THE EE EPP TEAM

ENGINEERING PRINCIPLES & PRACTICE

ELECTRICAL ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

NATIONAL UNIVERSITY OF SINGAPORE

Contents

Part I

EPP I

How to Work Safely?

Keeping the workplace safe is the legal and moral responsibility of everyone who uses it. Each and everyone of us have to make sure that anyone who comes to the workplace leaves it and returns home to their loved ones unhurt. The safety culture at the workplace is of paramount significance and every engineer must strive towards it. Safety may seem like an inconvenience, but you will be glad that this inconvenience was endured when it really matters. Though accidents cannot be completely avoided, they can be very easily reduced to acceptable levels.

If our students do not see and experience safety and health best practices being implemented in the University, they will also most likely enter the workplace ignorant of safety and health issues.

Prof. Tan Chorh Chuan,
Former President of NUS

Student Preparation

There are two parts to preparing for this class. The first task is to read this set of materials and complete the General Laboratory Safety Quiz `https://online.ece.nus.edu.sg/quiz/epp` as required by the Department of Electrical and Computer Engineering. There are 20 questions in the quiz and the passing the quiz requires scoring 100%.

If you fail the Quiz, you will be barred from the laboratory and will not be able to participate in the laboratory-based studio sessions. If one passes the Quiz, a certificate of completion will be awarded. Sign and date the certificate and submit into the Luminus folder before your studio session for access to the laboratory space.

For the second part of the preparation, students are required to watch the following videos pertaining to safety. If you are not on NUS campus, connecting to the NUS network via nVPN is required.

1. **Unsafe Acts** `https://nus.safetyhub.com/embeddirect/3LQ9OOJG/1687966`

   This is a video about unsafe acts. Reasons for unsafe acts and how to avoid them are explained in this video.

2. **Basic Electrical Safety**
   `https://nus.safetyhub.com/embeddirect/98U7B5TU/102226`

   This video discusses why electrocution may occur and the steps required to prevent such an accident.

Basic Laboratory Safety Procedures

1. **What to wear?**
   Much of the answer to this question is common sense. The laboratory is a workplace where many different types of equipment will be stored and used, and there is a likelihood that some of these may fall onto the ground. To protect one's feet, it is required that all students wear **toe covered shoes**. When dealing with electrical appliances, **rubber-soled shoes** are recommended to increase insulation from the ground.

   Long hair and very loose clothing, hand and neck jewelry must be secured or removed prior to work in the laboratory. When working with equipment that apply heavy loads on samples or with machines that cut or shape metals, students are required to wear **safety goggles**.

2. **What to do in cases of Emergency?**

   (a) **Accidents**
   All accidents must first be reported to the Laboratory Officer as they are trained in safety and first aid. Please do not attempt to render first aid if you are not trained to do so. Always call for help. In this regard, the student is required to know by heart at least the telephone numbers tabulated in Table 1.

	Agency	Tel. Number
i	Police	999
ii	SCDF (Fire Brigade and Ambulance)	995
iii	Campus Security	6874 1616
iv	Faculty of Engineering Safety Office (24 hrs)	6601 3765

Table 1: Important telephone numbers.

   Please note that it is important to inform Campus Security if a call has been placed to the Police or SCDF. This is to help direct the Police and SCDF to the exact location within campus of the accident. The telephone number of the Faculty of Engineering Safety Office is manned 24 hours.

   (b) **Fire**
   Once again, in cases of fire, the student must first inform the lab officer. The lab officer will attempt to put out the fire with a fire extinguisher if it is manageable. He may also require the students to evacuate the laboratory. If the fire requires all other occupants of the building to evacuate, the fire alarm will be activated.

   Upon hearing a **fire alarm**, students should leave their belongings and evacuate from the laboratory immediately and gather at the Assembly Point. Calmly egress from the lab using the designated fire exits. Fire exits are guaranteed to be not blocked at all times. When out of the lab, use the staircase to leave the building and not the elevators.

Upon leaving the building, students will gather at the **Assembly Point**, where their attendance will be taken and tallied with the attendance register. This is very important. If for any reason, a student who attended the lab is not at the Assembly Point, it shall be assumed that the student is still inside the lab. Instead of mitigating the emergency, the SCDF Officers will be engaged in a mission of search and rescue for the missing student.

Every laboratory has an associated Assembly Point. Therefore, it is very important that each and every student is aware of the Assembly Point associated with the laboratory that he or she is working at. It is always a good practice to make sure that the first task that a student performs upon entering a laboratory is to ascertain the Assembly Point associated with the laboratory. This information is often found on the Laboratory's Safety Notice Board. If you are unable to locate the Safety Notice Board, you should speak to any laboratory Officer.

- Where is the Assembly Point associated with the laboratory where you will be doing your studio sessions?

Working in Laboratories

You may be exposed to a diverse range of physical, electrical, mechanical, fire, chemical, radioactive hazards while working in laboratories. A **hazard** is anything that can cause potential harm to you and to others. Hazards can arise from the equipment being used, the agents and materials being used or the procedure being followed. The **risk** level from the hazard is evaluated by taking the product of the likelihood and severity of the hazard in causing injury or ill health to you and others.

To keep yourself safe while working in such environments, you need to follow the safe work procedures which are derived from the **safety and risk assessment** process.

Hierarchy of Control

The hierarchy of control is a system widely used in industries, safety organizations and also in NUS to minimize or eliminate exposure to hazards. Following the hierarchy from top to bottom when considering hazards leads to the implementation of safer systems with reduced risk of illness or injury.

1. **Elimination** refers to the physical removal of the hazard and is the most effective form of control. However, this often is the most difficult to implement in an existing process and major changes in equipment and procedures may be required.

2. **Substitution :** Where elimination is not possible, consider substitution. Is it possible to replace something that produces a hazard with something that does not produce a hazard? For example, to reduce risk of cuts from breakage of glassware, plasticware can be used instead.

3. **Engineering Controls** are implemented to prevent a hazard from coming into contact with the worker or people. Examples include fume cupboards, smoke absorbers, electrical conduits, safety guards etc. Well-designed engineering controls may be of higher cost than administrative controls or PPE (see items 4 and 5 for explanation on these two controls) but may be cost-efficient over the long-term or provide cost savings in other areas of the process.

4. **Administrative Controls** limit or prevent people's exposure to the hazards by attempting to change the way people work. Examples include changes in procedure, instituting standard operating procedures, employee training, installation of signs and warnings, inspections and maintenance of equipment, etc. Though they are generally inexpensive to establish, they can be costly to sustain over the long term.

5. **Personal Protective Equipment** (**PPE**) is the least preferred form of control and refers to equipment that can be worn by the worker to provide protection when all other control measures are not practical. Examples include gloves, safety goggles, laboratory coats, respirators etc.

- Do you think these control measures are mutually exclusive?

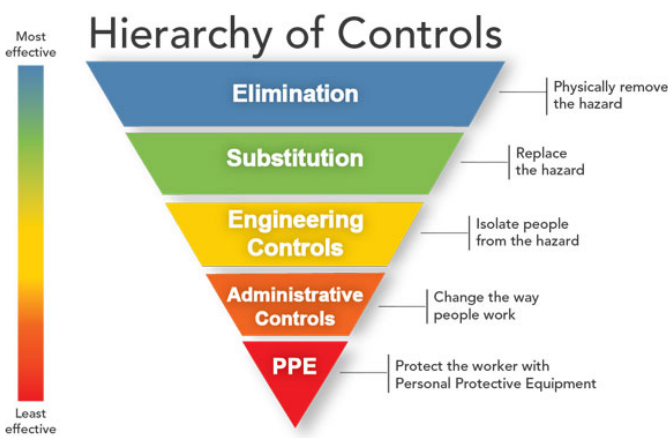


Figure 1: Hierarchy of control

Basic Electrical Safety

It is a common misconception that electrical incidents only happen when there are high voltages. However, the human body can detect currents at levels as low as 1 mA, or 0.001A, and a small current can have a dramatic effect on the human body. The consequences of current flowing through human body are illustrated in Figure 2.

- It only takes a very small current in the order of about ONE milliampere to feel it.

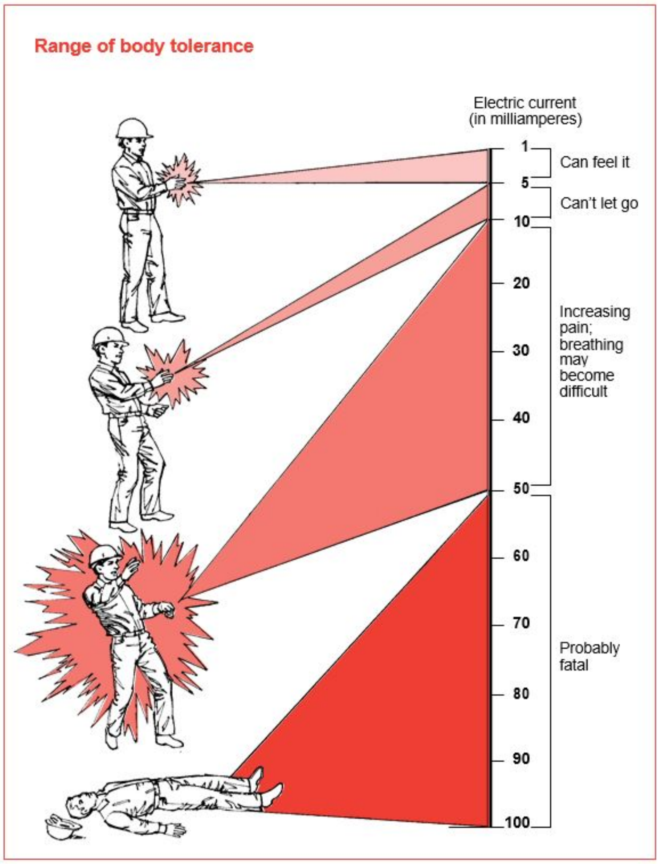


Figure 2: Range of tolerance of human body to electric current, (Source of the image: https://hsseworld.com/types-of-electrical-injuries/)

- If the current is increased to between FIVE to TEN milliamperes, the human experiences electrical shock and loses control of muscles. Thus, he will not be able to let go of the live object and the circuit cannot be broken by the human.
- Between TEN and FIFTY milliamperes, the human muscles fibrillate (i.e., contractual spasms) and the human experiences pain and has difficulty in breathing.
- Current exceeding FIFTY milliamperes for less than a second are fatal due to ventricular fibrillation in the heart and it causes burns to the skin and internal organs.

This means that voltages as low as 30V can present real hazards in situations where the person's resistance to electricity is low.

Major causes of electrical accidents

- Ignorance, Negligence and Forgetfulness
- Accidents resulting from the fault of persons other than the injured

- Working on live gear deliberately
- Disabling or working without safety devices such as grounding conductors, Ground Fault Circuit Interrupter (GFCI) etc.
- Misconception: "Low voltages are harmless." This is not true.

The left schematic in Figure 3 demonstrates electrocution by direct contact where the human accidentally comes into contact with the live wire. Current can only flow through the human if and only if there is a complete closed circuit. If the human is wearing proper soled shoes, this prevents electricity from flowing from the live wire to the ground. Unfortunately, he is in contact with an external conductor that closes the circuit and current flows through the human and causes harm.

Electrocution due to indirect contact is illustrated in the right-hand side schematic of Figure3. In this instance, the equipment is faulty and the live wire is in contact with the equipment chassis. If the human touches the equipment with proper soled shoes, he will not be hurt unless he closes the circuit by touching an external conductor.

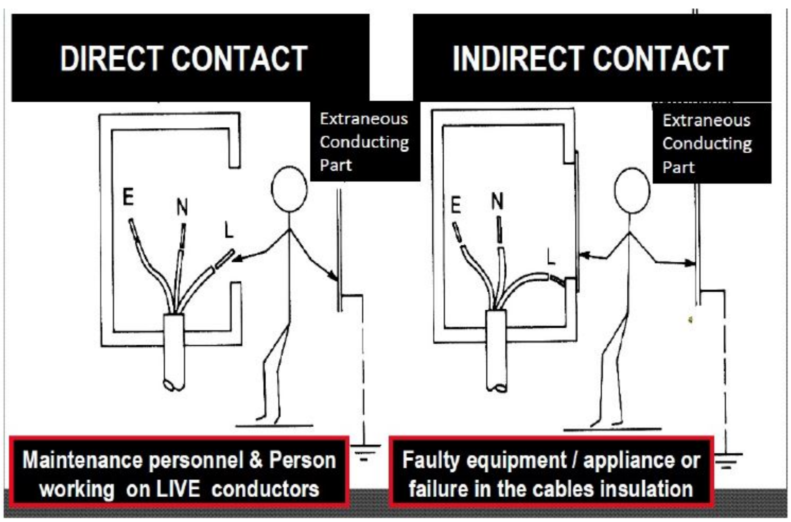


Figure 3: Direct electrocution and indirect electrocution

Practical Tips

When current flows through a human, it affects our muscles by contracting them. Hence when one grasps a faulty equipment, the current contracts the human's muscles and he will not be able to let go of the equipment. These brings us to the following practical tips when handling live voltage equipment.

1. Always treat all equipment as if they are live and avoid unnecessary contact.

2. Always wear proper soled shoes. The insulation provided by the soles prevents a closed circuit and current will not flow through the human to the ground.

3. Never grasp an equipment with your palm. Always use the back of your hand to contact the equipment. If it was live, the current will not tighten the grasp and he can let go of it.

4. On touching an electrical equipment for the first time, always put your non-major hand in your pocket. This will prevent your hand from touching an external conductor.

Figure 4 shows the possible current paths through the human body. During an electric shock, the greatest internal current density will be along the shortest path between two areas of contact. Circuits established from hand to hand, hand to leg and head to leg are therefore critical as they pass through most of the vital organs. Note:

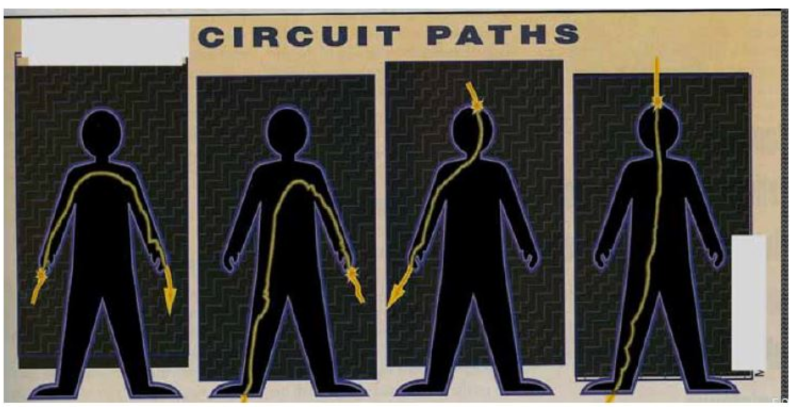


Figure 4: Various possible current path

- the current path through the leg can be prevented by wearing soled shoes.
- the current path through the head can be prevented by wearing insulated hard hats.
- by keeping a hand in one's pocket, the likelihood of a hand to hand current flow can be minimized.

Hazards in Appliances

All users of electrical appliances should conduct a visual check of equipment before use. Many faults with electrical appliances can be detected with a visual inspection.

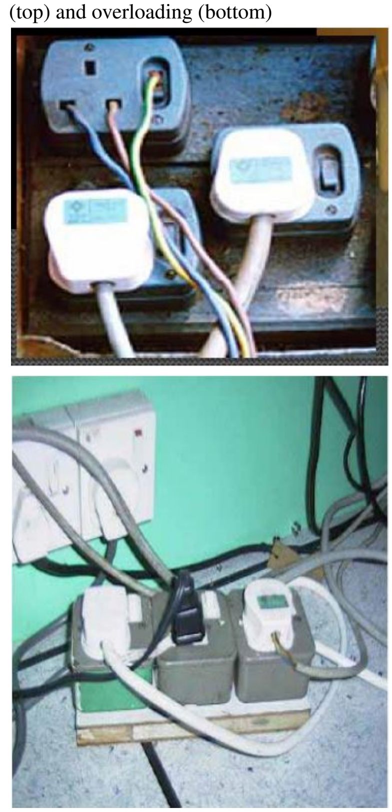
Portable appliances (Figure 5) account for 25% of electrical accidents at work and should be inspected before use and tested periodically depending on the frequency of use.

Figure 5: Portable appliances - grinder (left) and drill (right)

Typical electrical hazards in appliances

- Defective equipment
- Damaged electrical cords or failure of protective insulator
- Exposed electrical wires (See Figure 6 (top))
- Damaged electrical plugs
- Overloading of electrical plugs or extension cords (See Figure 6 (bottom))
- Using electrical equipment in wet or damp conditions
- Flammable materials in the vicinity
- Signs of equipment overheating
- Ingress of foreign material
- Failure of cord grip
- Locked ventilation for equipment

Figure 6: Electrical hazard - exposed wire (top) and overloading (bottom)



Typical mechanical hazards in appliances

- Entanglement with machine (Figure 7 (top))
- Crushing, shearing or cutting (Figure 7 (bottom))
- Impact
- Stabbing or puncture
- Friction or abrasion
- Contact with material in motion
- Struck by ejected materials/parts of machines

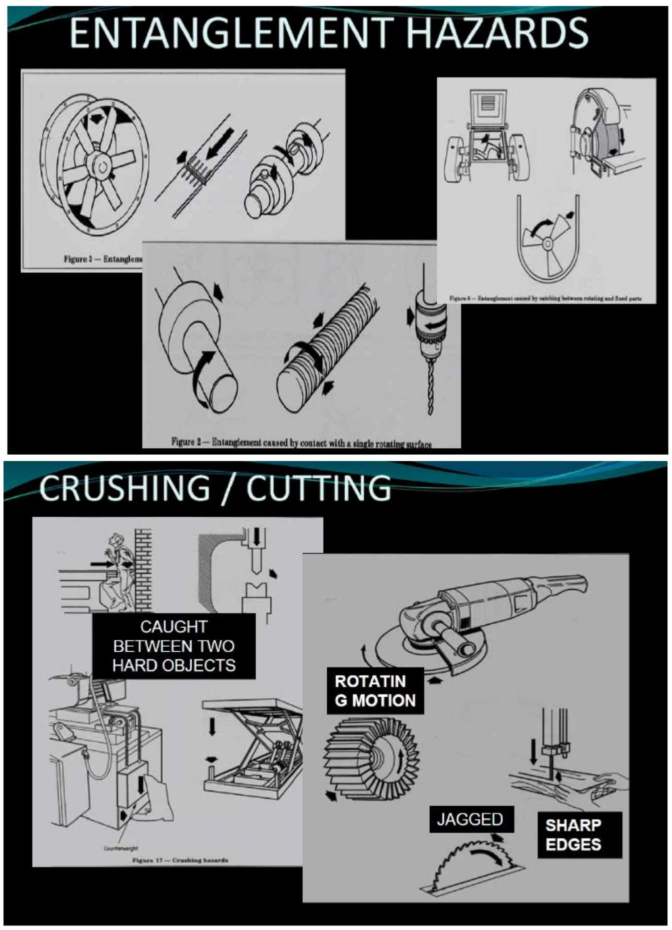


Figure 7: Examples of mechanical hazards - entanglement (top) and crushing/cutting (bottom)

How do Engineers Communicate?

Engineering Drawings and Diagrams

Communication is an important aspect is almost all professions. But the aspects to be emphasized on during communication vary from profession to profession. The mode(s) of communication has evolved over time differently for different professions. In addition to the usual categories of communication being oral and written, engineers also communicate through drawings, numbers and analyses.

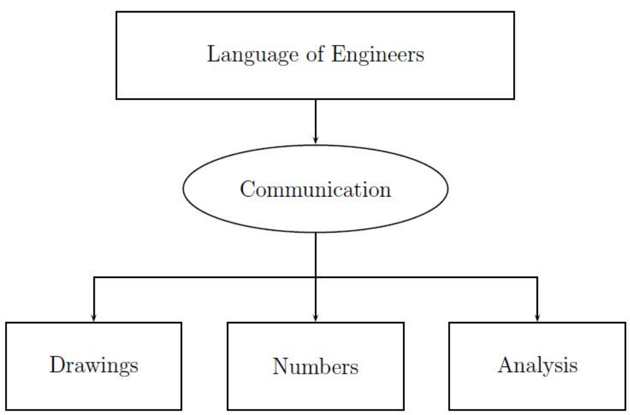


Figure 8: The language of engineers

The Language of Engineers allows us to communicate engineering information to others. Drawings or diagrams provide a clear method of communicating ideas and objects without ambiguity and limitations in language. Some examples are included in this chapter but the scope of this lesson does not permit us to get into the details of the many kinds of drawings. Explore the many resources referenced to learn more.

If engineering drawings are analogous to an art masterpiece, then surely, the individual brush strokes must be analogous to the numbers used in the engineering calculations. How do we present these numbers? How many decimal places must we use? Can our final answer have more precision than the numbers we used in our calculations? Engineers follow the rules of significant numbers while using number as tool of communication.

Finally, engineers are also masters at making quick estimates to engineering problems. We refer to these quick estimations as the **"back of envelope calculations"**. Experience and a good foundation of fundamental knowledge are pre-requisites for such estimates.

One of the methods of describing systems is using block diagrams. Block diagrams, as per ISO [1]]Software and System Engineering vocabulary, is a diagram of a system in which the principal parts or functions are represented by blocks connected by lines that show the relationships of the blocks.

[1][

In general ISO terminology, block diagrams are described as an overview diagram predominantly using blocks (rectangular). It provides a comprehensive view of object or system with a low degree of detailing. Block diagrams are used by engineers to communicate with others who may be engineers or non-engineering clients. Block diagrams are also used by engineers to describe how systems work.

Functional Block Diagram

Engineers can use block diagrams to describe systems in terms of their functions.

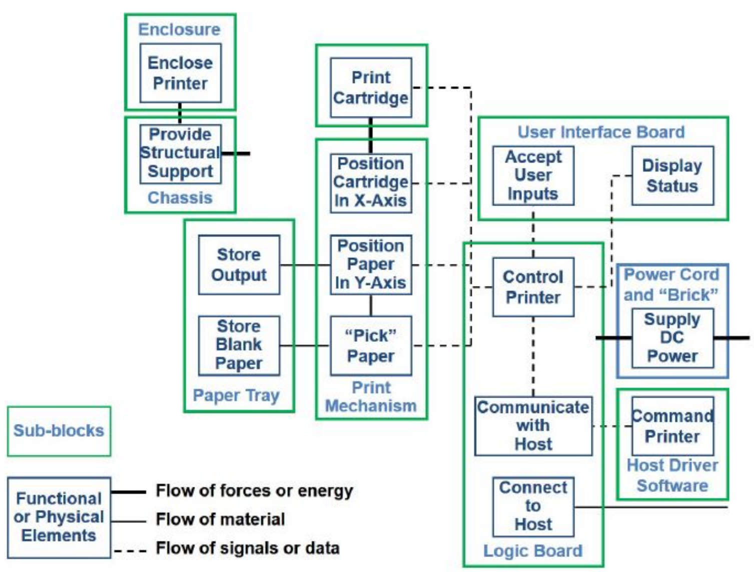


Figure 9: Functional block diagram of a printer, source: Ulrich and Eppinger, *Product Design & Development*

Figure 9 shows the functional block diagram of a deskjet printer. It illustrates the different parts of the printer and their interconnections. Each functional element which performs an action is represented in one rectangular block. There will be flow of electrical energy, material and data between different functional elements; these flows are represented by different types of line. Arrows and labels can also be added in to indicate the direction of this flow. To enhance the visualization of the main functions, different elements can be grouped together to form sub-blocks.

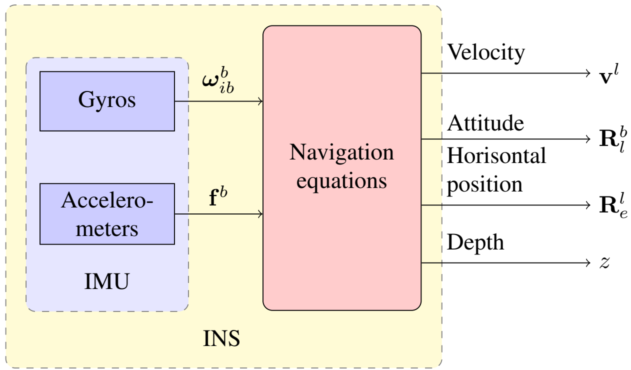


Figure 10: Functional block diagram of an Inertial Navigation System

Making Block Diagram

A simple method of preparing block diagrams is illustrated in this webpage `http://www.technologystudent.com/designpro/system1.htm`. We start by dividing the system into 3 parts:

1. Input,
2. Process, and
3. Output.

Take the case of a simple oven. List the input, the process and the output for the oven. The oven receives user input in turning on and in setting the temperature. The process includes turning on the heating elements to increase the temperature and turning them off when the desired temperature is reached. The output results in the internal environment of the oven having the desired temperature! Now that we understand this, we can break down the system into the necessary functions required to achieve the inputs, process and output above. For example, we need heating elements, a 'brain' or controller to control said heating elements and power for all these items! Are you able to think of any other required functional blocks?

Functional block diagram of a toaster oven is shown in Figure 11. The functional block diagram as its name suggests is not about the components of the toaster oven. But rather, it is about the functionalities of the toaster oven. However, there is a strong correlation between components and functionalities and in general there will be a strong resemblance between the functional and component block diagrams.

A complex system would also have 3 parts at a very high level of function. Take for example the MRT system. List the input, process and output functions for such a system. The high level block diagram can be split into lower

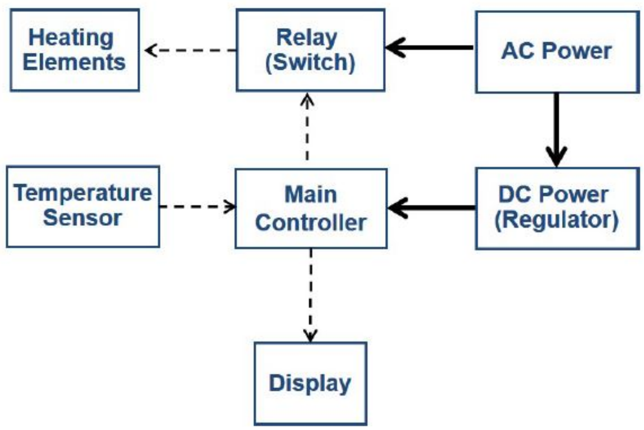


Figure 11: Functional block diagram of an oven

You may not be familiar with the Relay and DC Power Regulator shown in Figure 11.

- The oven generally consists of high power (240V AC Mains) and low power circuits (5V DC).
- The heating elements are high power circuits as they take into large amounts of power from the 240V AC mains.
- The rest of the toaster oven generally will be low power - the controller, the sensors and the display electronics.
- A relay is an electrically operated switch. The main controller is the 'brain' and controls the switching on and off of the heating elements. However, as it is a low powered device, it will not be able power the heating elements by itself. Through the relay, it connects the heating elements to the AC power whenever required.
- The main controller is usually implemented on a micro-controller that operates with a DC Voltage of about 5V o(r 12V). Connecting the AC power to it directly will not be feasible. We will have to convert AC voltage to DC voltage and to step down the voltage from 240V to 5V or 12V using the the DC regulator.

layers. Each layer can be seen as interaction between sub-systems, each sub-system having an input, process and output. For the system to work well, each of the sub-systems should be able to interact with each other without any problem. At the lowest level of detailing we will have schematic diagrams, mechanical drawings, blueprints or even layout diagrams. These other drawings are discussed in the following sections.

Schematic Diagrams

A schematic, or schematic diagram, is a representation of the elements of a system using abstract, graphic symbols rather than realistic pictures.

Most electrical circuits are expressed by schematic diagrams that illustrate the components used and the connections between them. A schematic of an electrical circuit that implements the function of a speaker is shown in Figure 12 and a schematic of a pressurized water reactor is shown in Figure 13.

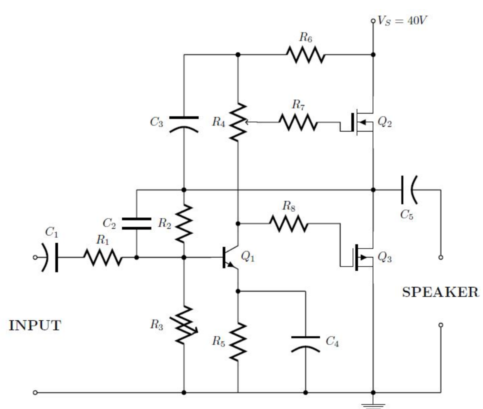


Figure 12: Schematic diagram of a speaker circuit

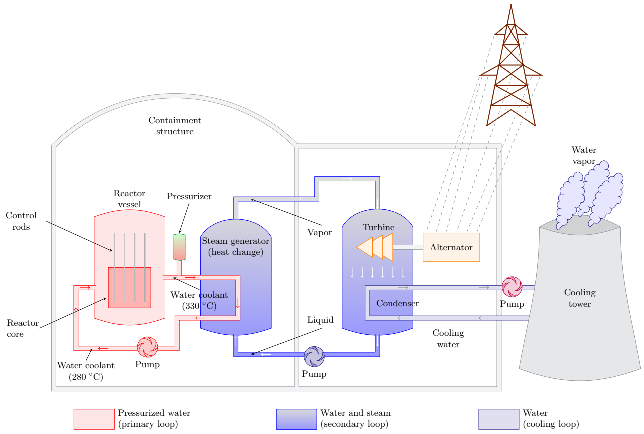
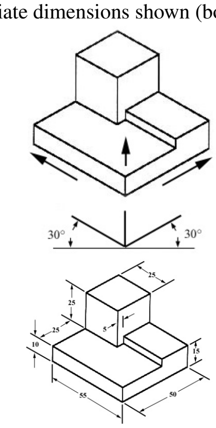


Figure 13: Schematic of a pressurized water nuclear reactor

Engineering Drawings of Structures

An engineering drawing is a type of technical drawing used to define engineering designs, products or components. Typically, the purpose of an engineering drawing is to clearly and accurately capture all geometric features of a product or component so that a manufacturer or engineer can produce the required item. It may also describe the process of making the item, may be used to convey engineering ideas during the design process, or may provide a record of an existing item. Drawings can be created in different forms such as oblique, isometric or orthographic.

Figure 14: The horizontal lines in the width and depth planes are drawn at 30° to the horizontal (top); the drawing comes with appropriate dimensions shown (bottom)



Isometric Drawing

Isometric drawing is way of presenting designs/drawings in three dimensions. In order for a design to appear three dimensional, the object's vertical lines are drawn vertically, and the horizontal lines in the width and depth planes are shown at 30° to the horizontal. When drawn under these guidelines, the lines parallel to these three axes are at their true (scale) lengths. Lines that are not parallel to these axes will not be of their true length.

Orthographics Drawing

Orthographic or orthogonal projection is a method of representing 3D objects via two dimensional drawings. Typically, the top (plan) view, front (elevation) view and side (end elevation) view is provided.

First-angle projection and third-angle projection are two different ways to select the plane of projection while creating orthographic drawings. The main difference between the two lies in the placement of the projection plane with respect to the viewer and the object.

