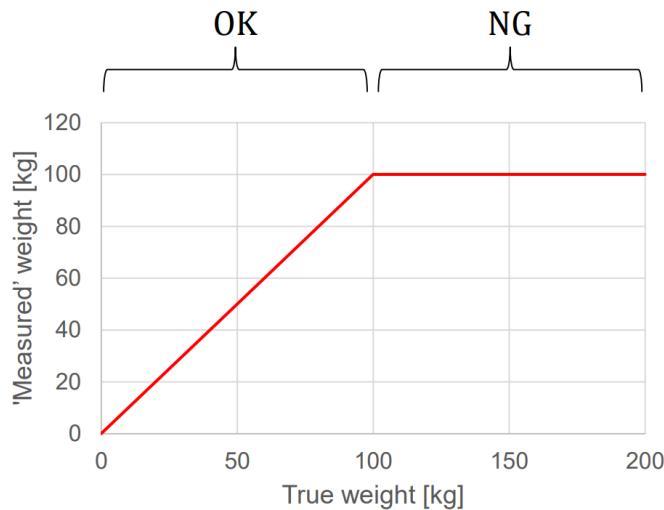


Figure 8.1: The scales work well up to 100kg, however it may not be able to measure heavier weights effectively.

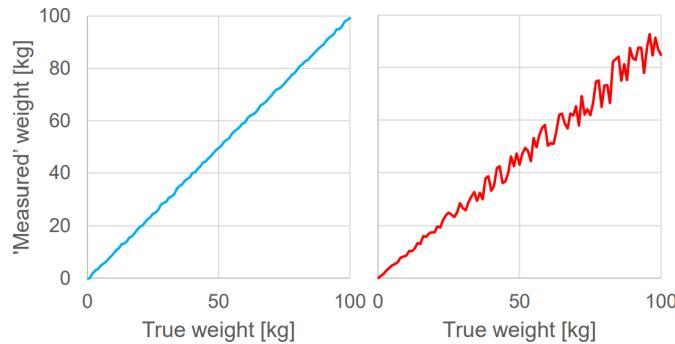


Figure 8.2: On the red graph we can see the effect that noise has on our measurement - a source of an inaccuracy. This must be dealt with to have an effective measurement system.

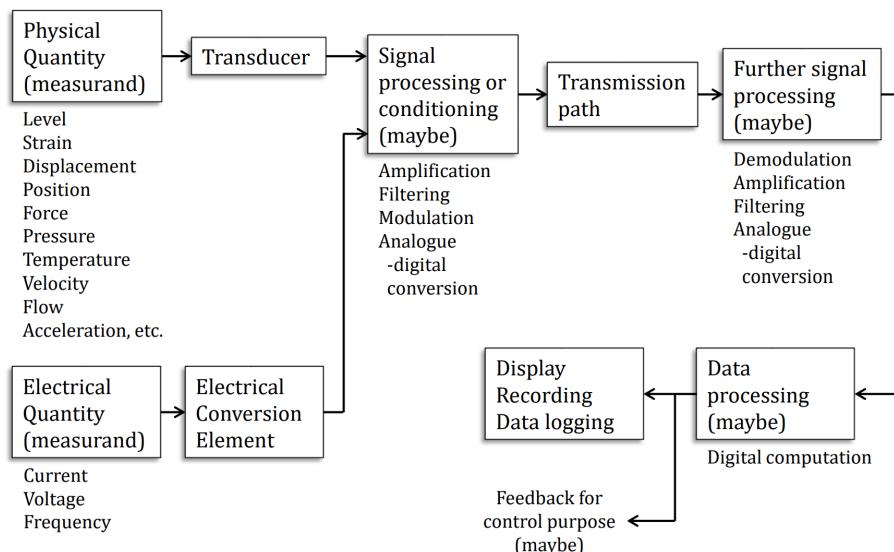


Figure 8.3: Block diagram to show a general measurement system.

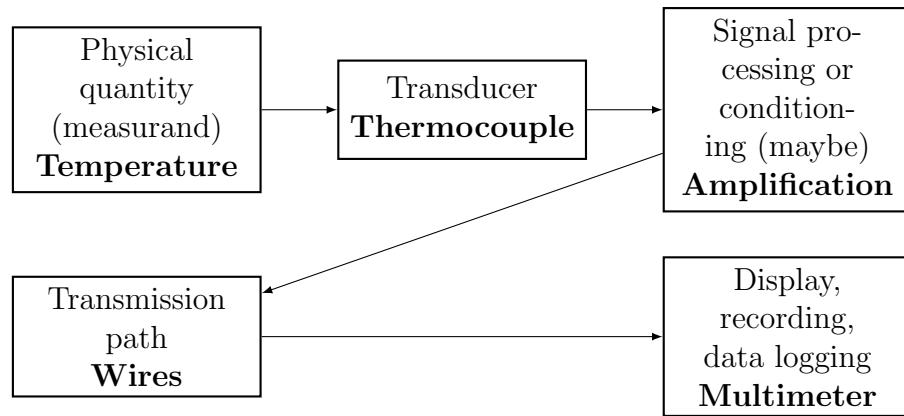


Figure 8.4: How the workflow of a measurement can be constructed.

8.3 Amplifiers

They are required in systems when the output from the transducer-signal conditions is small. Gains of 10 - 1000 are used to increase the levels of the signal, typically a millivolt or less, to what is compatible with the voltage-measuring devices used in the system. A negative-feedback amplifier circuit is an example of an amplifier.

8.4 Output stage

Generally, it is a voltage measurement device that is used to display the measurement in a form that can be read and interpreted. For example, digital voltmeters, self-balancing potentiometers, oscilloscopes, chart recorders and magnetic tape recorders.

8.5 Feedback-control stage

Used when the measurement system is employed in process control. The signal from the measurement system is compared with the command signal that reflects the required value of the quantity in the process. The process-controller forms the difference between these two and produces an error signal. The error signal is then used to automatically adjust the process. For example, the float system in a toilet (called a ballock) to control the water supply in the tank.

8.6 Calibration

How do we know whether one metre is one metre? We compare it with a reference, which is supposed to be reliable. This act of checking is called calibration. Our

references are the measuring **standards**.

8.7 Standards

There are seven fundamental quantities of the International Measuring System.

- Length m - metre
- Time s - second
- Mass kg - kilogram
- Temperature K - kelvin
- Electric Current A - ampere
- Luminous Intensity - cd - candela
- Amount of Substance - mol - mole

Units and standard for all other quantities are derived from these.

We can look at how the definition of the metre has changed over the past 2 centuries for some insight into the accuracy of our reference.

- 1792 - a ten millionth of the quarter of a meridian ($0.5 - 0.1\text{mm}$ of uncertainty)
- 1889 - a platinum-iridium bar at melting point of ice ($0.2 - 0.1\mu\text{m}$ of uncertainty)
- 1983 - the length of the path travelled by light in vacuum during a time interval of $1/299792458$ of a second (0.1nm of uncertainty)

Here are how the other fundamental quantities are defined.

- Metre - Length of the path travelled by light in a vacuum in $1/299792458$ of a second
- Time - Duration of 9192631770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the caesium 133 atom
- Mass - Equal to the mass of the International Prototype of the Kilogram
 - Equal to $h / (\text{Metre standard}^2 / \text{Time standard}) \dots$ since 2019 h : Plank constant (in $e = hf$, $h = 6.626 \times 10^{-34}\text{kg m}^2\text{s}^{-1}$)

- Temperature - $1/273.16$ of the thermodynamic temperature of the triple point of water (exactly $0.01\text{ }^{\circ}\text{C}$)
- Electric current - The flow of electric charges through a surface at the rate of one coulomb per second

The fundamental standards are protected by such organisations such as the National Institute of Standard and Technology (NIST) in the US. They set up the National Reference Standards and below these in accuracy, the Working Standards (and more below these).

8.8 Units and dimensions

- Dimension - a physical variable that is used to describe some aspect of a physical system
- Unit - a measure of a dimension. Primary standards are used to form the exact definition of a unit

SI units are now commonly used worldwide. The fundamental dimensions are:

- Length [L]
- Mass [M]
- Force [F] ($[\text{MLS}^{-2}]$)
- Time [τ]
- Temperature [T]

Chapter 9

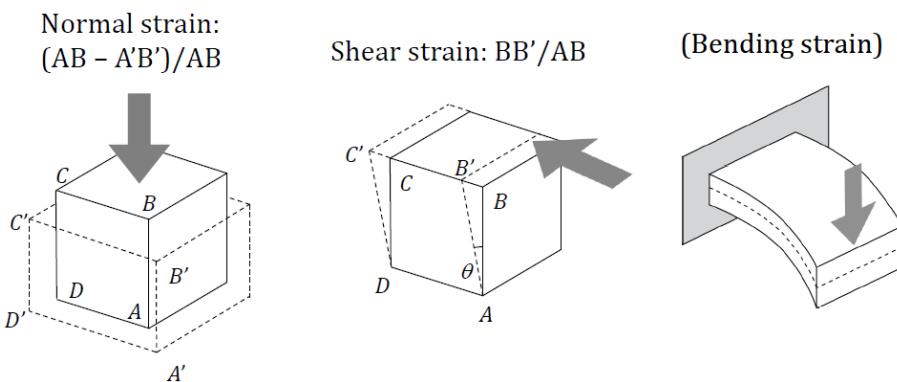
Strain Gauges

12/10/2020

9.1 Strain gauges

9.1.1 Strain

The changes in the value of a dimension of a body divided by the original value of the dimension is the relative change in the dimensions.



Information about strain is required in many engineering situations: an aircraft in flight, support pillars and spans of bridges, etc.

9.1.2 Stress

Density of the reactive (internal) forces distributed throughout the body (force per unit area), in response to external force. Types of stress:

- Tensile/compressive stress
- Shear stress
- Bending stress

Elastic modulus:

$$E = \frac{\text{stress}}{\text{strain}} \quad (9.1)$$

- Young's modulus - normal strain (and if relationship is linear)
- Shear modulus - shear strain
- Bulk modulus - volumetric strain

9.1.3 Strain gauges

A device which experiences a change of electric resistance when it is strained. A strain gauge is combined with other electrical components to obtain an electric voltage or current representing tensile, compressive or bending strain, together with the means of displaying or recording its value. The total resistance of a block of conducting material of uniform cross section A and length l :

$$R = \rho \frac{l}{A} \quad (9.2)$$

Where ρ is resistivity with units $\Omega \text{ m}$.

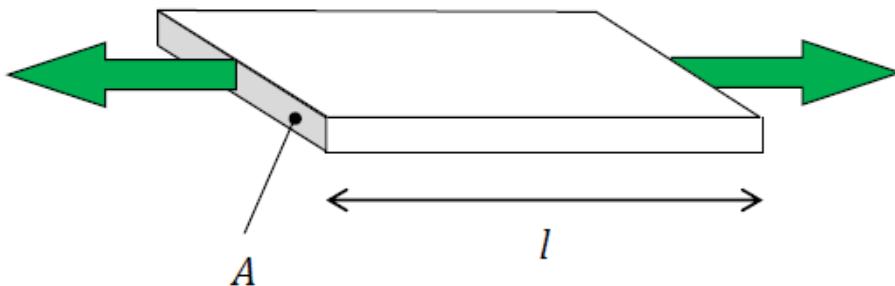


Figure 9.1: When the block is subjected to a tensile stress, its length will increase and cross-sectional area will decrease, both of which cause the resistance to increase according to equation (9.2)

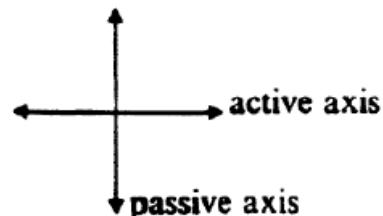
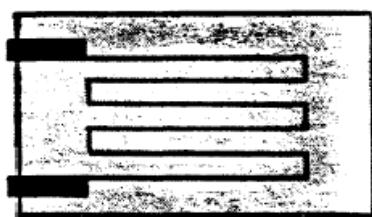
The resistivity of the material will also change because of the **piezo-resistive effect** (increase with tension and decrease with compression). The resistance of the block can thus be written as:

$$R' = (\rho + \delta\rho) \frac{l + \delta l}{A - \delta A} \quad (9.3)$$

Note: compressive stress will result in a decrease in total resistance.

9.1.4 Simple wire strain gauge

A long fine wire is folded to fit in a small area and then mounted on a flexible backing sheet, usually paper. When **firmly stuck** to the surface of a much more rigid body, any deformation of this body will cause an identical fractional change of the strain gauge wire. The change of resistance for any strain along the active axis will be much greater than for an equal strain occurring along the passive axis.



Single-active-axis strain gauge

9.1.5 Gauge factor

Defined as the fractional change of resistance of the gauge, divided by the fractional change in the length of the gauge along the active axis:

$$G = \frac{\frac{\Delta R}{R}}{\frac{\Delta l}{l}} \quad (9.4)$$

Since $\frac{\Delta l}{l}$ is the strain e in the body to which the gauge is fixed, this can be rewritten as:

$$\frac{\Delta R}{R} = eG \quad (9.5)$$

Most gauges have a G between **1.8 to 2.2**, depending on the gauge material and the magnitude of the piezo-resistive effect.

9.1.6 Foil strain gauge

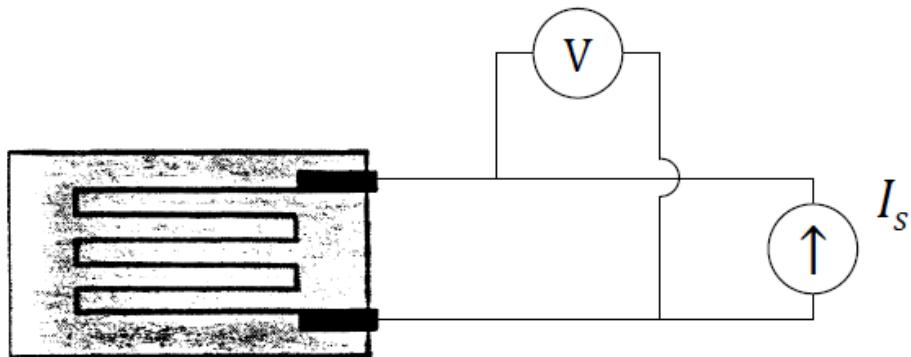
Most modern strain gauges are formed by rolling out a thin foil of the resistive material and then cutting away parts of the foil by a photo-etching process, to create the required grid pattern.

- Usually supplied with think backing paper as electrical insulation
- Adhesive layer for the fixation to the specimen should be creep-free and allow heat dissipation

Advantages over simple wire strain gauges:

- Larger surface area results in a larger area of adhesion
- Accurate reproducibility due to photo-etching technique
- Small dimensions mean they are good for localised strain, fit well to curvature

9.1.7 How can we measure strain



$$\Delta R/R = eG$$

Figure 9.2: Is this good enough?

Typically $e = 10^{-3}$, $G = 2.1$, $R = 120\Omega$, $I_S < 50\text{mA}$.

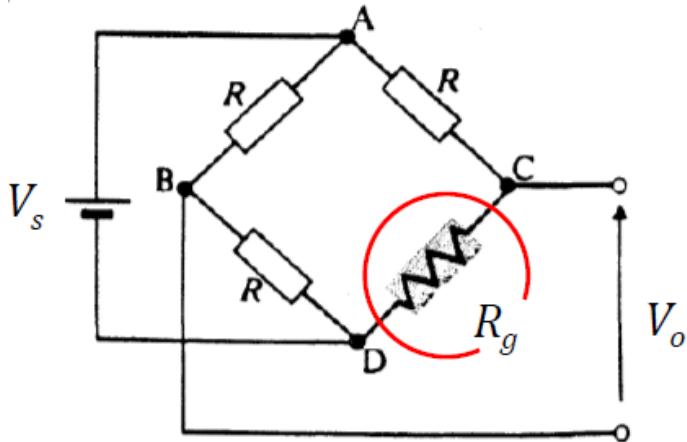
$$V = 6\text{V}, \Delta R = 0.252\Omega \quad (9.6)$$

$$\Delta V = 0.013\text{V} \quad (9.7)$$

The fractional change of the voltage is too small to detect comparing to the 'baseline' voltage ($= RI_S$).

9.2 Wheatstone bridge

Used to convert the change of resistance in the strain gauge into an electrical signal which could be used to indicate the value of strain. With this, **zero strain \leftrightarrow zero output voltage**.



Here,

$$V_C - V_D = \frac{V_S R_G}{R + R_G} \quad (9.8)$$

and then,

$$V_0 = V_C - V_B = \left(V_D + \frac{V_S R_G}{R + R_G} \right) - \left(V_D + \frac{V_S}{2} \right) = \frac{V_S R_G}{R + R_G} - \frac{V_S}{2} \quad (9.9)$$

R is chosen to have the same value as the unstrained gauge, i.e. $R = R_G$ when zero strain $\rightarrow V_0 = 0$ when zero strain. When the gauge is strained, $R_G = R + r$ and the output voltage is

$$V_0 = V_S \frac{R + r}{2R + r} - \frac{V_S}{2} = V_S \frac{r}{2(2R + r)} \quad (9.10)$$

Now using the gauge factor, $\frac{r}{R} = eG$ and the equation can be rewritten:

$$V_0 = V_S \frac{ReG}{2(2R + ReG)} = V_S \frac{eG}{2(2 + eG)} \approx \frac{1}{4} V_S eG \quad (9.11)$$

Because typically $eG < 0.02 \ll 2$

Sensitivity:

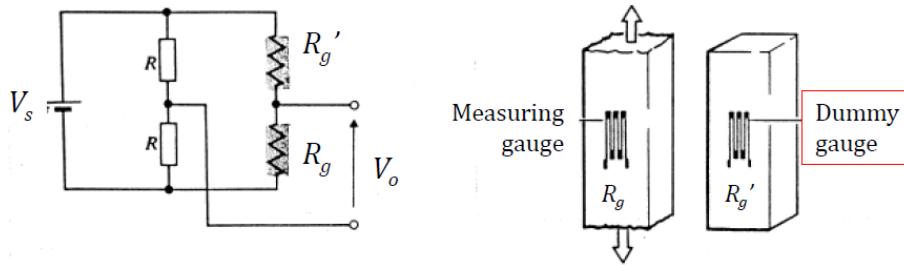
$$\frac{V_0}{e} = \frac{1}{4} V_S G \quad (9.12)$$

9.2.1 Temperature compensation

The output voltage from the strain gauge and the bridge also depends on the temperature. Changes in temperature will cause changes in:

- Dimensions of the specimen and hence the gauge due to thermal expansion
- Dimensions of the gauge itself (particularly thickness)
- Resistivity of the gauge material

Compensated with a dummy gauge:



The dummy gauge is:

- Identical to the measuring gauge
- Mounted on an unstressed specimen of the same material

The unstressed specimen should be placed as close to the strained member, in order to make it work accurately. Now, consider resistance of a strain gauge R_g :

- Changes due to tensile strain: $R_g \rightarrow R(1 + x) = R(1 + eG)$
- Changes due to temperature: $R_g \rightarrow R(1 + y)$

The measuring gauge and dummy gauge resistances are then:

$$R_g = R(1 + x)(1 + y) \quad (9.13)$$

$$R'_g = R(1 + y) \quad (9.14)$$

The output voltage of the bridge is:

$$V_0 = V_s \frac{R_G}{R_G + R'_G} - \frac{V_s}{2} \quad (9.15)$$

Substituting for R_G and R'_G , we have:

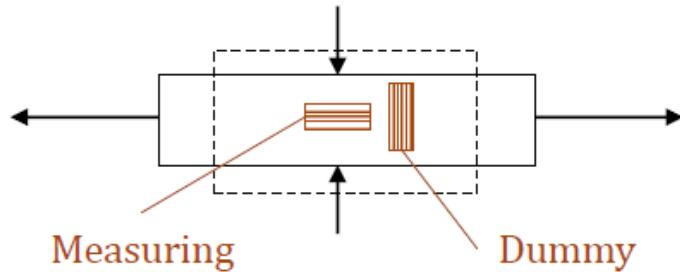
$$V_0 = V_s \frac{R(1 + x)(1 + y)}{R(1 + x)(1 + y) + R(1 + y)} - \frac{V_s}{2} \quad (9.16)$$

$$\frac{V_0}{V_s} = \frac{R(1 + x)(1 + y)}{R(1 + x)(1 + y) + R(1 + y)} - \frac{1}{2} \leftarrow \text{no influence by } y \quad (9.17)$$

However, for this you must find the specimen of the same material and keep them under the same temperature - are these easy enough? The dummy gauge could be placed on the same member as the measuring gauge with its active axis in the direction normal to that of the strain.

9.2.2 Effect of Poisson's ratio on a dummy gauge

Consider both measuring and dummy gauges are attached to the same specimen but perpendicular to each other.



Deformation of the specimen elongates the measuring gauge and shortens the dummy gauge, and the ratio of strains in the two directions is Poisson's ratio, $v = e_{\text{lateral}}/e_{\text{longitudinal}}$, which is between 0.25 to 4 for most materials. Ignoring temperature changes, the measuring gauge and dummy gauge resistances are then

$$R_g = R(1 + x) \quad (9.18)$$

$$R'_g = R(1 - \nu x) \quad (9.19)$$

The bridge output is:

$$V_0 = V_S \frac{R_g}{R_g + R'_g} - \frac{V_S}{2} \quad (9.20)$$

Substituting for R_g and R'_g , we have:

$$V_0 = V_S \frac{R(1 + x)}{R(1 + x) + R(1 - \nu x)} - \frac{V_S}{2} \quad (9.21)$$

$$\frac{V_0}{V_S} = \frac{R(1 + x)}{R(1 + x) + R(1 - \nu x)} - \frac{1}{2} \quad (9.22)$$

$$\frac{V_0}{V_S} = \frac{(1 + \nu)}{2[2 + (1 - \nu)x]} \approx \frac{1}{4}(1 + \nu)eG \quad (9.23)$$

Because typically $(1 - \nu)x \ll 2$. Comparing to the single gauge, $\frac{V_0}{V_S} = \frac{1}{4}eG$, sensitivity is increased by a factor $(1 + \nu)$ due to the effect of Poisson's ratio on the dummy gauge.

9.3 Practical Aspects of Strain Gauge Measurement

9.3.1 Strain gauge rosette

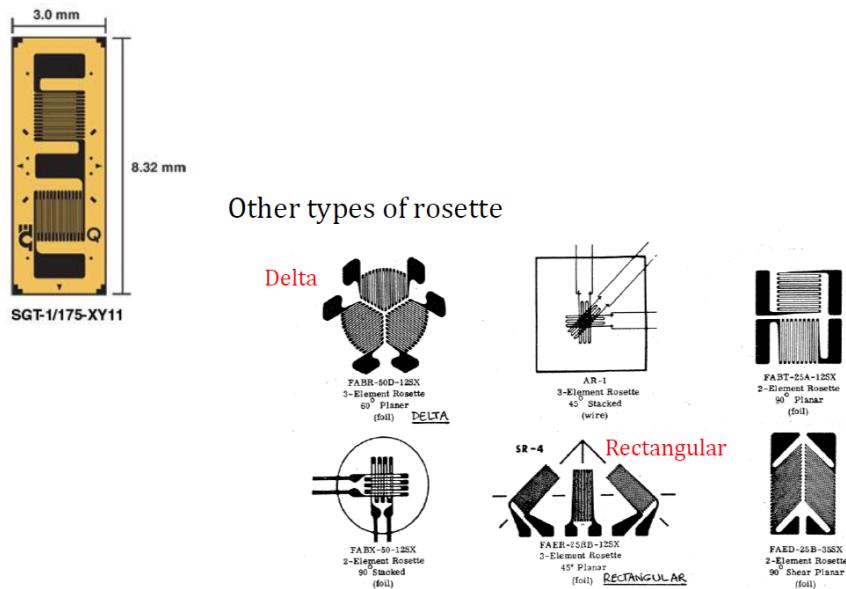
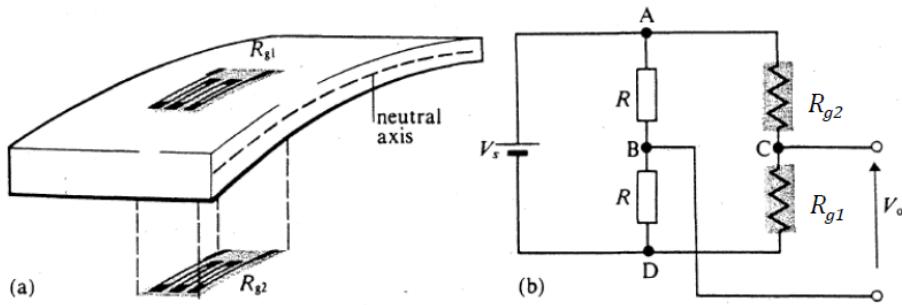


Figure 9.3: An example of a foil strain gauge having two elements with their active axis perpendicular to each other is shown in the top left (Omega Engineering Inc. USA). Some other types of rosette are also shown (figures courtesy of BLH Electronics, Waltham, Mass.)