- In the first-angle projection, the object lies between the observer and a non-transparent plane of projection. It is commonly used in almost all

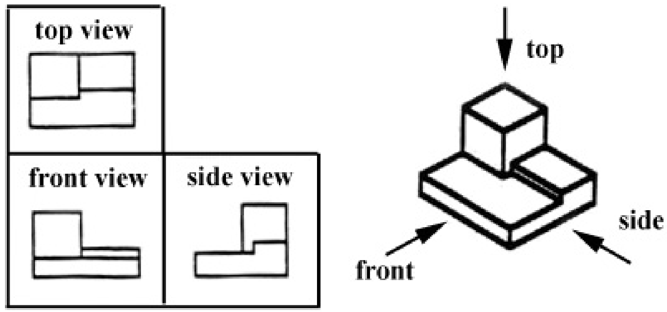


Figure 15: Example of orthographic drawing

countries other than the US.

- In the third-angle projection, a transparent plane of projection is imagined between the observer and the object. It is commonly used in the USA.

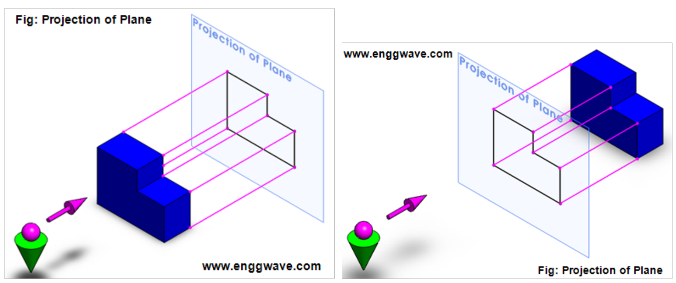


Figure 16: Plane of projections: First-angle projection (Left) and Third-angle projection

They also differ in the way the views are positioned on the orthographic drawing as shown in Figure 17.

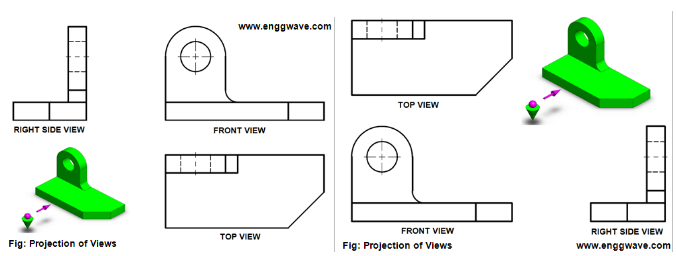


Figure 17: Left: First-angle projection with top view below the front view and right-side view on the left, Right: Third-angle projection with top view above the front view and right-side view on the right. Source: `www.enggwave.com`

Here are some general guidelines:

- The drawing should capture all detail of the object without ambiguity.
- Scales, units and dimensions should be included.
- The number of views that is required is dependent on the object itself. Certain objects may only require 2 views for full representation.
- Figure 9 is an example of third-angle projection, where we simply draw what we see from that view. A common notation used to indicate third angle projection is shown below.

- Sometime the object may have features that are not visible from the current viewing angle. These are hidden details, and dotted lines (or dashed lines with short dashes) are used to depict these edges or voids. Figure illustrates how a mug can be represented.

- Cente-lines are used to represent lines of symmetry in symmetric objects and also for the center of circles and holes. These lines are represented using chain-dotted (dot-dash-dot-dash) lines.

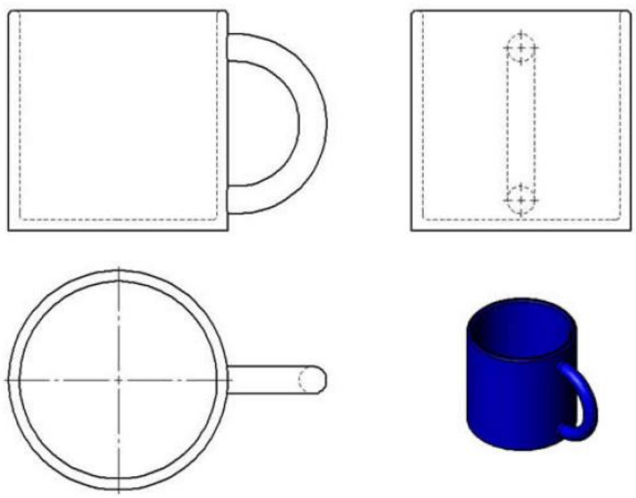


Figure 18: Orthographic drawing of a mug

Layout Diagrams

Layout diagrams show the position of different parts of the system with respect to each other. This could be a large system like a aircraft, see Figure 19.

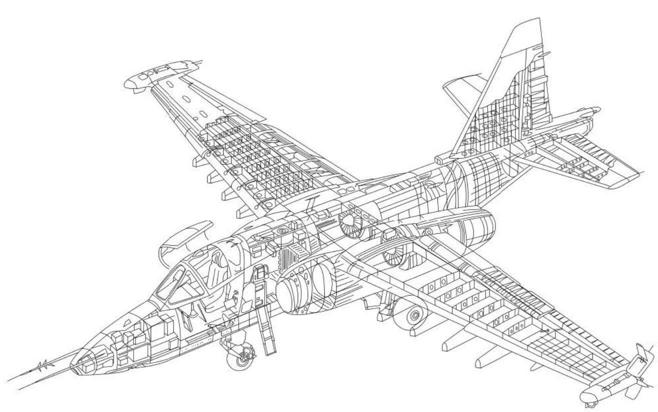


Figure 19: Layout diagram of a fighter aircraft; Courtesy: `commons.wikimedia.org`

Or it could be a printed circuit board layout. Printed circuit boards are used to make electronic and computer systems.

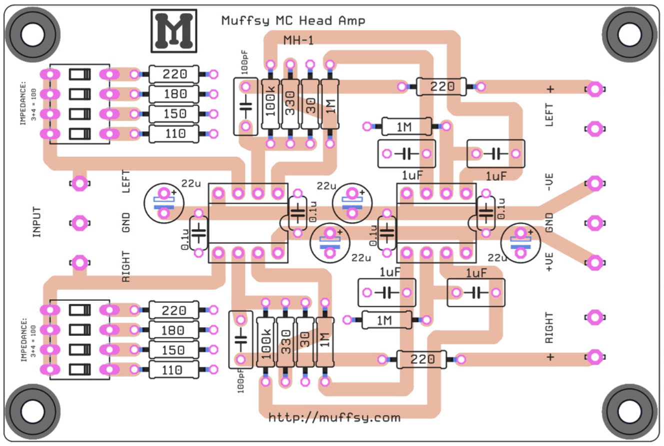


Figure 20: Layout diagram of a printed circuit board

How do Engineers Interpret Numbers?

Numbers play an important role is engineering communication. Dimensions in drawings, measured values from experiments, specifications of components etc are communicated through the use of numbers.

Engineers are frequently engaged in measurements. They may use sensors and/or measuring devices for these measurements. Such measurements are never perfect - there is always uncertainty associated with these measurements. This chapter deals with how engineers cope with such uncertainties.

Suppose you measured the length of a line using a ruler that has smallest division of 1 mm or 0.1 cm. If the reported measurement is 4.2 cm, it does not mean that the line is exactly 4.2 cm long. It could be bigger than 4.1 cm but smaller than 4.2 cm, or bigger than 4.2 cm but smaller than 4.3 cm. That means the reported measurement has uncertainty of 0.1 cm.

If you measure the length of the same line using a ruler that has much finer scale graduations, and suppose that you have perfect eyesight, the uncertainty in the measured value will be smaller. You may be able to report the length to be 4.20 cm, implying an uncertainty of 0.01 cm.

When we report measurement results, the complete statement of a measured value should include an estimate of the level of confidence associated with the value. Properly reporting an experimental result along with its uncertainty allows other people to make judgments about the quality of the experiment, and it facilitates meaningful comparisons with other similar values or a theoretical prediction. Without an uncertainty estimate, it is impossible to answer the basic question: "*Does my result agree with a theoretical prediction or results from other experiments?*"

Before explaining the standard methods of reporting and interpreting measured values, let's review a few terminologies that are widely used in this context.

Precision and Accuracy

Accuracy measures how close a measured value is to the true value or accepted value. Since a true or accepted value for a physical quantity may be unknown, it is sometimes not possible to determine the accuracy of a measurement.

Precision measures how closely repeated measurements agree with each other. Precision is sometimes referred to as *repeatability* or *reproducability* A measurement which is highly reproducible tends to give values which are very close to each other.

The most commonly used analogy to describe these two terms is the shooting of a target with arrows. The smallest circle located at the center of the target is called the bull's eye and represents the true value of the measurement. Shooting an arrow at the target is analogous to making a measurement.

If you make repeated attempts to hit the target (repeated measurement), you will end up with one of the four scenarios shown in Figure 21.

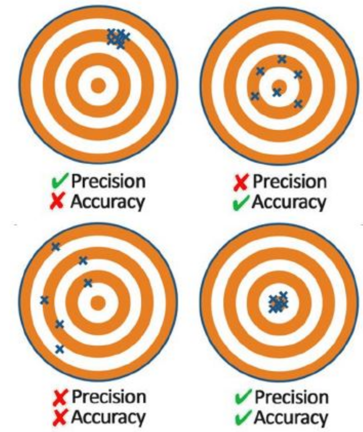


Figure 21: Accuracy and precision

1. **Low precision and low accuracy:** Repeated attempts land scattered all over the board (bottom left image).

2. **Low precision but high accuracy:** Repeated attempts land near the bull's eye but they are scattered (top right image).

3. **High precision but low accuracy:** Repeated attempts land close together but away from the bull's eye (top left image).

4. **High precision and high accuracy:** Repeated attempts land close together and on the bull's eye (bottom right image).

Types of measurement errors

When engineers and scientists refer to experimental errors, they do not refer to what are commonly called mistakes, blunders, or miscalculations. Examples of these mistakes are measuring width when length should have been measured, misreading the scale of an instrument, forgetting to divide the diameter by 2 before calculating the area of a circle using the formula $A = \pi r^2$, etc.

Experimental errors are inherent in the measurement process. There are two types of experimental errors: **random errors** and **systematic errors**.

Random errors are statistical fluctuations (in either direction) in the measured data due to the precision limitations of the measurement device. Random errors can be evaluated through statistical analysis and can be reduced by averaging over a large number of observations. Random errors affect the precision of measurement.

Systematic errors are reproducible inaccuracies that are consistently in the same direction. Some flaw in the measuring instrument (for example, the end of a ruler is worn down) will introduce systematic error in the measurement. In such case, reading is affected consistently and the error cannot be reduced by increasing the number of repeated measurements. These errors are difficult to detect and cannot be analyzed statistically. Systematic error affect the accuracy of measurement.

Calibration

When an instrument records inaccurate readings, it has to be calibrated to improve its accuracy. During calibration, a quantity whose value is quite accurately known is measured with the instrument. The measured value is compared with the known value and the difference between the two is used to adjust the measuring instrument.

Most of us are familiar with bathroom weighing scales. We have often zero-adjusted the weighing scale by turning a knob (or an adjusting button on digital scales) to zero. Alternatively, we can also calibrate the weighing scale with someone standing on it. This person's weight must be known accurately (perhaps by weighing this person on a different weighing scale that is accurate).

Reporting measurement result

When an engineer or a scientist reports the results of measurement, the report must describe the accuracy and precision of the measurements.

Accuracy refers to the closeness of the measured value to the true value. You may never know the true value exactly. For example, when you measure the resistance of a conductor how do you know its true value? The value you get from the measurement can be erroneous having both random error and systematic error.

Significant Figures

The uncertainty of a single measurement is limited by the precision and accuracy of the measuring instrument, along with any other factors that might affect the ability of the experimenter to make the measurement.

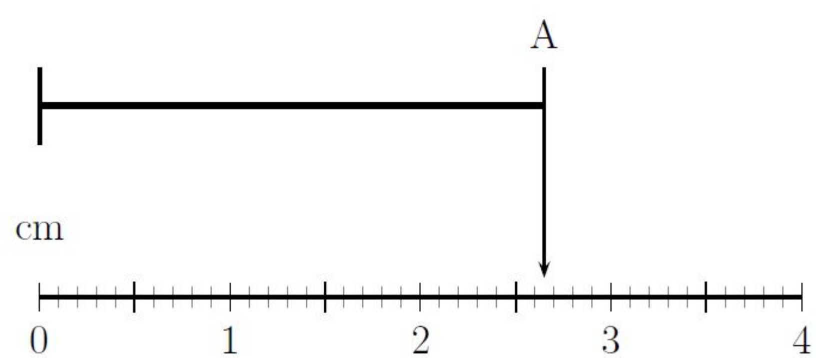


Figure 22: Uncertainty of a single measurement is limited by the precision of the measuring instrument

Suppose you are measuring the length of line with a ruler that has the smallest division of 1 mm or 0.1 cm. Then you cannot measure the length any more precise than to the nearest millimeter. However, you can estimate one past the smallest marking to the next decimal place in the measurement.

In Figure 22, the length of the line is between 2.6 cm and 2.7 cm. The length of the line can be estimated to be half of the distance between 2.6 cm and 2.7 cm, and the result can be reported with three significant figures as 2.65 cm.

The number of significant figures in any measurement is the number of digits that are known with some degree of reliability.

The case of another length measurement is shown in Figure 23; this time the length is spot on the 3.5 cm mark. As in the previous example, the measurement can be reported with three significant figures by estimating the additional digit to be 0. So the reported measurement is 3.50 cm.

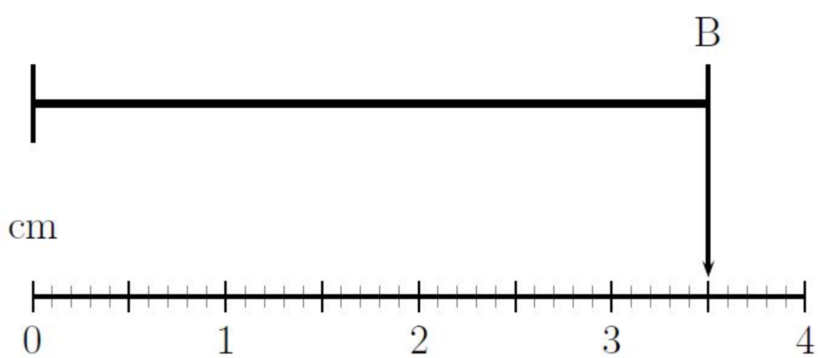


Figure 23: The measured value is reported using three significant figures to reflect the precision of measurement process.

Estimating uncertainty in repeated measurements

If a measurement is repeated several times, it is observed that the measured values are grouped around some central value. This grouping or distribution can be described with two numbers:

- the **mean** tells us the central value.
- the **standard deviation** describes the spread or deviation of the measured values about the mean.

If the measurement is repeated N times while measuring some quantity, then the mean value is

$$\bar{x} = \frac{1}{N}(x_1 + x_2 + x_3 + \cdots + x_N).$$

and the standard deviation is

$$\sigma_x = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_N - \bar{x})^2}{N - 1}}$$

The best estimate of a set of measurement is usually reported as the mean of the measurements, and the variation in the measurements is usually reported by the standard deviation.

The measurement result can then be reported as

$$x = \bar{x} \pm \sigma_x.$$

For measurements which have only random errors, the standard deviation means that 68% of the measured values are within σ_x from the mean, 95% are within $2\sigma_x$ from mean, and 99% are within $3\sigma_x$ from the mean.

Use of significant figures for simple propagation of uncertainty

Sometimes we determine a quantity by using two or more measured quantities. For example, the area of a rectangle is found by multiplying the length of two sides of the square. How does the uncertainty in the value of the area depend on the uncertainties of the measured values?

This can be analyzed using calculus with the help of partial derivatives of $f(x, y)$ with respect to x and y, where x and y are the measured quantities. However, by following a few simple rules, the significant figures can be used to find the appropriate precision for a calculated result for the four most basic math functions, all without the use of complicated formulae for propagating uncertainties.

Rules for deciding the number of significant figures

1. All non zero digits are significant. For example
    - 1.234 m has four significant figures.
    - 12.3 m has three significant figures.
2. Zeros between non-zero digits are significant. For example,
    - 502 K has three significant figures.
    - 2.3 kg has two significant figures.
3. Leading zeros to the left of the first non-zero digit are **NOT** significant. These zeros indicate the position of the decimal point. For example,
    - 0.0005 m has only one significant figures, even if it has four decimal places.
    - 0.015 m has two significant figures.
    - 0.1014 m has four significant figures.
4. Trailing zeros to the right of a decimal point are significant. For example,
    - 0.0540 kg has three significant figures.
    - 0.30 m has two significant figures.

5. If a number is expressed in scientific notation, for example

$$112,000 = 1.12 \times 10^5,$$

   where 1.12 is called the **mantissa** and 5 referred to as the **exponent**, then

   - the number of significant figures is identified only by the mantissa. Therefore, 1.12×10^5 has three significant figures.
   - if the same number is written as 1.120×10^5, then the number has four significant figures.

Rules for deciding the number of significant figures in a mathematical operation

1. For addition and subtraction

   - When one number is added to or subtracted from another number, the precision of the result will be the precision of the number with the least precision. For example, consider the sum

$$210\ m + 6\ m = 216\ m.$$

     The first number, 210, has three significant figures implying that we can measure it up to the nearest 10 m. The second number, 6, is one significant figure and we can measure it to the nearest m. Since the precision of the nearest 10 m is less than to the nearest m, the result should be shown to the nearest 10 m. Therefore, 216 m should be rounded to the nearest 10 m and written as 220 m.
   - for adding or subtracting two or more numbers with decimal places, the final answer will have decimal places that are the least of the decimal places of the input numbers. For example,

$$5.27\ kg + 2.1\ kg = 7.37\ kg$$

     should be reported as 7.4 kg. Since 2.1 kg has only one decimal place, the result should be rounded to the nearest number with one decimal place.

2. For multiplication and division

   - When one number is multiplied to (or divided by) another number, the number of significant figures in the result will follow the number with the least significant figures. For example,

$$1.35\ m \times 2.62\ m = 3.537\ m^2$$

     should be rounded off to 3.54 m^2 as both 1.35 m and 2.62 m have three significant figures.
   - Another example:

$$3.27 \times 1.1 = 5.597$$

     should be rounded off to 5.6.

Example

How many tiles will be required to fit a rectangle of sides 10.5 m and 1.3 m, if the area of a tile is 1.01 m^2?
Answer:
First we calculate the area of the rectangle -

$$10.5\ m \times 1.3\ m = 13.65\ m^2.$$

We do not round off the result to the most significant figures at intermediate stages of calculation. Doing so will introduce errors. We will round it off only in the final stage.

Number of tiles required -

$$N = \frac{13.65}{1.01} = 13.5148514851\ldots.$$

Numbers used in the calculations are -

- 10.5 : three significant figures,
- 1.3 : two significant figures, and
- 1.01 : three significant figures.

So the result should be two significant figures. Rounding it off we get **14** tiles are required.

Resolution : the smallest measurement an instrument can detect or measure.

Precision : indicator of the consistency of measurements (spread of mea-sured value around the average measured value)

Accuracy : indicator of indicator of a measurement from its true value

Guestimation - An Art of Approximation

Much of the material in this chapter are extracted from the book *Guesstimation* by L. Wienstein and J. Adam, Princeton University Press, 2008.

Guesstimation as one will notice immediately is made up of two root words: guess and estimation. Wikipedia suggests that a guesstimate may be a first rough approximation pending a more accurate estimate. Proponents of guesstimation do not claim accuracy, rather it provides a quick estimate that is **an order of magnitude** from the answer.

In engineering, an order of magnitude, is referred to within a factor of 10. For example, estimating 100 for a value of 1000 is within a factor of 10 and therefore also within an order of magnitude.

In engineering, more often than not, an estimate that is an order of magnitude away from the actual value is adequately reasonable. On many occasions, engineers need to make approximations or "guesstimate" for a particular quantity. For example, what is the distance from one point to another? What is the approximate density of a given item? How large a current will there be in a circuit?

In this chapter, we would like to give you an introduction to the skill of guesstimation. In the course of this module, we will use guesstimation often to arrive at ballpark values.

The first step in guesstimation is to make plausible guesses for the value. We shall label them as lower and upper bounds.

- The lower bound is the minimum guess of the value, beyond which the guess will seem unreasonable.

- Likewise, the upper bound is the maximum guess of the value, and similarly, if one guesses beyond this value, the guess may seem absurd.

The important point to note here is that we are making an estimate only. Guesstimation is about making informed guesses from very general information. You will be quite surprised at how close the guesstimate is from the actual value.

As you develop problem-solving skills, you will also develop skills at approximating. You will develop these skills through thinking more quantitatively, and by being willing to take risks. As with any endeavor, experience helps, as well as familiarity with units. These approximations allow us to rule out certain scenarios or unrealistic numbers. Approximations also allow us to challenge others and guide us in our approaches to our scientific world.

Few simple techniques

Before we look at some examples, it will be useful to appreciate a few simple techniques:

1. **Arithmetic mean or Geometric mean?** Once we have established the lower and upper bounds of an estimate, we will use the mean of these values for the estimate. If we use the arithmetic mean for this purpose, it will provide an answer that is not the same factor above and below the bounds. For example, suppose our lower bound is 1 and our upper bound is 100. The arithmetic mean is
$$\frac{1+100}{2} \approx 50.$$
The arithmetic mean is 50 times larger than the lower bound but only two times smaller than the upper bound. This is not a good practice. The alternative is to use the geometric mean, *i.e.*, the square root of the product of the lower limit and the upper limit. For the same example as above, the geometric mean will be
$$\sqrt{1 \times 100} = 10,$$
which is 10 times larger than the lower bound and also 10 times smaller than the upper bound. It lies in the sweet spot in terms of factors of the bounds.

Formula for geometric mean: (x1 * x2 *... xn)^(1/n), where n = number of periods and x1/x2/xn is a specific period.

2. To facilitate the calculation of the geometric mean, it is advisable to express the bounds in either scientific or engineering notations. In this tutorial we will use scientific notation exclusively, but the discussion will also apply to engineering notations.

3. **Estimating geometric mean:** Suppose the lower bound is $x_l \times 10^{y_l}$ and the upper bound is $x_u \times 10^{y_u}$; x_l and x_u are referred to as the coefficients in the scientific notation, while y_l and y_u are referred to as the exponents. The geometric mean is calculated by finding the arithmetic means of the coefficients and exponents. For example, if the lower bound is 1×10^0 and the upper bound is 1×10^2 then the geometric mean is calculated as
$$\frac{1+1}{2} \times 10^{\frac{0+2}{2}} = 1 \times 10^1.$$
This will always work for as long as the sum of the exponents is even. If the sum of the exponents are odd, then we subtract 1 from the sum of exponents by 1 and multiply the result by 3. For example, if the lower and upper bounds are 1×10^0 and 1×10^3, respectively then the geometric mean is estimated as
$$3 \times \frac{1+1}{2} \times 10^{\frac{0+3-1}{2}} = 3 \times 10^1.$$
Note that the square root of 1000 is ≈ 31.62 which is close to our approximation of 30. Remember, we are not trying to be accurate - we are trying to arrive at a reasonably approximate value.

4. **The advantage of using scientific or engineering notations:** It helps in reducing error when using very large and very small numbers. For example, consider the number 10000000. Imagine using this number for mathematical operation, the chance of counting the number of zeros is quite large. But the scientific notation 1×10^7 is nifty and requires us to count the number of zeros only once for the conversion. The same argument could be applied to a very small number like 0.0000000001. Counting the number of decimal places after each mathematical operation is very likely to lead to errors.

5. It is important to note that, for numbers expressed in the scientific notation as $x \times 10^y$, the exponent, y, is more important than the coefficient x. The error in x is not as significant as the error in y. For example, for the number 2×10^8, if 2 is changed to 3, this will lead to a change of the result by a factor of 1.5 times, but if 8 were changed to 9, the result will be change by a factor of 10.

6. To multiply two numbers using the scientific notation, we simply multiply the coefficients and add the exponents. For example,

$$(2.5 \times 10^5) \times (4 \times 10^6) = (2.5 \times 4) \times 10^{5+6} = 10 \times 10^{11} = 1 \times 10^{12}.$$

   Likewise, for division, we divide the coefficients and subtract the exponents:

$$\frac{4 \times 10^5}{2 \times 10^6} = 2 \times 10^{-1}.$$

7. Addition and subtraction in scientific notation can only be performed if the exponents are of the same value. Therefore, we express both numbers in a way that their exponents are of the same value. For example,

$$3 \times 10^7 + 4 \times 10^8 = 0.3 \times 10^8 + 4 \times 10^8 = 4.3 \times 10^8.$$

8. Since we are estimating values to a factor of 10, we only need to keep the coefficient to a single digit. There are two reasons for this practice. First, since we are estimating values, these are not accurately known and carrying precision in the numerals is not necessary. Second, when dealing with single digits, it is easier to perform mathematical operations, such as multiplication, division, addition or subtraction.

With preliminary considerations discussed above, we can now discuss examples of guesstimation. In this tutorial we will consider two examples.

Example 1

What is the surface area of a typical bath towel? Include the surface area of the fibers of the towel in your estimation. We will solve this problem by following the steps given below.

1. Estimate the dimensions of the towel.
2. Calculate the physical area of the towel. Remember to include the area of both sides of the towel.
3. Estimate the number of fibers per square cm of the towel.
4. Estimate the dimensions of a typical fiber.
5. Establish a reasonable model for the fiber and calculate the surface area of a typical fiber.
6. Multiply surface area of a fiber by the number of fibers per square cm and then multiply this value by the area of the towel in square cm.

Let's assume that a typical bath towel is 1×2 m in dimension. Therefore, the surface area of the towel without considering the fibers of the towel is

$$(1\ m \times 2\ m)\ per\ side \times 2\ sides = 4\ m^2.$$

How many fibers are there in a square cm? The accurate answer is not possible unless we manually mark out squares of 1 cm^2 and count the number of fibers in these squares and establish an average value. We are not going to do that. We will perform a guesstimation instead. We can safely say that the low limit of the number of fibers in a square cm is 10 and the upper limit is 1000. The geometrical mean is 100. So the guestimate of the number of fibers per square meter is

$$fibers\ per\ sq\ m \approx \frac{100}{(1 \times 10^{-2}\ m) \times (1 \times 10^{-2}\ m)}.$$

Let's guesstimate the surface area of each fiber. A towel at my home had fibers that were about 5 mm long and 1 mm wide. We may assume that a fiber can be modeled as a rectangular box of sides 1 mm each and height of 5 mm. We may ignore the surface area of the top of the box as being small compared to the surface area of the sides of the box. The bottom of the box is flush with the towel and should not be included in our calculations. Since the box has 4 sides, we estimate the surface area of a fiber as

$$4 \times 1\ mm \times 5\ mm = 20\ mm^2,$$

which is equivalent to $20 \times 10^{-6}\ m^2$ or $2 \times 10^{-5}\ m^2$.

Therefore, the surface area of the towel is

$$\begin{aligned} A &= (area\ of\ towel) \times (fibers\ per\ area) \times (surface\ area\ per\ fiber) \\ &= (4\ m^2) \times (100 \times 10^{-4}) \frac{fibers}{m^2} \times (2 \times 10^{-5}\ m^2) \\ &= \frac{800 \times 10^{-5}}{1 \times 10^{-4}}\ m^2 \\ &= 80\ m^2. \end{aligned}$$

Example 2

How much energy is needed to get a spaceship from Earth to Alpha Centauri, which is about 4 light years away. The trip must be made before the passengers die of old age. We are also required to estimate the fuel that is required. Assume 1 ton of rocket fuel provides 4×10^9 joules.

We shall solve this problem using the following hints.

1. How large should the spaceship be?

2. The mass of an aircraft carrier is 1×10^5 tons.

3. The speed of light is $c = 3 \times 10^8\ m/s$. A light year is the distance that light travels in a year. A year is approximately $3 \times 10^7\ s$.

4. What is the bound on the travel time before the passengers die of old age?

5. What will be the required speed of the spaceship?

6. What is the kinetic energy of the spaceship at that velocity?

7. Estimate the fuel required.

The kinetic energy of the spaceship will be provided by the rocket fuel. To calculate the kinetic energy, we need the mass of the spaceship and its velocity. To calculate the speed, we need the distance to be travelled and the duration - the round trip should be such that the passengers do not die of old age.

Let us assume that 40 years will be a reasonable estimate for the round trip considering the lifespan of a typical human.

If light takes 4 years to travel from earth to Alpha Centauri, the return trip will take 8 years if we travel at the speed of life. We will need to travel at $\frac{1}{5}th$ the speed of light if we want to reach Alpha Centauri in 20 years. The desired speed of the spaceship is

$$\begin{aligned} v &= \frac{1}{5} \times 3 \times 10^8 \\ &= 6 \times 10^7\ m/s. \end{aligned}$$

The spaceship should be of reasonable size to carry the supplies necessary for a 40 year round-trip. Let us assume that the spaceship will be the size of 10% of an aircraft carrier. Note that, the larger the mass, the larger is the kinetic energy and therefore the fuel required. Let's start with the assumption that mass of the spaceship is 10% of the mass of an aircraft carrier, that is,

$$m = 0.1 \times 1 \times 10^5\ tons = 1 \times 10^7\ kg.$$

1 ton = 1000 kg.

Table 2: Length order of magnitude

Length in meters (m)	Object
10^{11}	Earth-to-sun distance (1.5×10^8 km)
10^7	Earth's diameter (1.3×10^4 km)
10^6	Distance from New Orleans to Detroit (1600 km)
10^5	Length of Lake Michigan
10^4	Height of Mt Everest
10^3 (1 km)	George Washington Bridge
10^2	Length of a football field
10^1	Length of a Tennis court
10^0	A tall man's stride
10^{-1} (10 cm)	Width of a person's hand
10^{-2} (1 cm)	One side of a sugar cube
10^{-3} (1 mm)	Thickness of a coin
10^{-4}	Thickness of human hair
10^{-5}	Diameter of a human cell
10^{-6} (1 μm or 1 micron)	Thickness of a soap bubble film
10^{-9} (1 nm)	Small molecule
10^{-10}	Atom

The kinetic energy of the spaceship is

$$\begin{aligned} KE &= \frac{1}{2} \times (1 \times 10^7\ kg) \times (6 \times 10^7\ m/s)^2 \\ &= 18 \times 10^{21}\ J. \end{aligned}$$

Now we can estimate the fuel required

$$\begin{aligned} FR &= \frac{KE}{Energy\ per\ ton\ of\ rocket\ fuel}\ tons \\ &= \frac{18 \times 10^{21}\ J}{4 \times 10^9}\ tons \\ &\approx 5 \times 10^{12}\ tons. \end{aligned}$$

The fuel is more than the mass of the spaceship by an order of 8. With the power provided by the rocket fuel, we will not be able to reach our destination in the required time. Note that in our estimation we have not accounted for the losses in the rocket engines and the energy required for accelerating and decelerating the spaceship. Space travel will require a big leap in development before travel to the stars is a reality.

Finally, we end this tutorial with four tables that offer you a glimpse about the variation of quantities in terms of magnitudes. It provides you with a framework to make references to various sizes and dimensions. These tables are summarized in Tables 2 to 5.