How can be these useful? Strain (stress) in real-world applications is, not uniformly distributed and/or the principal stress axis is unknown. We use strain gauge rosettes to measure strain in three orientations. Understand complex stress/strain state (principal stress calculation is a topic of MECH0013 in term 2).

9.3.2 Bending Strain

To measure the strain of a member in which only bending is occurring, i.e. no additional tensile or compressive stresses present, we attach two strain gauges one on either side of its neutral axis with their active axes along the length of the member.



If the gauges are equidistant from the neutral axis, the tensile strain imposed on one will be equal to the compressive strain in the other, and the change of the resistance is the same magnitude but in opposite directions. Similarly to the previous cases, consider the fractional change of resistance \$x\$ and then output voltage.

$$V_0 = V_s \frac{R_{g1}}{R_{g1} + R_{g2}} - \frac{V_s}{2} \quad (9.24)$$

$$= V_s \frac{R(1+x)}{R(1+x) + R(1-x)} - \frac{V_s}{2} = \frac{1}{2} V_s x \quad (9.25)$$

Since \$x = eG\$, \$V_0 = \frac{1}{2} V_s eG\$ is a measure of bending occurring in the body.

9.3.3 Tensile or compressive strain in a bending member

The arrangement shown below allows to measure any tensile or compressive strain in the member, while ignoring any bending strain.

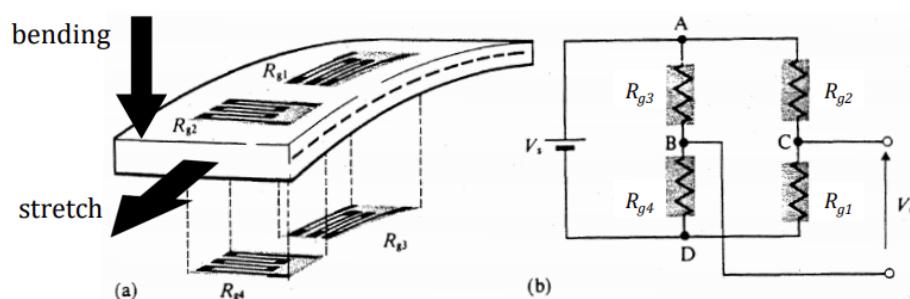


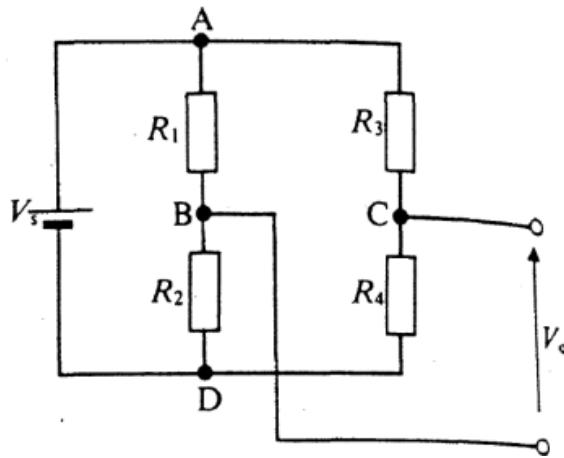
Figure 9.4: Variation of gauge resistance and output voltage in the configuration above.

	\$R_{g1}\$	\$R_{g2}\$	\$R_{g3}\$	\$R_{g4}\$	\$V_0\$
Under pure stretch	++	-	++	-	++
Under pure bending	++	-	--	+	\$\pm 0\$

9.3.4 Bridge balancing

In reality, resistance in the arms of the Wheatstone bridge slightly differ in value because of:

- Manufacturing variations
- Temperature differences between gauges and resistors
- Static strain occurring in a member

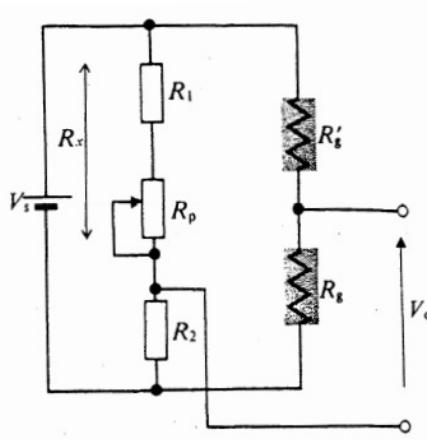


If the bridge is perfectly balanced:

$$\frac{R_4}{R_3 + R_4} = \frac{R_2}{R_1 + R_2} \quad (9.26)$$

$$\rightarrow \frac{R_3}{R_4} + 1 = \frac{R_1}{R_2} + 1 \rightarrow \frac{R_3}{R_4} = \frac{R_1}{R_2} \quad (9.27)$$

The imbalance needs to be adjusted before the measurement (adjustment of the output voltage to zero) → bridge balancing.



One possible way to achieve this is to connect a pot (potentiometer) in series with R_2 . The total resistance $R_x = R_1 + R_p$ should be adjusted such that:

$$\frac{R_x}{R_2} = \frac{R'_g}{R_g} \quad (9.28)$$

9.3.5 Semiconductor strain gauge

In some crystalline materials such as germanium and silicon, the piezo-resistive effect is very large. If a slice of such a crystal is used as a strain gauge, very large gauge factors can be obtained, which could range from 100 to 300.

$$\text{Gauge factor: } G = \frac{1}{e} \frac{\Delta R}{R} \quad (9.29)$$

The magnitude of the piezo-resistive effect, which determines the sensitivity of the gauge, depends on the minute quantities of impurity introduced into the base material before formation of the crystal.

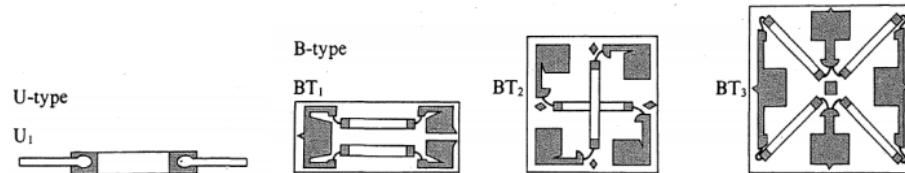


Figure 9.5: Some examples of semiconductor strain gauges

Characteristics of semiconductor strain gauges:

- The material of the strain gauge reaches its elastic limit at about 4000 microstrain (4000×10^{-6}), much smaller than that of metals (typically 20000 microstrain)
- The gauge factor G varies with high strain at high strain levels. i.e. the gauge is not linear
- The gauge factor varies with temperature
- The temperature coefficient of resistivity is large

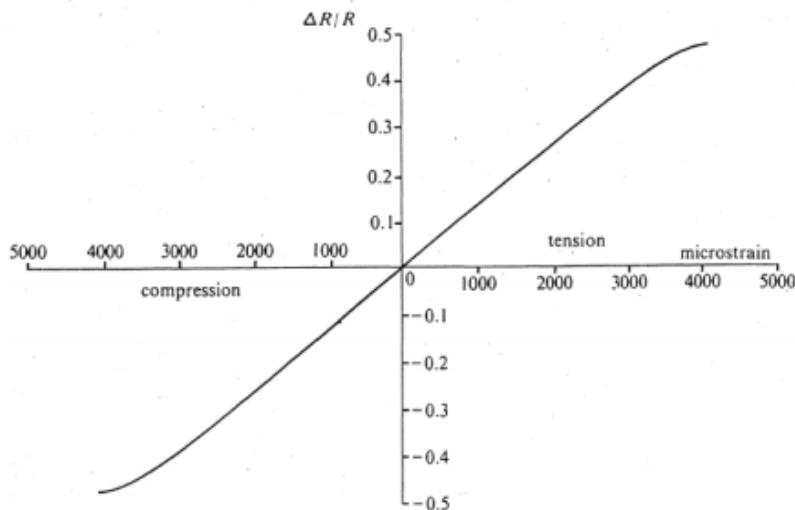
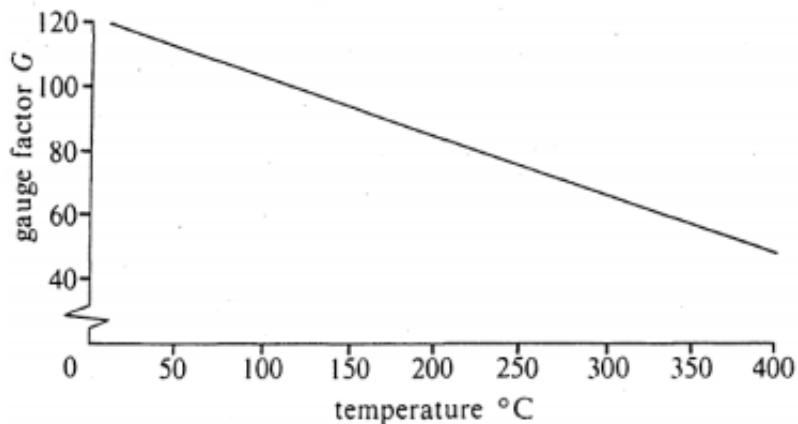


Figure 9.6: Change of resistance ratio with strain for a semiconductor strain gauge



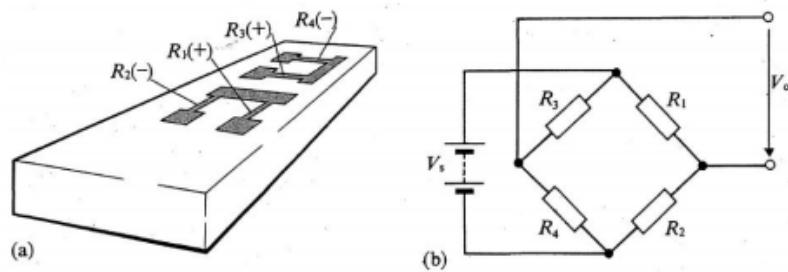
Negative effect of the temperature sensitivity of this type of strain gauge can be significantly reduced by using two gauges each consisting of two crystals connected in series. When two crystals are connected in series (but aligned in parallel), tensile or compressive strain applied to a gauge will produce an increase in the resistance of one crystal and a decrease in the other. Because the base material of each crystal is the same, their temperature coefficients of resistivity are approximately equal. In the configuration below, if the body experiences a tensile strain,

$$R_1 = R(1 + x) \quad (9.30)$$

$$R_2 = R(1 - x) \quad (9.31)$$

$$R_3 = R(1 - \nu x) \quad (9.32)$$

$$R_4 = R(1 + \nu x) \quad (9.33)$$



The bridge output voltage will thus be:

$$V_o = \frac{1}{2} V_s x (1 + \nu) \quad (9.34)$$

Even if temperature increases, that affects all crystals equally. Also, changes in dimensions will cause $R_1 \uparrow$, $R_2 \downarrow$, $R_3 \uparrow$, $R_4 \downarrow$. Hence, no unbalance occurs.

9.4 Summary

1. Strain gauges utilise variability of resistance due to deformation
2. Strain gauges are used in Wheatstone bridges with various gauge arrangements
3. Bridge balance calculation is an essential part of the module

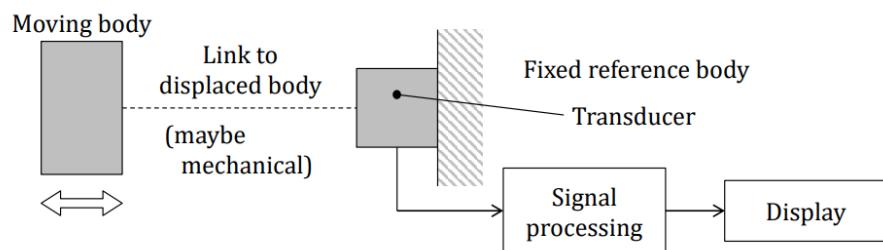
Chapter 10

Measurements of Displacement

26/10/2020

Displacement measurements:

- Most frequent in manufacturing and process-control applications
- Work as a secondary transduction process in many transducers for other measurements such as force, torque, pressure, flow and density.
- Mostly requires a stationary reference point.
- Utilises a wide variety of techniques and principles.



10.1 Capacitance-Based Displacement Transducers

10.1.1 Capacitive Displacement Transducers

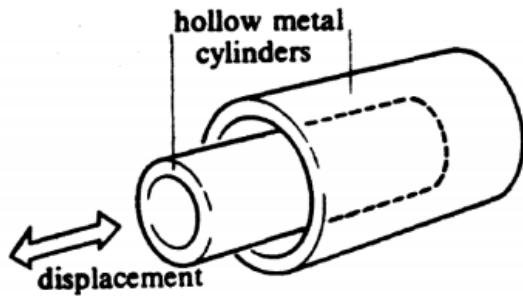
Displacement changes the capacitance of the transducer. The capacitance of a parallel-plate capacitor is given by:

$$C = \frac{\epsilon_0 A}{d} \quad (10.1)$$

Where:

- A is the area of the plates
- d is the separation between the plates
- ϵ_0 is the permittivity of free space ($8.854 \cdot 10^{-12} Fm^{-1}$)

Example: The capacitance of the two metal cylinders is proportional to the area of overlap.



Variable capacitance can be introduced by an insulator between the plates. If the insulation is as thick as the plate separation, the overall capacitance is:

$$C = \frac{\epsilon_0 A_1}{d} + \frac{\epsilon_0 \epsilon_r A_2}{d} \quad (10.2)$$

Where:

- ϵ_r is the relative permittivity of the insulator, or dielectric as it is commonly known.

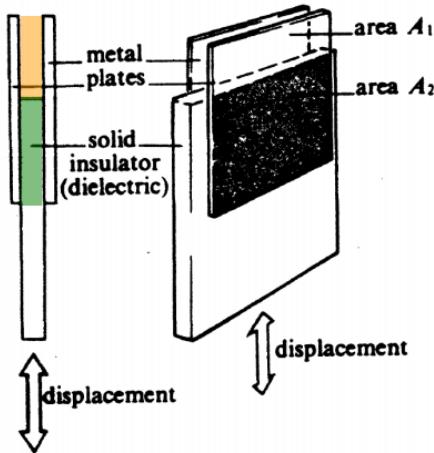


Figure 10.1: The orange part of the figure represents the 1st bit of the capacitance equation, when no insulator is inserted. The green part of the figure shows the 2nd part of the equation, where an insulator is introduced

10.1.2 Measurement of Capacitance

Suppose the capacitance of a capacitive transducer is a function of a displacement. When a constant voltage v is applied, charge is given as:

$$q = Cv \quad (10.3)$$

$$C = f(x) = \frac{q}{v} \quad (10.4)$$

It is impossible to measure q without disturbing it. If however, the voltage is made to vary, then q must vary too, and its rate of change, current i , can be measured:

$$\frac{dq}{dt} \left[\frac{C}{s} \right] = i[A] \quad (10.5)$$

Differentiating $q = Cv$ with respect to time gives:

$$\frac{dq}{dt} = i = C \frac{dv}{dt} \quad (10.6)$$

If the applied voltage is sinusoidal, i.e. $v = V_0 \sin(\omega t)$, then:

$$i = C \frac{dv}{dt} = V_0 \omega \cos(\omega t) \quad (10.7)$$

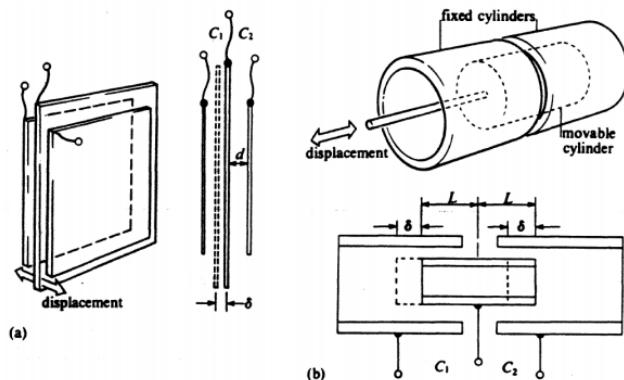
The amplitude of the current, i , equals to $V_0 \omega C$ (proportional to the capacitance).

Changes in capacitance caused by changes in displacement can be measured by determining the AC current flowing through the capacitor. However, in some cases in practice, the capacitance change due to the displacement is only a small fraction of the total capacitance and difficult to be measured:

$$C_{total} = C_{transducer} + C_{cable} + C_{stray} \quad (10.8)$$

10.1.3 Differential Capacitance Displacement Transducers

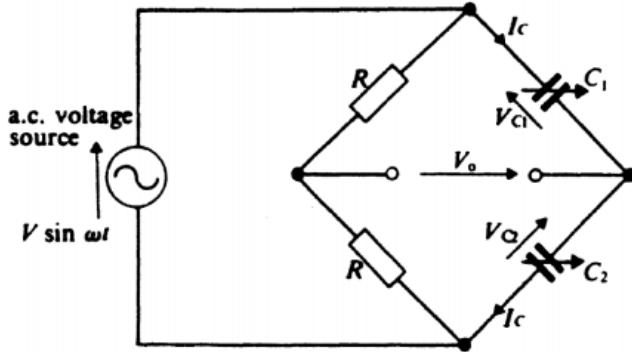
One way to get around the problem is to use either one of the differential forms of variable-separation or variable-area type transducer.



In each of the cases, two capacitors are formed which have equal capacitances when the transducer is geometrically centred.

10.1.4 Capacitive Bridge

The two capacitors are connected to a capacitive bridge. Here, difference in capacitance caused by the displacement unbalances the bridge and an AC output voltage V_o occurs.



To derive V_o , we start with the voltages across each capacitor:

$$V_{C1} = \frac{I_C}{\omega C_1} \quad (10.9)$$

$$V_{C2} = \frac{I_C}{\omega C_2} \quad (10.10)$$

This is because, if we consider voltage and current across/through a capacitor to be $v_C = V_C \sin(\omega t)$ and $i_C = I_C \cos(\omega t)$,

$$i_C = I_C \cos(\omega t) = C \frac{dv_C}{dt} = C \frac{d(V_C \sin(\omega t))}{dt} = \omega C V_C \cos(\omega t) \quad (10.11)$$

$$\therefore V_C = \frac{I_C}{\omega C} \quad (10.12)$$

Now, knowing the voltages across each capacitor V_{C1} and V_{C2} , the voltage across C2 as a fraction of the bridge energising voltage, V is:

$$\frac{V_{C2}}{V} = \frac{V_{C2}}{V_{C1} + V_{C2}} \quad (10.13)$$

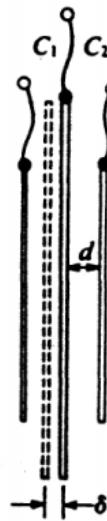
$$= \frac{\frac{I_C}{\omega C_2}}{\frac{I_C}{\omega C_1} + \frac{I_C}{\omega C_2}} \quad (10.14)$$

$$= \frac{C_1}{C_1 + C_2} \quad (10.15)$$

In the variable-separation type, the capacitances must be equal if the plate is centred. If the balance capacitance is C_0 , then:

$$C_0 = \frac{\epsilon_0 A}{d} \quad (10.16)$$

where d is the balance separation. When the plate is displaced by a distance δ to the left as in the figure, we have:



$$C_1 = \frac{\epsilon_0 A}{d - \delta} = C_0 \frac{d}{d - \delta} \quad (10.17)$$

$$C_2 = C_0 \frac{d}{d + \delta} \quad (10.18)$$

Substituting these for the equation of the voltage across C_2 yields:

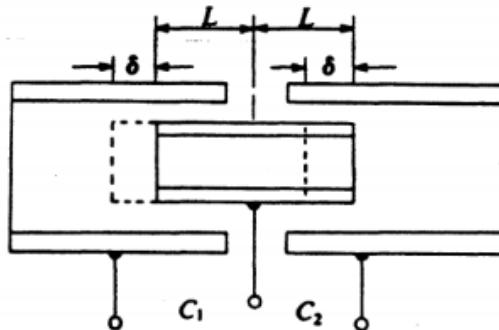
$$\frac{V_{C2}}{V} = \frac{d + \delta}{(d - \delta) + (d + \delta)} = \frac{d + \delta}{2d} \quad (10.19)$$

The voltage across each resistor is half the energising voltage, so the amplitude of the output voltage is:

$$V_o = V_{C2} - \frac{1}{2}V = \left(\frac{d + \delta}{2d} - \frac{1}{2} \right) V = \frac{\delta V}{2d} \quad (10.20)$$

There is a linear relationship between the voltage output and the displacement, despite the non-linear capacitance-displacement relationship.

For the variable-area type, the capacitances are proportional to their active areas, so if the movable cylinder is displaced to the left, we have:



$$C_1 = \frac{L + \delta}{L} C_0 \quad (10.21)$$

$$C_2 = \frac{L - \delta}{L} C_0 \quad (10.22)$$

Substituting these for the equation of the voltage across C_2 yields:

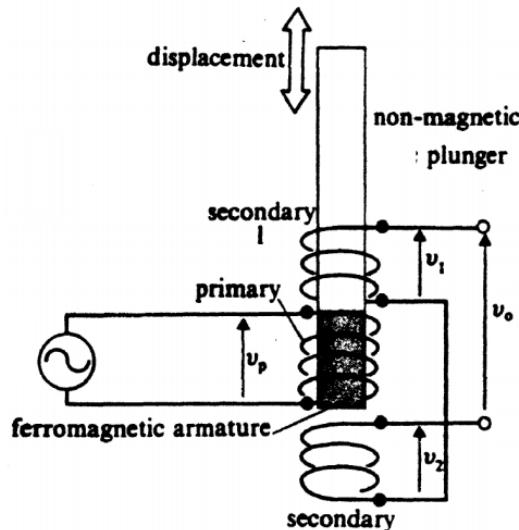
$$V_o = V_{C2} - \frac{1}{2}V = \frac{\delta V}{2L} \quad (10.23)$$

In both cases, the output voltage can be expressed as:

$$V_o = \frac{1}{2}xV \quad (10.24)$$

10.2 Inductance-Based Displacement Transducers

10.2.1 Linear Variable Differential Transformer (LVDT)



LVDT is one of the magnetic displacement transducers. An LVDT consists of:

- Primary coil
- Two secondary coils
- Ferromagnetic armature (iron core)
- An AC energising source
- Non-magnetic plunger

- and AC voltages are induced in the secondary coil by transformer action.