Area in square meters (m^2)	Typical Object
10^{14}	Land area of the earth
10^{12}	Area of Egypt
10^{11}	Area of New York State
10^{9}	Area of Los Angeles
10^{8}	Area of Manhattan
10^{6} (1 km^2)	City of London
10^{4}	Area of a football field
10^{2}	Area of a volleyball court
10^{0}	Top of a small office desk
10^{-4} (1 1 cm^2)	One face of a sugar cube
10^{-6} (1 1 mm^2)	Head of a pin
10^{-8}	Pixel on a computer display

Table 3: Area order of magnitude

Density in ($kg\ per\ m^3$)	Item
10^{18}	Neutron star;
	atomic nucleus
10^{9}	White dwarf star
10^{4}	Lead; iron
10^{3}	Water; human body
10^{0}	Earth's atmosphere at sea level

Table 4: Density order of magnitude

Mass in kilogram (kg)	Object
10^{30}	The sun
10^{27}	Jupiter
10^{25}	The earth
10^{15}	World coal reserves (estimated)
10^{12}	World oil production in 2001
10^{11}	Total mass of human world population
10^{10}	Great Pyramid of Giza
10^{9}	Matter converted into energy by the Sun each second
10^{8}	An aircraft carrier
10^{7}	RMS Titanic
10^{6}	Launch mass of the space shuttle
10^{5}	The blue whale (the largest animal)
10^{4}	Large elephant
10^{3}	Automobile (small)
10^{2}	Lion
10^{1}	A kitchen oven
10^{0}	One litre of water
10^{-1}	An average apple; human kidney
10^{-2}	Large coin
10^{-3} (1 gm)	A sugar cube
10^{-6} (1 mg)	Mosquito
10^{-9} (1 μg)	Sand grain (small)
10^{-12}	Human cell
10^{-27}	Neutron; proton; hydrogen atom
10^{-30}	Electron

Table 5: Mass order of magnitude

How do Systems Work? - Part I

Mechanical Systems

Engineering systems are highly complex and multi-disciplinary in nature. Analyzing such a system is a daunting task and the only way forward is to decompose the system into subsystems. Each one of these subsystems interacts with one or more subsystems, and the interactions occur at interfaces. If the subsystems and the interfaces are carefully defined and specified, then the design and analysis of the subsystems can be independently pursued without conflicts.

You have learnt about functional block diagram - a method used for decomposing a complex system into subsystems according to their functions.

In this chapter, you will learn about engineering principle related to mechanical subsystems, specifically, systems involving force. There are many other engineering principles, *e.g.*, in the areas of thermodynamics, fluid mechanics, fluid dynamics, and aerodynamics, to explain different aspects of a mechanical subsystem. But those topics are out of the scope of this module. Only force-related fundamental principles are covered in this chapter.

Force and Force-related Principles

In this section, we first review some concepts that you have learnt in high school physics before illustrating them in the context of engineering applications.

Different Types of Force

1. **Push or Pull:** When you push or pull an object, the interaction between your hand and that object is a force,
2. **Gravitational Force**,
3. **Magnetic Force**,
4. **Friction Force**, etc.

All forces, whether they represent the interaction of two bodies in direct contact or the interaction of two bodies at a distance (gravitational force), are defined by their magnitudes, their directions, and the points of application.

Force is a vector quantity.

Units of Force

The SI unit of force is **Newton** (N). One Newton force is equal to the force that would accelerate a 1 kg mass at 1 $\frac{m}{s^2}$.

If you keep an average apple (approximately 100 gm) on a table, the gravitational force exerted by apple on the table is approximately equal to 1 N.

Newton's Laws

Newton's laws form the foundation of mechanics and analysis & design of many engineering problems including structures, airframes (fuselage and wings), car frames, medical implants for hips and other joint replacements, machine parts, and orbit of satellites.

- **First Law:** *A body at rest remains at rest, or, if in motion, remains in motion at a constant velocity unless acted on by a net external force.*

- **Second Law:** *The acceleration of system is directly proportional to and in the same direction as the net external force acting on it, and inversely proportional to its mass.* If $\bar{F}_{net}$ is the force acting on a system of mass m, then its acceleration is
$$\bar{a} = \frac{\bar{F}_{net}}{m}.$$

Only external forces affect the motion of a system. You must define the boundaries of the system before you can determine which forces are external.

- **Third Law:** *Whenever one body exerts a force on a second body, the first body experiences a force that is equal in magnitude and opposite in direction to the force that it exerts.*

Normal Force

Force of gravity (also called the weight) is a pervasive force that acts at all times.

When you hold a bag, it remains stationary because the net force acting on it is zero. The counteracting force comes from your hand. The bag is also stationary when it is placed on a table (unless the table is too fragile) implying that the net force acting on the bag is zero. The counteracting force in this case is produced by the table. The question is: how an inanimate object (unlike human hands) support the bag?

In this case, the table actually sags slightly (not noticeable by eyes) under the weight of the bag. All objects deform when a force is applied to them. Unless the object is deformed beyond its limit, deformation exerts a restoring force much like a trampoline or a compressed spring. The greater the deformation, the larger is the restoring force.

We conclude that whatever supports an object, be it animate or not, must provide an equal and opposite force to keep the object stationary. If the force is perpendicular to the surface of contact between the object and its support, the force is called **normal force**. The normal force can be less than the object's weight if the object is on an incline.

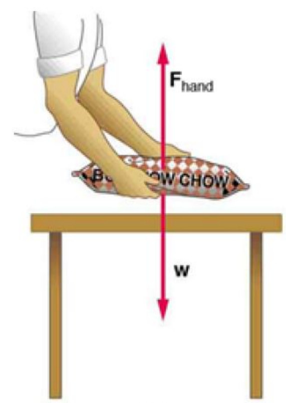


Figure 24: Suppose you are holding a bag of food with your hands, the force of gravity acting downward is counter-acted by the upward force applied by your hand making the net force acting on the bag equal to zero. And the bag remains stationary.

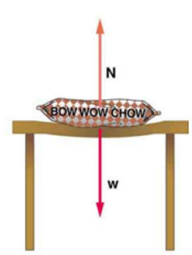


Figure 25: The table sags until the restoring force becomes as large as the weight of the bag. At this point the net external force on the bag is zero.

Friction Force

There are two types of frictional forces that are important in engineering design: dry frictional forces and viscous friction (or the fluid friction). Friction force allows us to walk or to drive our cars.

Dry friction exists because of irregularities between surfaces in contact. Imagine an experiment described below.

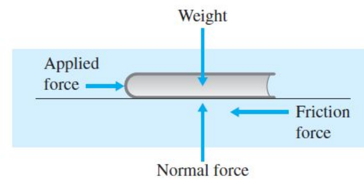


Figure 26: When friction force is equal to the push, the net external force is zero.

- Place a book on a table and start pushing the book gently (Figure 26).
- If the push (a force) is very small, the book will not move though a force is applied. The push is balanced by the friction force generated at the contact surface. The book will remain stationary for a range of magnitude of the push. That means, the static friction force is not constant but increases with increasing magnitude of the applied force (Figure 27).
- If the push is increased beyond a limit, then the object starts to move. It implies that friction force becomes smaller than the applied force creating a non-zero resultant force in the direction of the applied force. When the object is in motion, the friction is called *dynamic* or *kinetic* friction.

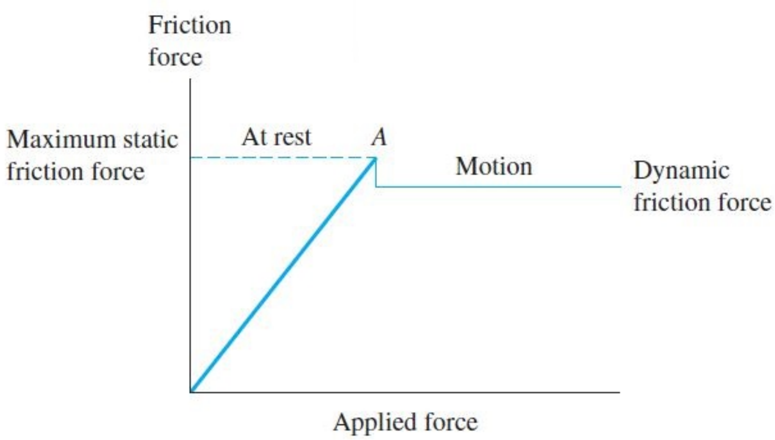


Figure 27: Friction force is equal to the applied force when the object is at rest. The point A represents the maximum static friction force.

We know from our experience that larger the weight of the object, larger is the force required to make it move. Larger weight means larger normal force. We can say that the static friction force is proportional to the normal force -

$$F_{static-friction} = \mu_S N,$$

where, μ_S is called the **coefficient of static friction** between the two surfaces involved.

The point A in Figure 27 can be identified experimentally by finding the condition at which an object starts to slide.

Another form of friction which must be accounted for in engineering analysis is **fluid friction** or viscous friction, which is quantified by the property of a fluid called viscosity. The value of viscosity of a fluid represents a measure of how easily the given fluid can flow. The higher the viscosity value is, the more resistance the fluid offers to flow.

Tendencies of Force

The effects of forces acting on an object:

1. Translation motion
2. Rotational motion
3. Elongation
4. Shortening or squeezing
5. Bending
6. Twisting.

Sometimes deformation (elongation, bending etc.) is easily observed, *e.g.*, in a squeezed tennis ball or a stretched rubber band. But pieces of rock or metal have deformation under force which is invisible and sometimes hard to imagine.

When deformations are not of much consequence engineers usually wish them away. Study of the effect of force with deformation neglected is called **rigid body mechanics**.

Assumption of rigidity greatly simplifies many calculations while still generating adequate predictions for many practical problems. For understanding the steering dynamics of a car we may treat the car as a rigid object, whereas rigidity will be a poor approximation for crash analysis and the car must be treated as a highly deformable object. **You will learn about deformations when studying mechanical failure of a system.**

Moment, Couple, Torque - Force Acting at a Distance

When you open a door you apply a pulling or a pushing force on the door-knob (or handle). The application of this force will make the door rotate about its hinges. In mechanics, this tendency of force is measured in terms of a **moment** of a force about an axis or a point. Moment has both direction and magnitude.

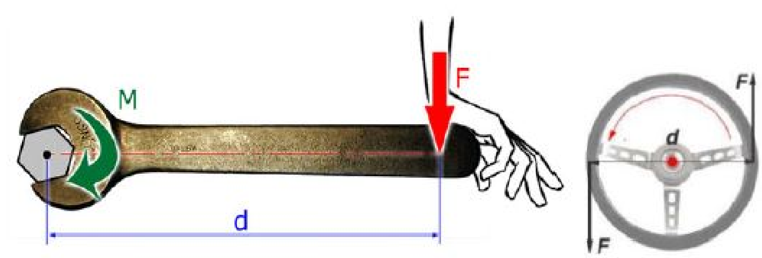


Figure 28: Left: the force F produces a moment in the clockwise direction. Right: Two parallel forces produces a couple with counterclockwise direction.

For the spanner or wrench shown in Figure 28, the magnitude of the clockwise moment is

$$M = F \times d.$$

The SI unit of moment is Newton-meter (Nm).

In case of the steering wheel example in Figure 28, two forces have same magnitude, opposite direction, and parallel lines of action. They too create rotational effect and are defined as **couple**. The net force due to a couple is zero ($F - F = 0$). If the perpendicular distance between the two forces is d then the moment of this couple is

$$M = F \times \frac{d}{2} + F \times \frac{d}{2} = F \times d.$$

Both forces create clockwise rotation and therefore, they are added.

Torque is the moment of a set of force vectors with resultant force equal to zero. So the moment of a couple is torque.

Equilibrium of Forces and Moments

If the net force acting on an object is zero and also the net moment is zero, then the system is said to be in **equilibrium**. According to Newton's laws, there are two possibilities:

- The object remains stationary. This is **static equilibrium**. Civil constructions, such as, bridges and buildings, work under static equilibrium.
- The object continues to move with constant velocity. This is **dynamic equilibrium**. A vehicle cruising at constant speed is under dynamic equilibrium.

When an object is not in equilibrium (of forces or moments):

- If the net force (F_{net}) is not zero but the net moment (M_{net}) is zero, the object will experience translational acceleration
$$\frac{dv}{dt} = \frac{\bar{F}_{net}}{m},$$
where v is the linear velocity of the object and m is its mass.
- If $F_{net} = 0$ but M_{net} is not zero, the object will experience angular acceleration
$$\frac{d\omega}{dt} = \frac{\bar{M}_{net}}{J},$$
where ω is the angular velocity (rad/s) and J is the moment of inertia.
- if both F_{net} and M_{net} are non-zero, the object will experience both translational and rotational acceleration.

Free Body Diagram

Free body diagram (FBD) is a widely used technique for analyzing the effects of forces and moments on a system.

FBD is the sketch of an object of interest with all the surrounding objects stripped away and all of the forces and moments acting on the object shown.

Drawing the FBD is an important step in solving force-related problems. It helps to visualize all the forces acting on a single object. The concept is explained using a few examples.

Example 1:

A simplified schematic of an elevator **accelerating** upward is shown in Figure 29. Forces acting on this system are:

1. The force of gravity (downward) acting on cabin with people,
2. The force of gravity (downward) acting on the counterweight,
3. The force acting upward on the cabin to accelerate it, and
4. The force acting downward to accelerate the counterweight.

The FBD is shown in Figure 30 separately for the cabin and the counterweight.

Some other forces, *e.g.*, friction at the contact between the cable and the pulley, and the force acting on the support are not shown in this simplified schematic though they exist in real world. Those forces are assumed here to be negligibly small for simplicity.

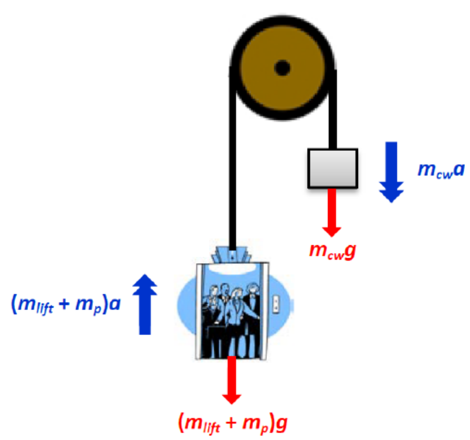


Figure 29: Forces acting on an elevator system

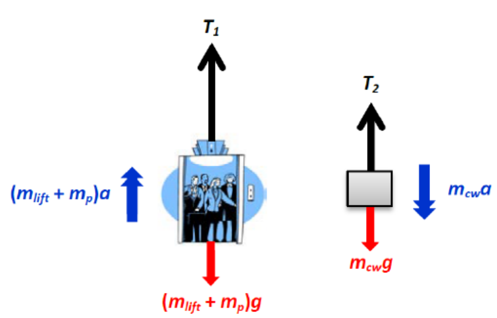


Figure 30: Free body diagrams for the elevator system. The forces T_1 and T_2 shown are internal forces known as **tension**.

Internal Force: Tension and Compression

When an object is subjected to an external force and the object is not free to translate in the direction of force, *internal forces* are created inside the material to hold the object together and to resist the change of shape. These forces can be

- either **tension** when the object is being pulled outward,
- or **compression** when the object is being compressed.

Example 2:

Consider a pole protruding out of a wall (Figure 31). One end of the pole is fixed to the wall while the other end is not restrained. Such a structure is known as *cantilever beam*. In this figure, a point force is applied at the free-end of the beam. Let's assume that the weight (force due to gravity) of the beam is negligibly small and hence is not shown. If the weight is not negligible, then that force would be acting through the center of mass of the beam.

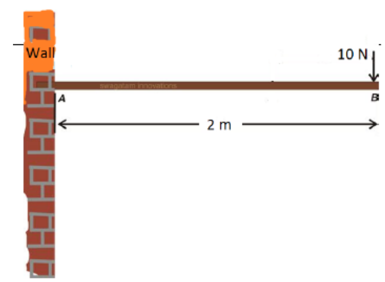


Figure 31: Cantilever beam

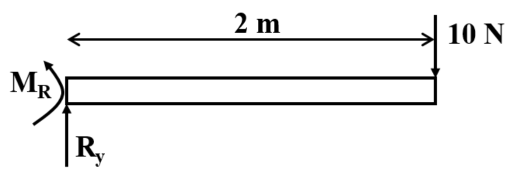
This is a stationary structure and, the conditions for static equilibrium

$$F_{net} = 0 \quad and \quad M_{net} = 0$$

must be satisfied. There must be a 10 N force acting upward to ensure equilibrium of forces. Where does that force come from? This is the **reaction force** (R_y) at the support of the cantilever.

Now though the equilibrium of forces is ensured, there will be a moment as the two forces do not act at the same point. To ensure equilibrium of moments, the support also gives a reaction moment (M_R).

FBD of the beam is shown below.



In order to ensure static equilibrium:

$$\begin{aligned} R_y &= 10\ N \\ M_R &= 20\ Nm \end{aligned}$$

Guidelines for determining reaction forces at the support

- Can the object undergo translational motion at the support? If NO, there must be a reaction force at the support. For the case of the cantilever beam, the support creates reaction forces in all three directions. However, R_x and R_z are not shown in the FBD as there is no applied force in those two directions. In other words, $R_x = 0$ and $R_z = 0$.
- Can the object be rotated at the support? If NO, then there must be a reaction moment at the support.

- Consider the case of a door. It can be rotated about the vertical axis; the reaction moment about the vertical axis at the door hinge is zero.

Forces acting on a system in motion

You have seen examples of equilibrium of forces for statinary systems. For a system in motion, if the net force acting on it is zero then the system continues to move at constant velocity. If the net force is non-zero, then the system accelerate in the direction of the net force and it is not in equilibrium.

Consider the forces acting on a quadcopter in motion.

1. Qaudcopters use four rotors. Spinning blades of rotor push air down creating a force in the opposite direction. The faster a rotor spins, the greater is the force. Depending on the pitch of the quadcopter, this force can be split into two components:
   - An upward force knows as **Lift**, and
   - A horizontal force that pushes the quadcopter forward during cruising.
2. **Force of gravity** acts downward through the center of mass.
3. **Drag** force acts on a quadcopter when it moves and the force acts in the direction opposite to the direction of motion.

The drag force is a nonlinear function of velocity (v) and can be approximated as

$$F_d = \frac{1}{2}\rho C_D A_{eff} v^2 \ \ Newton,$$

where, ρ (kg/m^3) is the density of air, C_D is the drag coefficient which is dimension-less, and A_{eff} is the effective frontal area of the quadcopter in motion.

How is the motion of a quadcopter controlled?

A pair of diagonally opposite rotors rotate clockwise while the other pair rotates counterclockwise as shown in Figure 32. While one pair of rotors produces positive angular momentum, the other pair produces negative angular momentum. If they are spinning at the same speed, overall angular momentum will be zero.

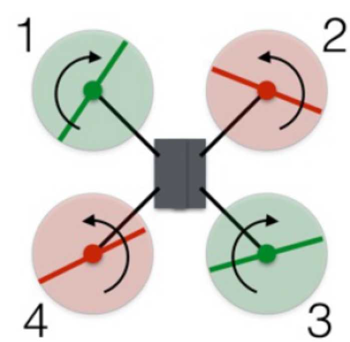


Figure 32: Two pairs of rotors of the quadcopter rotate in opposite direction. Source of the image: www.wired.com/2017/05/the-physics-ofdrone/

Vertical motion:

A quadcopter ascending from ground to a hover height goes through different phases of motion. For simplicity, we assume that it is ascending straight upward without any horizontal displacement. So, there should not be any horizontal component of the net force produced by the propellers. It requires all propellers producing equal thrust. And the net lift is equal to

$$F_L = \Sigma_{i=1}^{4} F_i, \quad F_1 = F_2 = F_3 = F_4.$$

Phase	Forces	$\frac{dv_v}{dt}$
Take-off	$F_L > mg, \ \ F_d = 0$	>0
Ascend under acceleration	$F_L > (mg + F_d), \ \ F_d \propto (v_v)^2$	>0
Ascend with constant v	$F_L = (mg + F_d), \ \ F_d \propto (v_v)^2$	>0
Ascend under deceleration	$F_L < (mg + F_d), \ \ F_d \propto (v_v)^2$	<0
Hover	$F_L = mg, \ \ F_d = 0$	0

Table 6: Different phases of vertical motion of a quadcopter ascending from ground to a hover height

In all phases of the vertical journey, the net force acting is

$$F_{net} = F_L - (mg + F_d).$$

The force due to gravity (mg) is constant, the lift (F_L) can be varied by varying the rotational speed of the propellers, and the drag force (F_d) varies with velocity.

The conditions on forces and the acceleration in different phases of this motion are shown in Table 6. The velocity in the vertical direction is represented by the symbol v_v in this table; it is positive in vertically upward direction.

Horizontal motion:

The horizontal component of the force is produced by tilting the quadcopter, which is achieved by spinning two pairs of propellers at two different angular speeds.

Let the propellers #1 and #2 in Figure 32 be on the front side of the quadcopter. Assume that these propellers are spun at N_f revolutions per minute (RPM) while the propellers #3 and #4 are spun at N_b RPM such that $N_f < N_b$.

Higher RPM means higher thrust. For the conditions given above,

$$(F_1 + F_2) < (F_3 + F_4).$$

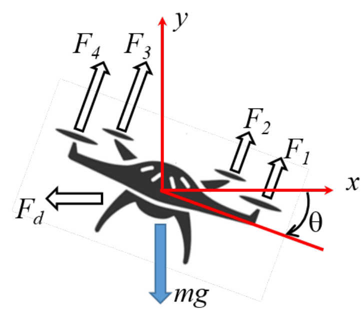


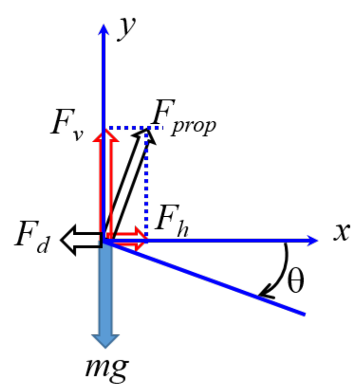
Figure 33: As $F_1 + F_2 < F_3 + F_4$, the resulting moment will cause the quadcopter to pitch forward.

As a result, the quadcopter will be pitched forward, i.e., tilted with front end at a lower height than the back end. This produces a horizontal force. The resultant force

$$F_{prop} = F_1 + F_2 + F_3 + F_4$$

is still perpendicular to the body of the quadcopter but is at an angle with the horizontal line.

To maintain the quadcopter at the crusing altitude, the vertical component of F_{prop} must be equal to the the force due to gravity and the horizontal component will push the quadcopter in the horizontal direction *i.e.*,



$$\begin{aligned} F_{prop} \cos\theta &= F_v = mg, \\ F_{prop} \sin\theta &= F_h. \end{aligned}$$

The equation of horizontal motion is

$$\begin{aligned} \frac{dx}{dt} &= v_h \\ \frac{dv_h}{dt} &= \frac{F_h - F_d}{m}, \end{aligned}$$

where the drag force $F_d \propto (v_h)^2$.

Yaw motion:

While hovering, the quadcopter can be turned or rotated about the z-axis (yaw motion) by changing the angular momentum of two pairs of rotors.

For example, if you decrease the spin of rotor 1 and rotor 3 in Figure 32 and increase the speed of rotor 2 and rotor 4 by the same amount, no moment is created to induce pitch. However, the sum of angular momentum will not be zero and, therefore, the quadcopter body must rotate. Total lift produced by four rotors can be maintained to the level required to counter the gravitational force and yet the body is turned.

There are many other engineering principles that are relevant in understanding how a system works, *for example,*

- Conservation principles (partially covered in the study of energy conversion), and
- Principles of flow (partially covered later in the context of electric current).

How do Systems Work? - Part II

Electrical System

Modern engineering systems are becoming more electric. The MRT in Singapore as well as many railways, rapid transit systems, trams, trolleys etc around the world use electrical traction system. Electric vehicles are gradually taking place in the road transportation systems.

Electrical power can be easily

- transmitted over long distance,
- converted into other useful form more efficiently, and
- controlled.

These are some of the reasons for designing systems to be powered by electricity. In addition, electricity can also be obtained from clean and renewable sources like solar, wind and hydro, and is thus in line with the global effort in becoming more "green".

Besides being driven by electrical power, engineering systems also use electrical technologies for sensing the environment, for communication, and for control and automation. Factory automation, modern manufacturing, internet of things etc. all are dependent on electrical engineering.

This chapter covers the fundamental principles associated with electrical systems.

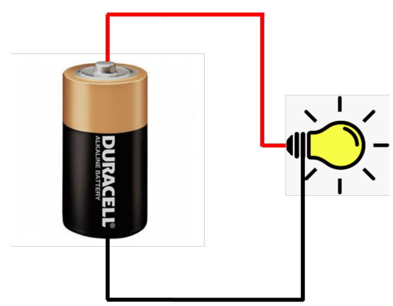


Figure 34: A simple electrical system

Lumped Circuit Model

Electrical engineering is the purposeful use of Maxwell's Equations describing electromagnetic phenomena. Consider the case of a light bulb. When it is connected by a pair of wires to a battery, it lights up. Someone, who is interested in fundamental laws of physics, can employ Maxwell's equations and carefully analyze the physical properties of the bulb, the battery, and the cables to understand what happens in this system.

But that is a complicated process. Adopting this approach for analyzing a practical system with many different components, is a daunting task. So electrical engineers have created an abstraction layer on top of Maxwell's equations called the lumped element model. By doing so, models of discrete components are defined first. Then a lumped circuit model is obtained by interconnecting different lumped element models.

Electrical engineers can create and exploit successive layers of abstraction (models of amplifier, filter, logic gate etc) to manage the complexity of useful electrical systems. This chapter will remain focused only on circuits with lumped elements.

Electric Charge

- Charge is at the core of all electrical systems.
- Electric charge is the physical property of matter that causes it to experience a force when placed in an electromagnetic field.
- There are two types of charge: positive or negative.
- The smallest amount of charge is the charge carried by an electron q_e which is equal to -1.602×10^{-19} C.
  - Electrons are the most common charge carriers.
- In materials known as **conductors**, large number of free electrons exist at room temperature which move randomly inside the conductor. But under the influence of an electric field, these free-to-move charges move in a particular direction giving rise to a flow of charge.

The unit of charge is Coulomb.

Electric Current

- Electric current is the rate of flow of electric charge similar to fluid flow. Charge carriers can be electrons in a conductor or ions in an electrolyte.
- Conventional direction of positive current is the direction of flow of positive charge.

The SI unit for current is Ampere (A), which is equivalent to Coulomb/sec (C/s). A current of 1 A means that one coulomb of charge is going past a given point per second.

Voltage or Potential Difference

- Voltage is a measure of the energy transferred per unit charge when the charge is moved from one point to another. The unit of voltage is **volt**, which is equivalent to
$$\frac{Joules}{Coulom} \longrightarrow \frac{J}{C}.$$
- Voltage is always measured between two points with labeled polarities, $+$ and $-$.
- Chemical reaction inside a battery creates a potential difference (voltage) between its terminals.

– When a conducting materials is connected externally between two terminals of a battery, current flows through the circuit.
– A battery with higher voltage has higher "pressure" to drive electricity through the circuit.

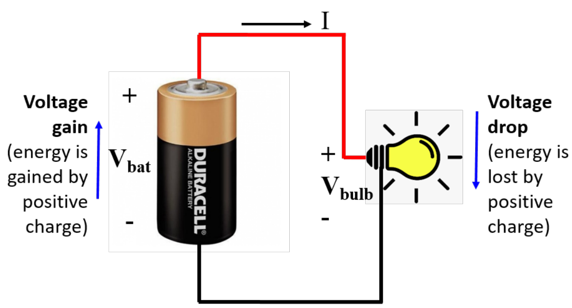


Figure 35: Energy is lost by a positive charge when it flows from higher voltage to lower voltage and energy is gained by positive charges flowing from lower voltage to higher voltage.

The charge flows through different components in a circuit. Energy is lost while it flows through some components but gained while flowing through some other components. Both types of components are required for any electrical system to work. In case of a light bulb connected to a battery (Figure 35), positive charges gain energy as they flow through the battery but energy is lost (converted into light energy) as they flow through the light bulb. The voltage polarity must conform to the gain and loss of energy by the flowing charges. The positive (+) polarity indicates higher potential or higher energy for the positive charge.

Electrical Power and Energy

- Voltage is a measure of energy transferred per unit charge (Joules/-Coulomb).
- Current is the rate of flow of electric charge (Coulomb/sec).
- Product of current and voltage is the rate of energy transfer
$$\frac{Joules}{Coulomb} \times \frac{Coulomb}{sec} = \frac{Joules}{sec}.$$
- Rate of energy transfer if power, P,
$$P = V \times I. \tag{1}$$

Consider the connections shown in Figure 36. Two different bulbs are connected to one single battery.

- $P_{bulb\ 1} = (1.5\ V) \times (20\ mA) = 30\ mW.$
- $P_{bulb\ 2} = (1.5\ V) \times (10\ mA) = 15\ mW.$

For the connections shown in Figure 37, you need to know the voltages across individual bulbs and the current in order to determine the power transferred to these bulbs.

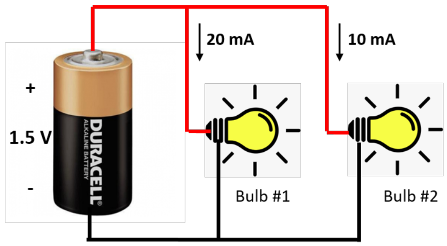


Figure 36: Two bulbs are connected in parallel. In parallel connection, same voltage exists across all individual components.

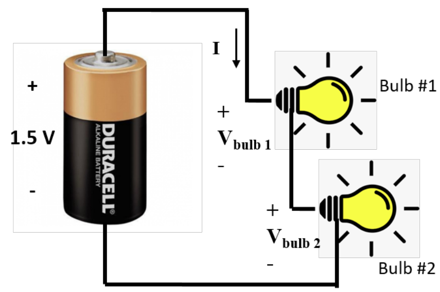


Figure 37: Two bulbs are connected in series. Same current flows through all components in a series connection.

Resistance and Ohm's Law

In the example of the bulb given above, we want to know how much current would flow through it. We do not care about many other physical parameters of the bulb, *e.g.*, shape, size and color. We are also not interested in other parameters such as density of the filament material. We can ignore all these internal and external properties of the bulb while finding the current and model it as a discrete component which conducts electricity.

We define the resistance of the bulb R to be the ratio of the voltage applied to the bulb and the resulting current through it. In other words,

$$R = \frac{V}{I}.$$

Relationship between the voltage and current for a bulb may be more complicated than this. For example, the effect of temperature is ignored.

All materials present some opposition to the flow of charges through them. The primary cause of this resistance is the collisions among electrons, which cause electrons to lose their energy. The loss of energy along the path leads to a voltage drop. This is true for the bulb as well as for the connecting wires, but the drop is negligibly small for the connecting wires made of good conductor like copper.

Ohm's Law states that the voltage drop across an ideal resistor is proportional to the current flowing through it. The constant of proportionality is formally called "resistance".

The unit of resistance is Ohm (Ω)

.

$$V \propto I,$$

$$\frac{V}{I} = R \quad or \quad V = IR. \tag{2}$$

It is important to note that Ohm's Law is an empirical relationship (i.e., observed through experiments). It is only an approximation, and does not hold at very high/low voltage and current values.

Ideal Voltage Source

In the example of the battery-and-bulb, the battery used to create a potential difference or voltage across the bulb is a source of voltage.

An **ideal voltage source** maintains a specified voltage across its terminals regardless of the magnitude of the current flowing through it. An ideal voltage source usually has a closed-loop control system which maintains the voltage output by some sort of internal compensation. In practice, electrical energy sources like batteries keep the voltage output within a tolerance of the nominal value for a range of output current. Battery is known as **practical voltage source**.

You will learn more about practical voltage source in the chapter titled *Where do Systems get Energy?*

.

Ideal Current Source

An ideal current source is also a source of electrical power that forces a specified magnitude of current to flow through it regardless of the voltage across its terminals. Again, such a source can be obtained by using a closed-loop control mechanism within the source.

The laboratory DC power supply can operate in **controlled voltage (CV)** and **controlled current(CC)** mode which are equivalent to ideal voltage source and ideal current source, respectively.

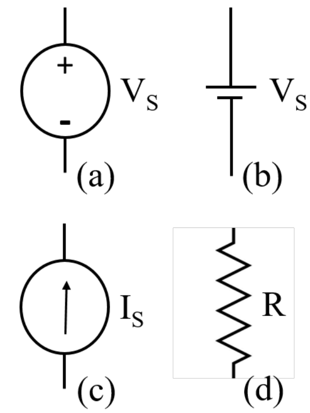


Figure 38: Symbols of DC voltage source (a)-(b), DC current source (c) and resistor (d) used to draw circuit schematic. Note that some books may use different symbols. V_S is the magnitude of the source voltage; polarity must be indicated for DC source. I_S is the magnitude of source current; the arrow indicates the direction of current. R is the resistance value.

Symbol for Two-terminal Lumped Circuit Elements

Resistors and sources are the primitive building blocks of electronic circuits. Electronic access to an element is made through its terminals. These components are called *two-terminal elements*. Some electronic components have more than two terminals, *e.g.*, a transistor comes with three terminals.

When we use circuit diagram for communicating to others, we do not draw the physical components. Instead, we use symbols. The symbols used for (1) ideal voltage source (2) ideal current source, and (3) fixed resistor are shown in Figure 38.

Can you draw using these symbols the circuit schematics of the systems shown in Figures 36 - 37? Assume the battery as an ideal voltage source.

I-V Characteristics for Two-terminal Elements

From the viewpoint of analysis, the most important characteristic of a two-terminal element is the relation between the voltage across and the current through its terminals, or the *V-I* or *I-V* relationship for short. This relation, called the *element law*, represents the lumped-parameter summary of the electrical behavior of the element.

For an ideal resistor, the relation between current and voltage is given by the Ohm's law

$$I = \frac{V}{R}.$$

This is a straight line of slope $\frac{1}{R}$ that passes through the origin ($V = 0, I = 0$) as shown in Figure 39 (a).
The element law of an ideal voltage source with magnitude V_S is

$$V = V_S,$$

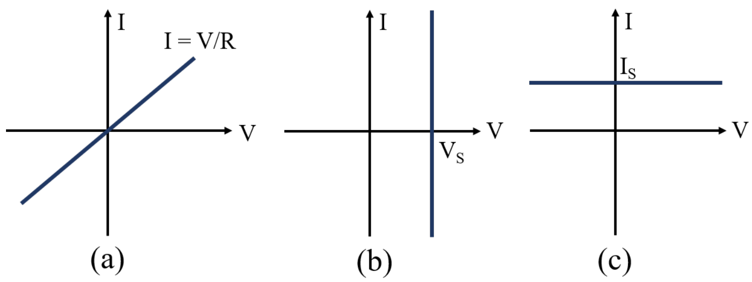


Figure 39: I-V plot for two terminal elements: (a) ideal resistor, (b) ideal voltage source, and (c) ideal current source.

regardless of the magnitude of current flowing through it. See Figure 39 (b) for the corresponding *I-V* plot. Similarly, for an ideal current source

$$I = I_S,$$

regardless of the magnitude of the voltage across the element (Figure 39 (c)).

Can you sketch the I-V plot for a piece of perfect conductor?

Losses: What happens when current flows through a resistor?

When a current flows through an element that has resistance, the collisions among electrons produce heat. The temperature of the element will rise if the heat is not dissipated fast enough. The rate of heat energy produced or the power can be calculated by $V \times I$.

If V is the voltage across a resistor (R) and the resulting current magnitude is I, then the power,

$$\begin{aligned} P_R &= V \times I \\ &= I^2 R. \end{aligned}$$

This relation tells us that, for an element with a given resistance R, the power loss increases with the current, and it is directly proportional to the square of the current!

Modeling Work and Losses using Resistance

- When a current (I) flows through resistance (R), heat is produced at a rate of I^2R.
- In some applications like in a heating coil to heat water or air, the heat is intentionally produced, hence is a useful output.
- However, most of the time the heat is a waste product and is considered as loss.
- Often, an equivalent resistance is connected to a power supply to represent the useful work or heat loss in a system. This is known as modelling and is used for analysis of the system as an 'electric circuit'.

Basic Concept of Circuits

In very simple terms, a circuit is a circular journey that begins and ends at the same place. It is also used in car racing - the F1 circuit.

In electrical systems, a circuit is the closed path through which current flows. An example of an electrical circuit is shown in Figure 40 that consists of an ideal voltage source, and two resistive elements.

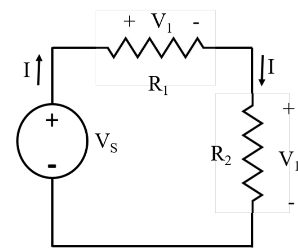


Figure 40: An electrical circuit

- Each element in a circuit must have a voltage V and a current I defined for its terminals.
  - The ratio of V to I is constant for ideal resistor that obey's Ohm's law. But it is not constant for other elements.
- The circuit must also have a voltage defined between any pair of points, and a current defined into any terminal.
- The elements must not interact with each other except through their terminal currents and voltages.
  - The internal physical phenomena that make an element function must interact with external electrical phenomena only at the electrical terminals of that element.

Measurement of Voltage and Current in a Circuit

The voltage across a circuit element can be measured by connecting a voltmeter (V) across it (parallel) as shown in Figure 41.

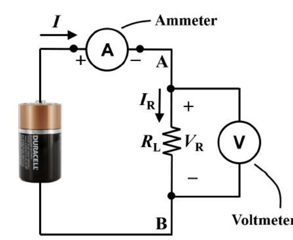


Figure 41: Measurement of voltage and current using voltmeter and ammeter, respectively.

An ideal voltmeter has infinite resistance making the current flowing through the meter to be zero:

$$I_{VM} = \frac{V_{AB}}{\infty} = 0.$$

So the voltage measured using an ideal voltmeter would be exactly the voltage between the points where the meter probes are connected. Nevertheless, all practical voltmeters have finite though very large resistance. The effect of voltmeter resistance on the measurement is called *instrument loading*.

Current is measured using ammeter. To measure current through a circuit element, the ammeter is connected in series with that element as shown in Figure 41. An ammeter also has two terminals. One of them is labeled as either the common or negative terminal (usually colored black), and the other terminal is the positive terminal (usually colored red). The measurement shown by the ammeter is for the current in the direction from its positive terminal (red) to its negative (black). If the current actually flows in this direction, then the reading will show positive value. Otherwise, it will show negative value.

- If you use a practical voltmeter with not very high resistance, will the measurement be smaller than or larger than the true value?
- Can you eliminate this measurement error by repeating the measurement?

An ideal ammeter has zero resistance making the voltage drop across the meter to be zero,

$$V_{AM} = I \times 0 = 0.$$

So the current reading with an ideal ammeter would be exactly the same current that would flow when the ammeter is not there. Nevertheless, all practical ammeters have finite though very small resistance causing *instrument loading*.

Circuit Laws

You have learnt about the Ohm's law that governs the characteristics of one of the circuit elements - an ideal resistor. You will learn about a few more circuit elements later in this module. Each one of them is governed by an element law, *i.e.*, a specific relation between the voltage and current associated with the element.

Besides these element laws of individual elements, there are two important laws for circuit:

1. Kirchhoff's Current Law (KCL) and
2. Kirchhoff's Voltage Law (KVL).

Fluid Flow Analogy of Circuit

Imagine of water flowing through a pipe that gets bifurcated. Since water is an in-compressible fluid, the net flow of water into a region of fixed volume (called the *control volume*) must be zero. This means the rate of water flowing into the control volume must be equal to the rate of water flowing out of that control volume.

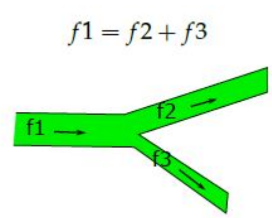


Figure 42: Fluid flow.

- The charge is analogous to fluid. Current is rate of flow of charge.
- A voltage source is like a pump that forces the fluid through the rest of the a network.
- The resistance can be thought of as a constriction in the pipes that causes turbulence of flow and produces heat.
- As the fluid is forced past each constriction, the pressure decreases just like the voltage drop caused by current passing through resistance.

What is steady-state?

Conservation of Charge: Kirchhoff's Current Law

In steady-sate, the net charge in a closed volume remains zero. Hence algebraic sum of currents into and out of the closed surface is equal to zero.

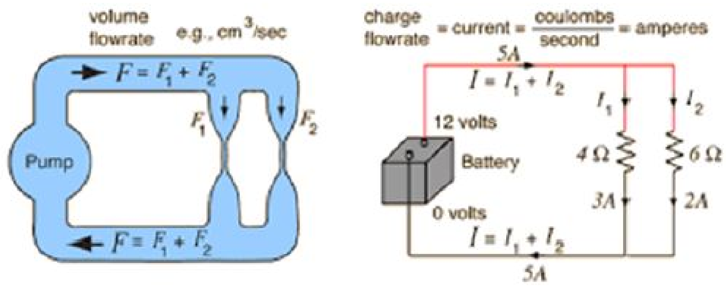


Figure 43: Fluid flow analogy of electrical circuit

The Kirchhoff's Current Law (KCL) is based on the conservation of charge. Recall that current is the rate of flow of charge,

$$i = \frac{dq}{dt}.$$

Since charge can be neither created nor destroyed, **the net flow of charge into or out of a control volume (node or super node) must be zero.**

What is a node? Node in a circuit is where multiple components are connected.

If we represent the current coming into the node as negative, and current flowing out of the node as positive then, for the section of circuit shown in Figure 44, we can write

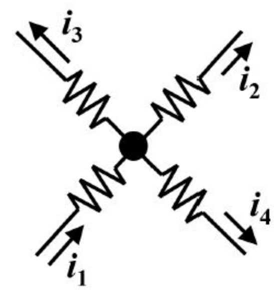


Figure 44: Kirchhoff's current law states that the sum of all currents entering a node is equal to the sum of all current leaving that node.

$$-i_1 + i_2 + i_3 + i_4 = 0.$$

Alternative statement for the Kirchhoff's Current Law (KCL): **"sum of all currents entering the node must be equal to the sum of all currents leaving the node"**.

$$i_2 + i_3 + i_4 = i_1.$$

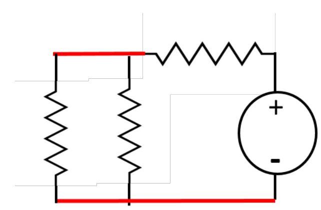


Figure 45: Each of the two sections drawn using red line represents one node.

In circuit schematics, a node may not necessarily be shown as a point. In the schematic of Figure 45, the top red line is one node, and the bottom red line is another node.

KCL is valid for any closed surface in the circuit which is also known as **super node** as shown in Figure 46. The use of super node in circuit analysis may not be obvious now. However, this is often used in Node voltage analysis method which you will learn later. Super node is equivalent to the control volume in fluid flow systems.

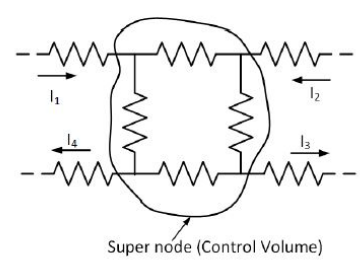


Figure 46: Kirchhoff's current law states that the sum of all currents entering a super node is equal to the sum of all currents leaving that super node.

Conservation of Energy: Kirchhoff's Voltage Law

Voltage is energy per unit charge. In steady state, the net change in energy of a unit charge around a closed path is zero. Hence algebraic sum of voltages around a closed path is equal to zero.

The Kirchhoff's Voltage Law (KVL) can be stated as **"around any closed loop, at any instant of time, the sum of voltage drops must equal to the sum of voltage rises"**.

In Figure 47, as charges move around the circuit, they receive energy from the voltage source and supplies energy to the resistances. By conservation law, the total energy consumed by the resistances is supplied by the source. As voltage is energy per unit charge, the voltage rise across the battery must be equal to the sum of voltage drops across the three resistances. Taking the voltage rise as positive and voltage drop as negative, the KVL equation for this circuit can be written as:

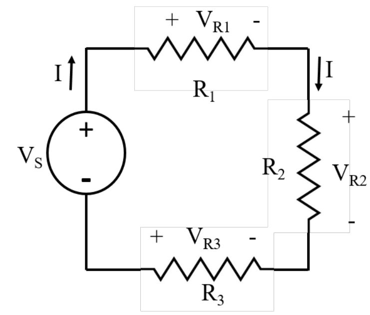


Figure 47: Kirchhoff's voltage law: sum of voltage rises around a loop is equal to the sum of voltage drops around that loop.

$$V_S - V_{R1} - V_{R2} - V_{R3} = 0.$$

All terms in a KVL equation are voltages. They may represent the magnitude of a voltage source (*e.g.*, V_S in this case), $I \times R$ in case of a resistor, or voltage that exists between two open terminals.

Figure 48 shows part of a circuit where voltages are labeled across different elements. We can write KVL equations around different loops as:

$$\begin{aligned}
-V_1 + V_5 - V_6 + V_7 &= 0, \quad Loop\ ABEFA \\
-V_1 - V_2 + V_3 - V_4 - V_6 + V_7 &= 0, \quad Loop\ ABCDEFA \\
-V_2 + V_3 - V_4 - V_5 &= 0, \quad Loop\ BCDEB
\end{aligned}$$

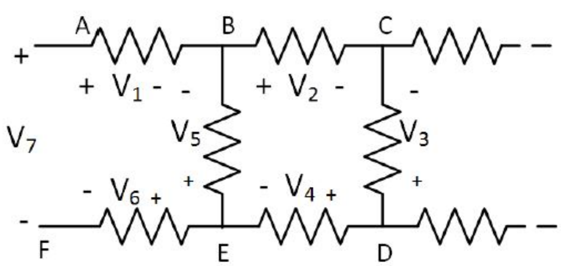


Figure 48: Illustration of writing KVL. There must be a voltage source or current source in the circuit which is not shown. Without any source, there will be no current in the circuit and voltage drops in all resistors will be 0.

Circuit Analysis

(a) Node Voltage Method, (b) Principle of Superposition and (c) Circuit with Nonlinear Element

You have learnt in the previous chapter how basic circuit laws can be used to find voltage and current in simple circuits, *e.g.*, a voltage source supplying energy to a couple of bulbs. However, the real world electrical systems and circuits are more complicated than what you have seen in the previous chapter. Electrical system in an automobile (Figure 49), connection of appliances in a residential unit, electronic devices, power generation and distribution systems are examples of electrical systems involving many interconnected components. Finding the current and voltages in these systems is not as simple as in the case of battery and bulb. So, there is a need for a systematic method of analyzing a circuit that can be applied to any circuit - simple or complicated.

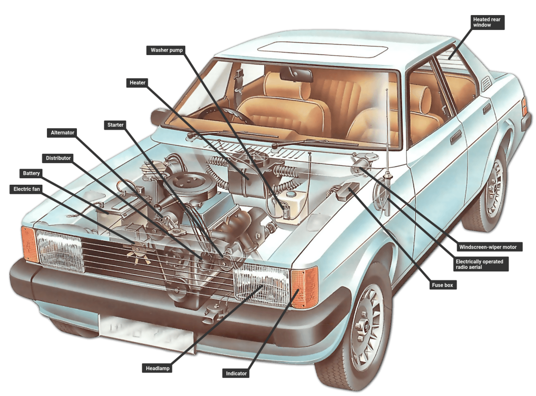


Figure 49: A practical electrical systems that involves many different interconnected components. (source of the image howacarworks.com)

Not-so-Elegant Method of Circuit Analysis

Before introducing a systematic approach called **Node Voltage Analysis**, two other not-so-elegant approaches are briefly introduced here.

Not-so-Elegant Method #1

We can convert a network of resistors with many nodes and branches into a simpler network using the concept of **equivalent resistance**.

When resistors are connected in series, they can be replaced with a single resistor with equivalent resistance value that is equal to the sum of individual resistances. For N resistors connected in series:

$$R_{series} = R_1 + R_2 + R_3 + \cdots + R_N.$$

When resistors are connected in parallel, they can be replaced with a single resistor such that equivalent conductance (reciprocal of resistance) is equal to the sum of the individual conductance. For N resistors connected in parallel:

$$\frac{1}{R_{par}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots + \frac{1}{R_N}.$$

The following example illustrates how to use the concept of equivalent resistance to simplify a complicated network resistors.

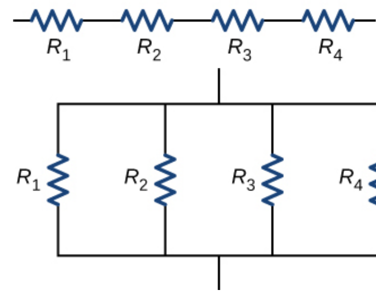
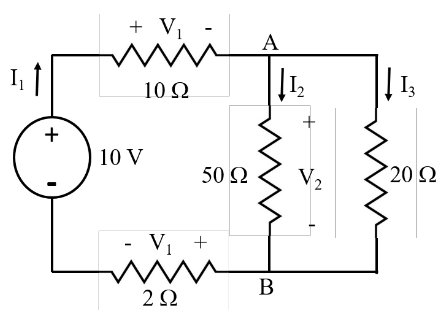


Figure 50: Four resistors connected in series (top) and in parallel (bottom)

Example 1:

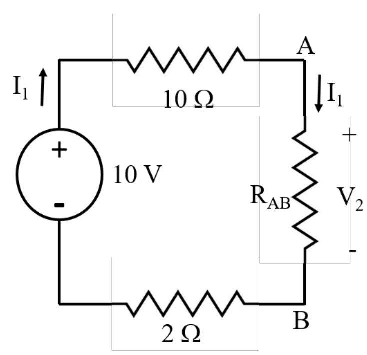
Find currents I_1, I_2 and I_3 in the following circuit.



Solution:

The 50 Ω and 20 Ω resistors are connected in parallel between the nodes A and B. Replace them with an equivalent resistance

$$\frac{1}{R_{AB}} = \frac{1}{50\,\Omega} + \frac{1}{20\,\Omega}, \quad \Rightarrow \quad R_{AB} = \frac{100}{7}\Omega.$$



In the new simplified circuit, three resistors are connected in series. So the overall equivalent resistance is

$$R_{eq} = 10\,\Omega + \frac{100}{7}\Omega + 2\,\Omega = \frac{184}{7}\,\Omega.$$

Current I_1 can be easily found:

$$I_1 = \frac{10}{\frac{184}{7}} = \frac{70}{184}\,A.$$

Then

$$V_2 = I_1 \times R_{AB} = \frac{70}{184} \times \frac{100}{7} = \frac{125}{23}\ V.$$

Referring to the original circuit,

$$I_2 = \frac{V_2}{50\ \Omega} = \frac{125}{23 \times 50} = \frac{5}{46}\ A,$$

$$I_3 = \frac{V_2}{20\ \Omega} = \frac{125}{23 \times 20} = \frac{25}{92}\ A.$$

Other derived formulae such as voltage divider rule, current divider rule etc could be applied. But we'll stop here as this method is not our main focus.

- This method is intuitive but not elegant.
- It is tedious.
- Its limitations will be exposed when you try to use this for more complicated circuit.

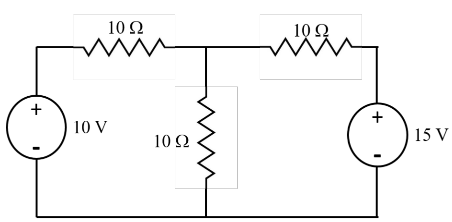
Not-so-Elegant Method #2

This method is also known as **branch current method.** The key ideas are -

- Assign a symbol to all unknown branch currents.
- When we do this, we do not know the true direction of current. Assign the directions arbitrarily, in other words, directions are also treated as unknowns with two possible values.
- If the solution for a particular current gives positive value, then the true direction is same as the assigned direction. Otherwise, true direction is opposite to the assigned one.
- Apply KCL to minimize the number of unknown current variables.
- Once you know the minimum number of unknown currents, write as many KVL equations.
- Solve the set of simultaneous equations to find the values of unknown currents.

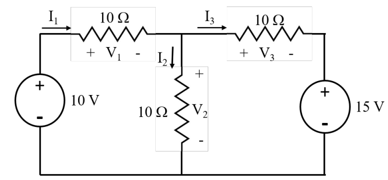
Example 2:

Find currents through all three resistors in the following circuit.



Solution:

First we assign the branch currents I_1, I_2 and I_3 to three resistors with arbitrarily chosen directions.



Polarity of the voltage across a resistor must conform to the direction of the current assigned. In a resistor, current flows from the positive terminal to the negative terminal.

There are three branch currents. However, they are related to each other by KCL. If we know the currents through any two branches, we automatically know the third branch current. Applying KCL, for the directions assigned,

$$I_3 = I_1 - I_2.$$

In effect, there are only two unknowns and we need two equations to solve for these unknowns.

Write KVL for the left loop:

$$10 - V_1 - V_2 = 0.$$

But $V_1 = 10 \times I_1$ and $V_2 = 10 \times I_2$. Therefore,

$$10I_1 + 10I_2 = 10. \tag{3}$$

Write KVL equation for the right loop

$$V_2 - V_3 - 15 = 0.$$

But $V_2 = 10 \times I_2$ and $V_3 = 10 \times I_3 = 10I_1 - 10I_2$. Therefore,

$$-10I_1 + 20I_2 = 15. \tag{4}$$

By solving simultaneous equations 3 and 4, we get

$$I_1 = \frac{1}{6}\,A, \quad I_2 = \frac{5}{6}\,A, \quad I_3 = -\frac{2}{3}\,A.$$

The negative sign of I_3 implies that the true direction of the current is opposite to what was assigned.

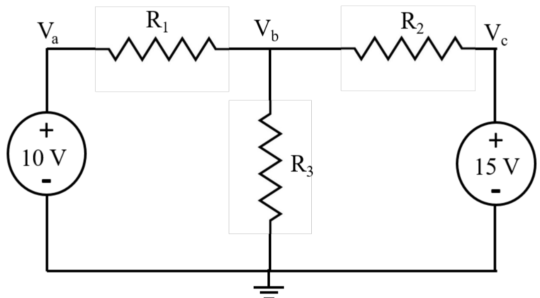
Node Voltage Analysis Method

Node voltage analysis method is the most commonly used method for analysis of electrical circuits.

What is Node Voltage?

Node voltage is the voltage of a particular node with reference to a common node.

You know by now that voltage is defined between any two points. In *Example 1*, we have assigned voltages V_1, V_2 etc to be voltage across a resistor, *i.e.*, the voltage between two ends of the resistor. When we define node voltage, we take one of the nodes as the reference node ($0\ V$), and all other nodes' voltages are with respect to this reference. In the schematic below, the conductor at the bottom is taken as the reference node with respect to which the node voltages V_a, V_b and V_c are assigned. Then the voltages across R_1, R_2 and R_3 are $(V_a - V_b)$, $(V_b - V_c)$ and $(V_c - 0)$, respectively.



In this schematic, the node voltages $V_a = 10\ V$ and $V_c = 15\ V$. Only unknown node voltage is V_b. We need only one equation to find this unknown voltage.

Node Voltage Method uses KCL only

In this method, Kirchoff's current law (KCL) is used to write equations involving the node voltages. If the voltage of a node is already given, then we do not need to write a KCL for that node. For all nodes for which the node voltage is unknown, write the KCL equation. So the number of KCL equations is equal to the number of unknown node voltages. We solve these simultaneous equations to find the node voltages. Once node voltages are determined, all other voltage and current in the circuit can be obtained there from.

The other common method is mesh current analysis method which uses KVL in writing the required equations.

Steps for Node Voltage Analysis

- Choose a reference node (Voltage = 0). Common practice is to choose the negative terminal of a voltage source as the reference node.

- Mark all the node voltages, and determine the number of independent variables (unknown voltages).

- Apply KCL at the nodes to obtain as many independent equations as the number of independent variables.

- Express current in each resistive branch in terms of the adjacent node voltages. (see the explanation in the next sub-section.)

- Solve the linear equations to determine the node voltages.

- Then all branch currents can be found from the node voltages.

Expressing Resistive Branch Currents in terms of Node Voltages

According to Ohm's law, current through a resistor is $I = \frac{V}{R}$ and it flows from higher voltage to lower (as shown for a single resistor at the bottom of Figure 51).

Following this rule, the three branch currents in Figure 51 can be expressed as

$$
\begin{aligned}
I_1 &= \frac{0 - V_2}{R_1}, \\
I_2 &= \frac{V_1 - V_2}{R2}, \\
I_3 &= \frac{V_1 - 0}{R_3}
\end{aligned}
$$

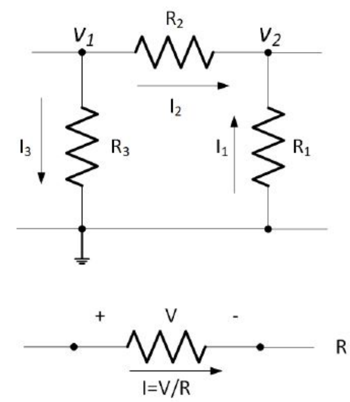
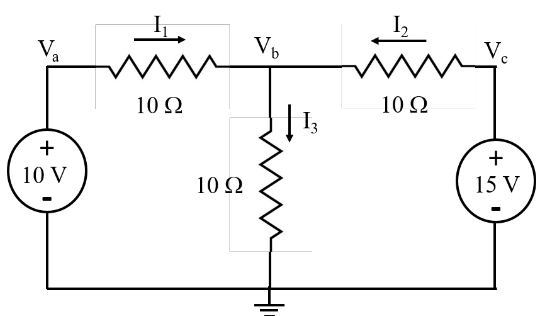


Figure 51: How to express branch current in terms of node voltages?

Example 3

Find the currents I_1, I_2 and I_3 in the circuit below using node voltage analysis.



Solution:

1. Node voltages $V_1 = 10\ V$ and $V_2 = 15\ V$. (these voltages are known because of the corresponding voltages sources.)

2. Write KCL for the only node with unknown node voltage

$$I_1 + I_2 - I_3 = 0.$$

3. Express the currents in the KCL equation in terms of node voltages -

$$\frac{V_a - V_b}{10} + \frac{V_c - V_b}{10} - \frac{V_b - 0}{10} = 0,$$

$$\frac{10 - V_b}{10} + \frac{15 - V_b}{10} - \frac{V_b - 0}{10} = 0.$$

4. Solving this equation we find

$$V_b = \frac{25}{3}\ V.$$

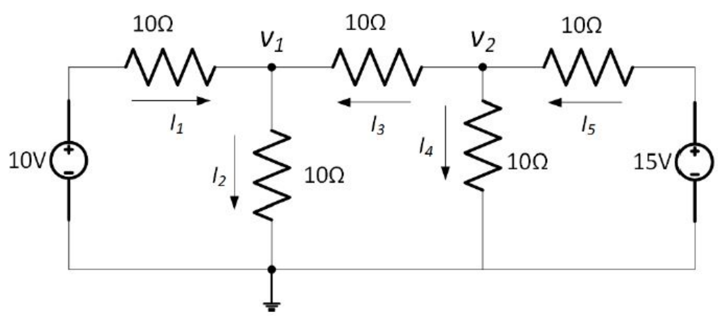
5. Then the branch currents are

$$I_1 = \frac{10 - \frac{25}{3}}{10} = \frac{1}{6}\, A,$$

$$I_2 = \frac{15 - \frac{25}{3}}{10} = \frac{2}{3}\, A,$$

$$I_3 = \frac{\frac{25}{3} - 0}{10} = \frac{5}{6}\, A.$$

Example 4



Solution:

1. Write KCL for nodes V_1 and V_2 -

$$\frac{10 - V_1}{10} + \frac{V_2 - V_1}{10} - \frac{V_1}{10} = 0,$$

$$\frac{15 - V_2}{10} - \frac{V_2 - V_1}{10} - \frac{V_2}{10} = 0.$$

2. Solve these two simultaneous equations to find the unknown node voltages V_1 and V_2.

3. Find the branch currents.

The Principle of Superposition

In a linear circuit with a number of independent sources, the response can be found by summing the responses to each independent source acting alone, with all other independent sources set to zero.

Although we are explaining the superposition principle in the context of circuit, the principle is applicable to any linear system (mechanical, chemical etc.) with multiple excitation.

What is a linear circuit?

A linear circuit consists of only elements that are governed by an element law that is linear, for example, Ohm's law for ideal resistor.

$$I_R = \frac{V_R}{R}.$$

The I-V characteristics for a resitor is a straight line. Two other circuit elements that you will learn later, capacitor and inductor, are also linear elements.

A diode, on the other hand, is not a linear element. The I-V plot of a forward-biased diode can be modeled using an exponential function

$$I_d = I_S(e^{\frac{V_d}{nV_T}} - 1),$$

where I_d and V_d are diode current and diode voltage, respectively, I_S is the saturation current (magnitude in the range of nA to μA), V_T is temperature dependent voltage (approximately 26 mV at 300 K), and n depends on how the diode is constructed (its value varies between 1 and 2).

In mathematics, a linear function $f(x)$ has the following two properties

$$f(x_1 + x_2) = f(x_1) + f(x_2),$$
$$f(\alpha x) = \alpha f(x).$$

Can you show using these two properties that a resistor is linear element and a diode is nonlinear?

What is an independent source?

The voltage source and the current source that you have learnt so far are independent sources, *i.e.*, the terminal voltage (current) of the voltage source (current source) is determined by the source itself and does not depend on any other voltage or current in the circuit.

On the contrary, a dependent voltage (current) source supplies a voltage (current) as commanded by a voltage or current from within the circuit of which the source is a part. Dependent sources are most commonly used to model elements having more than two terminals, *e.g.*, a transistor.

The dependent source is not covered in the scope of EPP1.

If a dependent source is controlled by a voltage within the circuit, it is called a voltage-controlled source (voltage-controlled voltage source or voltage-controlled current source). If current in a branch within the circuit determines the voltage (current) supplied by a dependent source, it is called current-controlled voltage source (or current-controlled current source).

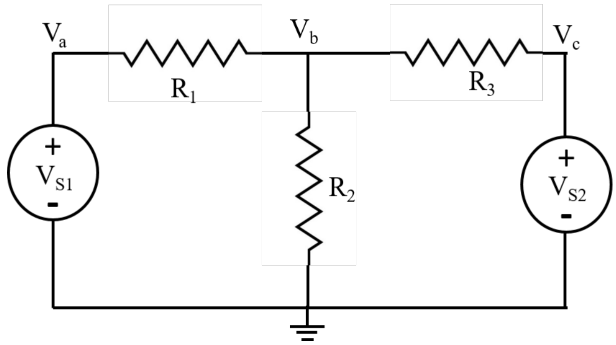
Superposition and Linearity

Superposition principle holds for linear circuits only, *i.e.*, circuits having linear elements - resistor, capacitor, inductor, independent sources and linearly controlled sources.

Circuits containing any non-linear elements do not follow superposition principle.