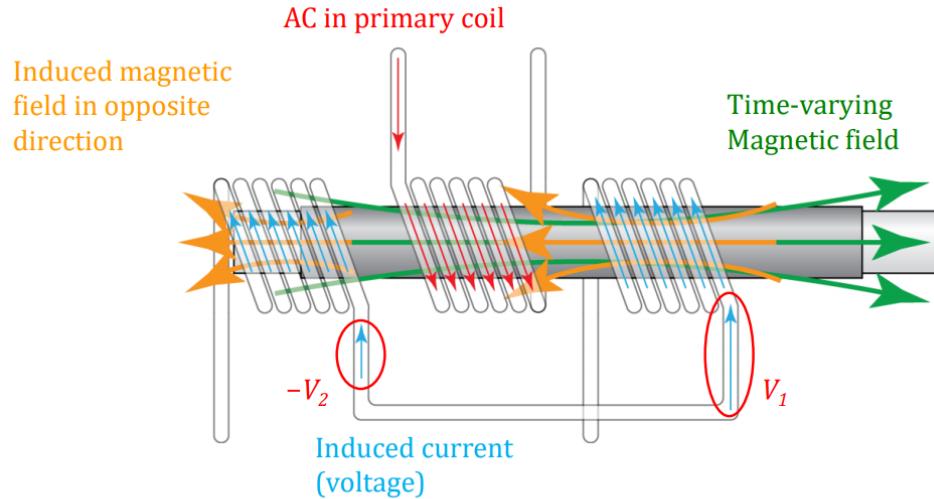


Figure 10.2: The working principle (with displacement to the right) of LVDTs

The working principle follows these steps:

1. AC in primary coil
2. Time-varying magnetic field
3. Induced magnetic field in opposite
4. Induced current (voltage)

If there is no displacement, the output voltage will be 0. When a displacement occurs, the electro-magnetic field around one of the secondary coils (the one the ferromagnetic core moves towards) will be stronger. Hence, the induced current and voltage on that coil will be larger than the other one.

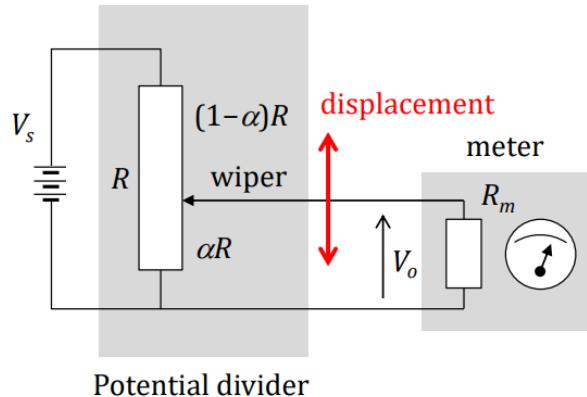
$$V_o = V_1 - V_2 \neq 0 \rightarrow \propto \text{displacement} \quad (10.25)$$

The positive or negative value of V_o can indicate the displacement of the core.

- $V_o > 0 \rightarrow$ Displacement to right
- $V_o < 0 \rightarrow$ Displacement to left

10.3 Resistance-Based Displacement Transducers

A very common and simple form of displacement transducer.



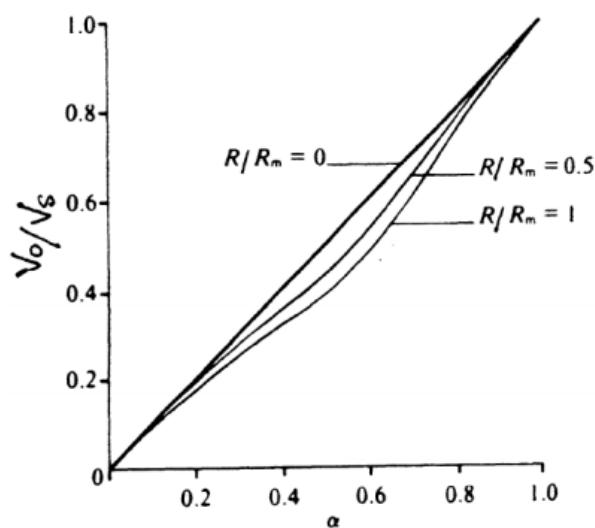
Resistor is typically a **coil winding**. Wiper displacement \rightarrow change of α ($0 < \alpha < 1$). We assume the pot is wound uniformly with resistance wire and α varies linearly with the position of the wiper. With the internal resistance of the meter, the effective resistance R_{eff} to determine V_o is given as:

$$R_{eff} = \alpha R \parallel R_m = \frac{\alpha R R_m}{\alpha R + R_m} \quad (10.26)$$

The two vertical lines \parallel indicate that αR and R_m are in parallel. Using potential divider theory, the meter voltage V_o is then:

$$V_o = \frac{R_{eff}}{(1 - \alpha)R + R_{eff}} V_s \quad (10.27)$$

$$\frac{V_o}{V_s} = \frac{\alpha}{1 + \alpha(1 - \alpha) \frac{R}{R_m}} \quad (10.28)$$



- If $R_m \gg R$, then $\frac{V_o}{V_s}$ is approximated as linear.
- If $R_m > 10R$, then the maximum error is approximated as $Err_{max} = 15 \frac{R}{R_m} (\%)$

Increasing V_s will increase the sensitivity ($= V_o / \text{wiper travelling distance}$), but it also increases the power dissipated by the pot.

$$P = \frac{V_s^2}{R} \rightarrow V_{s,max} = \sqrt{P_{max}R} \quad (10.29)$$

P_{max} is the rated maximum power dissipation (given in spec sheet). To increase the spatial resolution of pot:

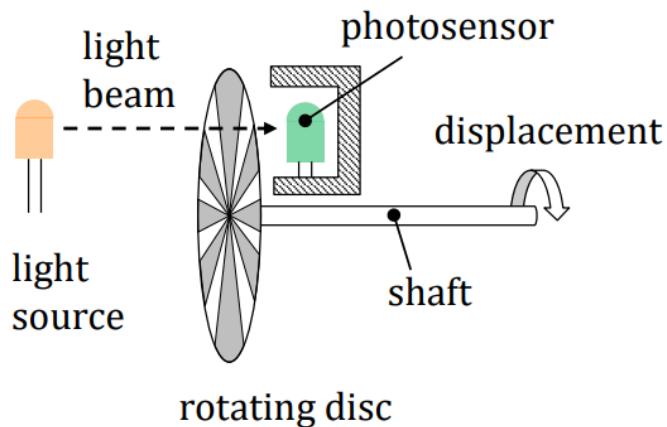
- carbon film
- conductive plastic
- ceramic-metal mix

could also be used. These also reduce the friction between the wiper and the resistive element. Linear resistance pots are made with spans ranging from about 10 mm to almost 1m.

10.4 Measuring Rotational Displacement

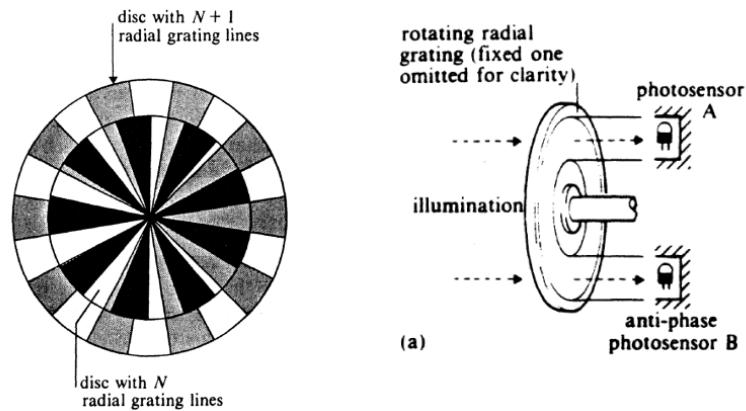
10.4.1 Optical Incremental Shaft Encoder

The output signals here, i.e. voltage output from the photosensor, depend on the angular displacement of the shaft.

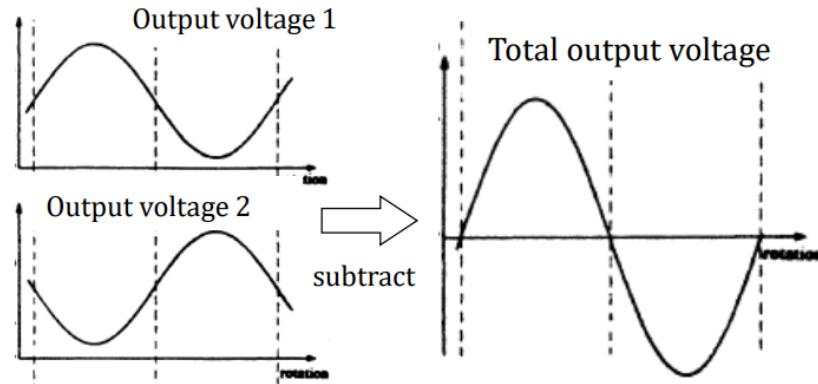


The pattern of the grating allow/disallow the light to reach the photosensor. The signal output varies depending on the **relative position** of the grating. Sinusoidal signal outputs are expected.

More elaborated version of incremental shaft encoder includes one with two discs (Moiré fringes) and two sensors.

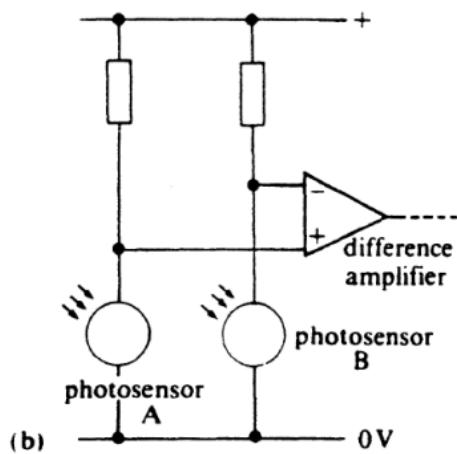


Differential output can be taken for clear signal.

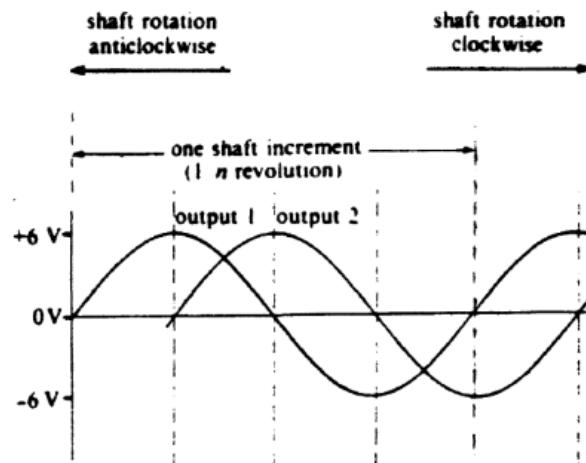


Advantage of the differential systems:

- Reduced influence by light variability
- Reduced influence by the environment

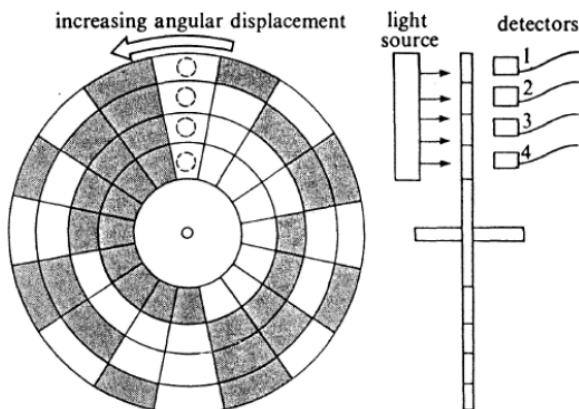


The pairs of light source and sensor may be placed 90° apart, in order to sense the direction of rotation .



10.4.2 Optical Absolute Shaft Encoder

Similar mechanism can be used as to measure an **absolute** rotary displacement by using binary code on the disc.



The four track disc above can have $2^4 = 16$ patterns \rightarrow 16 segments around the disc (resolution: $\frac{360}{16} = 22.5^\circ$). In general, the resolution is:

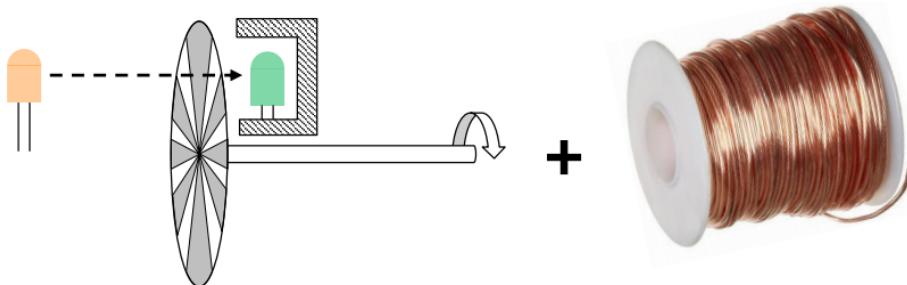
$$\frac{360}{2 \text{ number of tracks}} (\text{degrees}) \quad (10.30)$$

and the detector output ranges from 0 to $2^{(\text{number of tracks})} - 1$.

10.5 Measuring Large Linear Displacement

10.5.1 Measurement of Large Linear Displacements

The linear displacement transducers studied so far measure displacements of up to 1m. In order to measure larger displacements, other methods are used. The simplest form of these is:

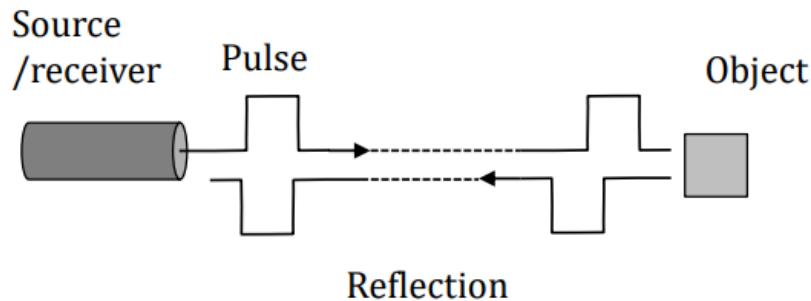


Commercial transducers of this type are available for displacements ranging from 50 mm to over 50m.

10.5.2 Wave Propagation Methods

Non-contact displacement measurements over long distances are possible by using ultrasonic waves, radio waves and light waves.

Pulse Reflection (Time of Flight)



$$D = V_{wave} \cdot \Delta t \quad (10.31)$$

Type of the pulse:

- Sonic/ultrasonic - 330 m s^{-1} (in the air)

- Radio/light wave - $3 \cdot 10^8 \text{ m s}^{-1}$

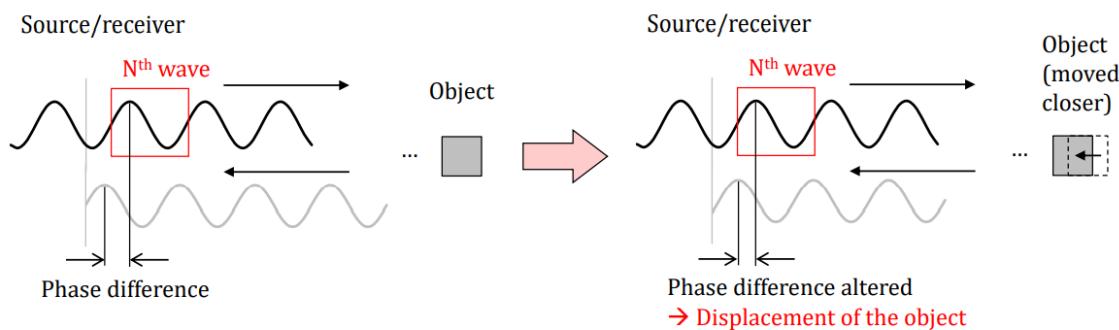
Accuracy of the measurement hinges on the measurement accuracy of time.

Other points to be considered:

- What is the medium
- Obstacles in the middle
- Diffraction
- ...

Phase Measurement

A continuous radio wave is emitted instead of pulses. The phase difference between the transmitted sine wave and the reflected sine wave is used to measure displacement.



Some Examples of Laser Distance Measurement

Laser range finders

- Distance: 15-400 m
- Error: 0.1% at 1 m

Laser scanner for 3D modelling

- Scan a 3D object with 4 lasers at about $127 \mu\text{m}$ accuracy to obtain its surface coordinates (used with trinangulation).

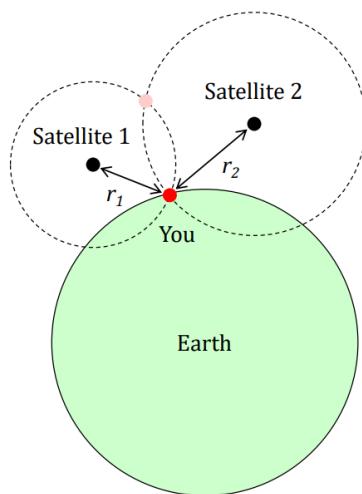
10.6 Other Methods of Measuring Position and/or Displacement

10.6.1 Projection Mapping

- Working principle of object/motion detection in Microsoft Kinect
- Detects objects shape or motion
- The projection from a light source is projected onto the object, and a number of cameras are used to detect and capture the image, based on the angle difference of the captured images.
- Applicable Range: 1.2-3.5 m (50 cm to 5 m)
- Depth Resolution:

10.6.2 Global Positioning System (GPS)

Location service with GPS



The distances between GPS satellites, r_1 and r_2 , are measured by time of flight of radio waves. The location of the person (x, y, z) is determined by solving:

$$r_i^2 = (X_i - x)^2 + (Y_i - y)^2 + (Z_i - z)^2 \quad (10.32)$$

with known locations of the satellites (X, Y, Z) . To identify the position in 3D, 3 equations are needed, i.e. you need to communicate with 3 satellites.

10.7 Datasheet

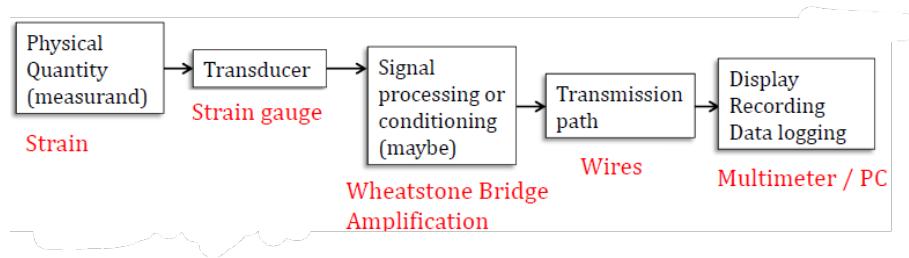
Datasheets are documents providing the specifications for a particular product. Some relevant examples from datasheets are:

- **Linearity:** how much the voltage-displacement relationship could go off the linear.
- **Temperature coefficient of sensitivity:** change of sensitivity due to the operating temperature
- **Zero shift with temperature:** how much the null position (location) could be shifted due to the operating temperature
- **Output ripple:** ‘Hint’ of AC input in DC output. Does not affect the mean DC output value.
- **Resistance range:** Range of available (total) pot resistance
- **Output smoothness:** Maximum resolution error
- **Tolerance:** Uncertainty in manufacturing of the resistance element

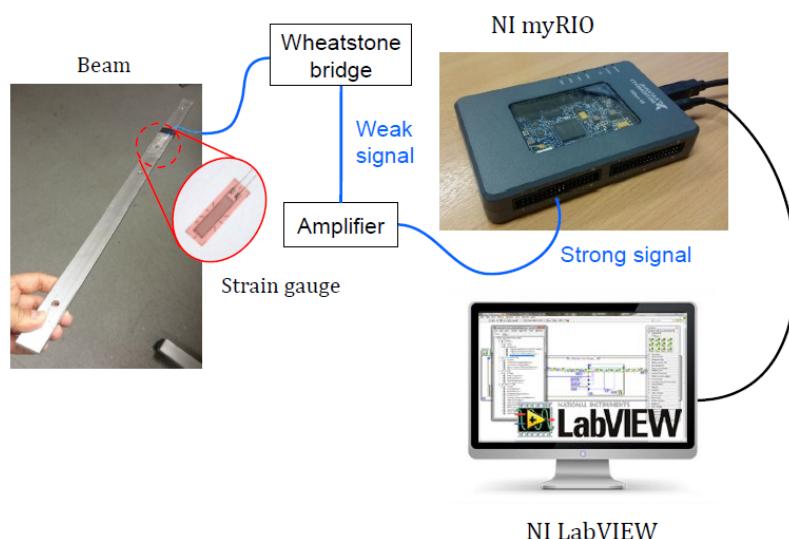
Chapter 11

Laboratory measurement

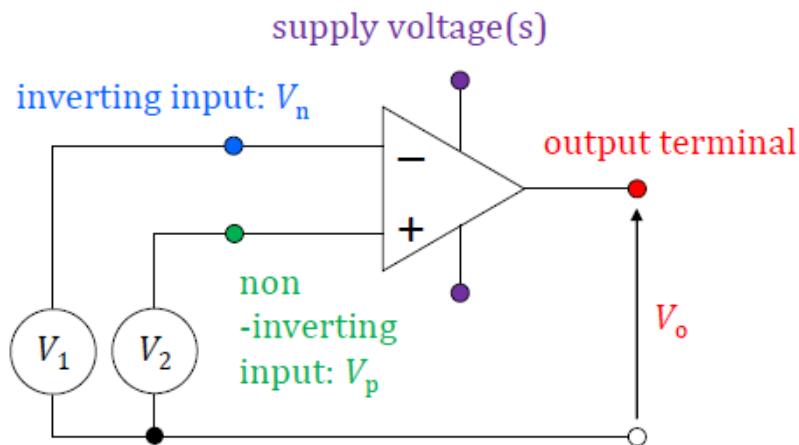
Stages in a measurement system



11.1 Amplifiers

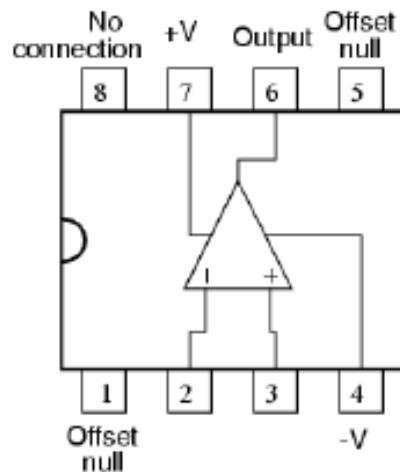


The voltage output from a transducer tends to be weak and this needs to be amplified. Operational amplifiers are normally around 1cm in size.

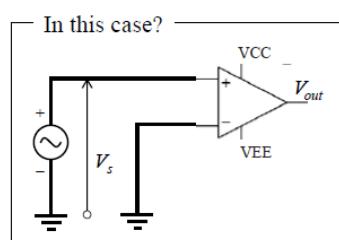
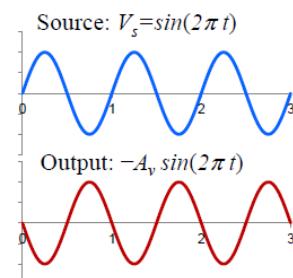
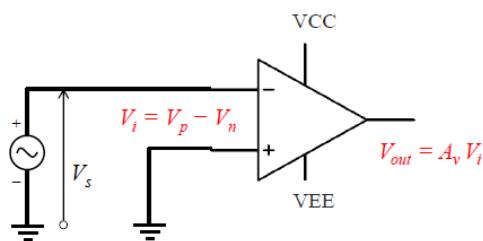


$$V_o = A_v(V_p - V_n) = A_v(V_2 - V_1) \quad (11.1)$$

Where A_v is the gain and $(V_p - V_n)$ can be from a Wheatstone bridge. Operational amplifiers are quite an old technology (since WW2) and are still used because they are cheap and convenient.



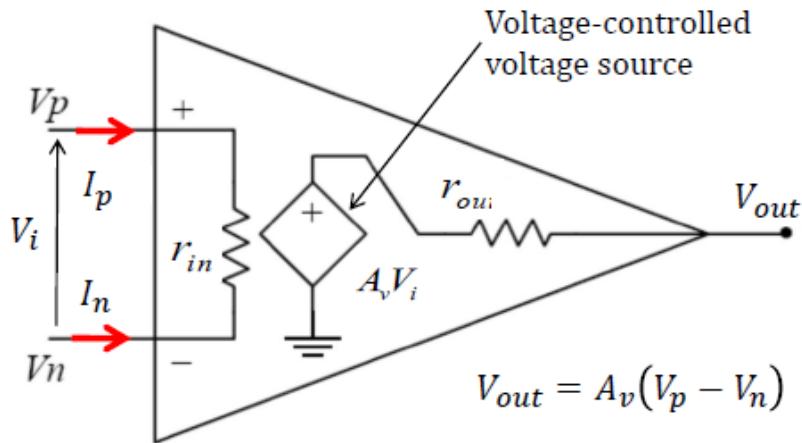
Pin-out for a 8-pin op-amp



$V_i = V_p - V_n$ hold well!

$V_{out} = A_v V_i$

11.1.1 Equivalent op-amp circuit and conditions of an ideal op-amp

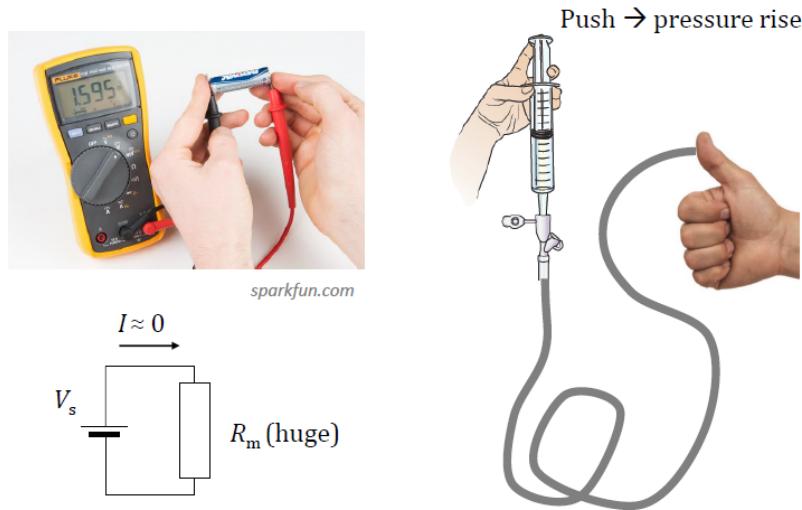


Terminology:

- Differential input voltage: $V_i = V_p - V_n$
- Input resistance: r_{in}
- Output resistance: r_{out}
- Open-circuit output voltage: V_{out}
- Differential voltage gain: A_v

Conditions of an ideal op-amp	In reality
No current into input terminals: $I_p = I_n = 0$	
Infinite input resistance: $r_{in} \rightarrow \infty$	$r_{in} > 200\text{k}\Omega$
Zero Output resistance $r_{out} = 0$	$r_{out} < 1\text{k}\Omega$
Infinite differential (or open-loop) gain $A_v \rightarrow \infty$	$A_v > 100000$
Zero common-mode voltage gain $A_{cm} = 0$	A_v also frequency dependant

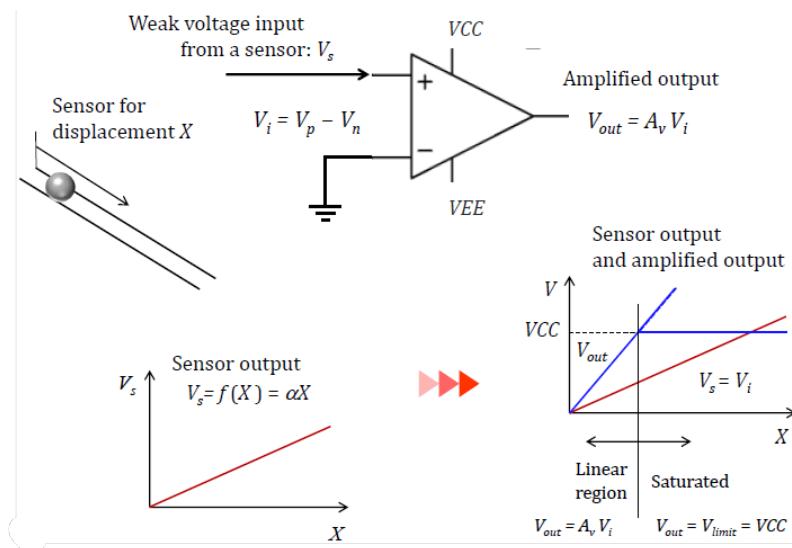
11.1.2 Non-zero voltage with zero current?



A question that is often asked is if the current is zero, how can we still have a voltage? When you connect a battery to a meter, where the meter resistance is normally huge. We can still measure the voltage. This is due to Ohm's law $I = \frac{V}{R_m}$. A good analogy to make is a plunger system, where the other end is covered. When you push down, increasing the pressure, this can still be felt on the tip of the thumb, despite there being no flow. In the same way, we can still 'feel' the voltage with 'no' current flow.

11.1.3 Saturation

In a practical scenario, we want to strengthen the weak signal from the sensor. Let us connect an input to the non-inverting terminal and the inverting input to the ground.

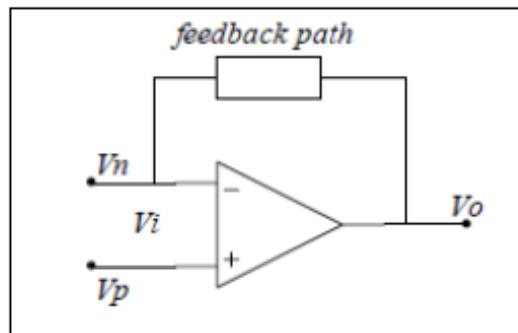


We can see that the signal is not infinitely amplified, as we are limited by the energy provided by our supply voltage. This is reflected in the graph above as the straight portion of the amplified line. This is called saturation. Due to A_v being quite large, our linear region is normally tiny. This may be a problem if we want to measure the linear region.

11.1.4 Negative feedback

So op-amps seem to be a useful device but also seem to be a bit tricky. They are much more useful when used as part of a larger circuit. Negative feedback is achieved by feeding a fraction of the output signal back to the **inverting** terminal as shown. By definition, the closed-loop gain of such a device is:

$$G = \frac{V_o}{V_i} \quad (11.2)$$



Negative feedback trades a reduction in gain with one that is smaller, stable and predictable, whilst giving the designer control over operational characteristics.

11.1.5 Basic feedback amplifier configurations

Only 2 resistors are required to construct a negative feedback amplifier. Note that output signal is always fed-back to the inverting input terminal.

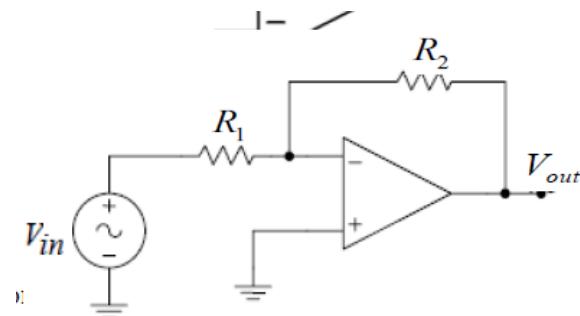


Figure 11.1: Inverting feedback amplifier arrangement.

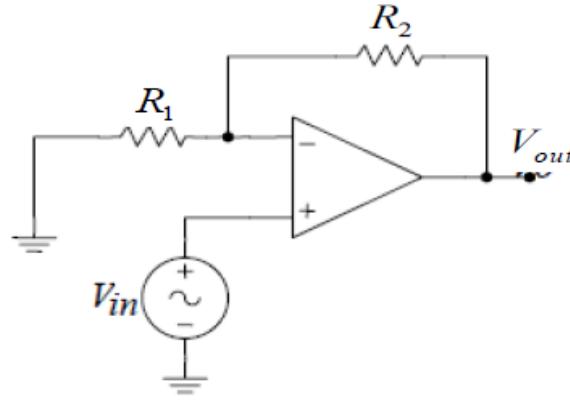
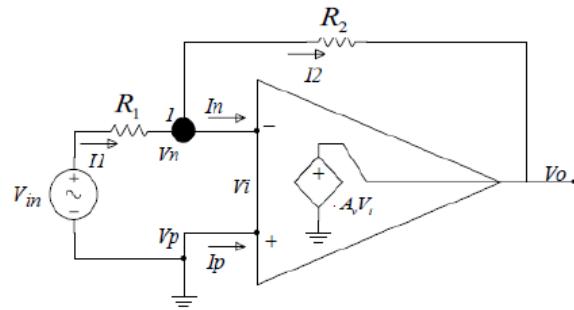


Figure 11.2: Non-inverting feedback amplifier arrangement.

We can see the difference between an inverting and non-inverting feedback amplifier is which input the source signal is connected to.

11.1.6 Basic inverting feedback amplifier



Let us derive its closed loop gain G . By definition:

$$V_0 = A_v(V_p - V_n) \quad (11.3)$$

$$V_p \rightarrow 0 \implies V_n = -\frac{V_0}{A_v} \quad (11.4)$$

Apply KCL to Node 1:

$$\frac{V_{in} - V_n}{R_1} - I_n - \frac{V_n - V_0}{R_2} = 0 \implies \frac{R_2}{R_1}V_{in} + v_n \left(1 + \frac{R_2}{R_1}\right) = V_0 \quad (11.5)$$

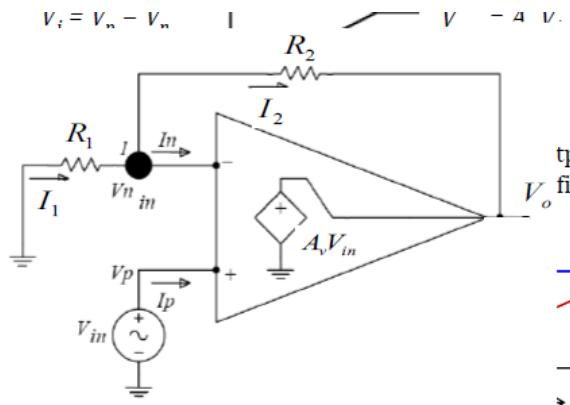
I_n is 0 from our model assumptions. Combining equations (11.4) and (11.5), we arrive at:

$$G \equiv \frac{V_0}{V_{in}} = -\frac{R_2}{R_1} \frac{1}{1 + \frac{1}{A_v} \left(1 + \frac{R_2}{R_1}\right)} \quad (11.6)$$

For an ideal op-amp, $A_v \rightarrow \infty$:

$$G = \frac{V_0}{V_{in}} = -\frac{R_2}{R_1} \quad (11.7)$$

11.1.7 Basic non-inverting feedback amplifier



Derive its closed-loop gain G . By definition:

$$V_0 = A_v(V_p - V_n) \quad (11.8)$$

Apply KCL to node 1

$$\frac{0 - V_n}{R_1} = \frac{V_n - V_0}{R_2} \rightarrow \frac{V_0}{R_2} = V_n \left(\frac{1}{R_1} + \frac{1}{R_2} \right) \quad (11.9)$$

Combining equation (11.8) and (11.9), we arrive at:

$$G \equiv \frac{V_0}{V_{in}} = \frac{1 + \frac{R_2}{R_1}}{1 + \frac{1}{A_v} \left(1 + \frac{R_2}{R_1} \right)} \quad (11.10)$$

For an ideal amplifier, $A_v \rightarrow \infty$:

$$G = 1 + \frac{R_2}{R_1} \quad (11.11)$$

This is positive and always greater than 1.

11.1.8 Summary

Ideal feedback amplifier characteristics

	Inverting amplifier	Non-inverting amplifier
$G = \frac{V_0}{V_{in}}$	$-\frac{R_2}{R_1}$	$1 + \frac{R_2}{R_1}$
R_{in}	R_1	∞
R_{out}	0	0

11.1.9 Voltage amplification summary

Operational amplifier:

- The gain usually is too large (typically $> 100,000$) and not well defined. (V_0 cannot be larger than supply voltage)
- The gains for two op-amps of the same type can differ by a factor of two or more.
- The gain is not stable, can change with temperature, can change with changes of supply voltage.

To reduce these effects, we use negative feedback.

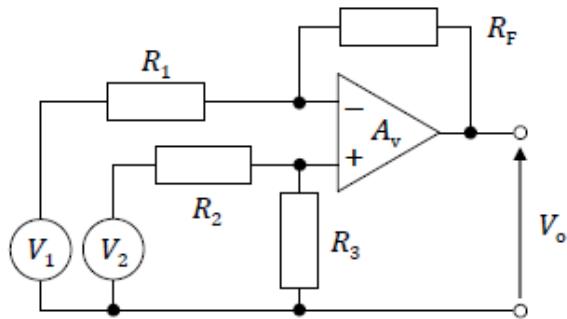


Figure 11.3: V_1 and V_2 could come from a Wheatstone bridge.

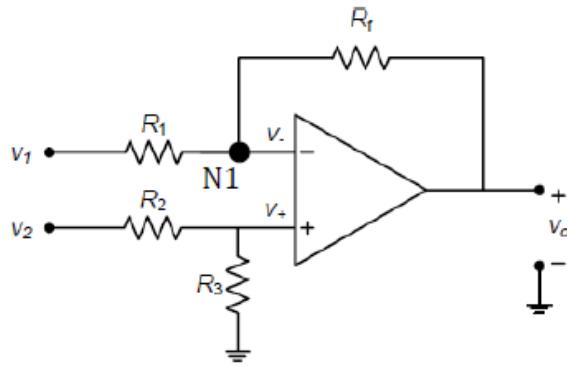
The output voltage V_0 is given as:

$$V_0 = \frac{R_F}{R_1}(V_2 - V_1) \quad (11.12)$$

Only when $\frac{R_3}{R_2} = \frac{R_F}{R_1}$. This is independent of A_v and the gain of this circuit, $G = \frac{R_F}{R_1}$ is well controllable.

11.1.10 Difference amplifier

Combining a non-inverting and inverting op-amp to take the difference between two inputs:



$$\frac{V_0}{A_v} = V_+ - V_- \rightarrow 0 \text{ as } A_v \rightarrow \infty \quad (11.13)$$

$$V_+ = \frac{R_3}{R_2 + R_3} \leftarrow \text{voltage divider} \quad (11.14)$$

KCL at node N1

$$\frac{V_0 - V_-}{R_f} = \frac{V_- - V_1}{R_1} \rightarrow \frac{V_0}{R_f} + \frac{V_1}{R_1} = \left(\frac{1}{R_f} + \frac{1}{R_1} \right) V_- \quad (11.15)$$

$$\frac{V_0}{R_f} + \frac{V_1}{R_1} = \left(\frac{R_f + R_1}{R_f R_1} \right) \left(\frac{R_3}{R_2 + R_3} \right) V_2 \quad (11.16)$$

Combining equation (11.14) and (11.16), we arrive at:

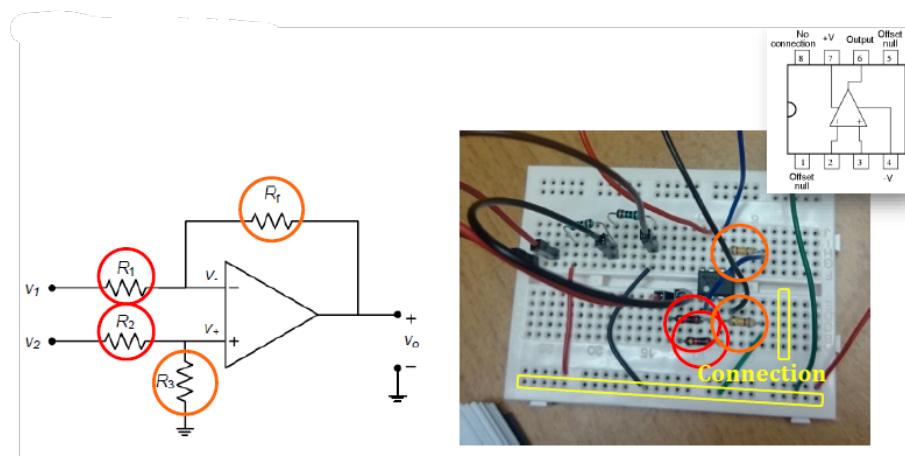
$$V_0 = R_f \left(\frac{R_f + R_1}{R_f R_1} \right) \left(\frac{R_3}{R_2 + R_3} \right) V_2 - \frac{R_f}{R_1} V_1 \quad (11.17)$$

$$V_0 = \frac{1 + \frac{R_f}{R_1}}{1 + \frac{R_2}{R_3}} V_2 - \frac{R_f}{R_1} V_1 \quad (11.18)$$

If we can make sure that $\frac{R_f}{R_1} = \frac{R_3}{R_2}$ is satisfied, we can further reduce our equation to:

$$V_0 = \frac{R_f}{R_i} (V_2 - V_1) \quad (11.19)$$

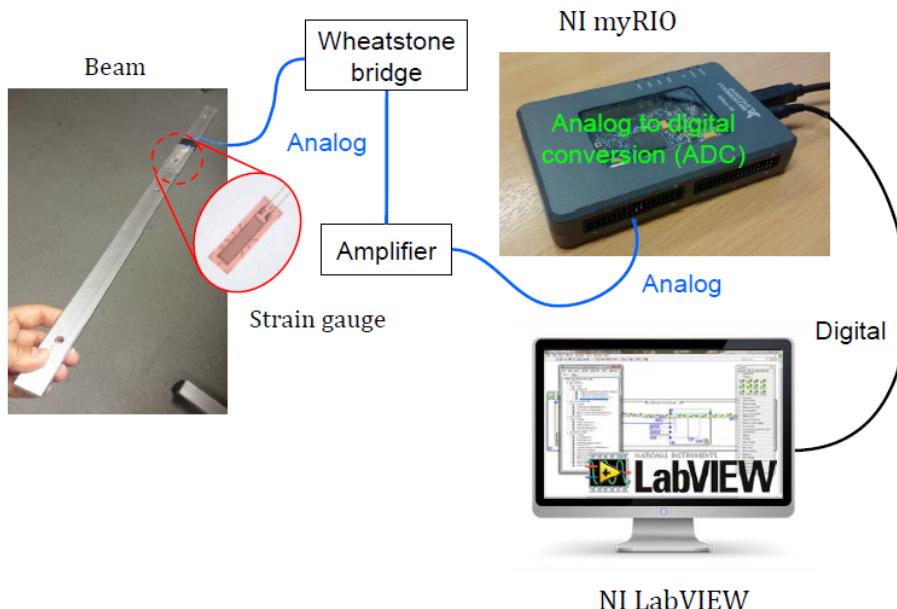
11.1.11 On a bread board



$$V_0 = \frac{R_f}{R_i} (R_2 - R_1) \quad (11.20)$$

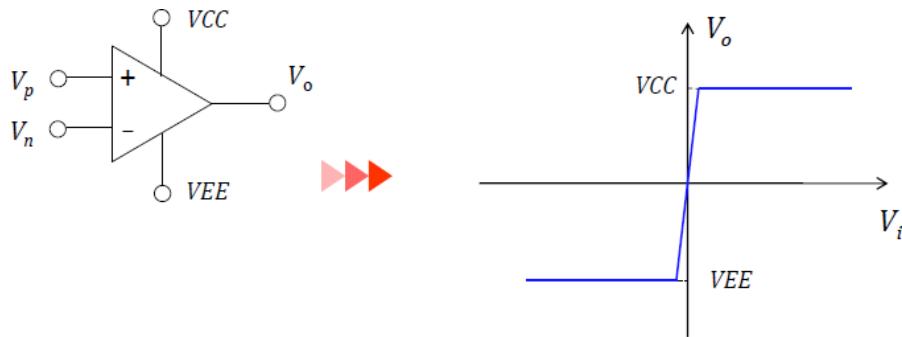
if $\frac{R_f}{R_1} = \frac{R_3}{R_2}$ can be satisfied.

11.2 Analogue-Digital (AD) conversion



11.2.1 Comparator

We are using an op-amp here, because when we have a high A_v , we have a very small linear region. Hence, when we look at a signal graph, we can see this as a sort of switching device:



With very large \$A_v\$, output \$V_0\$ would be: \$V_0 = A_v(V_p - V_n)\$. Hence:

$$V_0 \begin{cases} = VCC & \dots \text{if } V_p > V_n \\ = VEE & \dots \text{if } V_p < V_n \end{cases} \quad (11.21)$$

This essentially makes comparisons based on \$V_p\$ and \$V_n\$

$$V_p > V_n ? \rightarrow \text{YES/NO} \quad (11.22)$$

11.2.2 Example ADC mechanism using comparators

A circuit to compare voltages:

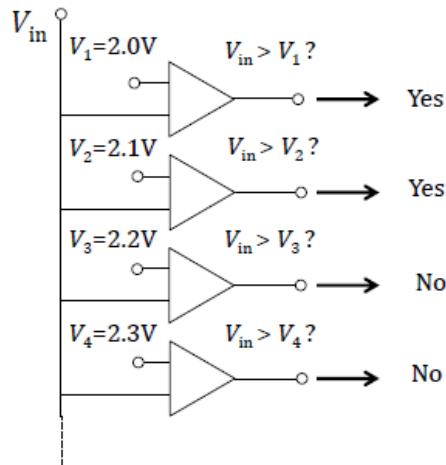


Figure 11.4: **Flash ADC**: number of comparisons done in parallel.

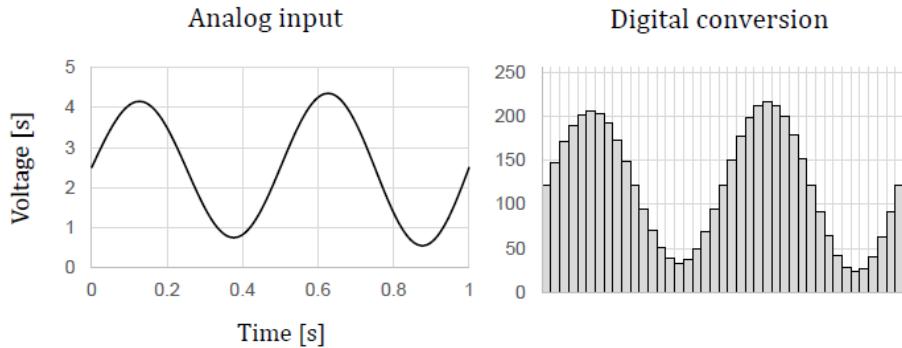
Setting up these comparators in parallel and setting a reference voltage for each, we can compare a single input and produce a digital binary signal, which tells us the range of voltages that \$V_{in}\$ may occupy. In this case we can see that:

$$2.1V < V_{in} < 2.2V \quad (11.23)$$

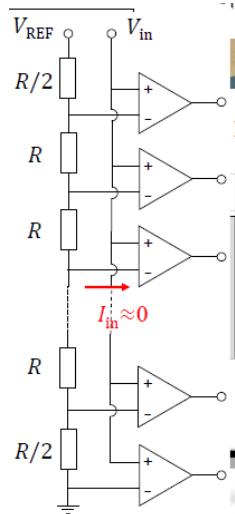
Hence, \$V_{in}\$ may be 2.1V, but other options are possible. When we store in a digital binary format this is called **quantisation**. When a voltage range 0-5V (\$V_{REF} = 5V\$)

is quantised in 8-bit,

$$V_{in} = 2.14V \rightarrow V_{in, \text{quantised}} = \text{int}\left(\frac{V_{in}}{5[V]/2^8}\right) = 110 \quad (11.24)$$



A more practical form of a flash ADC circuit can be achieved using voltage dividers, for the k-th comparator:



$$V_{o,k} = A_v \left(V_{in} - \frac{(k + \frac{1}{2}) R}{nR} V_{REF} \right), \quad k = 1, 2, \dots, n-1 \quad (11.25)$$

n comparators allow judgement of YES/NO per every V_{REF}/n volt → **resolution** of conversion.