Linear Circuits & the Principle of Superposition

Let's apply node voltage analysis method to the circuit shown below. The circuit is same as the one in Example 3, but with variables instead of numerical values.



There is only one unknown node voltage V_b. Applying KCL at this node,

$$\frac{V_{S1} - V_b}{R_1} + \frac{V_{S2} - V_b}{R_3} - \frac{V_b}{R_2} = 0.$$

Rearranging the equation to keep the terms containing unknown variables on the left side of the equation and term with known values on the right,

$$(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3})V_b = \frac{1}{R_1}V_{S1} + \frac{1}{R_3}V_{S2}.$$

Taking $G_1 = \frac{1}{R_1}$, $G_2 = \frac{1}{R_2}$ and $G_3 = \frac{1}{R_3}$,

$$(G_1 + G_2 + G_3)V_b = G_1 V_{S1} + G_3 V_{S2}.$$

$G = \frac{1}{R}$ is the **conductance** of a resistive element having resistance $R\ \Omega$. The SI unit of electrical conductance is **Siemens**. The archaic term for this unit is **mho** (Ohm spelled backward).

So the node voltage,

$$V_b = \frac{G_1}{(G_1 + G_2 + G_3)}V_{S1} + \frac{G_3}{(G_1 + G_2 + G_3)}V_{S2}.$$

Points to note:

- The node voltage V_b is the sum of source voltages multiplied by a constant coefficient.

- Coefficients depend on the values of the resistive elements only..

These points can be utilized to analyze a multi-source network. If the sources V_{S2} is set to zero, we get the contribution that V_{S1} makes to the node voltage V_b. Similarly we can find the contribution made by V_{S2}. Sum of contributions of individual sources give the final solution.

If a circuit contains multiple sources, We can find the contribution of one of the sources to a node voltage or branch current by setting all other sources to zero.

$$\begin{aligned}
V_{b,1} &= \frac{G_1}{(G_1 + G_2 + G_3)}V_{S1}; \quad V_{S2} = 0, \\
V_{b,2} &= \frac{G_3}{(G_1 + G_2 + G_3)}V_{S2}; \quad V_{S1} = 0, \\
V_b &= V_{b,1} + V_{b,2}.
\end{aligned}$$

What is meant by setting a source to zero

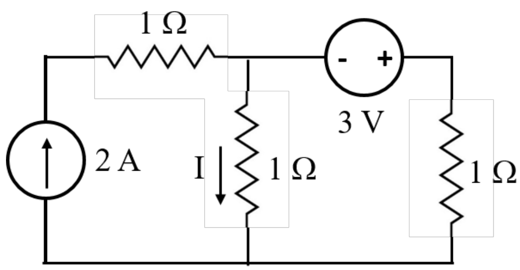
- If we want to set V_{S2} to zero regardless of the current flowing through that branch, it must be replaced by a short circuit.

- If we want to set I to zero regardless of the voltage appearing between the nodes where the source is connected to, it must be replaced by an open circuit.

Superposition principle is not a commonly used analysis technique, but rather a conceptual aid to visualize the behavior of the linear circuits containing multiple sources. Though sometimes, the analysis of a circuit is simplified by applying superposition principle i.e. by considering each independent source separately. Finally, the total response is obtained by adding the individual responses from all the independent sources together.

Example 5

Use the principle of superposition to find the current I in the circuit shown below.



Solution:

Step 1: Keep the 2A current source and set other source to zero (Figure 52). The circuit becomes a single source circuit. Find current I_1 using any method that you have learnt.

It is a simple problem. The 2A current is divided into two equal resistor in parallel (both 1 Ω). So,

$$I_1 = 1.0\ A.$$

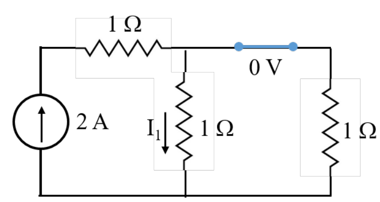


Figure 52: Circuit of Example 4 with all sources, except the 2A current source, set to zero.

Step 2: Keep the 3V source and set other source to zero (Figure 53). Find current I_2.

No current flows through the horizontal branch with 1 Ω resistor. So,

$$I_2 = -\frac{3}{1+1} = -1.5\ A.$$

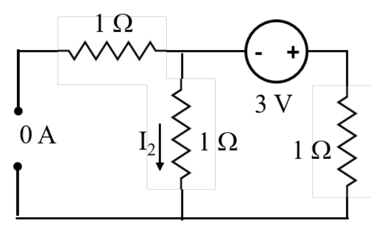


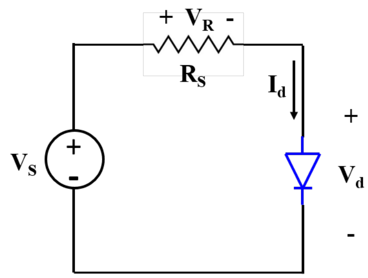
Figure 53: Circuit of Example 4 with all sources, except the 3V voltage source, set to zero.

Finally,

$$I \ = \ I_1 + I_2 = 1.0\ A - 1.5\ A = -0.5\ A.$$

Analysis of Circuit with Nonlinear Element

Consider the circuit shown below. The diode's I-V relation is nonlinear.



There are two unknown variables - current I_d and the voltage across one of the components (if you know the value of V_d, you can find the value of V_R and vice versa).

You need two equations to find the values of two unknown variables:

$$I_d = \frac{V_S - V_d}{R_S},$$

$$I_d = f(V_d).$$

The second equation, diode's I-V relationship is a nonlinear (exponential) function. You can solve these two simultaneous equations numerically using an iterative algorithm running on a computer.

An alternative method involves graphical analysis. The I-V characteristics curve of the diode is provided by the manufacturer. If we plot the graph of the first equation, then the point of intersection of the two graphs is the solution.

Let's rewrite the first equation as

$$I_d = -\frac{1}{R_S} V_d + \frac{V_S}{R_S}.$$

This represents a straight line with slope of $-\frac{1}{R_S}$ and vertical-axis intercept of $\frac{V_S}{R_S}$. It is called the **Load Line**.

Note that the slope and the intercept of the load line depend on V_S and R_S. Even if one nonlinear element is replaced with another, the load line remains the same for the given values of V_S and R_S.

If you draw the load line on the I-V characteristics graph of the nonlinear element, the values of current and voltage for the nonlinear device are given by the point of intersection of the load-line and the I-V graph of the nonlinear element as shown in Figure 55. The analysis also holds when the nonlinear element is replaced with a linear element.

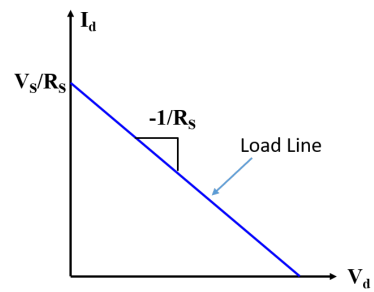


Figure 54: Load line is determined by the values of V_S and R_S only for any element connected to it.

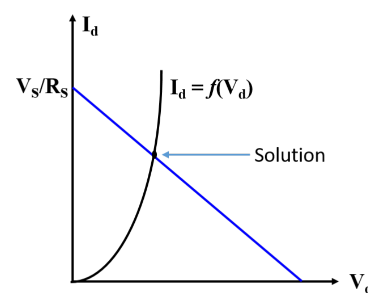


Figure 55: Diode voltage and diode current are the values of V_d and I_d at the point of interaction.

How do Systems get Energy? - Part I

Calculation of Energy Need

Almost all engineering systems need energy in some form. The energy is used to do work, which could be in form of

- motion of parts of the system or the whole system,
- changing temperature,
- producing light, etc.

There are many different ways to provide the energy required by a system to do useful work. Cars carry fuel (Petrol or diesel) with itself and convert the chemical potential energy of fuel into mechanical energy. Laptop computers have rechargeable batteries that store energy available from a charging point. Solar panels converts energy from solar radiation into electrical energy which can be used

- to supply energy directly to an appliance,
- to store in batteries or other storage mechanism for later use, or
- to supply energy to the utility grid.

Learning Outcome

Two learning outcomes from the theme "*How do Systems get Energy*":

1. How do we calculate the energy need of a system? (covered in this chapter)
2. How do we calculate the size of energy source for a given system? (covered in the next chapter)

In the next chapter (How do Systems get Energy? - Part II), you will learn about sizing of energy sources using the examples of battery and solar PV. However, the same concepts are applicable apply to any other sources.

In this chapter, you will learn

1. the basic principles of energy conversion and efficiency, and
2. how to apply these principles in determining energy need for a system.

Preliminaries

- Energy cannot be created or destroyed.
- It is converted from one form into another:
  - A wind turbine converts the kinetic energy (mechanical energy) of the wind into electrical energy,
  - A railway engine converts electrical energy obtained from the electric supply into mechanical energy.
- It could be transmitted from one part of the system to another.
  - a gear system transmits energy from one part of the machine to the other part (say wheels of the car).
  - electrical energy is transmitted through underground cables or overhead transmission lines over large geographical distance.

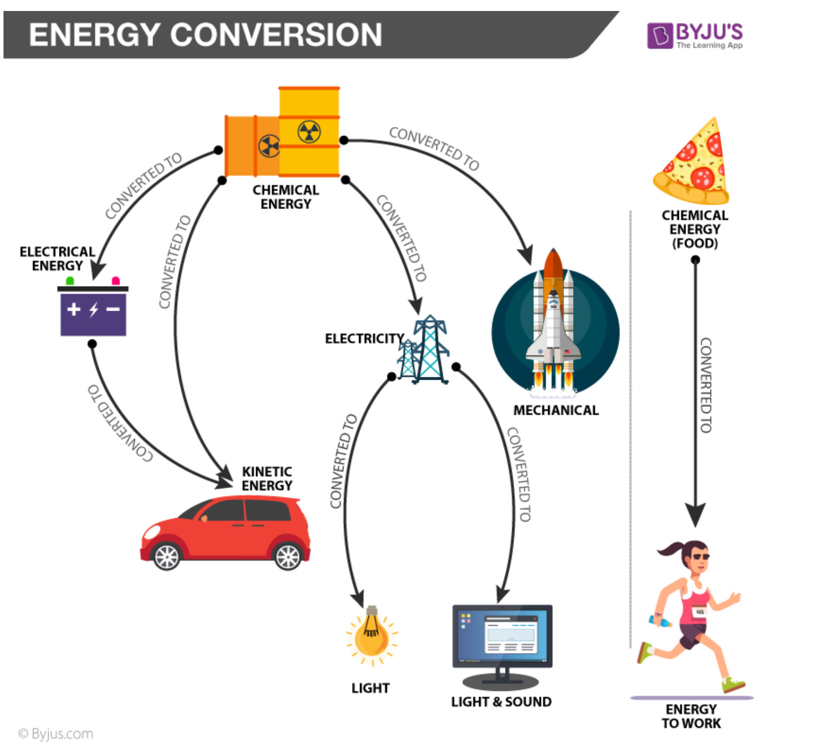


Figure 56: Energy conversion

One important specification associated with energy conversion/transmission is the instantaneous power that defines the rate of change of converted (or transmitted) energy:

$$p = \frac{dE(t)}{dt} \tag{5}$$

The SI unit of energy is **Joule** [J], and the unit of power is Joules per second [$\frac{J}{s}$], which is also called **Watt** [W]:

$$1\ W = \frac{1\ J}{1\ s}.$$

$1\ J = (1\ W) \times (1\ s) = 1\ Ws.$

Watt-second (W-s) is another unit of energy. It is the same as 1 J.

One Joule or 1 Ws is a small amount of energy in the context of many practical applications. It is the amount of energy that you may spend in lifting an average-sized apple to a height of 1 m; the energy you spend is converted into potential energy of the apple.

$$mgh \approx (0.1\ kg) \times (9.8 \frac{m}{s^2}) \times (1\ m) = 0.98\ J.$$

Larger units of energy are used in practical systems. For example, the utility companies use **Killowatt-hour** to quantify the electrical energy we consume.

$$\begin{aligned} 1\ kWh &= 1000\ W \times 1\ hour \\ &= 1000\ W \times 3600\ s \\ &= 3.6 \times 10^6\ J. \end{aligned}$$

The power consumed by various engineering systems have a relationship with the size and function of the system. In table 7, you can see how the power used in the system varies according to the size and function.

Table 7: Table of Appliances and their typical power rating

Appliance, Application	Power
Signal in TV Antenna receiver	10 nW
LED	10 mW
Laptop Notebook	50 ... 80 W
Refrigerator	150... 400 W
Washing Machine	500 W
Microwave	1500 W
Electric kettle	3000 W
Air-Con	1000 W
Continuous power drawn from 16 A power point	3.5 kW
Category A Car	97 kW
Electric locomotive	2 ... 6 MW
Wind Turbine	3 ... 8 MW
Synchronous Generator (power plant)	1500 MW

Efficiency

In practical systems, no machine (or transmission mechanism) can convert (or transmit) energy without any loss of energy. In a gear system, some part of energy is converted into heat due to friction. In electric trains, part of the energy drawn is lost in the contact resistance of the terminal and resistance of cables and wires.

The power that represents the intended outcome of the conversion/ transmission is called as output power P_{out} while the power received by the conversion/ transmission system from the source is the input power P_{in}. In all practical systems $P_{in} > P_{out}$, and the difference between the two is converted into some other form of energy which is not the intended outcome.

$$P_{in} = P_{out} + P_{Loss} \tag{6}$$

In the case of an electric bulb, the light energy is the intended output and the heat produced in the process is the loss.

The **efficiency** of conversion/ transmission is defined as

$$\eta = \frac{P_{out}}{P_{in}} = \frac{P_{out}}{P_{out} + P_{Loss}} < 1 \tag{7}$$

[1] BEWARE if any one claims to have a machine with efficiency of 1 or more than 1.

Efficiency of a machine is either expressed by a fraction (*e.g.* 0.8) or as percentage (80%).

Is efficiency η a constant?

The system efficiency is a measure of losses in the system. Higher the losses, lower will be the efficiency.

In any system, there are two types of losses:

1. Constant loss: loss that does not change with change in power output,
2. Variable load dependent loss: loss that changes with change in power output.

Consider a system with constant loss $P_{constant\ loss} = 0.1$ and variable loss $P_{variable\ loss} = 0.3P_{out}$. Assume that P_{out} varies from 0 to 1.

P_{out} may actually vary from 0 to 100W or from 0 to 100 MW. If we divide the actual power output by the maximum power output, we get the ratio varying from 0 to 1. This is called as **n**ormalization. Normalization is a good trick as you do not have to deal with large numbers that may be the value of the actual output of the system.

Then the efficiency of the system is

$$\begin{aligned} \eta &= \frac{P_{out}}{P_{out} + \sum P_{loss}} \qquad (8) \\ \eta &= \frac{P_{out}}{P_{out} + 0.1 + 0.3P_{out}}. \end{aligned}$$

Hence the efficiency is usually a function of the power output.

Energy and power expressed in terms of base SI units

Energy can exist in different forms (kinetic, potential, electrical *etc.*) and in an engineering system it is converted between different forms. All forms of energy should represent the same thing when they are expressed using the base units of SI system.

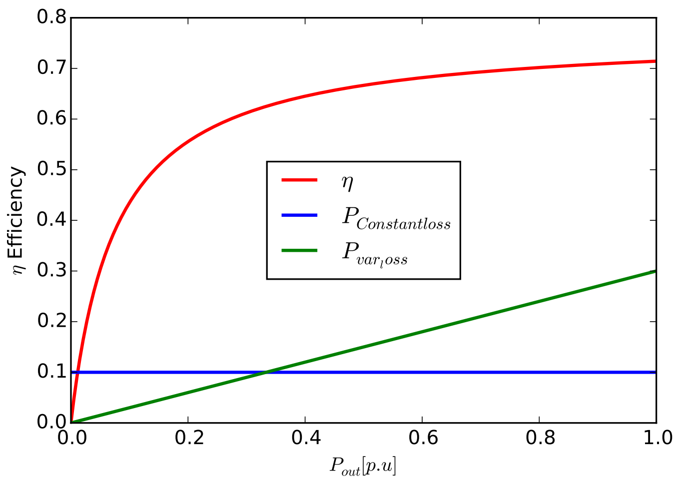


Figure 57: Efficiency in engineering systems

Table 8: Base SI units

Base Quantity	Name	Symbol
length	meter	m
mass	kilogram	kg
time	seconds	s
electric current	Amperes	A
Thermodynamic temperature	Kelvin	K
amount of substance	mole	mol
luminous intensity	candella	cd

There are 7 base units that are mutually independent (Reference: http://physics.nist.gov/cuu/Units/units.html).

Using these base units, you can derive the units of other physical quantities, such as, energy, force and power. For example, according to Newton's law,

$$
\begin{aligned}
Force &= mass \times acceleration \\
(Unit\ of\ force) &= kg \cdot \frac{m}{s^2} \\
&= kg \cdot m \cdot s^{-2}.
\end{aligned}
$$

You can calculate, following the same approach, the dimensions for the derived units for different quantities, some of which are shown in Table 9.

Table 9: Derived SI units

Derived Quantity	Name	Symbol
velocity	meter per second	ms^{-1}
acceleration	meter per $(second)^2$	ms^{-2}
current density	ampere per $(meter)^2$	Am^{-2}
magnetic field strength	Ampere per meter	$H = Am^{-1}$
electric charge	Coulomb	$C = As$
energy,work,quantity of heat	Joule	$J = Nm = m^2kgs^{-2}$
Power	Watt	$W = J/s = m2kgs - 3$
electric potential difference	volt	$V = W/A$=$m^2kgs^{-3}A^{-1}$

Kinetic Energy

$$
\begin{aligned}
K.E. &= \frac{1}{2}mv^2 \\
dim(K.E.) &= dim(mass) \times (dim(velocity))^2 \\
&= kg \times (\frac{m}{s})^2 \\
&= kgm^2s^{-2}.
\end{aligned}
$$

Potential Energy

$$
\begin{aligned}
P.E. &= Force \times distance \\
dim(P.E.) &= dim(mass) \times dim(acceleration) \times dim(distance) \\
&= kg \times \frac{m}{s^2} \times m \\
&= kgm^2s^{-2}
\end{aligned}
$$

Electrical Energy - Electric field

This is like potential energy. It is the amount of work done in moving a charge by the electrostatic force. This force acts on the charge due to an electrostatic field.

$$
\begin{aligned}
E_{ES} &= Q(charge) \times V(\text{electrical potential difference}) \\
dim(E_{ES}) &= dim(A \cdot s) \times dim(\frac{W}{A}) \\
&= dim(W\dot{s}) \\
&= dim(Joule) \\
&= dim(N \cdot m) \\
&= dim(kgm^2s^{-2})
\end{aligned}
$$

It can be shown in the same way that all forms of energy has the dimension of

$$kgm^2s^{-2}.$$

How to estimate energy need of a system

Consider a system that works at a constant power, *e.g.* P Watt for T seconds, then the energy consumed by the system is

$$E = P \times T \quad [J].$$

In a complex system with many energy conversion machines and devices, neither all have same power rating nor all operate for the same duration. For example, in a home, the water heater will consume electric power to heat water as long as the person is taking shower. Then it is turned off and consumes zero power.

You can find similar operation in a smart phone. For certain duration, the smart phone functions as a phone and consumes certain power. When it is used for surfing the internet, its power consumption is changed. When it is used for playing game, the power consumption is different from the previous two operations.

In engineering systems, we study the total energy consumption by first studying the load variation against time. To this end, we create a table of power consumed against time. Task-wise load variation as function of time is shown in Figure 58, where each task is represented by a column. Total energy (W-s) consumed is the total area under this plot.

Task	Power [W]	Time duration [s]
Task1	20	0-7.5
Task2	25	7.5-15
Task3	40	15-22.5
Task4	15	22.5-30
Task5	30	30-37.5
Task6	28	37.5-45
Task7	25	45-52.5
Task8	10	52.5-60

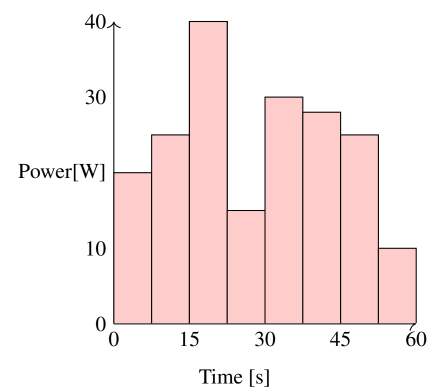


Figure 58: Load curve to estimate energy

Another way of representing the curve in Figure 58, is shown in Figure 59, which is called the **load duration curve** or **load profile**. The figure tells us that 10 W load is present for the entire duration (from different tasks). Similarly, 30 W load demand is present for 15 seconds (Task 3 and Task 5), and a load of 40 W exists for 7.5 seconds.

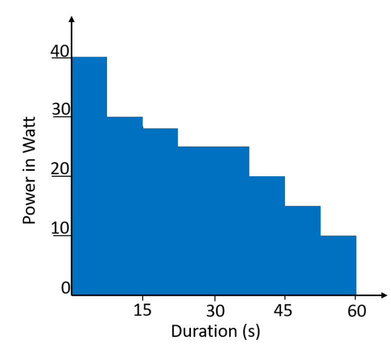


Figure 59: Load duration curve or load profile, another way of finding energy requirements

- We can get the load duration curve from load curve dividing the y-axis (Power) in certain number of steps.
- Then finding the time duration for each of the steps
- Then stacking the powers staring with lowest value and longest time duration.

- Thus the top most block will have the highest power for the shortest duration

Total energy (W-s) consumed is the total area under this plot.

Calculating energy needs using block diagram

If we know how energy flows through different subsystems and if we know the efficiency for each of the subsystems, then we can use them to calculate energy input required for a desired output.

Energy balance in engineering systems

For any engineering system, energy can be flowing in various forms, *i.e.*, electrical, mechanical, hydraulic, thermal etc. but the energy balance will be always maintained.

$$\sum E_{in} = \sum E_{out} + \sum E_{stored} + \sum E_{lost}$$

Since power is rate of energy flow $P = \frac{dE}{dt}$
We will also get power balance by differentiating on both sides

$$P_{in} = P_{out} + \frac{dE_{stored}}{dt} + P_{loss}$$

Using block diagrams to calculate energy requirements

After building a system block diagram we can calculate the power balance withing the system using simple calculations.

We will use

$$\eta = \frac{P_{out}}{P_{in}} < 1$$

Consider a simple system with all subsystems connected sequentially as shown in Fig.60. In this system the output delivered to the load is given. We need to estimate the power that need to be supplied at the input. Efficiency of each block is either given or estimated as η. The input power is simple to calculate

$$P_{in} = \frac{P_{out}}{\eta_1 \cdot \eta_2 \cdot \eta_3}$$

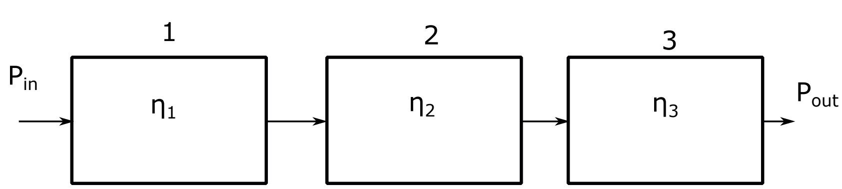


Figure 60: Simple system with series flow

If the subsystems are connected in parallel the power output will be the sum of the power output of each block:

$$
\begin{aligned}
P_{out} &= P_{out1} + P_{out2} + P_{out2} \\
P_{out} &= P_{in1} \cdot \eta_1 + P_{in2} \cdot \eta_2 + P_{in3} \cdot \eta_3
\end{aligned}
$$

Any system will be a combination of series and parallel blocks similar to the block diagram shown in Figure 61 . The output power is P_{out}; the input

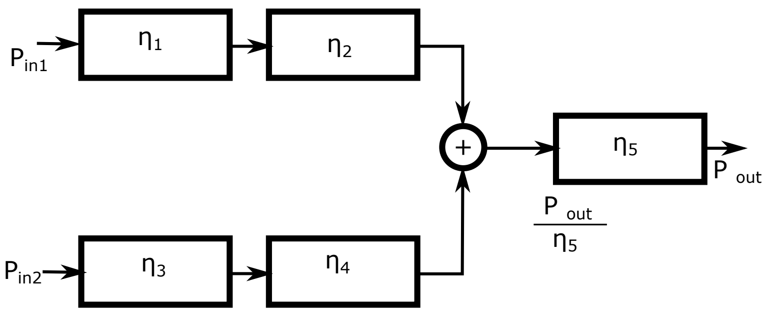


Figure 61: Series Parallel combination of block diagrams

power at the block 5 will be

$$P_{in5} = \frac{P_{out}}{\eta_5}$$

This input power is the sum of the output powers of block 2 and block 4. Since block 1 and block 2 are in series, we get the output powers of block 2 and block 4 as

$$
\begin{aligned}
P_{out2} &= P_{in1}\eta_1\eta_2 \\
P_{out4} &= P_{in1}\eta_3\eta_4
\end{aligned}
$$

Therefore,

$$P_{in5} = P_{in1}\eta_1\eta_2 + P_{in2}\eta_3\eta_4,$$

$$\frac{P_{out}}{\eta_5} = P_{in1}\eta_1\eta_2 + P_{in2}\eta_3\eta_4$$

If we have to represent only energy flows in the system, we can do so using what are called as Sankey Diagrams (Figure 62). Note that, there are multiple sources of energy input in the system represented by this Sankey diagram.

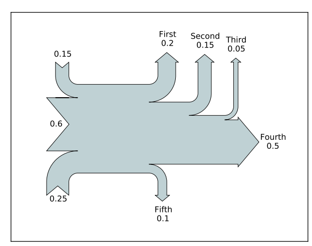


Figure 62: Sankey diagram shows flow in a system in terms of proportions.

Efficiency in Fossil Fuel based Power Generation

Electricity in most countries is generated by burning fossil fuel. Figure 63 shows a schematic drawing of how electricity is generated from fossil fuels.

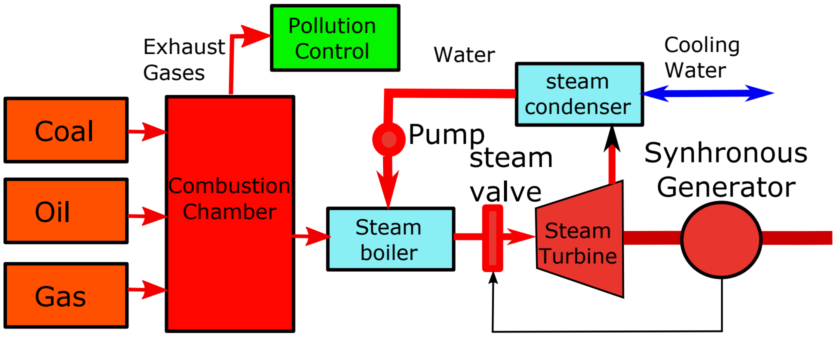


Figure 63: Fossil Fuels to Electricity

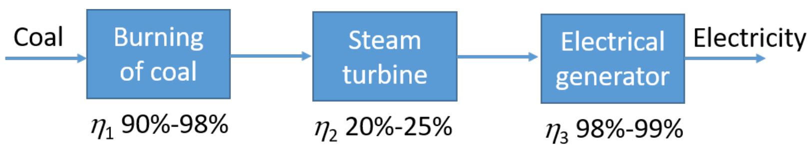


Figure 64: A simplified representation of the process of generating electricity from coal

You can use the information given in Table 10 to find the efficiency of generating electricity from coal.

The process:

Considering the maximum efficiency for each of the subsystems,

$$\eta_{coal-to-electricity} = 0.98 \times 0.25 \times 0.99 = 0.243.$$

In a process where subsystems are connected in series, the maximum efficiency will depend on the lowest efficiency in the chain.

System	Conversion	Efficiencies
Natural Gas Furnace	Chemical[MJ/L]$\rightarrow$ Heat	90-96%
Internal Combustion Engine	Chemical $\rightarrow$ Mechanical	15-20%
Power Point Boiler	Chemical (Oil, Coal) $\rightarrow$ Heat	90-98 %
Steam Turbines	Heat $\rightarrow$ Mechanical	25-30%
Electricity Generator	Mechanical $\rightarrow$ Electrical	98-99 %
Gas Turbine	Chemical (gas) $\rightarrow$	35-40 %
Hydro power plant	Grav. Potential $\rightarrow$ Mechanical	60-90 %
Wind	Kinetic $\rightarrow$ Mech $\rightarrow$ Electrical	30-60 %
Photo Voltaic Cell	Radiation $\rightarrow$ Electrical	10-20%
PV Inverter	Electrical $\rightarrow$ Electrical	97-98%
Electrical Transformer	Electrical $\rightarrow$ Electrical	98-99 %

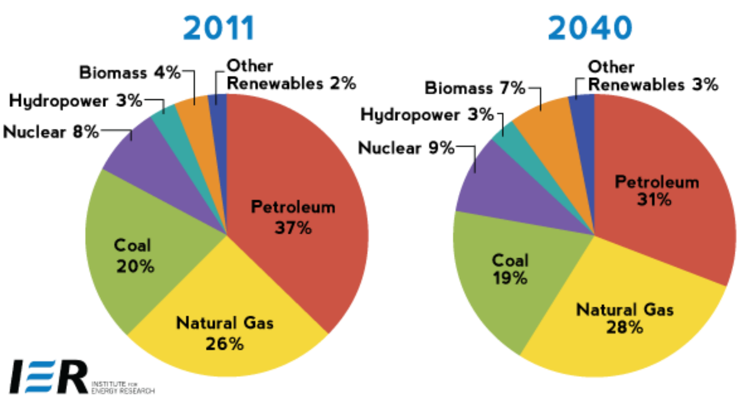
Table 10: Efficiencies of different energy conversion systems

How do Systems get Energy? - Part II

Battery & Photovoltaic (PV) cell

In the previous chapter, you have learnt about general concepts of energy conversion and conversion efficiency. This chapter focuses on the energy sources used in practical applications.

Energy Sources



An overview of energy sources is provided at http://www.mpoweruk.com/electrical_energy.htm.

The bulk of the energy used in today's world comes from fossil fuels. Detrimental effects on environment due to burning fossil fuel are also well understood. However, as the figure suggests, dependence on fossil fuel will remain there for some time.

Advantages of using Renewable Energy

- Renewable energy won't run out
- In general, renewable energy technologies require less maintenance.
- Operating cost is much lower as we don't need to buy fuel.

- They have health and environmental benefits.
- Less reliance on imported fossil fuel.

Disadvantages of using Renewable Energy

- High initial investment
- Many of these sources are not available 24/7, year-round.
- Because of intermittent nature, there is need for high storage.
- Geographic limitations: some countries are less suitable for renewable technologies than others.

Another reason for not much variation in energy sources between 2011 and 2040 is the life span of existing fossil fuel based energy sources. Most coal-fired power plants have a lifespan of 50-60 years. So the power plants commissioned in the 90s will still be functional in 2040.

Scientists and engineers are continuously searching for ways to make renewable sources economically viable and renewable energy more dependable. As that is achieved, renewable energy will get larger share of the energy industry.

A more in-depth discussion on various energy sources is beyond the scope of this module. You will learn about sizing of energy sources in two specific contexts.

1. Battery: source of energy for autonomous systems, electric vehicles, handheld devices, portable appliances, etc.
2. PV system: most promising renewable energy source.