- 8-bit quantisation requires $2^8 = 256$ comparators.
- 16-bit quantisation requires $2^{16} = 65536$ comparators.

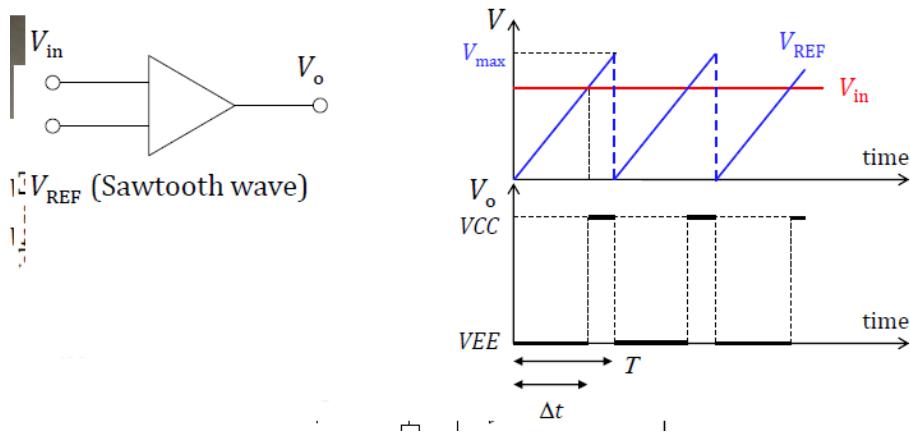
Flash ADC can work very quickly but becomes complex for high resolutions.

11.2.3 Other types of ADC

There are other types of conversion:

- Successive conversion - successively narrow down range of V_{in}
- Integrating - use input signal integration time to find V_{in}
- Ramp-compare

Ramp-comparison utilises a saw tooth wave and the time period of switching to work out what V_{in} is.



$$V_{in} = \frac{\Delta t}{T} V_{max} \quad (11.26)$$

Ramp-compare can be done with a simply system (only one comparator) and can be accurate but may required (relatively - you can make the frequency of the sawtooth very large) long processing time.

11.2.4 Typical issues to be considered

Resolution

Number of bits allocated to represent the input range (= interval of discrete data values).

Accuracy

How the digitised values are different from the original values. E.g. affected by accuracy of Δt measurement in ramp-compare. Though quantisation causes rounding error, it is not necessarily associated with accuracy.

Sampling rate

How many data points over time (how many ADC per time period). Affected by ADC time. Particular impact on time-dependent data/frequency (e.g. music) leading to aliasing/anti-aliasing.

Noise

Some methods (e.g. integrating) are good for noisy data.

11.3 Interface devices to PC and programming platform

11.3.1 Arduino

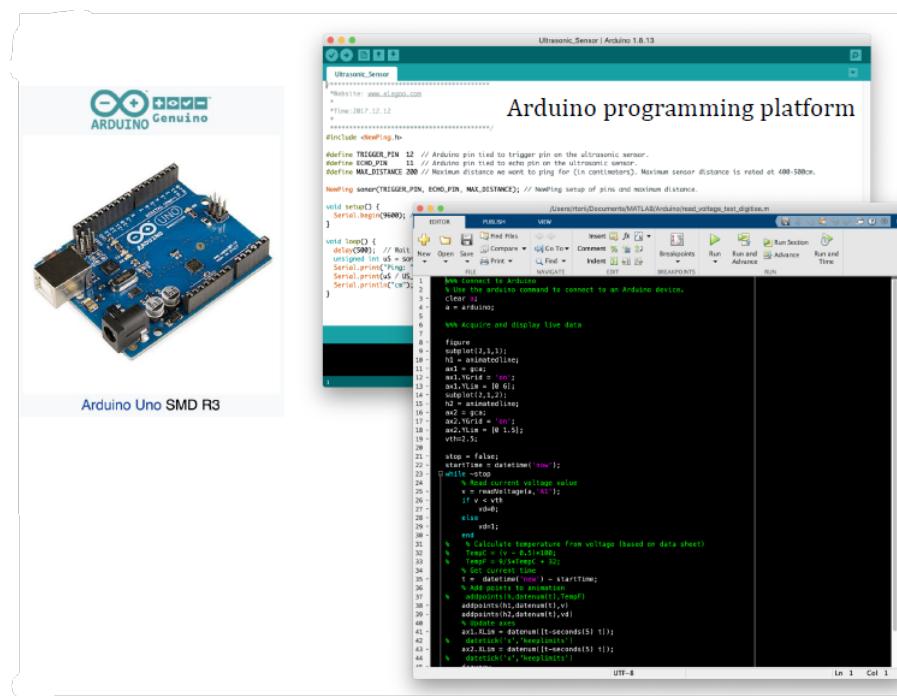
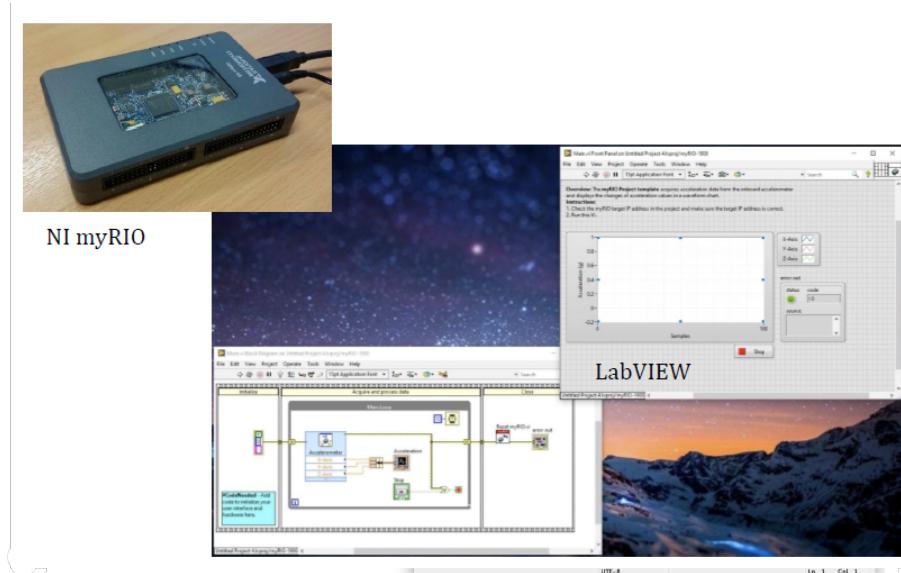


Figure 11.5: Can be linked up with MATLAB or any other language.

11.3.2 NI products and LabVIEW



11.4 Experimental Errors

Because of the inherent limitations of measurement equipment or techniques, there will always be some 'uncertainty' associated with experimental results. This uncertainty of error conveys significant information about the experiment and the result obtained, so you will be expected to assess the error in the measurements that you perform.

Example

We want to know the volume of a cylinder based on diameter and height measurement:

$$V = \pi r^2 h = \frac{\pi h d^2}{4} \quad (11.27)$$

We must ask how accurate are d and h , and also how they can be measured in the calculation of V ? Suppose you are measuring a variable x :

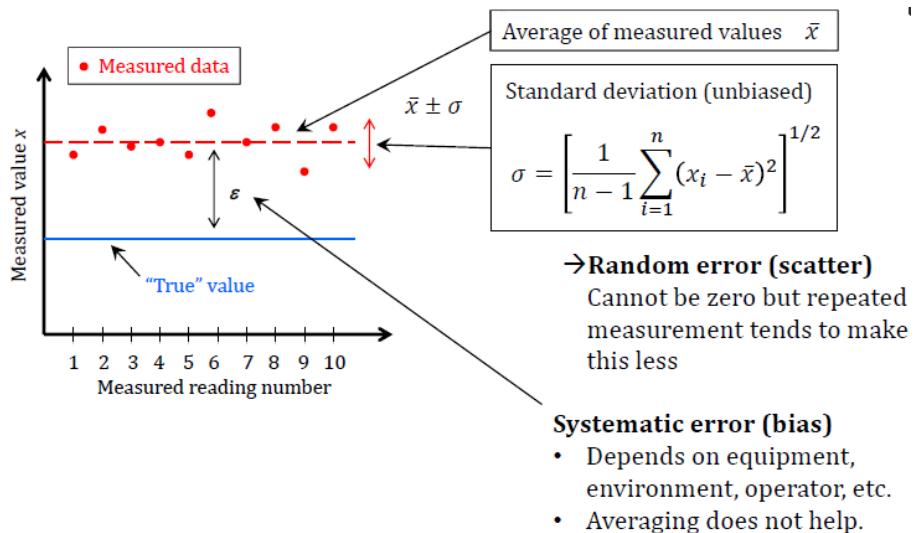


Figure 11.6: Actual error is mostly combination of both types of errors.

11.4.1 Systematic errors

Depends on:

- Equipment
- Environment
 - For example temperature may affect the measurement results, e.g. strain gauges. We deal with this by introducing temperature compensation.
- Operator
 - Consider measuring a sample with length L and mass M . These are well defined variables, however the thickness h will require sampling.

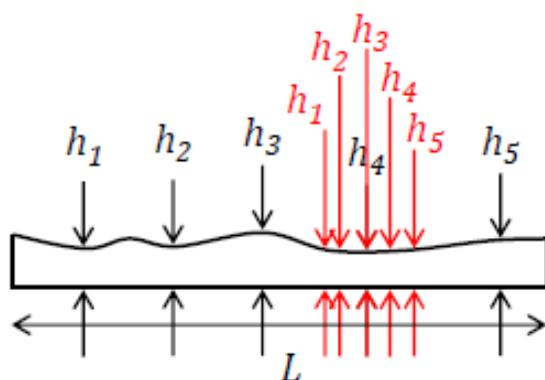


Figure 11.7: Sampling needs to be done carefully to ensure good representation.

Looping back to our cylinder example earlier, let us say:

$$d = 8.2 \pm 0.5\text{cm} \quad (11.28)$$

$$h = 9.5 \pm 0.2\text{cm} \quad (11.29)$$

11.4.2 Combining errors in a single quantity

Uncertainty analysis / error propagation - Estimating the overall error in the measurement system. For example, when two resistors are connected in series, what is the expected uncertainty of the total resistance?

$$R_1 = 1000 \pm 1.5\Omega \quad (11.30)$$

$$R_2 = 500 \pm 1.0\Omega \quad (11.31)$$

The straightforward option is to simply add them directly $\rightarrow 2.5\Omega$

$$R_{total} = R_1 + R_2 = 1500 \pm 2.5\Omega \quad (11.32)$$

The error in R_1 and R_2 are a statistical indication, whereby the error is assumed to follow a Gaussian (normal) distribution. However, simple addition of the errors is too pessimistic! The two errors are independent and unlikely to show their maximum at the same time. If the errors are **independent** and in the same **quantity**, they can be combined as:

$$E = \pm \sqrt{e_1^2 + e_2^2} \quad (11.33)$$

and for this resistor example, overall error is 1.8Ω . The rule above can be applied to any number of independent errors:

$$E = \pm \sqrt{e_1^2 + e_2^2 + e_e^2 + \dots} \quad (11.34)$$

11.4.3 Various combinations of quantities

Assuming independent measurements A and B , which have total errors ΔA and ΔB associated with them, are combined to give the result X , which has error ΔX , then:

$$\text{if } X = A + B \text{ or } X = A - B \longrightarrow \Delta X = \sqrt{(\Delta A)^2 + (\Delta B)^2} \quad (11.35)$$

$$\text{if } X = AB \text{ or } X = \frac{A}{B} \longrightarrow \frac{\Delta X}{X} = \sqrt{\left(\frac{\Delta A}{A}\right)^2 + \left(\frac{\Delta B}{B}\right)^2} \quad (11.36)$$

$$\text{if } X = A^n \longrightarrow \frac{\Delta X}{X} = n \frac{\Delta A}{A} \quad (11.37)$$

$$(11.38)$$

Note for equation (11.36), we always use plus in our square root because the error only increases with addition or subtraction. When more than two quantities are

involved, the equations are extended in a straightforward way:

$$\text{if } X = A + B - C + D \dots \rightarrow \Delta X = \sqrt{(\Delta A)^2 + (\Delta B)^2 + (\Delta C)^2 + (\Delta D)^2} \quad (11.39)$$

$$\text{if } X = \frac{AB}{CD} \rightarrow \frac{\Delta X}{X} = \sqrt{\left(\frac{\Delta A}{A}\right)^2 + \left(\frac{\Delta B}{B}\right)^2 + \left(\frac{\Delta C}{C}\right)^2 + \left(\frac{\Delta D}{D}\right)^2} \quad (11.40)$$

Adding errors in summation ($A+B$) and subtraction ($A-B$) are simpler because they must be the same physical quantity (mass, length, etc.) with the same unit. Products and divisions are usually in different quantity and the errors need to be added in normalised fashion. Let us use these in our cylinder example now:

$$V = \pi r^2 h = \frac{\pi h d^2}{4} \quad (11.41)$$

$$d = 8.2 \pm 0.5 \text{cm} \quad (11.42)$$

$$h = 9.5 \pm 0.2 \text{cm} \quad (11.43)$$

$$V = 387 \text{cm}^3 \quad (11.44)$$

Using two of the error estimation rules in combination,

$$\frac{\Delta V}{V} = \sqrt{\left(\frac{\Delta h}{h}\right)^2 + \left(\frac{2\Delta d}{d}\right)^2} \rightarrow \Delta V = 54 \text{cm}^3 \quad (11.45)$$

Therefore the volume is $V = 387 \pm 54 \text{cm}^3$

11.4.4 General rule for combining errors

Consider the problem of computing a quantity N , where N is a known function of the n independent variables $u_1, u_2, u_3, \dots, u_n$. That is:

$$N = f(u_1, u_2, u_3, \dots, u_n) \quad (11.46)$$

The u s are the measured quantity with errors $\pm \Delta u_1, \pm \Delta u_2, \pm \Delta u_3, \dots, \pm \Delta u_n$ which result in error in N . We could now write:

$$N \pm \Delta N = f(u_1 \pm \Delta u_1, u_2 \pm \Delta u_2, u_3 \pm \Delta u_3, \dots, u_n \pm \Delta u_n) \quad (11.47)$$

Expanding f in Taylor series we get:

$$\begin{aligned} f(u_1 \pm \Delta u_1, u_2 \pm \Delta u_2, u_3 \pm \Delta u_3, \dots, u_n \pm \Delta u_n) &= \\ f(u_1, u_2, u_3, \dots, u_n) + \Delta u_1 \frac{\partial f}{\partial u_1} + \Delta u_2 \frac{\partial f}{\partial u_2} + \dots + \Delta u_n \frac{\partial f}{\partial u_n} &+ \text{higher order terms} \end{aligned} \quad (11.48)$$

Where $f(u_1, u_2, u_3, \dots, u_n)$ is our N . We can neglect our higher order terms because $\Delta u_n \ll 1$ and what we are left with is our error.

$$\text{Error} = \Delta u_1 \frac{\partial f}{\partial u_1} + \Delta u_2 \frac{\partial f}{\partial u_2} + \dots + \Delta u_n \frac{\partial f}{\partial u_n} \quad (11.49)$$

Considering Δu s are not absolute limits of error but rather as statistical bounds of uncertainties the proper method of combining the error is according to the root-sum square formula:

$$E = \sqrt{\left(\Delta u_1 \frac{\partial f}{\partial u_1}\right)^2 + \left(\Delta u_2 \frac{\partial f}{\partial u_2}\right)^2 + \dots + \left(\Delta u_n \frac{\partial f}{\partial u_n}\right)^2} \quad (11.50)$$

11.4.5 Example

The resistance of a certain size of copper wire is given as

$$R = R_0[1 + \alpha(T - 20)] \quad (11.51)$$

Where $R_0 = 6\Omega \pm 0.3\%$ is the resistance at 20°C , $\alpha = 0.004^\circ\text{C}^{-1} \pm 1\%$ is the temperature coefficient of resistance, and the temperature of the wire is $T = 30 \pm 1^\circ\text{C}$. Calculate the resistance of the wire and its uncertainty.

The nominal resistance is:

$$R = 6[1 + 0.004 \times (30 - 20)] = 6.24\Omega \quad (11.52)$$

The uncertainty is from the equation earlier,

$$\Delta R = \sqrt{\left(\Delta R_0 \frac{\partial R}{\partial R_0}\right)^2 + \left(\Delta \alpha \frac{\partial R}{\partial \alpha}\right)^2 + \left(\Delta T \frac{\partial R}{\partial T}\right)^2} \quad (11.53)$$

Here,

$$\frac{\partial R}{\partial R_0} = 1 + \alpha(T - 20) \quad (11.54)$$

$$= 1 + 0.004 \times (30 - 20) = 1.04 \quad (11.55)$$

$$\frac{\partial R}{\partial \alpha} = R_0(T - 20) \quad (11.56)$$

$$= 6 \times (30 - 20) = 60 \quad (11.57)$$

$$\frac{\partial R}{\partial T} = R_0\alpha \quad (11.58)$$

$$= 6 \times 0.004 = 0.024 \quad (11.59)$$

And,

$$\Delta R_0 = 6 \times 0.003 = 0.018\Omega \quad (11.60)$$

$$\Delta \alpha = 0.004 \times 0.01 = 4 \times 10^{-5}\text{C}^{-1} \quad (11.61)$$

$$\Delta T = 1^\circ\text{C} \quad (11.62)$$

Thus, the uncertainty in the resistance is:

$$0.0305\Omega \text{ or } 0.49\% \quad (11.63)$$

11.4.6 What if we do not know the relationship between error sources and measured variables?

Example: load cell (force measurement)

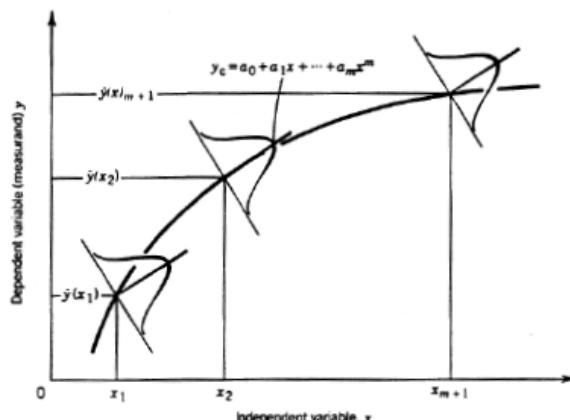
$$V_0 = f(\text{load, etc}) \quad (11.64)$$

Simply add relevant errors up based on the specification sheet as a safety measure.

11.5 Reference information about least-square curve fitting

11.5.1 Least square analysis

The least squares method is often used to establish the functional relationship between the dependent measured variable and the independent process variable. In most situations, the anticipated relationship is polynomial. The variation found in the measured variable is assumed to follow a normal (Gaussian) distribution about each fixed value of the independent variable.

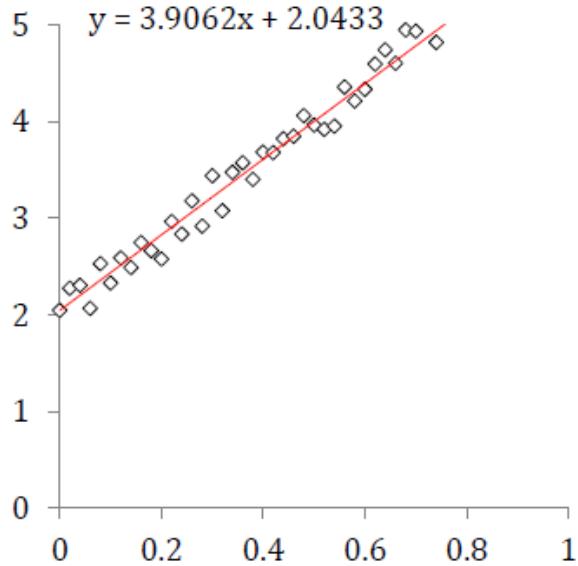


Suppose two variables x and y are measured over a range of values. We want to find their mathematical relationship. The simplest polynomial function is linear: $y = ax + b$. Least square analysis is done by minimising the quantity:

$$S = \sum_{i=1}^n [y_i - (ax_i + b)]^2 \quad (11.65)$$

This is the same as finding:

$$\frac{\partial S}{\partial a} = 0, \quad \frac{\partial S}{\partial b} = 0 \quad (11.66)$$



$$\frac{\partial S}{\partial a} = \sum_{i=1}^n (2ax_i^2 + 2bx_i - 2x_iy_i) = 0 \rightarrow a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i \quad (11.67)$$

$$\frac{\partial S}{\partial b} = \sum_{i=1}^n (2ax_i + 2b - 2y_i) \rightarrow a \sum_{i=1}^n x_i + nb = \sum_{i=1}^n y_i \quad (11.68)$$

$$a = \frac{n \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (11.69)$$

$$b = \frac{(\sum_{i=1}^n x_i^2)(\sum_{i=1}^n y_i) - (\sum_{i=1}^n x_i y_i)(\sum_{i=1}^n x_i)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (11.70)$$

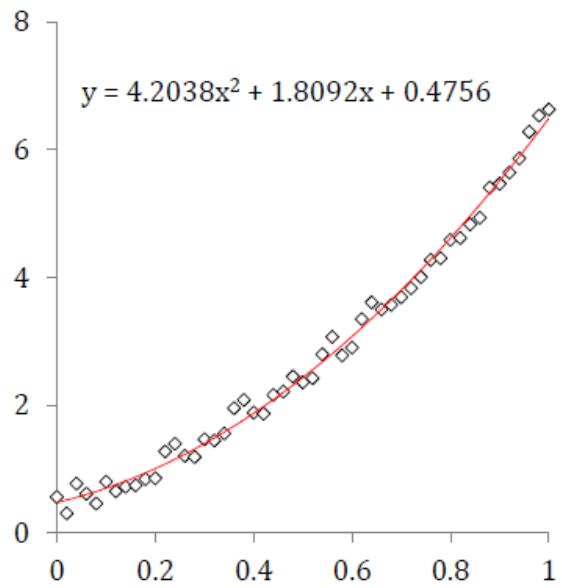
Even with higher order polynomials, for example

$$y = ax^2 + bx + c \quad (11.71)$$

We can define quantity

$$S = \sum_{i=1}^n [y_i - (ax_i^2 + bx_i + c)]^2 \quad (11.72)$$

for minimisation.



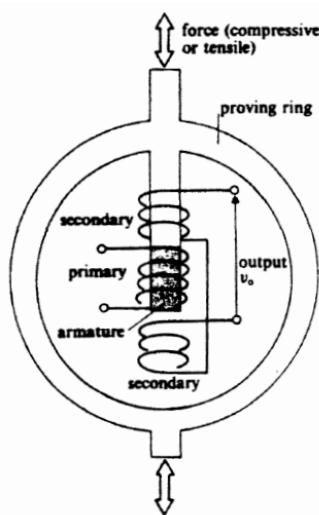
Chapter 12

Measurement of Force and Acceleration

16/11/2020

12.1 Measuring Forces

In the simplest form of force transducer, the force is applied to some form of **elastic member**, which compresses, expands or bends. The resulting displacement or strain is then sensed by a **secondary transducer** which converts it to an output signal.



The other forms of force transducers:

1. Using the force-balance principle
2. Using the piezoelectric effect

12.1.1 Elastic Sensing Elements

Elastic sensing elements:

- Are primary transducer (force → displacement)
- The force-displacement relationship should be definitive (mostly linear)

The deformation is then detected by either measuring the strain in the member using strain gauges or measuring the displacement of a point on the elastic member.

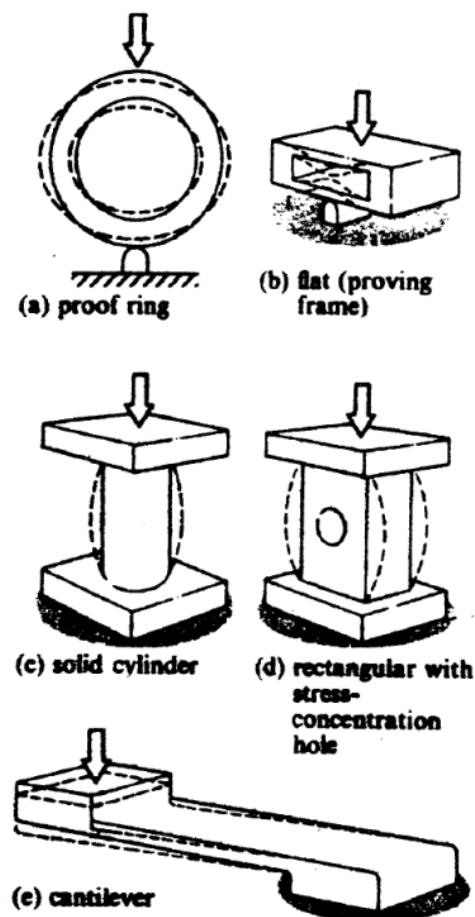


Figure 12.1: Common elastic sensing elements

Real World Example: Electronic Scale

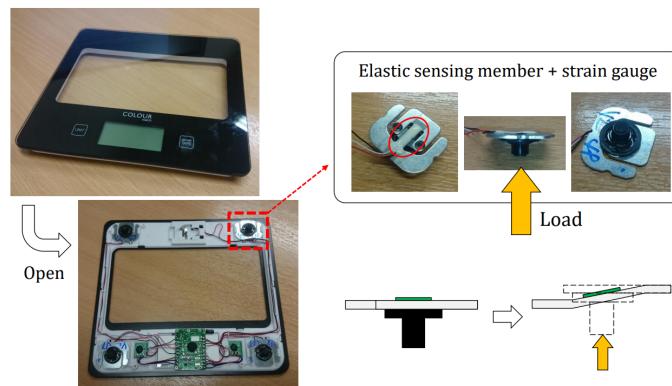
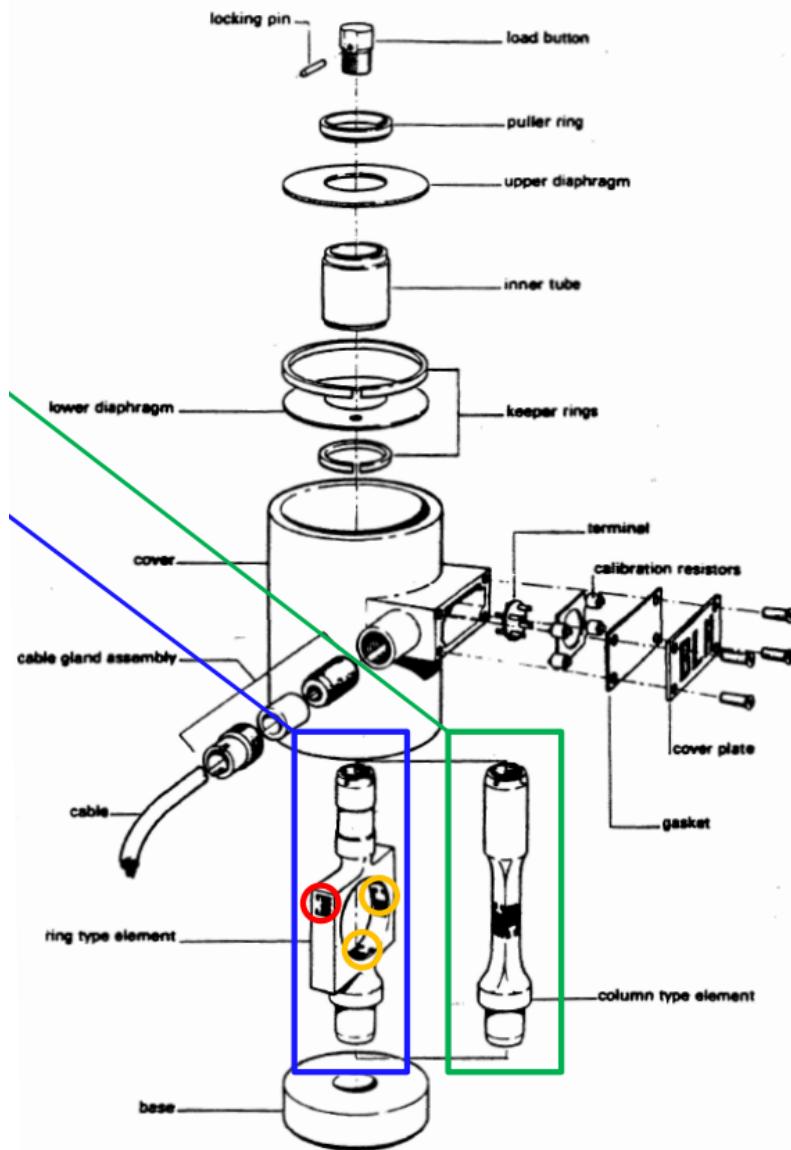


Figure 12.2: Electronic Kitchen Scale

Each transducer has a sensing member and a strain gauge. When a load is applied on the elastic member, it undergoes a displacement, which is sensed by the strain gauge. The output signal from all the transducers is used to determine a digital reading.

12.1.2 Strain Gauge Load Cell

The most common type of load sensing cell.



Sensing Element Types:

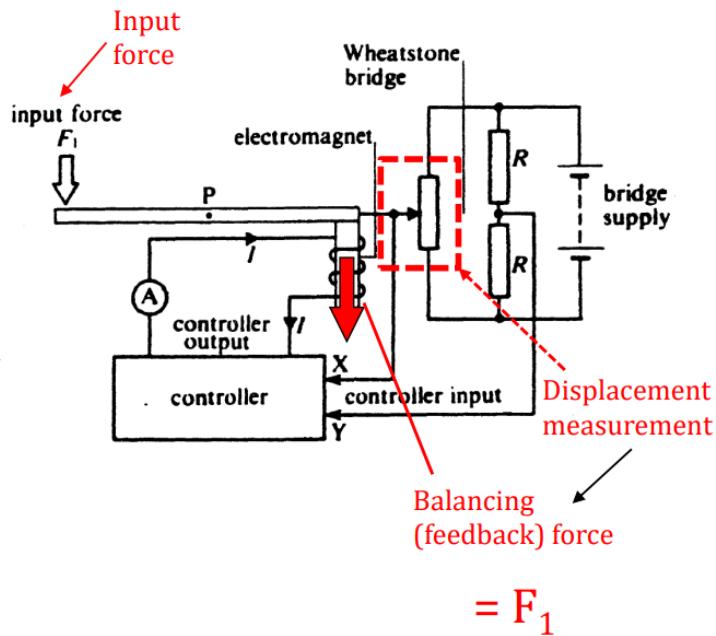
- Column type (green)
- Ring type (for smaller forces) (blue)

Each sensing element has multiple strain gauges mounted on it: typically, four measuring gauges (yellow) and two dummy gauges (red) for **span temperature compensation**.

- To account for the deformation of sensing element due to temperature.
- To compensate “load-displacement” relationship.

12.1.3 Force Transducer Using the Force-Balance Principle

Another type of force transducer makes use of a force balancing the input force provided by a feedback mechanism.



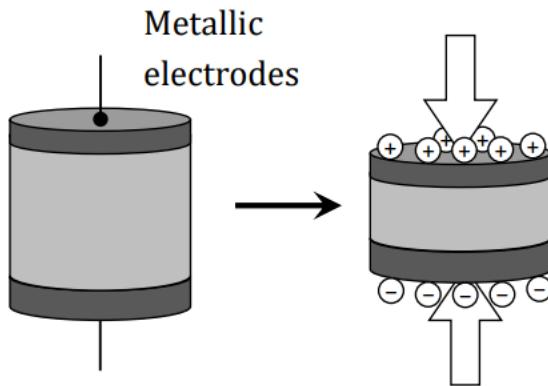
Advantages:

1. The relationship between the current I and the feedback force is linear.
2. The rotation of the beam is almost zero all the time and the force-deflection characteristics for the beam would not be present.

12.1.4 Piezoelectric Force Transducers

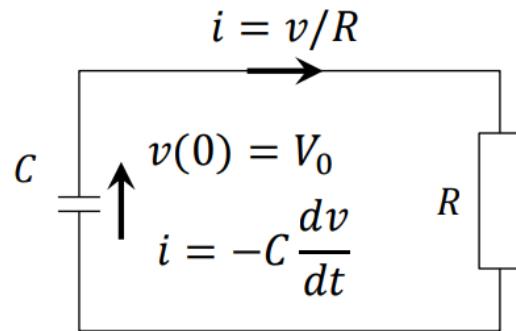
These transducers use a piezoelectric crystal (e.g. quartz) or ceramic (e.g. PZT – lead zirconate titanate [Has to be electrically polarised before acquiring piezoelectricity]) , where electric charges of opposite polarity appear on the parallel faces of the transducer when the faces are squeezed together.

The amount of charge depends linearly on the force applied, within a limited range, and its polarity depends on the directions of the crystallographic axes.



As a whole, this works as a capacitor as well. Step input force yields charge Q_o , and corresponding output voltage is $V_o = Q_o/C$.

If the force is held steady, no more charge is generated but the voltage across the transducer decreases because the charges are taken by the measuring circuit as a current (i.e. flow of charge).



From the current equations:

$$\frac{v}{R} = -C \frac{dv}{dt} \quad (12.1)$$

After rearranging:

$$\frac{1}{v} dv = -\frac{1}{RC} dt \quad (12.2)$$

Integration of the equation gives:

$$v = V_0 e^{-\frac{t}{RC}} \longrightarrow \text{Exponential decay} \quad (12.3)$$

Piezoelectric transducers are **not suitable** for static or quasi-static force but are used for measuring rapidly changing forces, such as those in mechanical vibrations. (Note: **Piezoresistive** transducers are used for static measurements)

12.2 Measuring Acceleration

Acceleration transducers are called accelerometer.

A measure g is widely used, indicating the acceleration experienced by a freely falling body due to gravity ($\approx 9.81 \text{ m/s}^2$).

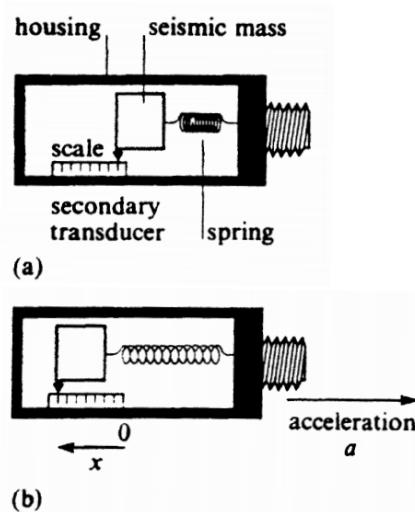
This is a convenient standard for the calibration and in many transducers specifications, the acceleration is stated as multiples of g .

Several different types of accelerometers (all of them are making use of seismic-mass):

- Seismic-mass accelerometer
- Strain-gauge accelerometer
- Potentiometric accelerometer
- Servo accelerometer

12.2.1 Seismic-Mass Accelerometer

A known mass, called seismic mass (or proof mass), is connected to some form of spring.



Ignoring friction, the seismic mass displacement along the sensing axis of the accelerometer is related to the force exerted by the spring as $F = kx$ (k is the spring constant in N/m).

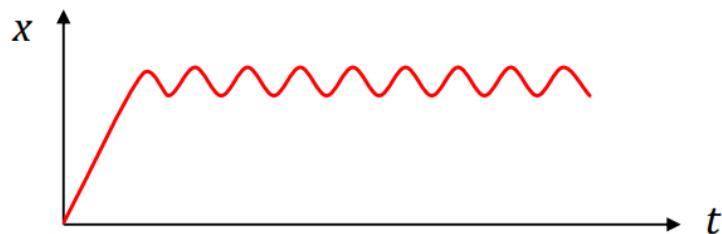
Considering also the acceleration force $F = ma$, we get:

$$ma = kx \quad (12.4)$$

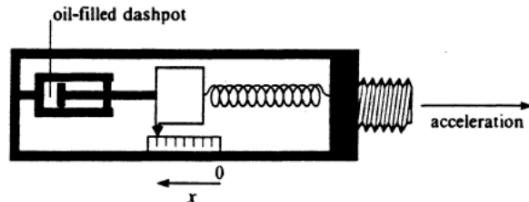
$$a = \frac{kx}{m} \quad (12.5)$$

However, this only applies to **static response** of accelerometer.

Question: What is the expected motion of the seismic mass when the accelerometer is turned upright (measuring rapidly changing / dynamic acceleration)? (assuming the motion is limited in the sensing axis)



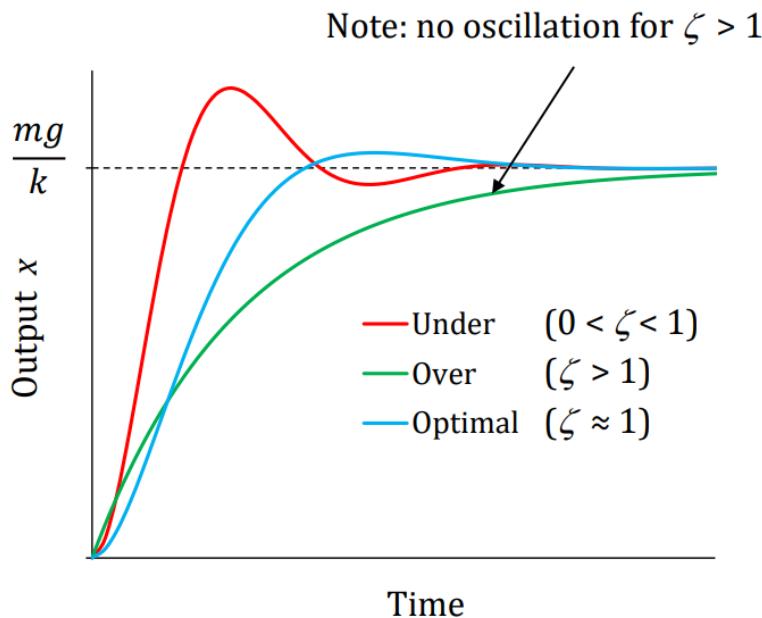
In order to obtain a more acceptable dynamic response, a controlled amount of extra damping is added via an oil-filled dashpot.



As the result, the position of the seismic mass x varies in time as shown in the figure below, in three different modes:

- Underdamped (Red)
- Overdamped (Green)
- Optimal (Blue)

where the steady state is reached relatively rapidly.



The amount of damping in a system of this sort is expressed by means of a damping factor ζ (zeta). Without damping, the system would oscillate at its natural frequency:

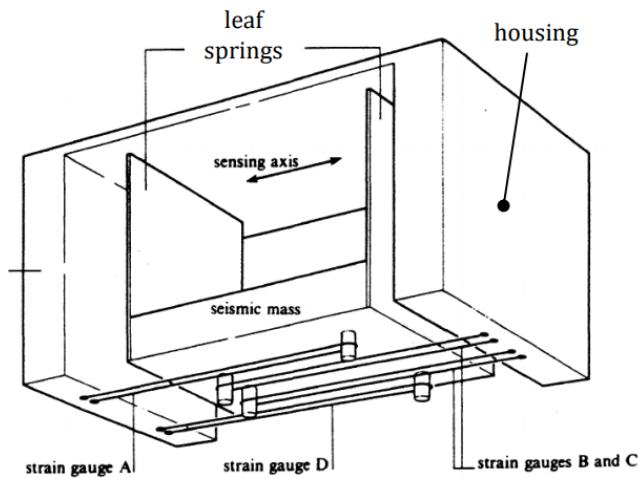
$$f_n = \frac{1}{2\pi} \sqrt{\frac{k}{m}} \quad (12.6)$$

As the damping increase, the oscillating frequency changes as:

$$f = f_n \sqrt{1 - \zeta^2} \quad (12.7)$$

12.2.2 Strain Gauge Accelerometer

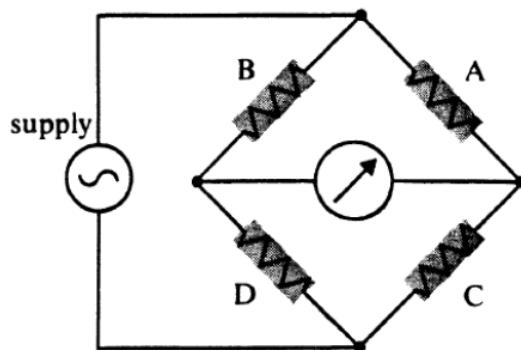
- The seismic mass is suspended from the housing by a pair of leaf springs.
- The closed housing contains silicon fluid for the damping.
- Four unbonded strain gauges are used (that is, no backing material).



For this type of accelerometers, the seismic mass can move in any direction, so an estimate for **cross-axis sensitivity** is important.

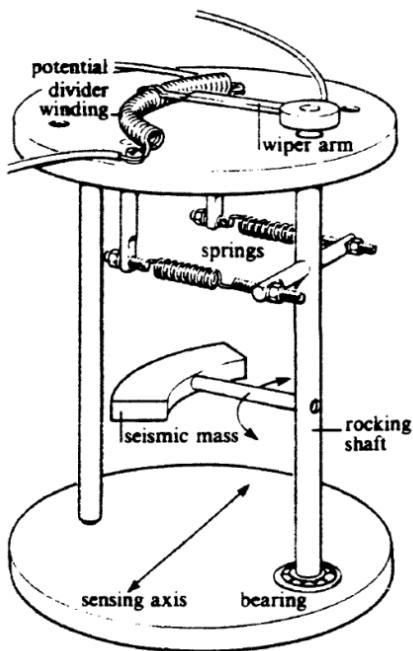
→ Sensitivity to the motion in the direction normal to the sensing axis.

→ Often mentioned in specifications, expressed in percentage of the full-scale output.

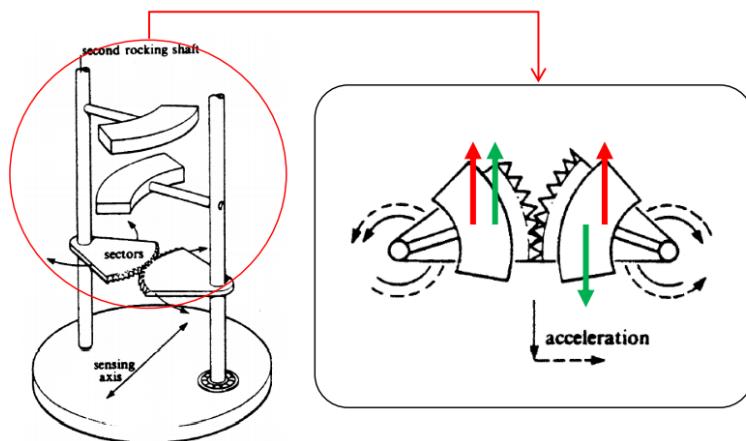


12.2.3 Potentiometric Accelerometer

This is again a type of seismic accelerometer, has been designed for use in the flight recorder of an aircraft. It measures the vertical accelerations.



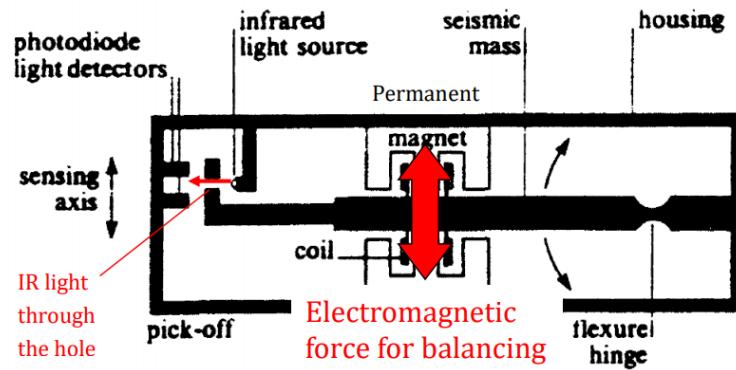
A large cross-axis sensitivity is expected and is a problem. → Can be avoided by a mechanism with a second shaft as shown below.



12.2.4 Servo Accelerometer

The spring force of the previously introduced accelerometers is replaced here by an electromechanical force provided by what is called a **forcer**.

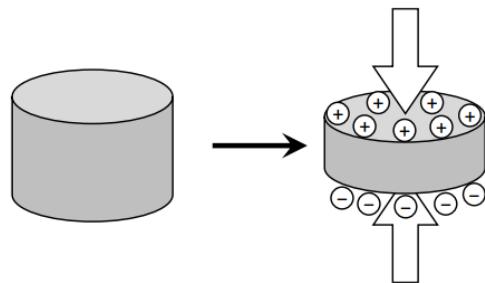
The system is similar to the force transducer using force-balance principle; the deflection of the sensing element (the beam) is detected by the light source and photodiodes (on the left) whose output is fed back to the current through the forcer coil to maintain zero deflection position via the electromechanical force generated between the coil and permanent magnet.



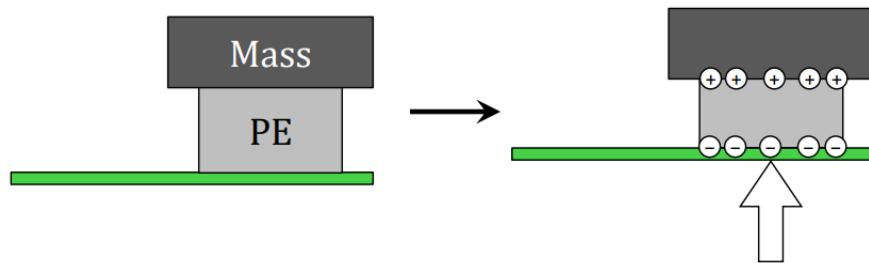
12.2.5 Piezoelectric Accelerometer

Piezoelectric effect

- Electric charge appearing on the surface of certain solid materials in response to mechanical stress applied



Use as accelerometer with a mass on top



Advantages:

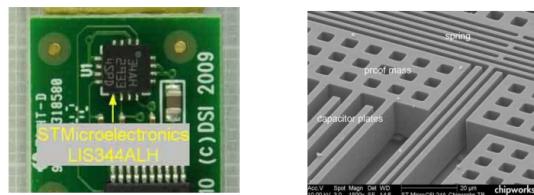
- Small, simple and cheap
- Durable (little degradation in time)

Disadvantages:

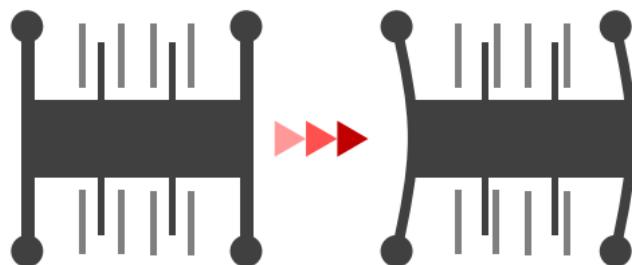
- Mass of the measurement system may alter the dynamic characteristics of whole system
- Difficult to measure static variables

12.2.6 MEMS Accelerometer

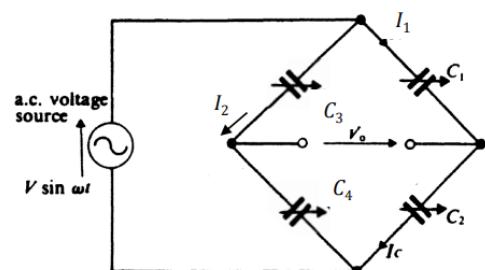
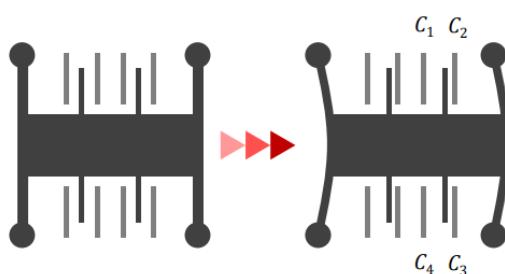
In many nowadays applications such as mobile phones, same principles as we learned are used but in a smaller scale, in MEMS (micro-electromechanical system).