Battery: a Practical Voltage Source

Batteries are widely used as the source of energy in portable appliances or autonomous systems. They are made in a variety of shapes, sizes, and ratings, from miniaturized button batteries capable of delivering only a few microamps to large automotive batteries capable of delivering hundreds of amps. There are different types of batteries using different chemistry, *e.g.*, lead-acid, Nickel, Lithium etc. But all of them use chemical reaction to separate charges that are deposited on different electrodes creating a voltage between them. A battery is a practical voltage source.

Battery Specifications

You can find some words and numbers written on the battery labels like those shown in Figure 65. These words and numbers are explained in this subsection. Remember that not all batteries give the same set of information.

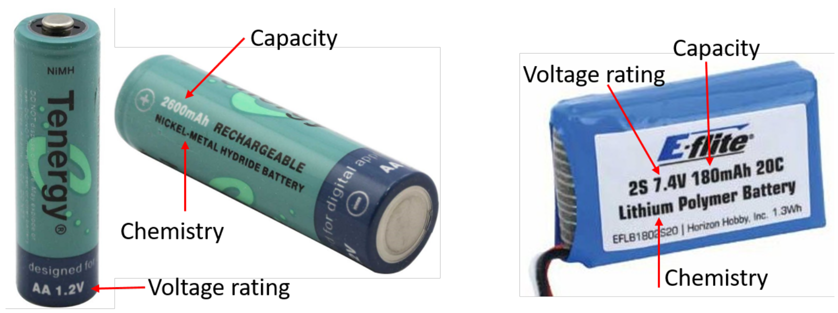


Figure 65: Battery specifications shown on the label: Nickel-Metal Hydride battery (left) and LiPo battery (right)

- **Chemistry:** Different batteries use different chemical reactions to produce voltage. Nickel Metal Hydride (Ni-MH) uses Nickel based chemical reaction but Lithium Polymer (Lipo) battery uses Lithium-ion technology with polymer electrolyte.
- **Voltage Rating:** This is the output voltage of the battery or battery pack. The Ni-MH batery in Figure 1 gives 1.2 V output but the LiPo battery pack shown in the figure gives 7.4 V.
  - Labels of battery modules or battery pack sometimes show the cell count or **S-rating**. The 2S written on the Lipo battery means there are 2 cells in Series.

Basic LiPo cell gives 3.7 V. By making a module with two such cells in series gives $2 \times 3.7\ V = 7.4\ V$.

- **Capacity** or **Ampere-hour** rating gives an indication of the amount of energy stored in the battery. The NiMH battery in the figure is rated 1600 mAh - that means constant current of 1600 mA will fully discharge the battery in one hour.

The current drawn out of a battery is determined by the resistance of the load connected to it. If the load connected to the 1600 mAh battery draws 100 mA current, the battery will endure for 16 hours.

- **C-rating** is used to specify the highest magnitude current that can be drawn out of the battery without damaging it.
  - For a given battery, the capacity (Ah) is fixed. So higher the discharge current magnitude, shorter it will take to be fully discharged.
  - For a battery of 1800 mAh capacity,
    * if the discharge current is 1800 mA, it will take 1 hour to be fully discharged. This discharge is rated as 1C.

$$discharge\ current\ for\ 1C = \frac{Capacity}{1\ h}.$$

    * if the discharge current is doubled, *i.e.*, 3600 mA, it is rated as 2C. In this case, battery will be fully discharged in half an hour.

$$discharge\ current\ for\ 2C = \frac{Capacity}{\frac{1}{2}\ h}.$$

  - Increased discharge current results in higher power loss making the battery hotter. Manufacturers use the C-rating to specify the safe level of discharge current.

You can determine safe discharge current from C rate:

$safe\ discharge\ current = capacity \times C\ rate.$

- Some batteries specify two C-ratings - (a) continuous C-rating for constant discharge current and (b) burst C-rating for short interval discharge current. Burst C-rating is always higher than the continuous C-rating.

Amount of Energy Stored

The product of the terminal voltage and the capacity is the amount of energy stored in the battery:

$$(V) \times (Ah) = (V \times A) \times h = Wh.$$

The Ni-MH battery shown in Figure 65 contains

$$(1.2\ V) \times (2600 \times 10^{-3}\ Ah) = 3.12\ Wh$$

$1\ Wh = 3600\ Ws = 3600\ J.$

energy.

Sometime, the amount of energy contained when in fully charged battery is specified on the label. Energy (1.3 Wh) contained in the LiPo battery shown in Figure 65 is indicated on the label.

Energy per Kg and Energy per Litre

While selecting battery for applications with constraints on volume and/or weight, we look for energy stored per unit mass and/or energy stored per unit volume. Relative performance for different battery chemistry is shown in a **Ragone chart**, see Fig.66.

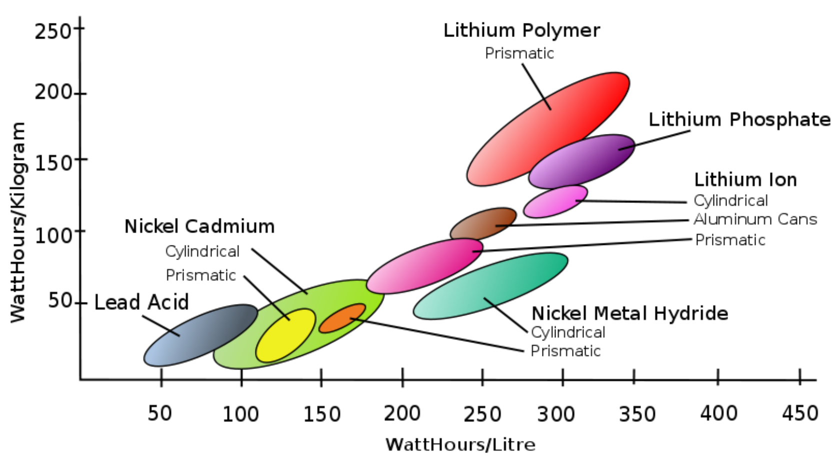


Figure 66: Ragone Chart for rechargeable batteries, By Original Author: Barrie Lawson [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons

Ideal Voltage Source versus Practical Voltage Source

Battery is a source of electrical energy that gives a voltage between its terminals. Current is drawn out of the battery when a load is connected to it. A battery converts chemical potential energy into electrical energy. A generator is another voltage source that it converts mechanical energy into electrical energy.

Battery specifications include a terminal voltage, *e.g.*, a car battery gives 12 V and an AA size Ni-Metal Hydride battery gives 1.2 V. **Does the terminal voltage remains constant while it provides energy to the load?**

A constant terminal voltage is what we expect in a **ideal voltage source**. We use the symbols shown in Figure 67 in circuit schematics to represent such a voltage source.

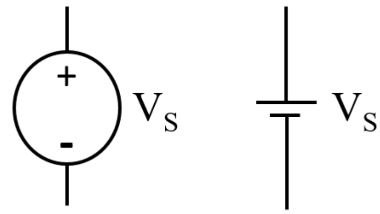


Figure 67: Ideal voltage sources (DC) are drawn in circuit schematics using one of these symbols. V_S is the value of the terminal voltage. For the circular symbol, we need + and - signs to indicate the polarity. For the other symbol, the longer line represents the positive terminal.

In reality, the terminal voltage drops to a level lower than the terminal voltage given in the specification due to power loss inside the battery as it converts chemical energy into electrical energy.

Battery Discharge Characteristics

Figure 68 shows the discharge characteristics of a battery. It shows how the voltage of the battery changes with the state of charge of the battery.

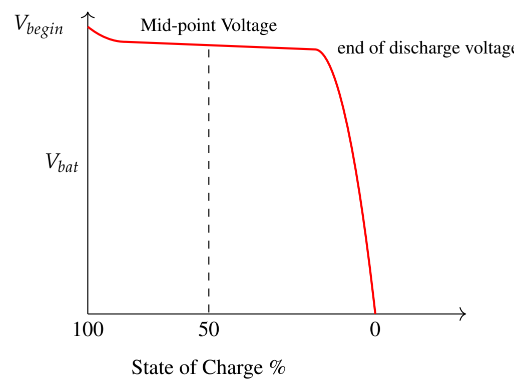


Figure 68: Discharge curve of a battery

What is **state of charge?**
The state of charge (SOC) of a battery denotes the capacity that is currently available as a function of the rated capacity. The value of the SOC varies between 0% and 100%. A 100% SOC means the battery is fully charged, whereas a SOC of 0% means it is completely discharged. With aging, the maximum SOC starts decreasing. This means that for an aged battery, a 100% SOC would be equivalent to a 75%–80% SOC of a new cell.

Now going back to Figure 68, When the battery is fully charged but not loaded, voltage is at the maximum. During operation the battery voltage will be slightly lower than the maximum voltage; this is called the Mid-point voltage. The lowest voltage the battery should be operated at is the end of discharge voltage V_{eod}. If we discharge the battery further, its voltage decreases rapidly and we may not be able to use the battery again.

What does the area under the discharge curve in Figure 68 represent?

Internal Resistance of Battery

When a battery supplies energy to the load, part of the energy is lost in the energy conversion process or in transmission through the terminals. The lost energy is converted into heat. Higher is the current drawn from the battery, larger is the amount of energy lost.

The power loss in the battery can be modeled using a resistance as shown in the figure below. The parameter R_{int} is the **internal resistance** of the battery and V_S is the terminal voltage given in the specification, *e.g.*, 1.5 V for alkaline AA battery.

The open circuit voltage

$$V_{OC} = V_S.$$

When a resistor (R_L) is connected across the battery terminals (Figure 70),

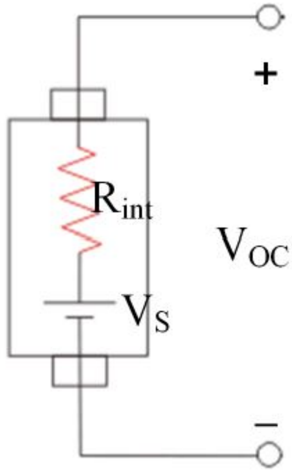


Figure 69: A practical battery is modelled as a series connection of an ideal voltage source and a resistor representing the internal resistance. If you use an ideal voltmeter to measure the terminal voltage of a battery, the voltmeter will show exactly the value of V_S. An ideal voltmeter has infinite internal resistance and hence does not draw any current. The terminal voltage measured using an ideal voltmeter is called the **Open Circuit Voltage**, V_{OC}.

current starts to flow, and the voltage across the terminal becomes lower than V_S. Applying KVL to the circuit shown in Figure 69,

$$V_L = V_S - I_L \times R_{int}. \tag{9}$$

So, the voltage versus current plot of a battery is a straight line with negative slope. The slope of this line is $-R_{int}$ and the intercept on the voltage-axis is equal to the source voltage V_S.

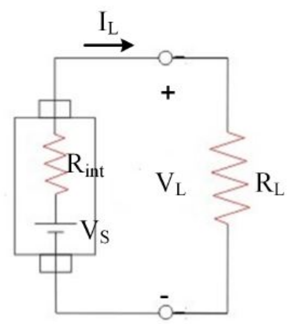


Figure 70: The terminal voltage V_L is smaller than the open circuit voltage V_{OC} when a load is connected across the battery.

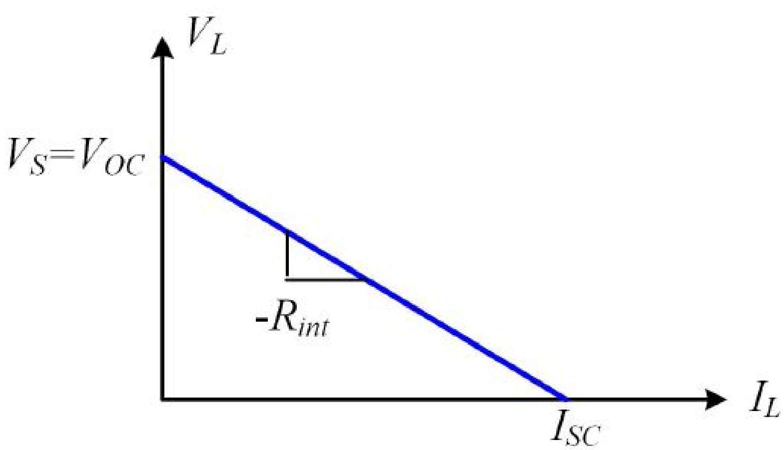


Figure 71: Voltage versus current plot of a battery.

In order to characterize a battery, *i.e.*, to determine the values of V_S and R_{int}, you need to experimentally determine the equation of this straight line. You can easily determine two points on this straight line by varying the value of R_L. For example, two values of load resistor R_{L1} and R_{L2} will result in

$$\begin{aligned} V_{L1} &= V_S - I_{L1} \times R_{int}, \\ V_{L2} &= V_S - I_{L2} \times R_{int}. \end{aligned}$$

You can solve these two simultaneous equations to determine the values of the unknown quantities V_S and R_{int}.

Internal resistance is an important parameter but is not mentioned on the battery cover. Internal resistance of battery changes over time. It may also change because of temperature.

In simple terms, internal resistance is a measure of the difficulty a battery has delivering its energy to the load. The higher the number, the harder it is for the energy to reach the load. The energy that does not go to the load is lost as heat.

While a LiPo battery is in use, a build up of Lithium Oxide is formed on the inside terminals of the battery and the internal resistance is increased. After prolonged uses, the battery will simply wear out and be unable to hold on to any energy you put in during charging - most of it will be lost as heat. If you have seen a supposedly fully charged battery discharge almost instantly, the reason is probably high internal resistance.

There is a correlation between the C-Rating of a battery and its internal resistance. In general, batteries with a higher C-Rating also have a low internal resistance. This is not always the case, as there are always variances in manufacturing, but the general idea seems to hold true.

Battery Pack

A **cell** is the smallest, packaged form a battery can take and is generally on the order of one to six volts. The voltage of a cell is different for different battery chemistry.Lead acid battery has an open circuit voltage of 2V per cell and Nickel-based batteries are 1.2V/cell. Lithium ion batteries give 3.6V/cell - 3.7V/cell depending on anode and cathode materials used.

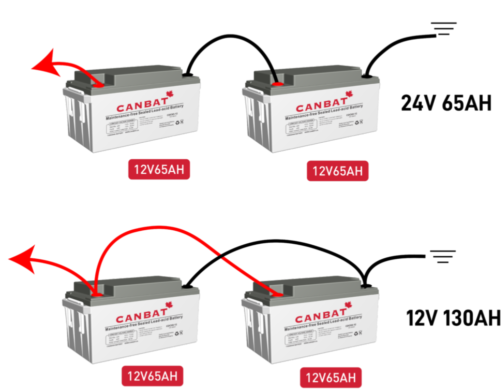


Figure 72: Series connection gives higher voltage with no change in the capacity (Ah) and Parallel connection gives higher Ah with no change in voltage. Source of image: canbat.com

A battery **module** consists of several cells connected in either series or parallel. A **battery pack** is assembled by connecting modules together, again either in series or parallel.

Connecting cells/modules in series increases the voltage with no change in capacity whereas parallel connection increases the capacity with no change in voltage.

Internal connection of cells is sometime written on battery labels. For example, **3S2P** means 3 cells in series forming one module and 2 such modules are connected in parallel.

Thevenin Equivalent Model of a Battery Pack

What is Thevenin equivalent?

The behavior of any arbitrary linear circuit, with respect to a pair of terminals can be represented with a Thevenin equivalent, which consists of a ideal voltage source in series with a resistor.

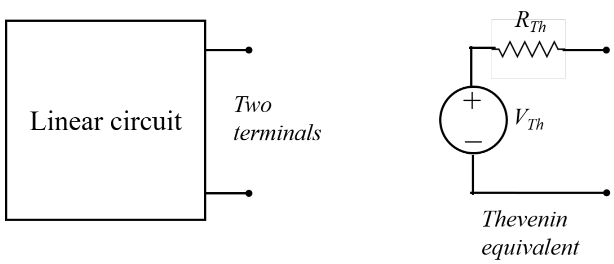


Figure 73: Thevenin equivalent of any arbitrary linear circuit

We need to determine the Thevenin voltage (V_{Th}) and the Thevenin resistance (R_{Th}) so that the model (the equivalent circuit) behaves just like the original circuit when a load is connected to the terminals.

Method 1:

There are two unknowns, V_{Th} and R_{Th}, so two measurements or calculations should suffice. We can do this by connecting two different values of load resistor (Figure 74) and measuring the voltage V_L using a ideal voltmeter.

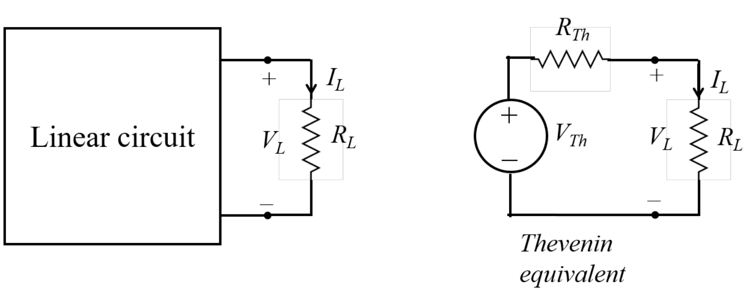


Figure 74: Finding Thevenin voltage and resistance using two measurements.

From the Thevenin equivalent circuit in Figure 74,

$$V_L = \frac{R_L}{R_{Th} + R_L} V_{Th}. \tag{10}$$

If we take this measurement for two different values of R_L -

$$
\begin{aligned}
V_{L1} &= \frac{R_{L1}}{R_{Th} + R_{L1}} V_{Th} \\
V_{L2} &= \frac{R_{L2}}{R_{Th} + R_{L2}} V_{Th}.
\end{aligned}
$$

The values of chosen resistors R_{L1} and R_{L2} are known, and the measured voltages V_{L1} and V_{L2} are also known. We can solve these two simultaneous equations to determine the two unknown quantities V_{Th} and R_{Th}.

An alternative measurement for this two-point method is to use an ammeter to measure the current I_L in addition to the measurement of V_L. In that case, we express the loop equation as

$$V_L = V_{Th} - R_{Th} I_L. \tag{11}$$

For two different values of the load resistor,

$$
\begin{aligned}
V_{L1} &= V_{Th} - R_{Th} I_{L1} \\
V_{L2} &= V_{Th} - R_{Th} I_{L2}
\end{aligned}
$$

The equation 11, which is identical to the equation 9, represents a straight line with negative slope. The two measurements with two different values of load resistance corresponds two points on this straight line.

We can solve these two simultaneous equations to find V_{Th} and R_{Th}.

<u>Method 2:</u>

The open circuit voltage (V_{OC}) and the short circuit current (I_{SC}) are the two ends of the straight line of equation 11. When we measure the open circuit voltage, $I_L = 0$ and so

$$V_{Th} = V_{OC}.$$

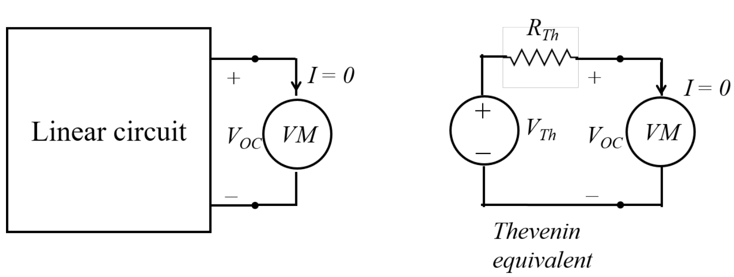


Figure 75: Measure open circuit voltage using a ideal voltmeter so that $I_L = 0$.

When we measure the short circuit current, $V_L = 0$ and

$$I_{SC} = \frac{V_{Th}}{R_{Th}}, \qquad R_{Th} = \frac{V_{Th}}{I_{SC}}.$$

The Thevenin equivalent can also be obtained by measuring the open circuit voltage (using a ideal voltmeter) and the short circuit current (using a ideal ammeter).

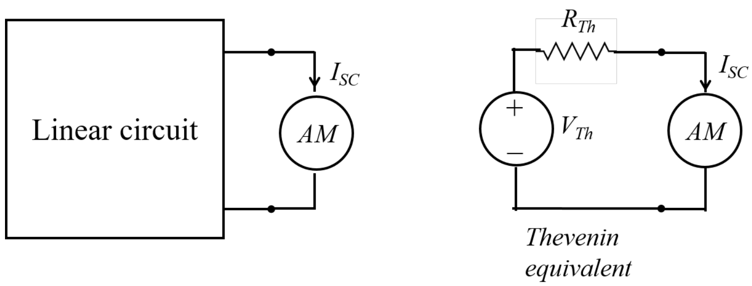


Figure 76: Measure short circuit current using a ideal ammeter so that $V_L = 0$.

Note that shorting the output may not always be practical. For example, some devices may have over-current protection circuitry that prevents large short-circuit currents from flowing. Or the device might not be able to handle the large current that might flow when the output is shorted without being damaged.

Alternative method for finding R_{Th} for Circuit Schematic

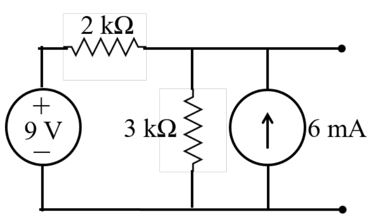
If the circuit consists of independent sources and resistors only, then the R_{Th} can be found by deactivating the independent sources and finding the equivalent resistance as seen between the terminals.

In electronic circuits, there are dependent voltage sources whose output is determined by either the voltage at some other node or current in another branch of the circuit.

How to deactivate an independent source?

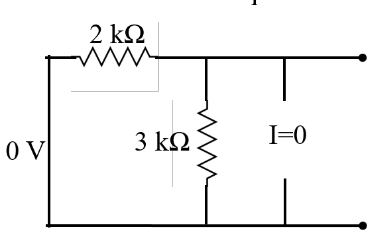
- If the independent source is a voltage source, replace it with a short circuit across the source. This will force the voltage to zero between the nodes where the independent source is connected.
- If the independent source is a current source, remove it from the circuit. This will force the current in that branch to be zero.

Consider the circuit below



So far, you haven't encountered a current source. A ideal current source is one that gives out the specified amount of current regardless of the terminal voltage. The circle with an inscribed arrow in the schematic is a ideal current source. You will later learn about PV panels which can be modeled as a practical current source.

After deactivating the independent sources, the circuit is left with only resistors. We can find the equivalent resistance between the terminals.



Thevenin resistance is

$$R_{Th} = (2\ k\Omega)||(3\ k\Omega) = 1.2k\Omega.$$

Solar PV - Renewable Energy Source

Most renewable energy comes either directly or indirectly from the sun. Sunlight can be used directly for heating and lighting homes and other buildings. It can also be used for generating electricity which can then be used for a variety of residential, commercial and industrial uses. The kinetic energy of wind is a consequence of heating of atmosphere - effectively the root source is the sun.

In this module, renewable energy using PV panels will be discussed in details and wind energy will be briefly mentioned.

Solar Energy is becoming increasingly more popular choice for generating electricity. The energy contained in solar irradiation converted into electricity in two main ways:

- **PV** (Photovoltaics) cells, also called solar cells, are electronic devices that convert sunlight directly into electricity. Today, PV is the fastest growing renewable energy technology.

- **Concentrated Solar Power** (CSP) uses mirrors to concentrate solar rays. These rays heat a fluid and this produce steam to drive a turbine of an electrical generator.

Amount of solar energy/power available

In any energy conversion process, the output energy is always less than the input energy. In order to evaluate how much energy can be generated, we must know how much solar energy is available for conversion and what is the conversion efficiency.

Irradiance:

Irradiance is a measure of solar power $\frac{W}{m^2}$ and is defined as the rate at which solar energy (**W-h**) falls onto a surface.

The solar irradiance on a cloudless sky during midday is 1000 $\frac{W}{m^2}$ on earth. In space above the atmospheric layer it is 1300 $\frac{W}{m^2}$.

An effective area of 100 $m \times 100$ m available for solar energy conversion receives

$$1000\ \frac{W}{m^2} \times 100\ m \times 100\ m = 10 \times 10^6\ W$$

power during midday.

The area in meter squared is the area that is facing the sun. At the equator, the rays of the sun are incident normally on a plane kept flat on the ground. As we move away from the equator, the area facing the sun (normal area) becomes smaller. So the latitude of the location is a factor affecting the amount of solar energy that can be converted into electricity.

In order to get the maximum power density at noon, the solar panels are tilted at an angle equal to the latitude of the location. In northern hemisphere, the panels are placed facing south, and in southern hemisphere facing north.

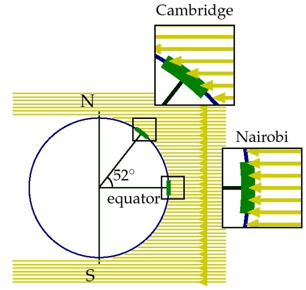


Figure 77: Angle of incidence varies along the latitude. So, in locations away from the equator, the solar panels are tilted to maximize the power received.

Insolation:

Insolation (comes from **IN**cident **SOL**ar radi**ATION**) of a place is the average solar power received per meter square area at that place.

If the average daily energy per m^2 received in a location is X kWh, then the insolation of that location is

$$\frac{X\,kWh}{m^2\,day}.$$

It actually quantifies average power square meter as the definition mentioned above as both the numerator and the denominator includes time.

Intensity of solar radiation varies as the day progresses with peak reached at noon (shown in the left of Figure 78). Intensity also varies round the year. Insolation is determined by taking all these variations into consideration.

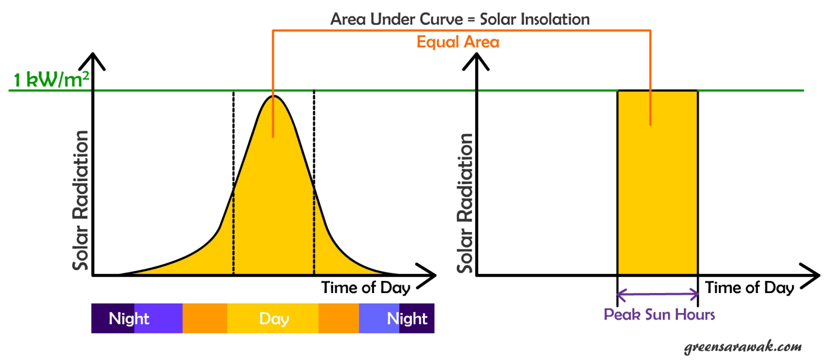


Figure 78: Insolation is the area under the curve. Peak Sun Hour of a location multiplied by 1 kW/m^2 is the average daily solar energy received per square metre area.Source of the image: greensarawak.com

Peak Sun Hour (PSH):

Peak Sun Hour (PSH) of any location is the number of hours required to get average daily energy received per square meter if a constant solar irradiance of 1 $\frac{kW}{m^2}$ would be available.

For example, a location that receives 8 $\frac{kWh}{m^2}$ daily can be said to have received 8 hours of solar energy per day at 1 $\frac{kW}{m^2}$.

In Singapore, the irradiation on a very sunny day can be as high as 6.7 $\frac{kWh}{m^2}$ while on a rainy day, it could be as low as 0.8 $\frac{kWh}{m^2}$.

The average annual irradiation of Singapore is 1663 $\frac{kWh}{m^2}$ and hence the average daily irradiation is

$$\frac{1663}{365}\,\frac{kWh}{m^2} \approx 4.56\,\frac{kWh}{m^2}.$$

So PSH for Singapore is

$$\frac{4.56\frac{kWh}{m^2}}{1\frac{kW}{m^2}} = 4.56 hours.$$

PSH is a useful parameter for calculating amount of solar energy received as the solar PV modules are often rated at an input irradiation of standard 1 $\frac{kW}{m^2}$.

Solar cell to solar array

Solar cell: Semiconductor device that converts sunlight into direct current (DC) electricity.
Solar module: PV modules consist of PV cell circuits sealed in an environmentally protective laminate and are the fundamental building block of PV systems.
Solar panel: Includes one or more PV modules assembled as a pre-wired, field-installable unit.
Solar array: A PV array is the complete power-generating unit, consisting of any number of PV modules and panels.

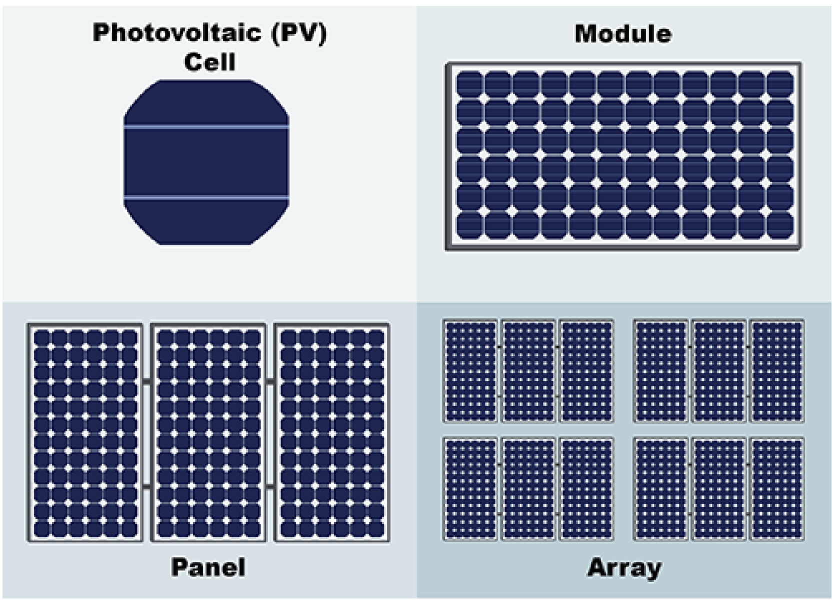


Figure 79: PV system: solar cell to PV array. Source: Florida Solar Energy Center http://www.fsec.ucf.edu/en/consumer/solar_electricity/basics/cells_modules_arrays.htm

Besides the energy conversion PV units, a PV system also require various other components. Some of these are briefly mentioned below.

- Battery: Storage of energy is a must have in a PV system. Energy stored during the day time can be used at night. Even during the day, the output power varies due to change in intensity of sunlight. Energy stoed in the battery can help to stabilize the system output.
- Charge controller: This subsystem controls the flow of energy into the battery while charging it.
- Power converters: For grid-connected PV array, an inverter is used to convert DC output of the PV into AC which is suitable for connecting to the grid. Even for stand-alone PV system, power converters are required as most appliances are designed for AC power.
- Mechanical mounting and cabling.

Conversion efficiency

Conversion efficiency of the solar modules may range from below 10% to about 20% depending on the technology used and the manufacturer of the module. Thin-film modules tend to provide the lowest output per square meter, poly-crystalline silicon modules are intermediate, and mono-crystaline silicon modules offer the highest efficiency.

Technology	Module Efficiency
Mono-crystalline Silicon	12.5-15%
Poly-crystalline Silicon	11-14%
Copper Indium Gallium Selenide (CIGS)	10-13%
Cadmium Telluride (CdTe)	9-12%
Amorphous Silicon (a-Si)	5-7%

Figure 80: Conversion efficiency of different PV technologies (Source: Handbook for Solar Photovoltaic (PV) System, Energy Market Authority, and Building & Construction Authority, Singapore)

Rated efficiencies are measured under the Standard Test Conditions (STC) where solar irradiation is maintained at its maximum value of 1 $\frac{kW}{m^2}$, temperature is kept at 25° C, and a few other constraints are maintained. Due to the variability of meteorological factors and technical performance of solar PV systems, under real conditions, efficiency is lower than the rated efficiency.

Nameplate power rating of PV panels

Nameplate rating of solar panels is given using the unit Watt-peak or Wp. It specifies the power the panel would produce under ideal conditions (radiation intensity of 1000 $\frac{W}{m^2}$, 25° C temperature, etc.). A panel with nameplate rating of 250 Wp will give 250 W power under ideal condition.

ELECTRICAL DATA @ STC	Product code*: RECxxxTP2M						
Nominal Power - P_{MAX} (Wp)	300	305	310	315	320	325	330
Watt Class Sorting - (W)	-0/+5	-0/+5	-0/+5	-0/+5	-0/+5	-0/+5	-0/+5
Nominal Power Voltage - V_{MPP} (V)	33.0	33.3	33.5	33.7	33.9	34.0	34.3
Nominal Power Current - I_{MPP} (A)	9.11	9.17	9.26	9.36	9.45	9.56	9.62
Open Circuit Voltage - V_{OC} (V)	38.3	38.8	39.1	39.6	40.0	40.3	40.8
Short Circuit Current - I_{SC} (A)	10.01	10.04	10.07	10.10	10.13	10.15	10.19
Panel Efficiency (%)	18.0	18.3	18.6	18.9	19.2	19.5	19.8

Values at standard test conditions (STC: air mass AM 1.5, irradiance 1000 W/m², temperature 25°C), based on a production spread with a tolerance of P_{MAX}, V_{OC} & I_{SC} ±3% within one watt class. At a low irradiance of 200 W/m² at least 95% of the STC module efficiency will be achieved.
*Where xxx indicates the nominal power class (P_{MAX}) at STC indicated above.

Figure 81: Electrical characteristics under standard testing condition (STC) shown in PV datasheet. (Source: recgroup.com)

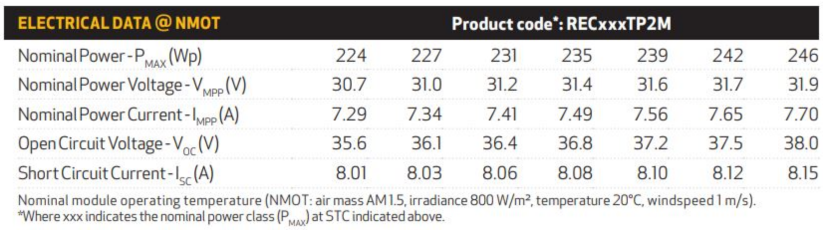
MECHANICAL DATA	
Dimensions:	1675 x 997 x 38 mm
Area:	1.67 m²
Weight:	18.5 kg

Figure 82: Mechanical dimensions of the panel whose electrical datasheet shown in Figure 81

Referring to the datasheet in Figure 81, the product mentioned in the first column has power output 300 W_p and efficiency 18% under standard testing condition. Surface area of the panel is 1.67 m^2 (Figure 82). The watt-peak rating is the power output of the panel is the input irradiance is 1000 $\frac{W}{m^2}$.

$$1000\frac{W}{m^2} \times 1.67\ m^2 \times 0.18 \approx 300\ W.$$

Some manufacturer also provide the electrical characteristics for nominal module operating temperature (NMOT). For the REC panel, the electrical characteristics @ NMOT is shown in Figure 83



ELECTRICAL DATA @ NMOT	Product code*: RECxxxTP2M						
Nominal Power - P_{MAX} (Wp)	224	227	231	235	239	242	246
Nominal Power Voltage - V_{MPP} (V)	30.7	31.0	31.2	31.4	31.6	31.7	31.9
Nominal Power Current - I_{MPP} (A)	7.29	7.34	7.41	7.49	7.56	7.65	7.70
Open Circuit Voltage - V_{OC} (V)	35.6	36.1	36.4	36.8	37.2	37.5	38.0
Short Circuit Current - I_{SC} (A)	8.01	8.03	8.06	8.08	8.10	8.12	8.15

Nominal module operating temperature (NMOT: air mass AM 1.5, irradiance 800 W/m², temperature 20°C, windspeed 1 m/s).
*Where xxx indicates the nominal power class (P_{MAX}) at STC indicated above.

Figure 83: Electrical characteristics under nominal module operating temperature (NMOT) shown in PV datasheet. (Source: recgroup.com)

PV sizing for stand-alone application

1. Estimate the energy demand of the application.

2. Find the peak sun hour (PSH) at the location of the application.

3. Determine the power required (P_{user}) using

$$P_{user} = \frac{Required\ energy}{PSH}.$$

4. Output of the PV system must be greater than the required power to compensate for power loss in various components in the system.

5. Efficiency (η_{sys}) of the system (excluding the energy conversion efficiency of PV panels) is typically in the range of 75% - 80%.The output of PV array should be

$$P_{array} = \frac{P_{user}}{\eta_{sys}}.$$

6. You may need multiple panels to meet the requirements.

Energy Storing Elements: Part I

First-order Systems

Energy flows around in every engineering system. For example, the car engine provides energy for it to move. The chemical potential energy of fuel is converted to kinetic energy of the car. Similarly any part of a machine stressed under loading will have elastic potential energy stored in it. You can find many other examples where energy conversion and energy storage takes place.

Rate of change of energy in a system is known as *power*. In a practical system, power is always finite. This means that the energy stored in any part of the system cannot be changed/transferred instantaneously. This property of energy often limits the rate of change of certain variables in a system,*e.g.*, the speed of a moving mass cannot be changed suddenly. Effect of changing force on the change of speed is not instantaneous, rather time-dependent; the relation between force and speed involves temporal integral.

This response rate of systems is of practical interest to engineers. For example, how much time is taken by a car from rest to full speed? How much time is required for a room to be cooled to the desired temperature after the air conditioner is turned ON? The response behavior is explained in this chapter using two energy storing elements in electrical systems: capacitor and inductor.

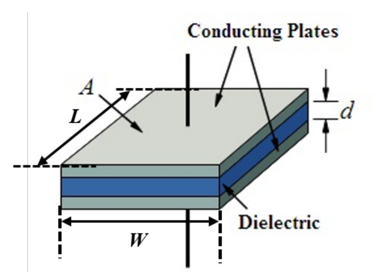


Figure 84: Capacitor is constructed by separating two sheets of conductor by a thin layer of insulator.

Capacitors

A capacitor is constructed by separating two sheets of conductors by a thin layer of insulating material. The insulating material is called *dielectric*. Air, paper, Mylar, polyester etc are commonly used insulators in capacitor.

When a DC voltage source is connected across a capacitor,

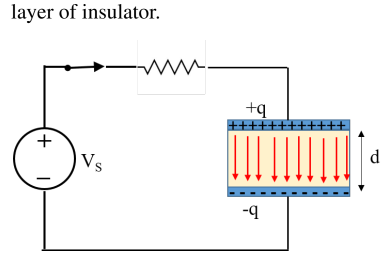


Figure 85: When a DC source is connected, an electric field is created in the dielectric. Energy is stored in the electric field.

- It transports the electric charge q onto the plate connected to the positive terminal, and simultaneously transports an equal amount of charge off the plate connected to the negative terminal.
- Although positive and negatives charges are separated, the net charge in

the capacitor is zero.

The charge q on the positive plate and its image charge $-q$ on the negative plate produce an electric field (E) within the dielectric material, and a voltage is developed between the terminals of the capacitor; the voltage V is proportional to the charge q.

Capacitance is a circuit parameter that represents the amount of charge required per unit voltage.

The unit of capacitance is **Farad**.

The electric field created by the separation of charge is

$$E = \frac{q}{\epsilon A},$$

where, A is the area of the capacitor plate and ϵ is the **dielectric constant** or **permittivity** of the insulating material.

The permittivity of free space is $\epsilon_0 = 8.85 \times 10^{-12}\ F/m$. The permittivity of other insulating material is $\epsilon = \epsilon_r \epsilon_0$, where ϵ_r is the relative permittivity of the material. **Relative permittivity is a material property.**

Voltage developed between the plates is

$$V = E \times d = \frac{qd}{\epsilon A},$$

where d is the separation between the two plates. Amount of charge per volt (by definition, the **capacitance**) is

$$C = \frac{q}{V} = \frac{\epsilon A}{d}. \tag{12}$$

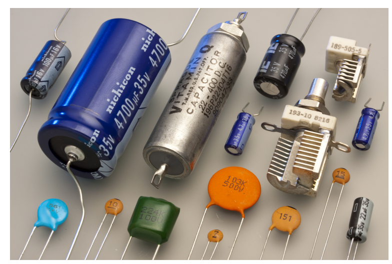


Figure 86: Capacitors with different capacitance values, different shapes and different dielectric materials are shown.

Capacitance depends on the overlapping area (A) of the plates, the distance (d) between the plates, and the permittivity (ϵ) of the dielectric material.

Most of the capacitors shown in Figure 86 have fixed capacitance value. Some applications (for example, radio tuner) require capacitors with variable capacitance. One such capacitor is shown in Figure 87.

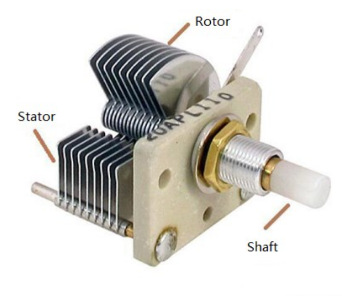


Figure 87: By turning the rotor, the overlapping area of the plates and, hence the resulting capacitance, can be varied. Point to note: the area A in the equation 12 is the overlapping area of the parallel plates. If we move one plate with respect to other in the direction parallel to the surface, the overlapping area is changed.

Linearity

Capacitor shows a linear relation between the terminal voltage (v) and the charge (q) on each plate

$$q = Cv.$$

Although we only consider linear and time-invariant (the capacitance doesn't change with time) capacitors in this text, we note that

- Some transducers such as electric microphones, and other electric sensors and actuators, are appropriately modeled with time-varying capacitors.
- Most capacitors used in electronic equipment (paper, mica, ceramic, etc.) are linear, but often vary by a small amount with temperature.
- There are practical examples of nonlinear capacitance. For example, the charge associated with a reverse-biased semiconductor diode varies as the 2/3 power of voltage, because the distance d between positive and negative charge is a function of voltage.

Relation between Voltage and Current

The rate at which the charge (q) is transported to or from the plate is capacitor current

$$i = \frac{dq}{dt}.$$

So the voltage-current relation for capacitor involves temporal derivative

$$i = \frac{dCv}{dt} = C\frac{dv}{dt}. \tag{13}$$

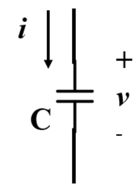


Figure 88: Circuit symbol of capacitor with positive current and corresponding voltage polarity shown.

Here it is assumed that the capacitance C remains constant as current flow into or out of the capacitor.

With DC source, when the capacitor voltage is in the steady-state (constant DC), the current becomes zero:

$$I_{ss} = C\frac{dV_{ss}}{dt} = 0.$$

Capacitor behaves like *open circuit* in steady-state with DC supply.

Instantaneous Change in Capacitor Voltage : Is it Possible?

If you want to change capacitor voltage from one value (v_1) to another (v_2) instantaneously, *i.e.*, in zero time duration, then

$$\frac{dv}{dt} = \lim_{\Delta t \to 0} \frac{v_2 - v_1}{\Delta t} = \infty.$$

That means infinite current will be required which is not practically possible. **Capacitor voltage can not be changed instantaneously.** It changes gradually (continuous) with finite gradient $\frac{dv}{dt}$.

Memory Property of Capacitor

$$C\frac{dv}{dt} = i$$

The expression for capacitor voltage $v(t)$ at a time t can be found by integrating the above,

$$v(t) = \frac{1}{C}\int_{-\infty}^{t} i(t)\, dt.$$

So the voltage across a capacitor depends on the entire past history of current, which is the essence of memory.

It is the memory property that allows the capacitor to be the primary memory element in all integrated circuits.

If the voltage $v(t_1)$ is known, it is necessary and sufficient to know the current for the interval $t_1 \le t \le t_2$ to determine the voltage $v(t_2)$ at time $t = t_2$,

$$\begin{aligned} v(t_2) &= \frac{1}{C}\int_{-\infty}^{t_2} i(t)\, dt \\ &= \frac{1}{C}\int_{-\infty}^{t_1} i(t)\, dt + \frac{1}{C}\int_{t_1}^{t_2} i(t)\, dt \\ &= v(t_1) + \frac{1}{C}\int_{t_1}^{t_2} i(t)\, dt. \end{aligned}$$

Energy Stored in Capacitor

Associated with the ability to exhibit memory is the property of energy storage, which is often exploited by circuits that process energy.

Have you observed that a charger light stays on for a little longer after the main switch is turned-off? This is due a capacitor inside the charger stores a little bit of energy and keeps the light on for the extra time after main supply is cut-off. Whenever a capacitor is connected to the battery, the plates will be oppositely charged until the battery voltage equals the capacitor voltage. Energy is stored in the electric field produced in the dielectric of a charged capacitor. The amount of energy stored is equal to the amount of work done by the battery in moving the charges to set up the electric field. Voltage represents the work per unit charge in moving it from the negative plate to the positive plate. If a voltage $v - C$ is produced by moving a small charge dq, then the amount of work done is

$$dw_{cap} = v_C dq. \tag{14}$$

If an initially uncharged capacitor (charge = 0) is charged to have Q on the positive plate causing voltage V across the capacitor, then the work done to achieve this is

$$\begin{aligned} W_{cap} &= \int_0^Q dw_{cap} = \int_0^Q v_C \, dq \\ &= \int_0^Q \frac{q}{C} \, dq \\ &= \frac{Q^2}{2C} \end{aligned}$$

But,

$$Q = CV,$$

and therefore, the energy stored in the capacitor (C) charged to a voltage V is

$$W_{cap} = \frac{1}{2} C V^2. \tag{15}$$

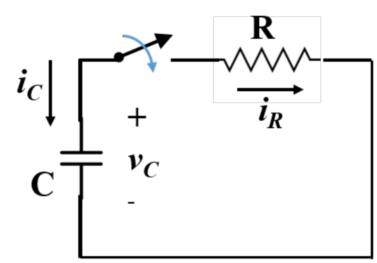


Figure 89: Capacitor discharging through a resistor. The direction of capacitor current and the polarity of capacitor voltage are chosen according to the convention shown in Figure 88 which means positive value of i_C gives positive rate of change in v_C.

Discharging of Charged Capacitor

Suppose a capacitor, charged to certain voltage V_0, is connected to a resistor R through a switch as shown in Figure 89. If the switch is closed at $t = 0$, then,

$$v_C(0) = V_0.$$

After the switch is closed, capacitor voltage will drive a current through the resistor like a battery would do but with a difference. As the capacitor releases energy to drive the current, its voltage is decreased (in technical terms, **capacitor is discharged**). Referring to the circuit of Figure 89 with the switch closed -

$$
\begin{aligned}
i_C + i_R &= 0, \\
C\frac{dv_C}{dt} + \frac{v_C}{R} &= 0.
\end{aligned}
$$

$$RC\frac{dv_C}{dt} + v_C = 0. \tag{16}$$

Discharging of an initially charged capacitor is described by a **homogeneous first-order ordinary differential equation** (ODE) (equation 16). Any system that can be described by a first-order ODE is known as a **First Order System**.

Reorganizing the first-order ODE,

$$\frac{dv_C}{dt} = -\frac{v_C}{RC},$$

$$\frac{dv_C}{v_C} = -\frac{1}{RC}dt.$$

By integration, we get

$$\ln v_C + K = -\frac{t}{RC}.$$

Therefore,

$$
\begin{aligned}
\ln v_C &= -\frac{t}{RC} - K \\
v_C &= e^{-K}e^{-t/RC}.
\end{aligned}
$$

For the initial condition

$$v_C(t=0) = V_0,$$

$$V_0 = e^{-K}.$$

Therefore, the capacitor voltage expressed as function of time is

$$v_C(t) = V_0 e^{-\frac{t}{RC}}. \tag{17}$$

The speed of discharging (also charging as you will find later) the capacitor is determined by the product of R and C. This is known as the **time constant**

The unit of time constant is second.

$$\tau = RC.$$

- Can you show using conformity of dimensions that the unit of time constant is second?

Capacitor voltages during discharging are shown in Figure 90 for different values of τ but all starting at the same initial voltage.

Time constant is a unique parameter of all 1st-order system (electrical, mechanical, chemical, thermal etc.) that determines the speed of response of the system.

- Larger the time constant, slower is the discharging of the capacitor.

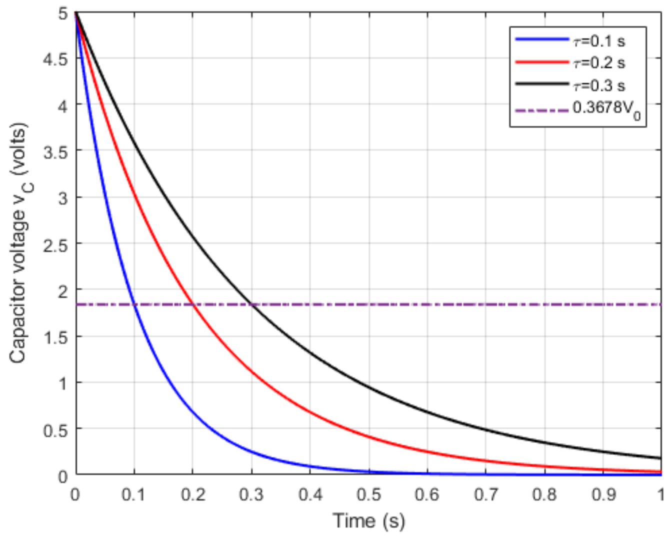


Figure 90: Discharging of a capacitor - initially capacitor is charged to 5 V. (a) blue line: $R = 1\,k\Omega, C = 100\,\mu F, \tau = 0.1\,s$ (b) blue line: $R = 2\,k\Omega, C = 100\,\mu F, \tau = 0.2\,s$ (a) blue line: $R = 3\,k\Omega, C = 100\,\mu F, \tau = 0.3\,s$. Capacitor voltage is reduced to $0.3678 \times 5\,V \approx 1.8\,V$ at $t = \tau$ for all three cases.

- At time equal to one time-constant ($t = \tau$), capacitor voltage drops to

$$v_C(t = \tau) = V_0 e^{-1} = 0.3678 V_0.$$

- At time equal to five times the time-constant ($t = 5\tau$), the capacitor voltage is so small that it is assumed to be fully discharged.

$$v_C(t = 5\tau) = V_0 e^{-5} = 0.0067 V_0.$$

Charging of a Capacitor

For charging of a capacitor, it should be connected to a voltage source (Figure 91). After the switch is closed at $t = 0$, the KVL around the loop gives

$$Ri + v_C = V_S$$
$$RC\frac{dv_C}{dt} + v_C = V_S.$$

This is also a 1st-order ODE but not homogeneous. The final solution is given below without going into the details.

$$v_C(t) = V_0 e^{-\frac{t}{\tau}} + V_S(1 - e^{-\frac{t}{\tau}}). \tag{18}$$

$\tau = RC$ is the time constant.

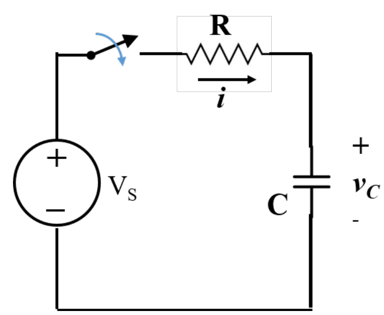


Figure 91: Charging of a capacitor.

Some Applications

- **Camera Flash Light:** The flash is a light source of very high intensity for a short duration. This requires a source that can supply large power for a short duration. The battery is capable of supplying smaller power for a longer duration. However, using a special circuit, a capacitor can be

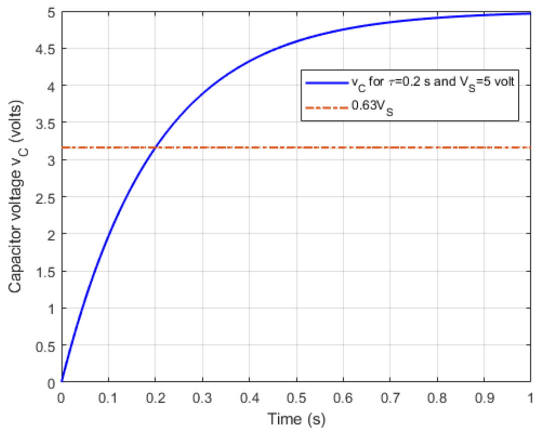


Figure 92: Charging of a capacitor: initial voltage of the capacitor is $V_0 = 0$ V. After the switch is closed at $t = 0$, the capacitor voltage gradually increases towards the steady-state voltage which is equal to the source voltage ($V_S = 5\ V$). The rate of change of capacitor voltage decreases with increasing time. At $t = \tau$, the capacitor voltage becomes $(1 - 0.3678)V_S \approx 3.1610\ V$.

charged to a few hundred volts and accumulate some energy, which can then be discharged through the flash light at a very high rate.

- **Battery charger:** Have you noticed that the indicator light of a charger stays ON a bit longer after you have turned off the charger input supply? This is due to the presence of a capacitor at the output of the charger. The capacitor stores some energy and supplies to the LED indicator even after input is cut-off. This property of capacitor also helps in removing the ripples in the DC voltage output.

Inductor

Inductor is a circuit element that works on the magnetic effect of electric current.

When current passes through a wire, a magnetic field is produced in the surrounding space. The magnetic field is represented by flux lines. **Magnetic flux density** is the amount of magnetic flux passing through unit area:

$$B = \frac{\phi}{A},$$

where, ϕ is the magnetic flux and A is the area through which the flux passes.

Often, the wire is made into a coil form to increase the strength of the magnetic field for certain current. The **flux-linkage** Ψ of the coil is flux (ϕ) multiplied by the number of turns (N) of the coil,

$$\Psi = N\phi.$$

The flux linkage is proportional to the current flowing through the coil, *i.e.*, $\Psi \propto i$ or

$$\Psi = Li. \tag{19}$$

The constant of proportionality L is known as the **inductance**. **The unit of inductance is Henry (H).**

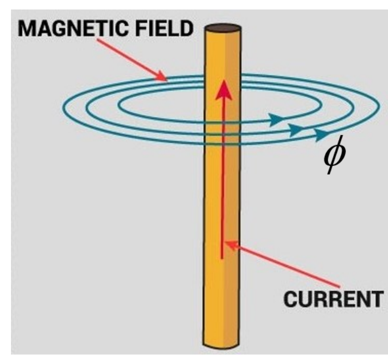


Figure 93: Electric current creates a magnetic field in the surrounding space.

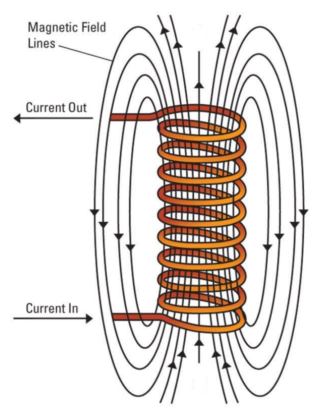


Figure 94: Magnetic field produced by a coil is much stronger than the field produced by a wire for the same magnitude of current.

Faraday's Law

When a coil is placed in a time-varying magnetic field, a voltage is induced in it.

The magnitude of the induced voltage is equal to the rate of change of flux linkage of the coil:

$$v_L = \frac{d\Psi}{dt}. \tag{20}$$

Element Law for Inductor

Combining Equations 19 and 20,

$$\begin{aligned} v_L &= \frac{dLi}{dt}, & (21)\\ &= L\frac{di}{dt}. & (22)\end{aligned}$$

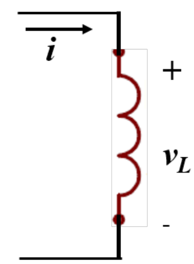


Figure 95: Inductor voltage is proportional to the rate of change of inductor current.

If a constant current flows through the inductor then

$$\frac{di}{dt} = 0$$

and, therefore, the inductor voltage is zero. Hence, in DC steady-state analysis, we treat the inductor as a short circuit.

Consider the circuit shown in Figure 96.The switch was left open for long time before closing it at $t = 0$. Then

$$i = 0 \quad for \quad t < 0.$$

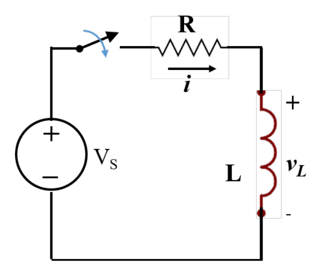


Figure 96: RL circuit

Inductor current cannot be changed instantaneously. So, the current remains at 0 immediately after the switch is closed. After the switch is closed, the KVL equation:

$$\begin{aligned} V_S &= Ri + L\frac{di}{dt},\\ \frac{L}{R}\frac{di}{dt} + i &= \frac{V_S}{L}.\end{aligned}$$

The solution of this first-order ODE with initial condition $i(0) = 0$ is

$$i(t) = \frac{V_S}{R}(1 - e^{-\frac{t}{\tau}}), \tag{23}$$

where, $\tau = \frac{L}{R}$ is the **time constant** of the RL circuit.

- If the switch is closed for long enough time (five times the time constant or more), the current will reach the steady state, *i.e.*, a DC value.
- In steady-state,

$$v_L = L\frac{di}{dt} = 0,$$

and, hence, $v_R = V_S$ and the steady state current is $i_{ss} = \frac{V_S}{R}$.

- So the current in the RL circuit can be expressed as

$$i(t) = I_{ss}(1 - e^{-\frac{t}{\tau}}).$$

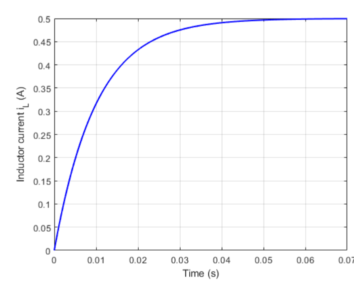


Figure 97: Inductor current versus time graph for $V_S = 5\,V$, $R = 10\,\Omega$ and $L = 1\,mH$. Inductor current cannot be changed instantaneously. After closing the switch, the current increases gradually and approaches the steady-state value $\frac{V_S}{R} = 0.5\,V$.

Energy Stored in Inductor

The instantaneous power of an inductor

$$\begin{aligned} p_L &= v_L i_L \\ &= L\frac{di_L}{dt} i_L. \end{aligned} \tag{24}$$

Energy required to increase the inductor current from 0 to a steady-state value of I can be found by integrating the instantaneous power

$$\begin{aligned} W_L = \int p_L dt &= L\int_0^I i_L di_L, \\ &= \frac{1}{2}LI^2. \end{aligned} \tag{25}$$

Conservation of Flux Linkage

Suppose you have connected a battery to an inductor. A current will flow so that flux-linkage increases with a slope equal to the applied voltage -

$$V_{bat} = L\frac{di_L}{dt}.$$

What will happen if you disconnect the coil from the battery. As the current drops from some finite current to 0 (the loop is broken by disconnecting the coil from battery), rate of change of flux linkage is

$$v_L = L\frac{di_L}{dt} = -\infty.$$

In practice, when you disconnect the wire as the flux-linkage collapses to zero a large voltage will be produced between the two points where the circuit is broken. The air around the two points will breakdown causing a flash arc. The flash arc is a current. Due to conservation of flux-linkage, this arc will try to keep the flux-linkage constant.

For a DC current, if the current is sufficiently large this arc current will continue to flow. It is dangerous to break DC current. If done, it will produce a flash arc that will continue to flow. If this arc is produced in an environment where there are inflammable gasses, it will cause a fire.
Watch this Youtube video https://www.youtube.com/watch?v=Zez2r1RPpWY.

Since every current produces a flux-linkage and inductance is the circuit parameter used to represent the flux-linkage, we can say that every closed current path (circuit) when broken or opened will potentially produce a large voltage across the points where the circuit is broken. This can lead to an arc and cause fire. So one has to be very careful when breaking or opening an electrical circuit.

If you are using an electric kettle and while the water is being heated, you turn off the switch, you will notice a greenish arc near the switch. This can be a serious problem if you have a cooking gas leak in kitchen. Hence during a suspected gas leak, one should not turn-on or turn-off any electrical appliance. The same can be observed with electric iron.

Applications of Inductor

- **Spark plug in internal combustion engine (car engine):** An ignition system generates a spark to ignite a fuel-air mixture in spark ignition internal combustion engines in petrol cars. First, a current is built up in the ignition coil (an inductor) storing magnetic field energy. When this current is interrupted, a large voltage appears across the electrodes of the spark plug, connected to the induction coil. With this high voltage across the two electrodes, the air becomes ionized and an electrically conductive channel is developed leading to the spark. The magnetic energy in the coil is converted to the spark which is light and heat energy.

  Watch this Youtube video on how spark plug works: https://www.youtube.com/watch?v=OMLSNwQiiKg

- **Inductor as traffic light sensor:** Most traffic lights in Singapore have a sensor embedded on the road (the black loopy lines) which is basically a coil of wires. A vehicle near the traffic light is positioned above this coil. The inductance of the coil increases due to the steel of the vehicle body. The change in inductance is picked by a circuit to indicate the vehicle's presence. The traffic lights are then changed accordingly to facilitate smooth traffic flow.

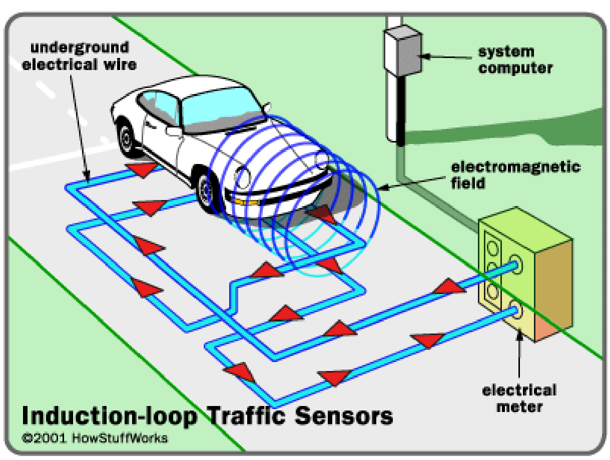


Figure 98: Inductive traffic sensor

Energy Storing Elements: Part II

(a) Electrical-Mechanical Analogues and (b) Second-order Systems

In the previous chapter, you learnt how energy can be stored in a capacitor and an inductor. Similarly, spring and mass store potential and kinetic energy, respectively. In this chapter, we shall understand the analogy between the energy storage elements in electrical and mechanical systems in the context of exchange of energy between two elements.

Inductor and Capacitor

- Inductor:

1. An inductor is made using a wire wrapped around a magnetic material.
2. Energy is stored in the magnetic field produced by current.
3. The energy is proportional to the square of current,

$$W_L = \frac{1}{2}Li^2. \tag{26}$$

Current is the rate of flow of charge.

- Capacitor:

1. A capacitor is formed by using two parallel conducting plates separated by a dielectric material.
2. It stores energy in the electric field.
3. The energy is proportional to the square of voltage

$$W_C = \frac{1}{2}Cv^2. \tag{27}$$

4. Voltage of a capacitor is also a measure of the charge ($q = Cv$).

Mass and Spring

- **Ideal mass:**
  - The ideal mass is rigid.
  - Its motion is unaffected by friction or other damping force.
  - Motion is governed by Newton's law:

    $$F = ma = m\frac{dv}{dt} = m\frac{d^2x}{dt^2},$$

    where F is net force acting on the mass (m). And a, v and x are its acceleration, velocity and displacement, respectively.
  - Kinetic energy proportional to the **square of velocity**,

    $$KE = \frac{1}{2}mv^2, \tag{28}$$

    and velocity is the rate of change of displacement.