A simplified example of MEMS accelerometer is shown below – this design is making use of a seismic mass and capacitance displacement transducer.



Damping is provided by **squeeze film damping**.



Following the same derivation as in the displacement measurement:

$$V_{c1} = \frac{I_1}{\omega C_1} \quad (12.8)$$

$$V_{c2} = \frac{I_1}{\omega C_2} \quad (12.9)$$

$$V_{c3} = \frac{I_2}{\omega C_3} \quad (12.10)$$

$$V_{c4} = \frac{I_2}{\omega C_4} \quad (12.11)$$

The voltage across C_2 as a fraction of the bridge energising voltage, V is:

$$\frac{V_{c2}}{V} = \frac{V_{c2}}{V_{c1} + V_{c2}} = \frac{\frac{I_C}{\omega C_2}}{\frac{I_C}{\omega C_1} + \frac{I_C}{\omega C_2}} = \frac{C_1}{C_1 + C_2} \quad (12.12)$$

$$\frac{V_{c4}}{V} = \frac{C_3}{C_3 + C_4} \quad (12.13)$$

The capacitance in balance:

$$C_0 = \frac{\epsilon_0 A}{d} \quad (12.14)$$

where d is the balance separation. This varies with displacement as:

$$C_1 = \frac{\epsilon_0 A}{d + \delta} = C_0 \frac{d}{d + \delta} \quad (12.15)$$

$$C_2 = C_0 \frac{d}{d - \delta} \quad (12.16)$$

$$C_3 = C_0 \frac{d}{d - \delta} \quad (12.17)$$

$$C_4 = C_0 \frac{d}{d + \delta} \quad (12.18)$$

Substituting these for the equation of the voltage across C_2 and C_4 yields:

$$\frac{V_{c2}}{V} = \frac{d - \delta}{(d - \delta) + (d + \delta)} = \frac{d - \delta}{2d} \quad (12.19)$$

$$\frac{V_{c4}}{V} = \frac{d + \delta}{2d} \quad (12.20)$$

The amplitude of the output voltage is thus:

$$V_o = V_{c4} - V_{c2} = \left(\frac{d + \delta}{2d} - \frac{d - \delta}{2d} \right) V = \frac{\delta V}{d} \quad (12.21)$$

Acceleration can then be calculated from (ignoring damping):

$$F = ma = kx \quad (12.22)$$

$$a = \frac{k\delta}{m} = \frac{kdV_o}{mV} \quad (12.23)$$

Chapter 13

Measurement of Velocity, Torque and other Variables

23/11/2020

13.1 Measurement of Velocity

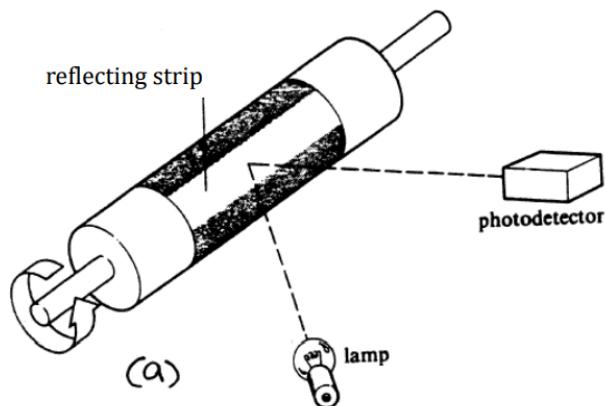
Linear- and angular-velocity transducers are available based on many different principles. Some of the more common types are introduced here. Linear velocity is often measured with an angular-velocity transducer by using a wheel to convert the linear motion into rotation. Angular-velocity transducers are given different names depending on their use:

- Revolution counters – measure rates of rotation of a body by counting the rotation
- Tachogenerators – generate an output (typically voltage) proportional to the angular velocity of an element of the system

Mechanical rotation is coupled to an angular-velocity transducer via ‘sensing shaft’ which rotates at the angular velocity to be measured.

13.1.1 Velocity transducers with a frequency output

Similar principle to the angular position transducers, such as optical incremental shaft encoders, can be used to measure rotational frequency.



Simple examples include a rotating shaft with a reflecting strip on non-reflecting background colour band (figure). Here, one electrical pulse per revolution is received by the detector.

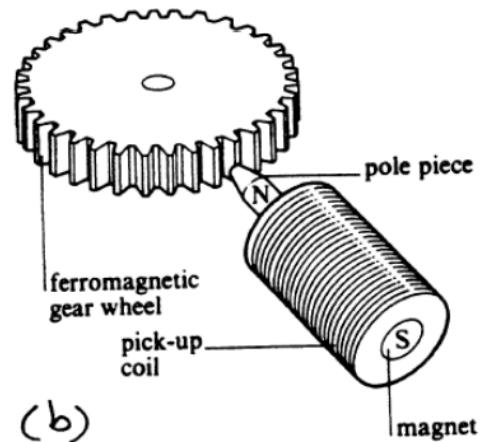
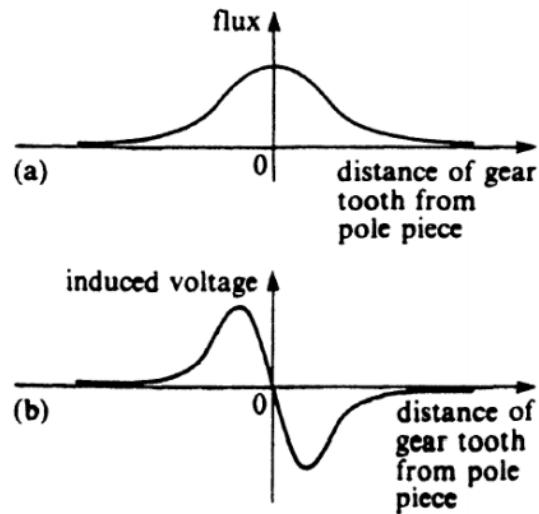


Figure 13.1: A variable reluctance tachometer. A pick-up coil surrounds a permanent magnet, one pole of which is placed close to a ferromagnetic gear wheel mounted on the sensing shaft.

The reluctance of the magnetic circuit varies depending on the distance between the pole and gear tooth, and hence the magnetic flux inside the coil changes which induces the electromotive force (voltage).



The induced EMF (electromotive force) and voltage is proportional to the rate of change of flux, hence the distance between the gear and pole is related to the induced voltage via:

$$\epsilon = -\frac{d}{dt}(\text{flux}) \quad (13.1)$$

13.1.2 Tachogenerator

The tachogenerator generates a voltage (usually DC) proportional to angular velocity, using electromagnetic induction. The transducer uses the energy of rotation of the sensing shaft to generate an electrical output and requires no external power supply.

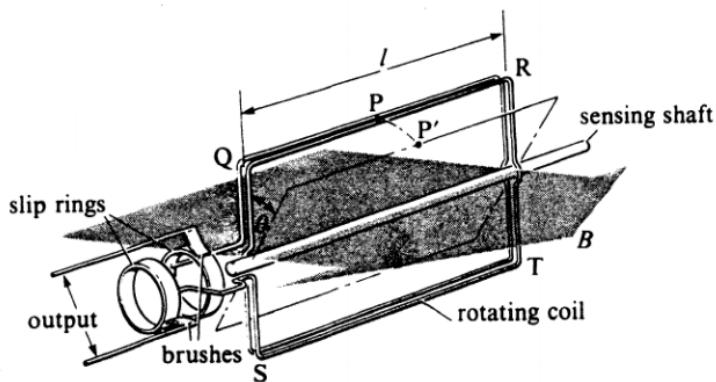


Figure 13.2: Basic AC tachogenerator

Any conductor moving through a magnetic field at an angle θ to the direction of the field has an EMF induced as:

$$\epsilon = Blv \sin(\theta) \quad (13.2)$$

Where:

- B is the magnetic flux density
- l is the length of the conductor
- v is the velocity of the conductor in the field

If the conductor is moving parallel to the field, $\theta = 0$ and no EMF is induced, whereas if it is moving normal to the field, $\theta = \pi/2$ and $\epsilon = Blv$.

In the tachogenerator, the induced voltage is picked up via two spring-loaded brushes, rubbing on two conducting rings called ‘slip rings’. The slip rings are connected to the two ends of the coil and rotate with it. The voltages generated in sides QR and ST are in the correct direction to add. The other two sides (QS and RT) do not contribute to the voltage.

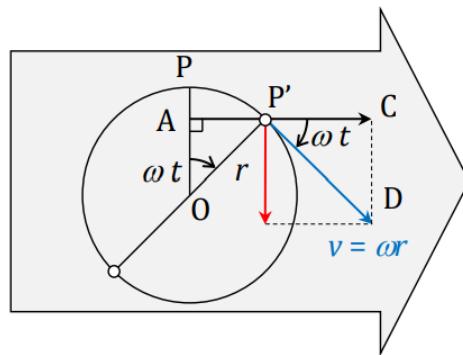


Figure 13.3: Red arrow: $v \sin(\omega t) \rightarrow$ Component crossing the magnetic field

From the figure, the induced EMF in the side QR can be calculated as:

$$\epsilon = Blv \sin(\theta) = Blwr \sin(\omega t) \quad (13.3)$$

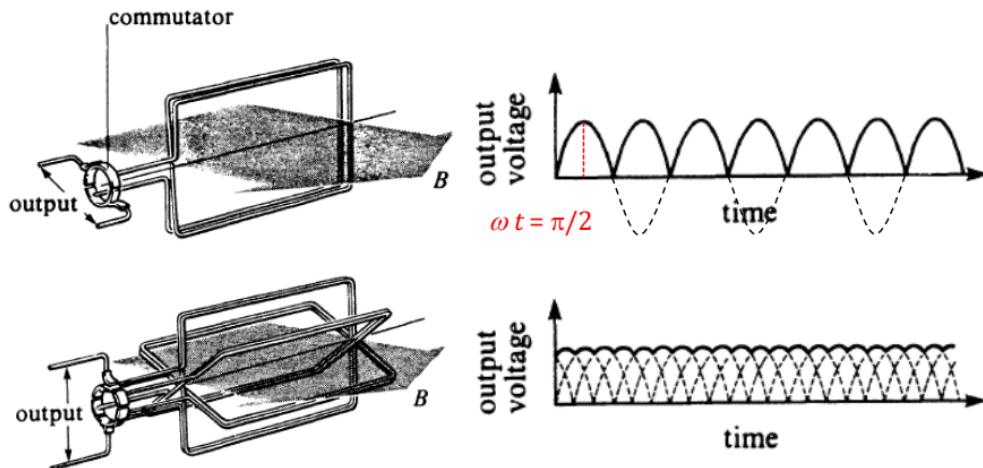
With the contribution from the other side (ST), for n turns of the coil:

$$\text{Total induced voltage} = 2Bl \cdot n \cdot wr \cdot \sin(\omega t) \quad (13.4)$$

From the equation, we can see that the output voltage is sinusoidal with amplitude proportional to the angular velocity ω (or frequency $f = \omega/2\pi$).

However, it is difficult to deal with an output where both the amplitude and frequency vary with angular velocity. It is more convenient to have DC output with its magnitude proportional to the angular velocity.

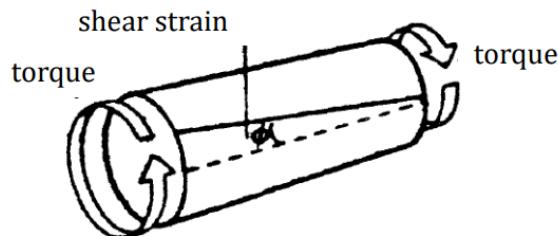
Modifications using **commutator** which switches over the two (n) output terminals from the coil every half ($1/n$) turn.



13.2 Measurement of Torque

13.2.1 Torque measurement with strain gauges

Torque, a twisting force, can be measured by detecting the angular displacement or surface strain of a shaft that is subjected to a torque.



For the use of strain gauge, it is necessary to know how much strain is present on the surface of the shaft for a given torque.

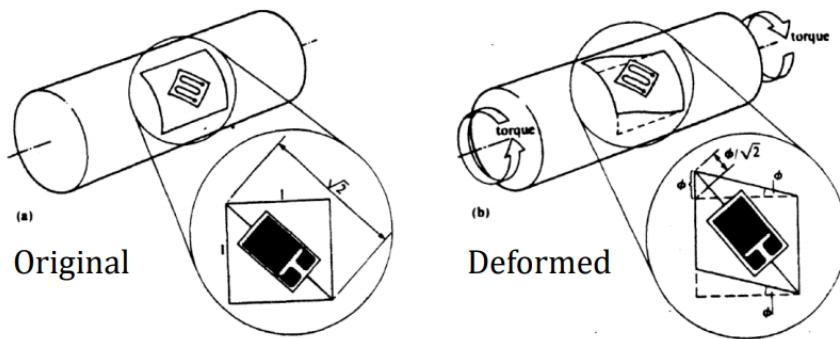
The relationship between torque T , shear strain (angle) ϕ , shear modulus of the shaft material G and radius r is given as:

$$T = \frac{1}{2}\pi Gr^3\phi \quad (13.5)$$

The shear sensitivity can be considered as the shear strain in the surface per unit applied torque:

$$\text{Shear sensitivity} = \frac{\phi}{T} = \frac{2}{\pi Gr^3} \quad (13.6)$$

A strain gauge will measure linear strain in the direction of the gauge axis. To maximise this for a given shear strain in the shaft, the gauge is mounted with its active axis 45° to the axis of the shaft.



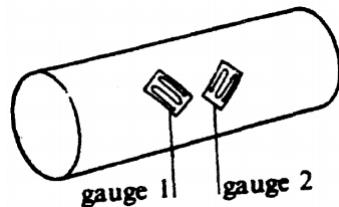
Considering a unit square area on the surface and when ϕ is small, the strain in its diagonal is given as:

$$e = \frac{\phi}{\sqrt{2}} = \frac{\phi}{2} \quad (13.7)$$

The gauge sensitivity can then be:

$$\frac{e}{T} = \frac{\phi}{2T} = \frac{1}{\pi Gr^3} \quad (13.8)$$

Temperature compensation can be achieved by mounting two strain gauges with their active axes perpendicular to each other.



13.3 Measurement of Other Variables

13.3.1 Voltage – voltmeter

Analogue voltmeter and ammeter are making use of electromagnetic force. Digital ones could be made with comparators.

- Analogue Voltmeter - A coil is wrapped around an electromagnet, which is placed inbetween 2 poles of a magnet. The current passing through the coil generates an EMF, which moves the dial across, giving a reading of the voltage.
- Digital Voltmeter - A comparator is used. The input voltage is passed through a series of comparators and through yes-no logic gates, the correct voltage is measured and displayed on a digital reader.

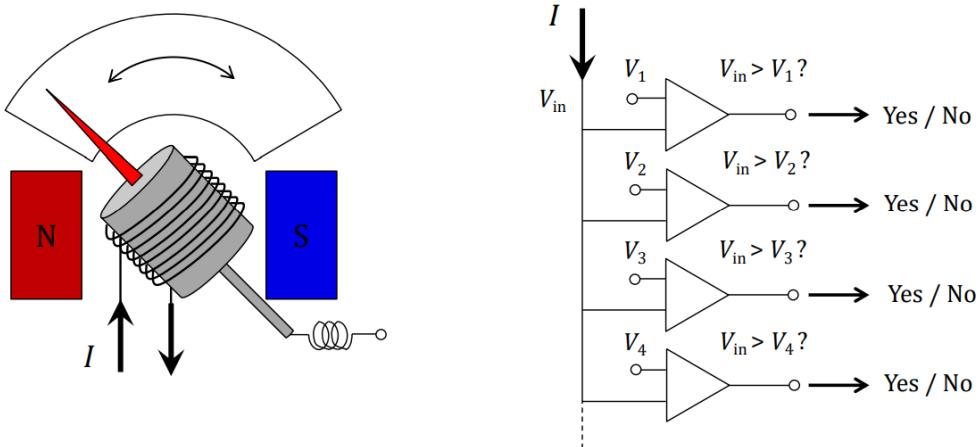


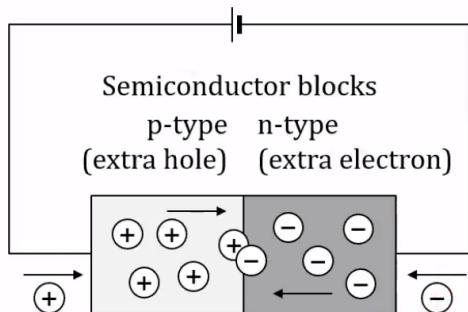
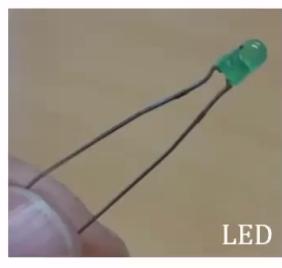
Figure 13.4: Left-Analogue Voltmeter, Right-Digital Voltmeter

Advantages of digital voltage meters (DVMs) over analogue:

- Read out of DVMs is easy as it eliminates observational errors in measurement committed by operators.
- Error on account of parallax and approximation is entirely eliminated.
- Reading can be taken very fast.
- Output can be fed to memory devices for storage and future computations.
- Versatile and accurate
- Compact and cheap
- Low power requirements
- Portability increased

13.3.2 Light – photodiode

LEDs or light emitting diodes make use of semiconductor blocks. Semiconductor blocks usually have a p-type and an n-type block. The n-type usually has an excess of electrons, whereas the p-type has extra capacity to hold electrons. When a current is passed through a circuit with semiconductor LED, the electrons move from the n-type to the p-type block, hence completing the circuit.

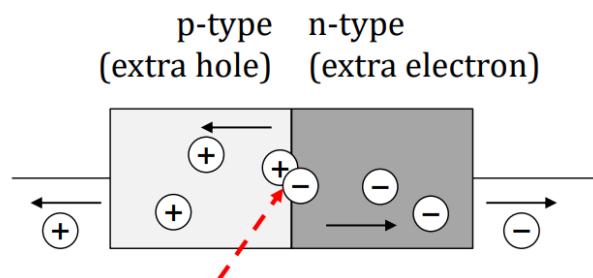
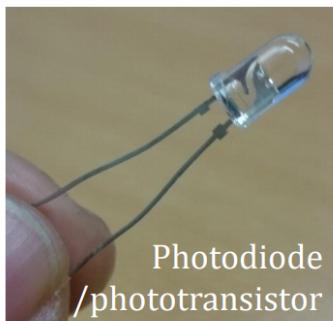


LED: Energy → Light

- Type of semiconductors determines the amount of energy
- The amount of energy determines wavelength of the light ($E = hv$)

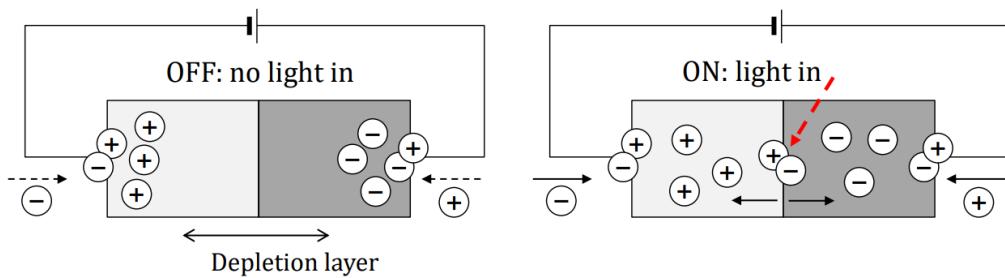
What is the difference between an LED and a Photodiode/Phototransistor?

A photodiode or a phototransistor work on opposite principles to an LED. When light is shone on the diode, it gets converted into energy and completes the circuit. So in this case, the light energy is used to push the electrons from the n-type semiconductor block to the p-type semiconductor block.



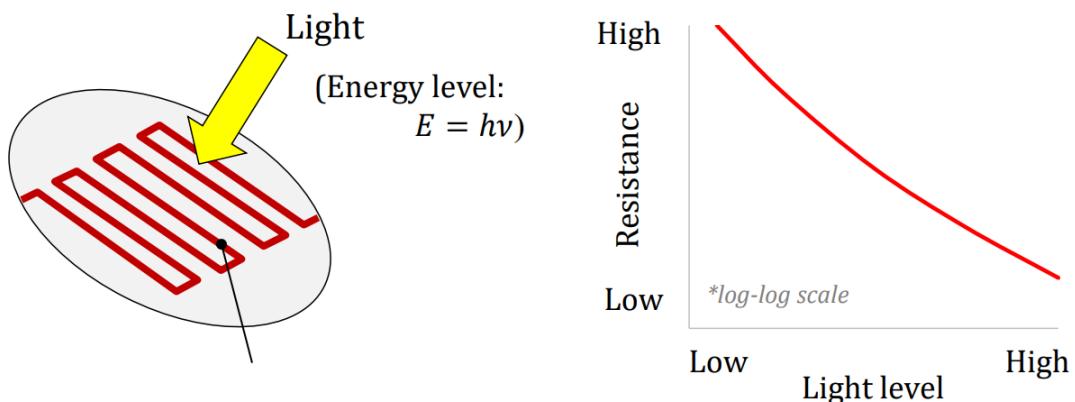
Photodiode: Light → Energy (photoelectric effect)

- Pattern 1: Working as a power source → photovoltaic mode – same principle in solar cells
- Pattern 2: Working as a switch → photoconductive mode



13.3.3 Light – LDR (light dependent resistor)

A light dependant resistor or an LDR changes its resistance depending on the intensity of light being shone on the semiconductor.



Semiconductor (CdS: cadmium sulfide) track:

- Light energy excites electrons from the valence band to the conductive band
 - Low resistance for high light level
- Required energy level to “free up” electrons depends on the track material
 - Semiconductor requires lower excitation energy than metal

13.3.4 Other Important Measurements

- Pressure Measurement
 - Pitot tube
 - Pressure probes (for healthcare applications)
- Velocity/Flow Measurement
 - Rotor
 - Ultrasound doppler

- Optical approach
 - * Dye visualization
 - * Particle image velocimetry

Chapter 14

Measurement of Temperature

30/11/2020

There are many ways to measure temperature, including changes of:

1. Pressure/volume (thermal expansion) – generic thermometer;
2. Contact potential between different metals – thermocouple;
3. Resistance of a metal or a semiconductor – thermistor;
4. Energy radiated by a hot object – radiation pyrometer.



Note: None of the properties of materials used for temperature sensing varies strictly linearly with temperature. Thus, calibration is very important for temperature transducers.

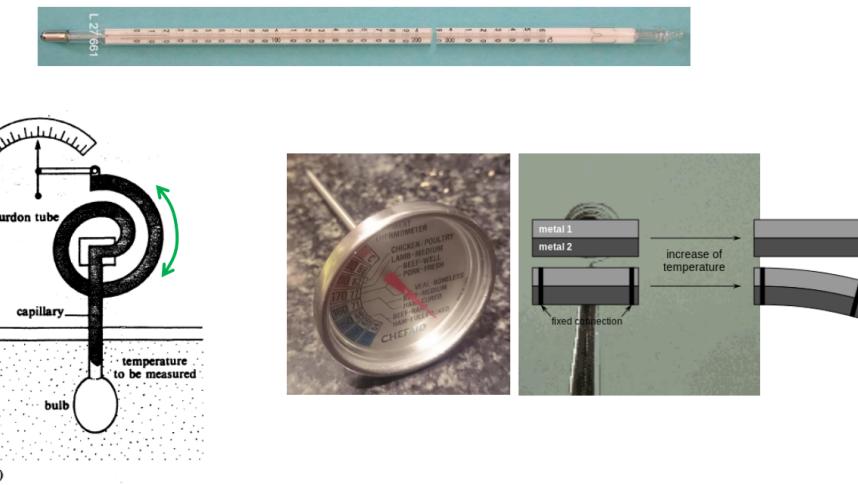
14.1 Standard Temperature for Calibration

In the IPTS (the International Practical Temperature Scale), values of temperature are assigned to the eleven reproducible fixed points shown below.

Fixed point	Assigned temperature	Standard instrument
		Optical pyrometer (above 1337.58 K)
Freezing point of gold	1337.58 K	Thermocouple
Freezing point of silver	1235.08 K	(90 – 1337.58 K)
Freezing point of zinc	692.73 K	419.58 °C
Boiling point of water	373.15 K	100 °C
Triple point of water	273.16 K	0.01 °C
Boiling point of oxygen	90.188 K	-182.962 °C
Triple point of oxygen	54.361 K	-218.789 °C
Boiling point of neon	27.102 K	-246.048 °C
Boiling point of equilibrium hydrogen	20.28 K	-252.87 °C
Equilibrium between the liquid and vapour phases of equilibrium hydrogen at 33330.6 Pa	17.042 K	-256.108 °C
Triple point of equilibrium hydrogen	13.81 K	-259.34 °C

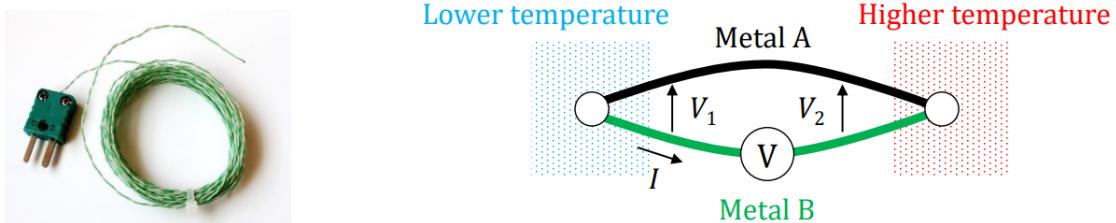
14.2 Measurement Using Changes in Volume/Pressure

Thermal expansion is the key.



14.3 Thermoelectric Temperature Sensors (Thermocouples)

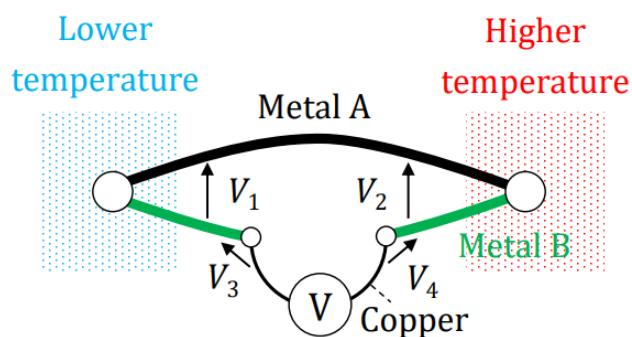
A thermocouple consists of two wires of different metals joined at both ends. If the junctions between the metals are at different temperatures, an electric current will flow around the circuit. This phenomenon is called **thermoelectric or Seebeck effect**.



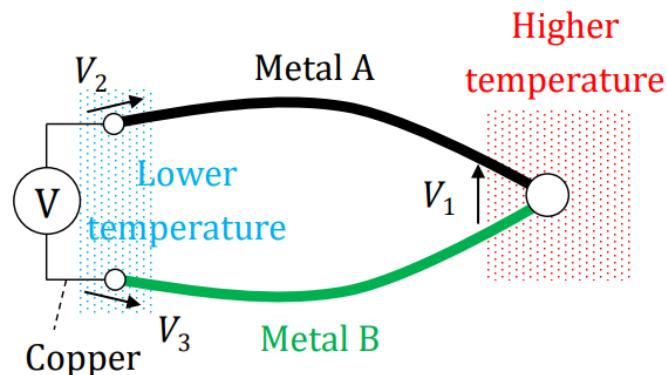
Across the junction of any two dissimilar metals, there always appear a difference in electric potential called the ‘contact potential’. The contact potential varies with the temperature at the junction between the two metals, increasing in magnitude with an increase in temperature.

14.3.1 Thermocouples in practice

Contact potential could also occur at the junction between metal B and the wires leading to the voltmeter, unless the wire is made of metal B (very unlikely), generating contact potential V_3 and V_4 .



In practice, it does not matter as long as the junctions are kept under the same temperature (i.e. $V_3 = V_4$).



An alternative configuration is the one in the bottom panel. Here, the net contact potential is $V_1 + V_3 - V_2$.

14.3.2 Thermocouple materials and construction

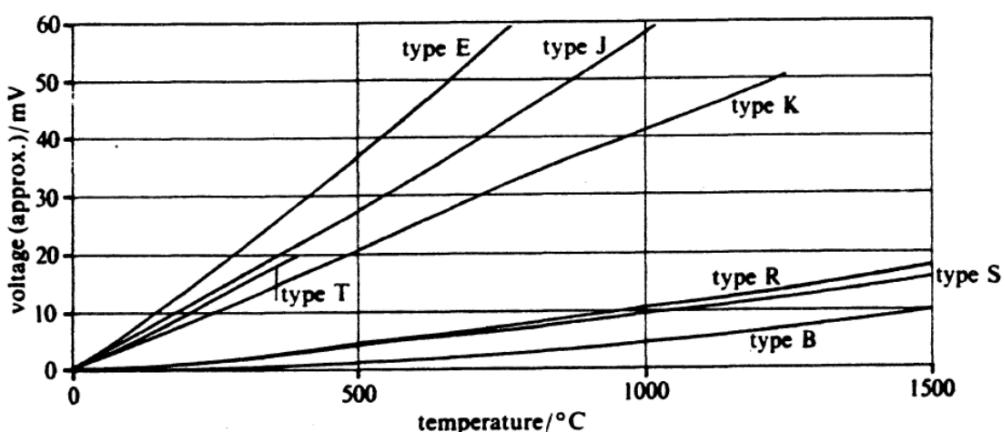
Various standards have been developed for thermocouples, with the relevant British Standard being BS4937, part of which is shown below.

Designated type	Metal or alloy For 1 st wire	For 2 nd wire	Useful temp range (°C)	Environment for use
Type S	Platinum	90% platinum, 10% rhodium	0 to 1400	A
Type R	Platinum	87% platinum, 13% rhodium	0 to 1400	A
Type J	Iron	Copper nickel	-200 to 850	B
Type K	Nickel chromium	Nickel aluminium	-200 to 1100	A
Type T	Copper	Copper nickel	-250 to 400	C
Type E	Nickel chromium	Copper nickel	-200 to 850	C
Type B	70% platinum, 30% rhodium	94% platinum, 6% rhodium	0 to 1500	A

A: In oxidising or inert atmospheres
B: In reducing or inert atmospheres
C: In most non-corrosive atmospheres

The choice of the right thermocouple for any application depends on:

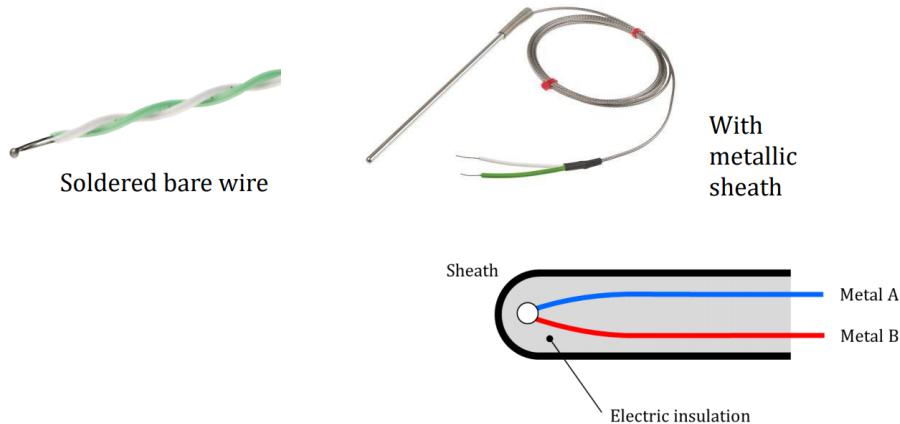
- The sensitivity as well as the temperature range (shown in the figure)
- Applicability in corrosive environments
- Whether it can withstand thermal and mechanical stresses
- Accuracy and reproducibility
- Price



14.3.3 Thermocouple structure

Thermocouples can have 2 different types of probe ends:

1. Soldered bare wire
2. With metallic sheath



The main differences between the two different types:

- With a metallic sheath, the probe can touch conductive surfaces
 - Sheath has electric insulation inside, so the reading won't be affected
- Longer response time
 - The metallic sheath and the electric insulation must be heated up first before reaching the joint inside
- Insulation material needs to be high electric resistance + high thermal conductivity

14.4 Basic Metallic Resistance Thermometers



The resistance of certain metals varies with temperature. The resistance R_T at temperature T is approximately shown relative to the resistance R_0 at 0°C , using the following equation:

$$R_T = R_0(1 + \alpha T) \quad (14.1)$$

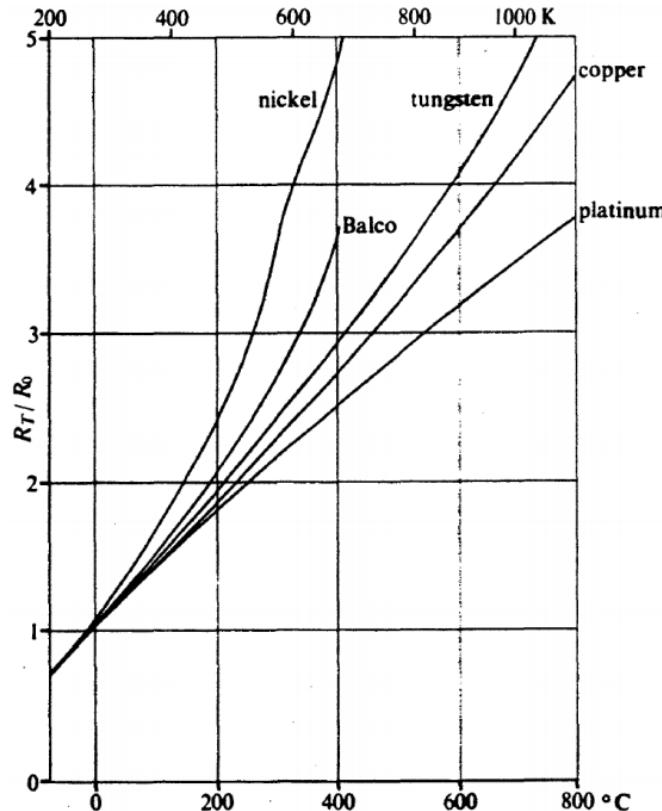
Where:

- α is called the temperature coefficient of resistance

This could be described more generally:

$$R_T = R_0(1 + \alpha(T - T_0)) \quad (14.2)$$

taking T_0 as the reference temperature from R_0 .

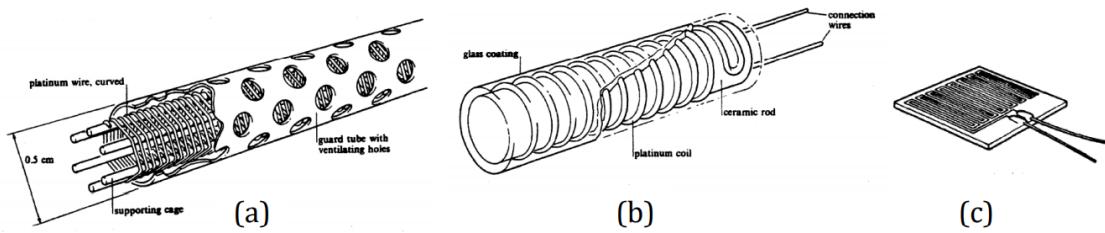


The figure shows the relative resistance variation of certain metals over temperature (using R_0 at 0°C). Although the sensitivity of platinum is the smallest, it has advantage of:

- Most stable and reproducible (linear),
- Chemically inert and resistant to contamination.

Typical resistance thermometer construction includes:

- A type with its platinum coil of wire directly exposed to the fluid whose temperature is to be measured.
- The resistance wire is isolated from the fluid, so that the resistance wire does not become contaminated with other materials which could change the relationship of its resistance to temperature. The ceramic rod is to help the sensor to withstand mechanical shock.
- The resistance metal is deposited as a thin film pattern on an insulating substrate. This type of sensor can be attached to a flat surface.

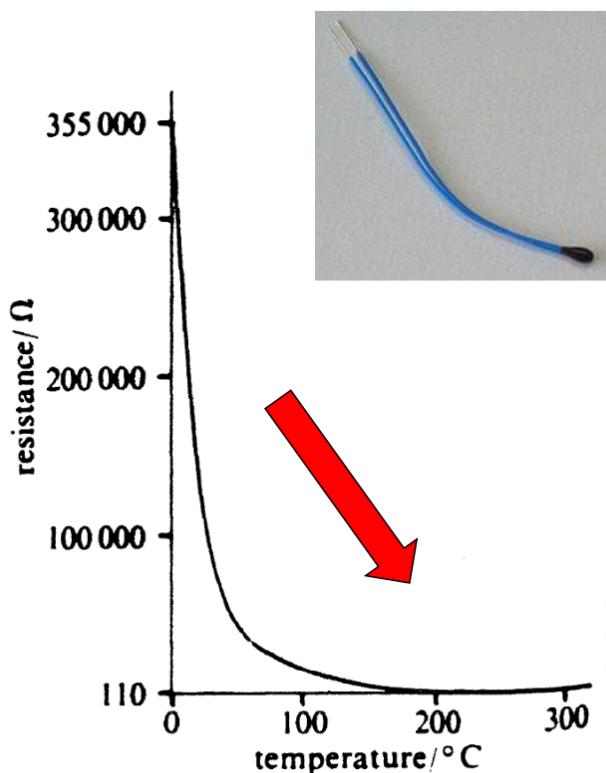


Errors can be caused by:

- Self-heating effect – electric current flows through the resistance wire heats the resistor up because of energy dissipation.
- Thermoelectric effect – likewise thermocouple, if the resistance wire and the lead wire are made of different metal, an additional electric current will flow due to thermoelectric effect.

14.5 Thermistors

THERM-ally sensitive res-ISTOR



The resistance against temperature curve for a thermistor (semiconductor) is very different from that for metallic resistance thermometers. It decreases with temperature rather than increase, and the negative temperature coefficient of resistance in this case is much larger than that of metals.

The relationship between R_T and temperature T for thermistor can be modelled as:

$$R_T = A \cdot e^{\frac{B}{T}} \quad (14.3)$$

Where:

- A and B are constants.
- B is called the characteristic temperature of the thermistor (material constant) whose value is normally between 2000 K and 4000 K.

If the resistance at a temperature T_0 is known (for example, base resistance at room temperature),

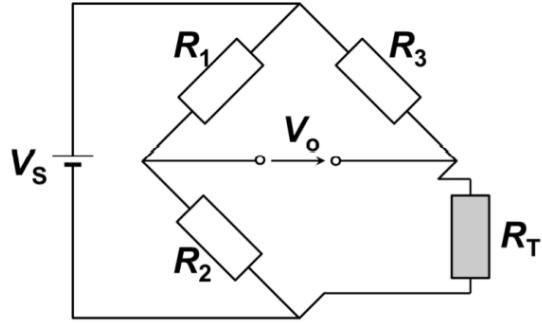
$$R_0 = A \cdot e^{\frac{B}{T_0}} \quad (14.4)$$

and combining the two equation:

$$R_T = R_0 \exp \left(B \left[\frac{1}{T} - \frac{1}{T_0} \right] \right) \quad (14.5)$$

Thermistors can be made **much smaller** than metallic resistance thermometers. This enables them to **respond to temperature variation more quickly**. However, this also means that the **self-heating effect is greater** for the same current than metallic sensors, and so they must be operated at smaller current levels.

14.5.1 Measuring circuits for resistive sensors



We use Wheatstone bridge for resistive sensors and the output voltage from the bridge is:

$$V_o = \left(\frac{R(1+x)}{R(1+x) + R} - \frac{1}{2} \right) V_s = \frac{x}{2(2+x)} V_s \quad (14.6)$$

Where x is increase in the resistance due to the measurement and all resistors are assumed to have the same resistance at reference temperature. **If x is small enough**, this can be approximated as:

$$V_o = \frac{x}{4} V_s \quad (14.7)$$

But this does not apply to resistive temperature sensors, that is thermistors (x is large) and the bridge output voltage becomes highly nonlinear against T .

A simple workaround is not to measure V_o but to vary R_2 (by use of potentiometer) such that the bridge is balanced. If $R_1 = R_3$, $R_2 = R_T$ allows calculation of T , based on the following relationships.

Resistance thermometer:

$$R_T = R_0(1 + \alpha(T - T_0)) \quad (14.8)$$

Thermocouple:

$$R_T = R_0 \exp \left(B \left[\frac{1}{T} - \frac{1}{T_0} \right] \right) \quad (14.9)$$

14.6 Pyrometer and Radiation

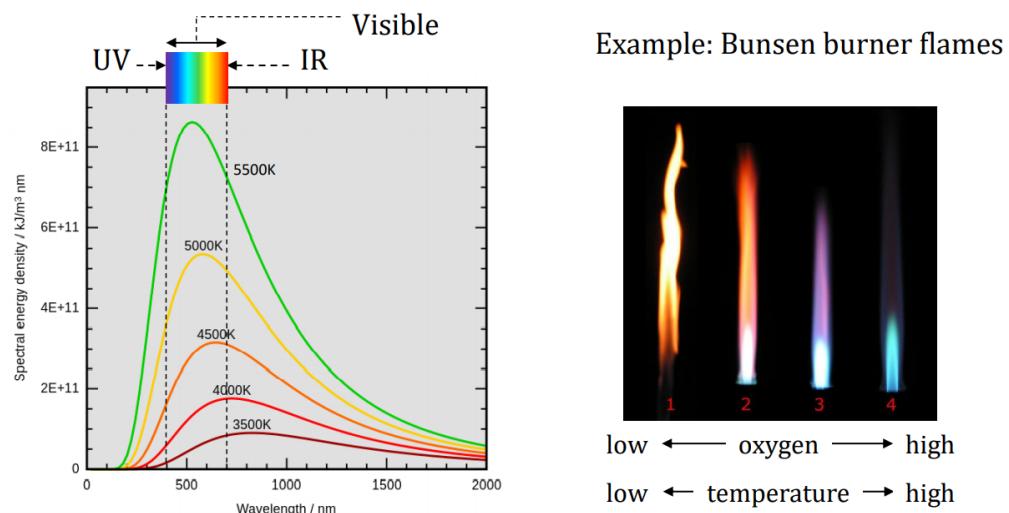
Radiation pyrometers detect the infrared radiation given off by the object whose temperature is being measured. Radiation pyrometers can be used to measure temperatures as low as -50°C , so infrared radiation is not only emitted by hot bodies. Any object whose temperature is not at absolute zero (0 K) emits radiation.

Radiation = travel of energy wave (like X-ray)

- Does not require medium to propagate
- \neq sound (which requires medium to travel)

14.6.1 Black body

A body at a certain temperature emits a certain amount of energy with a certain spectrum (see figure below). Note: An ideal body, blackbody, emits radiation in all wavelengths. This is used to measure temperature in approximate manner.



14.6.2 Radiation pyrometers



Working principle:

Temperature transducer is used in the detector (e.g. thermopile), the detected temperature is converted to the radiation energy using Stefan-Boltzmann law.

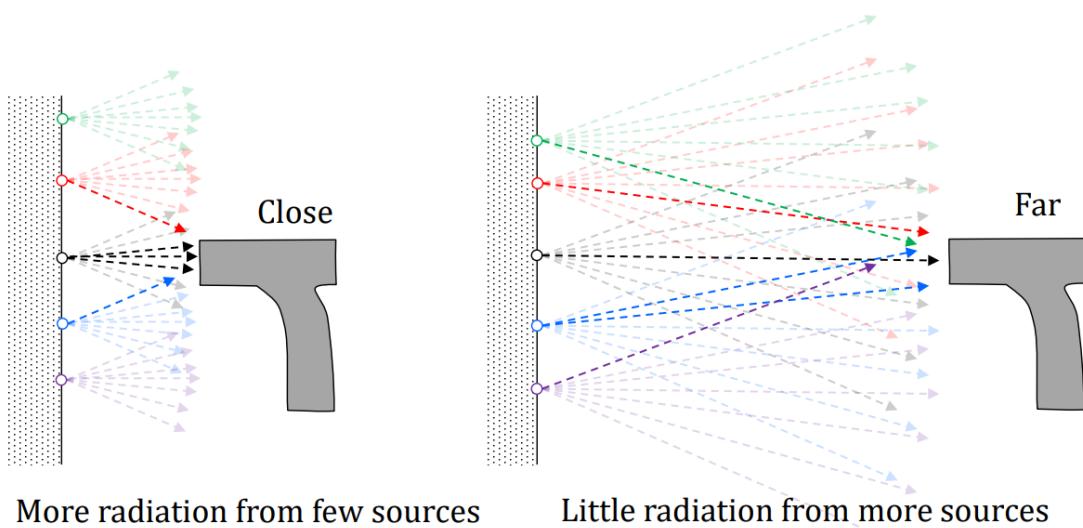
$$M = \sigma T_{detected}^4 \quad (14.10)$$

- Note: $T_{detected}$ is the temperature due to radiation and not the temperature of the surface

The energy is then converted to temperature of the measured surface using the black body energy plot.

Some features

- Very short response time
- Reflective or transparent/translucent surface is likely to lead to error
- Distance independent measurement



14.7 Temperature Response

A challenge of temperature measurement → **no thermometer can respond instantly to a change in temperature**. The response is expressed mathematically as:

$$\frac{dT}{dt} \propto (T_f - T) \quad \text{or} \quad \frac{dT}{dt} + KT = KT_f \quad (14.11)$$

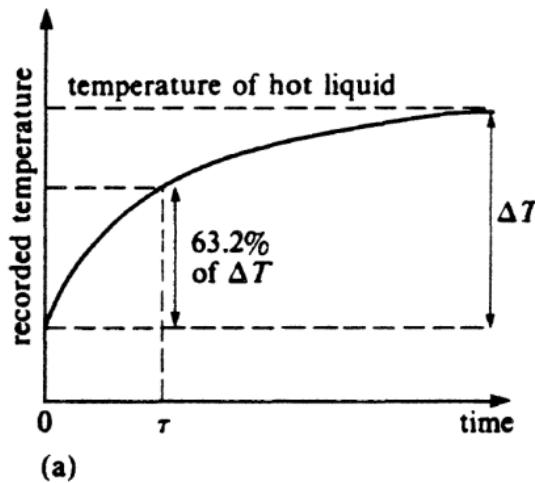
Where:

- K is a constant
- T_f is the temperature to be measured

The general solution is:

$$T = A \cdot e^{-Kt} + T_f = (T_0 - T_f) \cdot e^{-Kt} + T_f \quad (14.12)$$

The speed of change is characterised by the time constant $\tau (= 1/K)$ which is the time interval in which the temperature varies 63.2% of the difference between initial and final temperatures.



14.7.1 Transient temperature response

If the transducer is subjected to a sinusoidally varying input temperature, the differential equation becomes:

$$\frac{dT}{dt} + KT = KT_f = KT_a \sin(\omega t) \quad (14.13)$$

The solution of this equation in terms of the time-constant τ is:

$$y = A \cdot e^{\frac{-t}{\tau}} + \frac{T_a}{\sqrt{1 + \omega^2 \tau^2}} \sin(\omega t + \varphi) \quad (14.14)$$

Where:

- $\varphi = \tan^{-1}(\omega/k)$. The figure below shows the two temperature variations.