- **Ideal spring:**
  - Ideal spring has no mass or internal damping.
  - Hooke's law $F_{SP} = -kx$ is valid for small and non-distorting displacement.
  - Positive value of x produces negative restoring force F_{SP}.
  - Spring's equilibrium position is $x = 0$ when net force is zero.
  - Whenever a spring is compressed or extended by x (displacement from the neutral position), potential energy is stored in the spring proportional to the **square of displacement**.

    $$PE = \frac{1}{2}kx^2. \tag{29}$$

Mechanical-Electrical Analogue

Displacement and charge are two physical variables in two different fields - one mechanical and the other electrical. They don't have same dimensions. But mechanical-electrical analogue helps us to visualize the similarities between an electrical system and a mechanical system, both with energy storing elements.

Displacement-Charge Analogue

- The rate of change of displacement x is velocity, $v = \frac{dx}{dt}$.
- The rate of change of charge Q is current $I = \frac{dQ}{dt}$.

- The second derivative of displacement is acceleration, $a = \frac{dv}{dt}$ which is related to the net force (effort) by $F = ma$.
- There is no special name for the rate of change of current, but it is related to the voltage (effort) in case of an inductor according to $V_L = L\frac{dI}{dt}$.

Force acts on mass to accelerate it and inertia resists the change. Similarly, the voltage acts as an effort to increase the current and the inductance L resists this change.

Kinetic Energy and Magnetic Energy

We can say the inductance in an electrical circuit is analogous to mass in a mechanical system. If a mass (m) is moving at a velocity v, the kinetic energy in the mass is

$$KE = \frac{1}{2}mv^2.$$

The energy stored in the inductor L due motion of charge (rate of charge flow is current I) is

$$W_L = \frac{1}{2}LI^2.$$

Both mass and inductor store energy in the kinetic form.

Spring-Capacitor Analogue

A spring stores energy when it is compressed or expanded; energy is in the form of potential energy. A capacitor also stores energy in the potential form *i.e.*, electrostatic potential energy due to the charge.

In a spring, the displacement x (expansion or contraction) in the spring due to an effort (force F) is resisted by the stiffness k of the spring,

$$x = \frac{F}{k},$$

where k is the stiffness of the spring.

For a capacitor, voltage V is the effort that tries to change Q.

$$Q = CV = \frac{V}{1/C}.$$

Therefore, $\frac{1}{C}$ is analogue of the stiffness (k) in the mechanical system.

The mechanical-electrical analogue is summarized in Table 11.

Spring-Mass Oscillation

A spring stores potential energy and a moving mass possesses kinetic energy. When these two elements interact, energy is transferred between them.

Consider an ideal spring (spring constant k) hanging from a rigid structure. If a block of mass m is attached to the lower end of the spring, the spring will be stretched downward and an upward restoring force will be produced.

Table 11: Mechanical-Electrical Analoque

	Mechanical	Electrical
Kinematics		
Fundamental variable	Displacement x $[m]$	Charge q $[As]$
First derivative	Velocity $v = \frac{dx}{dt}$ $[m/s]$	Current $I = \frac{dQ}{dt}$ $[A]$
Second derivative	Acceleration $\frac{dv}{dt}$ $[m/s^2]$	$\frac{dI}{dt}$ $[A/s]$
Law in Physics	Newton's Law: $m\frac{dv}{dt} = F$	Faraday's Law: $L\frac{dI}{dt} = V$
Dynamics		
Effort	Force $[N]$	Voltage $[Volt]$
Resistance to effort:		
Kinetic	Mass m	Inductance L
Potential	Stiffness k	Inverse of capacitance $\frac{1}{C}$

For the illustration shown in Figure 99,

$$\begin{aligned} k\Delta L &= mg \\ \Delta L &= \frac{mg}{k}. \end{aligned}$$

If the system is not disturbed, it will remain in this position. However, if the system is perturbed by moving the mass upward/ downward and then the perturbing force is removed, the mass will start oscillating.

Let's assign a variable y to the displacement of the mass in the vertical direction and let it be positive in the upward direction with respect to the equilibrium position $y = 0$ and negative in the downward direction (Figure 100).

When y is non-zero, the net force is

$$F_{net} = k(\Delta L - y) - mg = k\Delta L - ky - mg.$$

But $k\Delta L = mg$ and so $F_{net} = -ky$. Applying Newton's law,

$$\begin{aligned} m\frac{d^2y}{dt^2} &= -ky \\ \frac{d^2y}{dt^2} + \frac{k}{m}y &= 0. \end{aligned}$$

We are not interested in the analytical solution of the ODE. Interested students may refer to the Appendix for the complete solution. To convince yourself, you may verify its correctness by substituting the solution into the ODE.

If the initial displacement (caused by perturbation) is $y(0) = A$ and the initial velocity is $\frac{dy}{dt}(0) = 0$, then the solution to this **homogeneous second-order ordinary differential equation** (ODE) gives

$$y(t) = A\cos(\omega_n t). \tag{30}$$

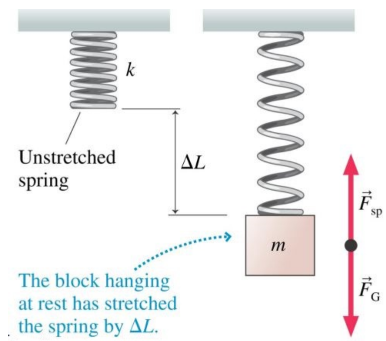


Figure 99: Spring-mass system will be in static equilibrium if the net force acting on the mass is zero. Two forces acting on the mass are (a) force of gravity $F_G = mg$ and

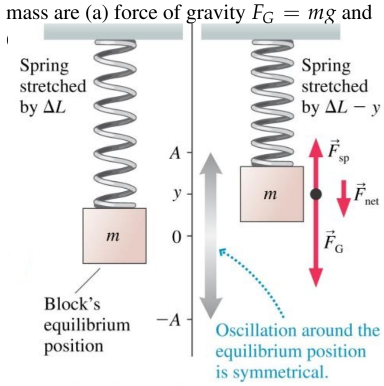


Figure 100: If the system is perturbed and then the perturbing effort is removed, the net force is non-zero and the mass accelerates in the direction of net force. Source: http://www.physics.umd.edu/courses/Phys260/agashe/S10/notes/lecture2

The parameter

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

is known as the **natural frequency** of the spring-mass system. The spring-mass will show a **simple harmonic oscillation** with frequency of oscillation ω_n rad/s.

Example 1

Consider a 10 kg mass hung from a rigid structure using a spring of stiffness $k = 4000\frac{N}{m}$. The natural frequency is

$$\omega_n = \sqrt{\frac{4000\ N/m}{10\ kg}} = 20\ rad/s.$$

If the mass is pushed upward by 2 cm from the equilibrium position and then released, it will start oscillating with a frequency of 20 rad/s. The displacement and velocity expressed as function of time are

$$y(t) = 2\cos(20t),$$

$$v(t) = \frac{dy(t)}{dt} = -40\sin(20t).$$

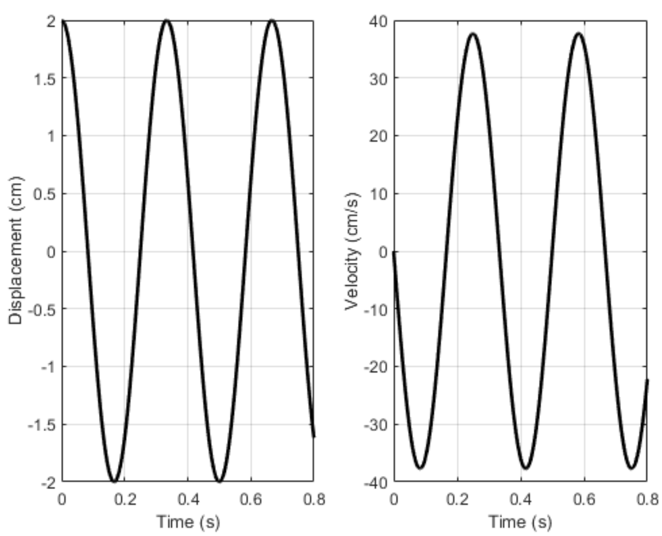
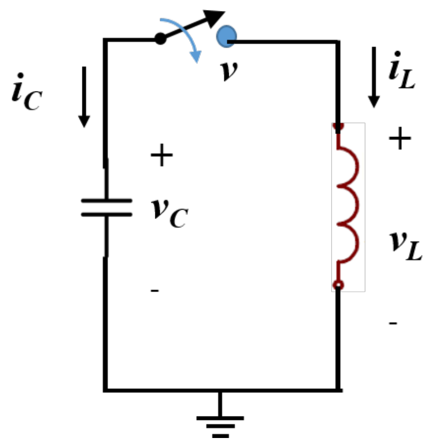


Figure 101: Displacement and velocity of a perturbed spring mass system in Example 1. Initial displacement $y(0) = 2\ cm$ and velocity $v(0) = 0\ cm/s$. Note that when the mass moves towards the equilibrium ($y0$) position, magnitude of velocity increases. That means, when the potential energy of the spring decreases, the kinetic energy of the mass increases. Similarly, when the potential energy of the spring increases (moving away from the equilibrium), the kinetic energy of the mass decreases. Energy is transferred periodically from spring to mass and vice versa.

Can you plot the graphs of potential energy versus time and kinetic energy versus time in this system?

Oscillation in L-C Circuit

Similar to the spring-mass system, an electrical circuit with L and C shows oscillation. In this case, energy is transferred periodically between electrostatic form in capacitor and electromagnetic form in inductor. Consider the circuit below with a capacitor and an inductor connected in parallel:



Note: If you take an uncharged capacitor and an inductor off the shelf and connect them, there will be no oscillation, as there is no energy in either of them. You need to perturb the system from the trivial solution, *i.e.*, $v_C = 0$ and $i_L = 0$, by starting with an initially charged capacitor. You put a switch (as shown in the figure) between them and close it at $t = 0$.

- The capacitor is charged to a specific voltage, *i.e.*, $v_C(0) = V_0$.
- Before the switch is closed, there is no current and, therefore, $i_C = i_L = 0$ for $t < 0$.
- Inductor current cannot be changed instantaneously; current will be zero immediately after the switch is closed. So the initial condition for the derivative of v_C
$$\frac{dv_C}{dt}(0) = \frac{i_C(0)}{C} = 0.$$
- Indicated direction of current and voltage polarity for inductor and capacitor are same as those used when we defined the element laws in the previous chapter. Therefore,
$$\begin{aligned} i_C &= C\frac{dv_C}{dt}, \\ i_L &= \frac{1}{L}\int_{-\infty}^{t} v_L dt. \end{aligned}$$
- After the switch is closed at $t = 0$
$$\begin{aligned} i_C &= -i_L, \quad (KCL) \\ v_C &= v_L. \end{aligned}$$

$$\begin{aligned} C\frac{dv_C}{dt} &= -\frac{1}{L}\int_{-\infty}^{t} v_C dt, \\ C\frac{d^2v_C}{dt^2} &= -\frac{1}{L}v_C, \end{aligned}$$

$$\frac{d^2v_C}{dt^2} + \frac{1}{LC}v_C = 0. \tag{31}$$

Refer to the Appendix for the analytical solution of this homogeneous ODE.

The solution of this second-order linear constant-coefficient homogeneous ODE with initial conditions

$$v_c(0) = V_0, \quad \frac{dv_C}{dt}(0) = 0$$

is

$$v_C(t) = V_0 \cos(\omega_n t), \tag{32}$$

where ω_n is the *natural frequency* of oscillation of the LC circuit and is given by

$$\omega_n = \sqrt{\frac{1}{LC}}. \tag{33}$$

Can you show using mechanical-electrical analogue that the formulae for the natural frequency are analogous?

Example 2

For the LC circuit shown earlier, let L = 100 μH and C = 1 μF. The capacitor is initially charged to 5 V prior to $t = 0$, and the switch is closed at $t = 0$.

The natural frequency (Equation 33) is

$$\omega_n = \sqrt{\frac{1}{(100 \times 10^{-6}) \times (1 \times 10^{-6})}} = 10^5 \; rad/s.$$

Initial conditions:

$$v_C(0) = 5 \; V, \quad \frac{dv_C}{dt}(0) = 0.$$

So the capacitor voltage (Equation 32) :

$$v_C(t) = 5\cos(10^5 t) \quad V,$$

and the inductor current

$$\begin{aligned} i_L(t) &= \frac{1}{L}\int v_L(t)dt \\ &= \frac{1}{100 \times 10^{-6}}\int 5\cos(10^5 t) \\ &= 0.5\sin(10^5 t) \quad A. \end{aligned}$$

Figure 102 shows the graphs of capacitor voltage versus time and inductor current versus time.

Energy Interpretation of LC Oscillation

There is an important energy interpretation to Figure 102. The oscillations in i_L and v_C carry out a repetitive exchange of energy between the inductor and the capacitor. The state of each element drives the growth of the other at the expense of the energy it stores. The energy stored in the capacitor

$$W_C(t) = \frac{1}{2}C(v_C(t))^2,$$

the energy stored in the inductor

$$W_L(t) = \frac{1}{2}L(i_L(t))^2,$$

and the total energy

$$W_T(t) = W_C(t) + W_L(t)$$

versus time are shown in Figure 103.

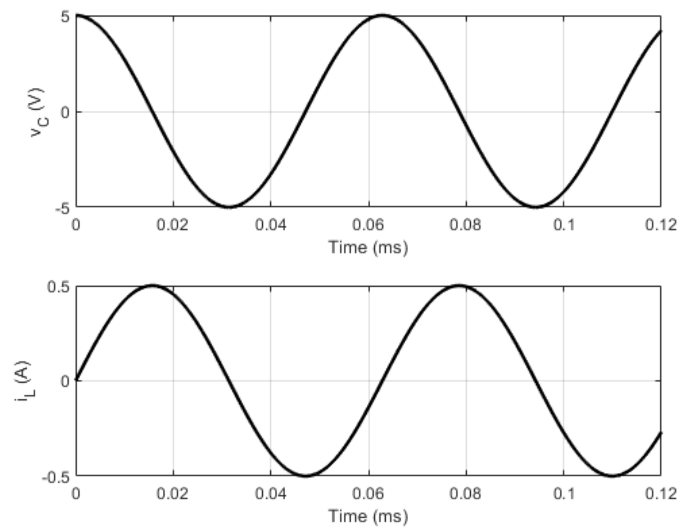


Figure 102: Immediately after the switch is closed, capacitor voltage starts to decrease while inductor current starts to increase. This continues until the capacitor voltage becomes zero implying that capacitor has no more energy stored. Then the inductor current starts charging the capacitor but with opposite polarity, and continues till inductor current becomes zero and capacitor voltage reaches the maximum negative. This oscillation will continue for ever as there is no loss of energy in ideal capacitor and ideal inductor.

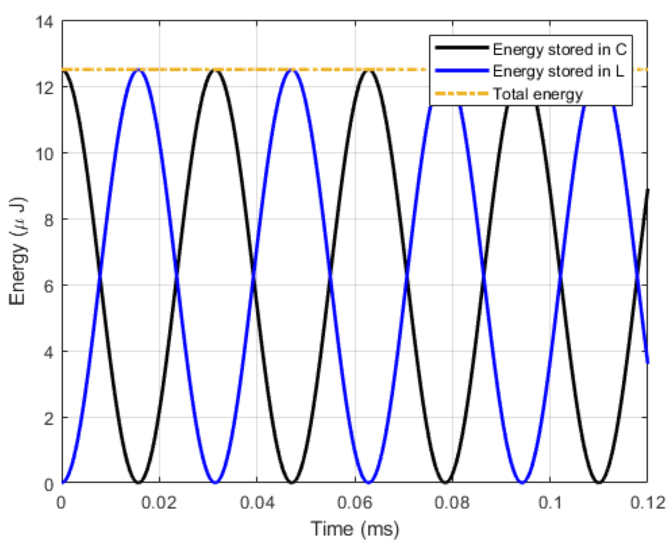


Figure 103: At t = 0, the energy stored in the capacitor is $\frac{1}{2}C(V_0)^2 = 12.5\mu J$ and the energy stored in the inductor is $\frac{1}{2}L(i(0))^2 = 0$. As there is no dissipating element in the circuit, energy is transferred back and forth between C and L. As there is no dissipative element in the circuit, total energy remains the same which is equal to $12.5\mu J$, *i.e.*, the energy stored in the capacitor at $t = 0$.

Damped Oscillation

While analyzing the oscillation in the previous section, no energy dissipating element is considered. All elements are assumed ideal. So the total energy in the system remains same and the energy oscillates forever between two elements. In reality, friction, drag etc. cause dissipation of energy in mechanical system, and energy is lost in resistive elements or in non-ideal inductor and capacitor in an electrical system. In this section, oscillation is explained when energy-dissipating elements are present.

Spring-Mass-Damper

Consider the mechanical system (Figure 104) with a body of mass m hung by a spring of stiffness coefficient k and a damper that resists the motion with a effort proportional to the velocity.

Dampers are widely used in mechanical systems, *e.g*, in car suspension.

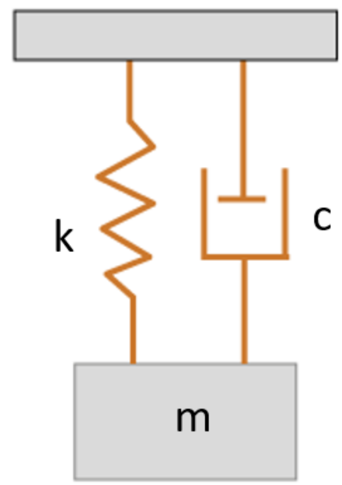


Figure 104: Spring-mass-damper system

In static equilibrium, when the mass is not moving, the force exerted by the damper is 0. So the force of gravity is cancelled by the spring force

$$mg = k\Delta L.$$

If the mass is perturbed from the static equilibrium, it will be set in motion. As the mass moves, a force (F_d) is exerted by the damper which is proportional to the velocity. The net force acting on the mass will be

$$F_{net} = -ky - c\frac{dy}{dt},$$

where y is the displacement with respect to the equilibrium position, k is the stiffness of the spring, and c is the damping coefficient. Force due to gravity is cancelled by spring force $k\Delta L$. Applying Newton's law,

$$m\frac{d^2y}{dt^2} = -ky - c\frac{dy}{dt}$$

$$\frac{d^2y}{dt^2} + \frac{c}{m}\frac{dy}{dt} + \frac{k}{m}y = 0. \tag{34}$$

The analytical solution of this second-order ODE is not our objective. Instead, the system will be explained through numerical solution for a system with the following parameters:

You can solve differential equations numerically using Python, MATLAB or many other computational tools.

$$m = 1\,kg, \quad k = 100\,\frac{N}{m}, \quad c = 2\frac{N}{m/s}.$$

Assume that the mass was hanging at an equilibrium position before being pushed upward by 2 cm or 0.02 m. Then the mass is released at $t = 0$. Initial conditions are

$$y(0) = 0.02\,m, \quad v(0) = \frac{dy}{dt}(0) = 0.$$

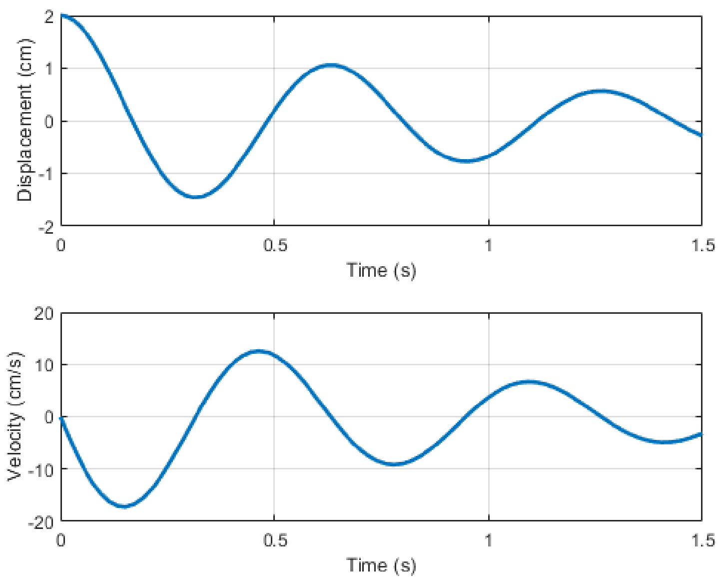


Figure 105: Displacement and velocity versus time for a spring-mass-damper with $m = 1\ kg$, $k = 100\ N/m$, $c = 2\ N/m/s$, and initial conditions $y = 2\ cm$, $v = 0$

Moving the mass upward by 2 cm increased the potential energy stored in the spring by

$$\frac{1}{2}k(y(0))^2 = \frac{1}{2} \times (100\ N/m)(0.02\ m)^2 = 20\ mJ.$$

After it is released, the potential energy (PE) in the spring is converted into kinetic energy (KE) causing the velocity of the mass to increase. When the PE is reduced to zero, the KE is at its maximum, however, the maximum kinetic energy is less than the potential energy (20 mJ) initially stored in the spring. This is because energy is also dissipated through the damper. The graphs in Figure 107 show PE and KE versus time.

Natural frequency, Damping ratio and Damped natural frequency

Instead of giving the analytical solution to equation 34, the final expression for the displacement as a function of time will be given. Before that, two important parameters that characterize the oscillation are defined.

1. **Natural frequency:**

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

This is the frequency of undamped oscillation, *i.e.*, when $c = 0$.

2. **Damping ratio:**

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

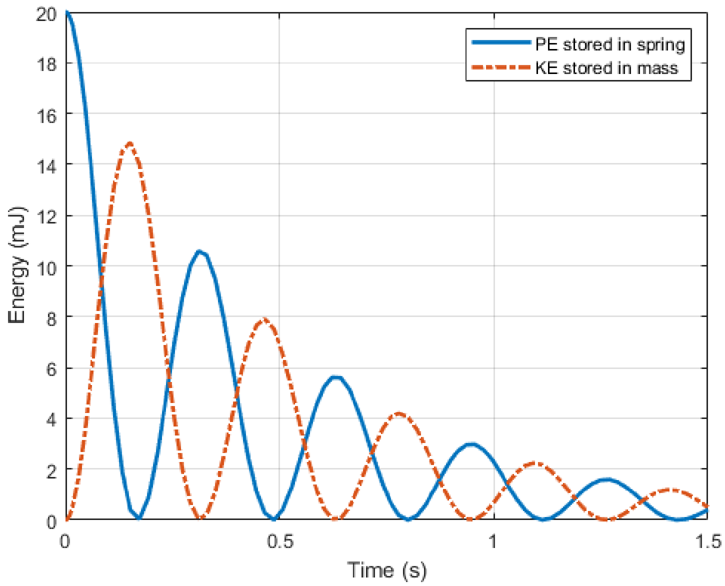


Figure 106: Energy versus time in spring-mass-damper system

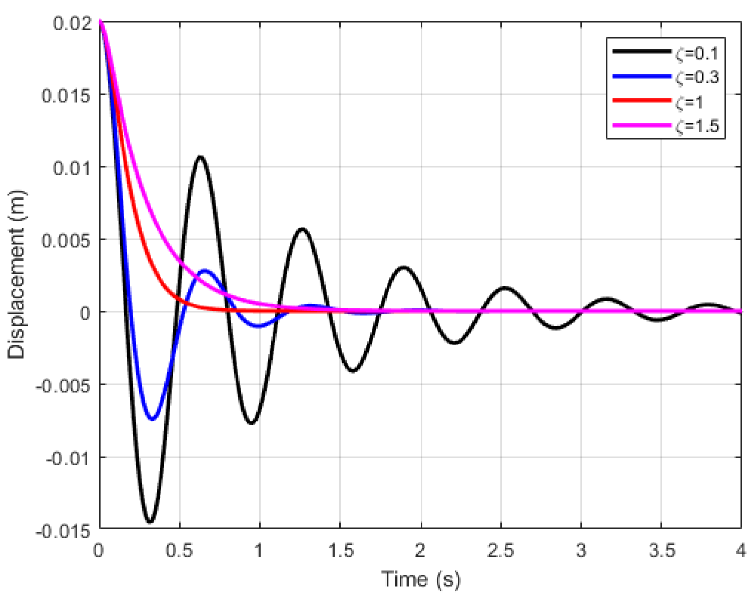


Figure 107: The displacement versus time graphs for spring-mass-damper system having the same natural frequency but different values of damping ration.

The value of ζ determines whether the system will oscillate and, if oscillates, how long it takes to dampen the oscillation. The displacements $y(t)$ for different values of ζ but same value of ω_n are shown in Figure 107.

- **Under-damped response** ($0 < \zeta < 1$): the system is oscillatory with decreasing amplitude of oscillation.
- **Critically damped response** ($\zeta = \cancel{0}$ 1): no oscillation - the mass exponentially goes to the equilibrium position.
- **Over damped response** ($\zeta > 1$): no oscillation - the mass goes exponentially to the equilibrium but the response takes longer than in the critically damped case.

Let's rewrite the equation 34 by replacing the component parameters (k, m and c) with the parameters (ω_n and ζ) characterizing the oscillation:

$$\frac{d^2y}{dt^2} + 2\zeta\omega_n\frac{dy}{dt} + \omega_n^2 y = 0. \tag{35}$$

RLC Circuit

To illustrate the phenomenon of damped oscillation in electrical circuits, we consider a series RLC circuit, *i.e.*, resistor, inductor and capacitor connected in series. There is no source, but the capacitor is charged externally prior to the closure of the switch at $t = 0$. We shall call the circuit in Figure 108 *undriven series RLC circuit.*

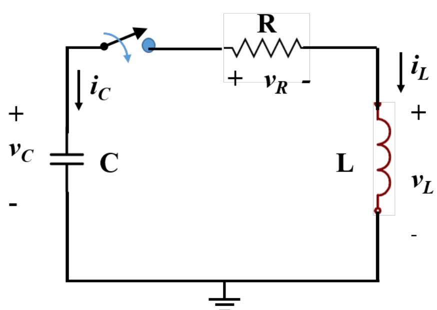


Figure 108: Undriven series RLC circuit

- Series RLC circuit driven by a voltage source is explained later in this section.
- There is a possibility of connecting the RLC circuit in parallel. More about series and parallel RLC circuits will be covered in EPP II.

The initial condition for capacitor voltage is

$$v_C(0) = V_0.$$

After the switch is closed, the energy stored in the capacitor will try to drive a current through the circuit but inductor current cannot be changed instantaneously. So $i_L(0) = i_C(0) = 0$, and the current will change gradually. Initial condition for the derivative of v_C:

$$\frac{dv_C}{dt}(0) = \frac{i_C(0)}{C} = 0.$$

Applying KVL and the element laws for resistor and inductor,

$$v_C = v_R + v_L,$$

$$v_C = Ri_L + L\frac{di_L}{dt}.$$

However, $i_C = -i_L$ and therefore,

$$v_C = -Ri_C - L\frac{di_C}{dt}.$$

Using $i_C = C\frac{dv_C}{dt}$,

$$v_C = -RC\frac{dv_C}{dt} - LC\frac{d^2v_C}{dt^2},$$

$$\frac{d^2v_C}{dt^2} + \frac{R}{L}\frac{dv_C}{dt} + \frac{1}{LC}v_C = 0. \tag{36}$$

Let's express this equation in terms of parameters associated with oscillation, *i.e.*, natural frequency and damping ratio. Comparing equation 36 with equation 35,

1. **Natural frequency** (ω_n):

$$\omega_n^2 = \frac{1}{LC} \quad \Rightarrow \quad \omega_n = \frac{1}{\sqrt{LC}}.$$

This is the frequency of undamped oscillation, *i.e.*, when $R = 0$ and the circuit becomes an LC circuit.

2. **Damping ratio:**

$$2\zeta\omega_n = \frac{R}{L} \quad \Rightarrow \quad \zeta = \frac{R\sqrt{LC}}{2L},$$

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

Rewriting the ODE of equation 36 in terms of the parameters ω_n and ζ,

$$\frac{d^2v_C}{dt^2} + 2\zeta\omega_n\frac{dv_C}{dt} + \omega_n^2 v_C = 0. \tag{37}$$

The solution of this ODE for under-damped case ($\zeta < 1$) is

$$v_C(t) = e^{-\zeta\omega_n t}\frac{V_0}{\sqrt{1-\zeta^2}}\cos(\sqrt{1-\zeta^2}\omega_n t - \phi), \tag{38}$$

Finding the solution of the second-order ODE is not a learning outcome of this module. Interested students may refer to the Appendix.

where

$$\tan\phi = \frac{\zeta}{\sqrt{1-\zeta^2}}.$$

Point to note: The frequency

$$\omega_d = \sqrt{1-\zeta^2}\omega_n$$

is known as *damped natural frequency*, which is smaller than the natural frequency, *i.e.*,

$$\omega_d < \omega_n.$$

Damped natural frequency (ω_d) is smaller than the natural frequency (ω_n) of the corresponding loss-less system.

You may use mechanical-electrical analogues to determine the natural frequency and damping factor of the spring-mass-damper system from those of RLC, and vice versa.

$$\begin{array}{rcl} m-s-d & \leftrightarrow & RLC \\ m & \leftrightarrow & L \\ k & \leftrightarrow & \frac{1}{C} \\ c & \leftrightarrow & R \end{array}$$

$$\omega_{n,RLC} = \frac{1}{\sqrt{LC}} \quad \leftrightarrow \quad \frac{1}{\sqrt{m(1/k)}} = \sqrt{\frac{k}{m}} = \omega_{n,msd}$$

$$\zeta_{RLC} = \frac{R}{2}\sqrt{\frac{C}{L}} \quad \leftrightarrow \quad \frac{c}{2}\sqrt{\frac{1/k}{m}} = \frac{c}{2\sqrt{km}} = \zeta_{msd}.$$

How does a System Fail?

Mechanical Failure

When an engineering system cannot fulfil the intended purpose, we say that the system has failed. The beams and columns of a building are meant to support the dead load (the weight of the building itself) and the live loads. When they fail, they cannot do so and the building or part of it collapses. An audio amplifier is an electronic system intended to amplify audio signals. If it doesn't do that, it is said to have failed.

The factors to be considered in analyzing whether a structure element would fail or not:

1. The force or torque acting on the structural element,

2. The dimensions and geometry of the element, and

3. Strength of the material used to make the element.

Figure 109: The intended purpose of the rod in this case is to keep the bulb hanging from the ceiling.

A structure may consist of many components of different shapes and sizes, *e.g.*, cylinder, push rod, cam, and gear in a car engine. But they can be categorized in to a few standard shapes, *e.g.*, massive body, plate, bar, rod, and string.

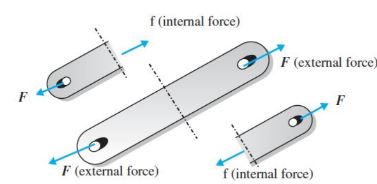


Figure 110: Internal force (Source: Saeed Moaveni, *Engineering Fundamentals - An Introduction to Engineering*)

Internal Force & Reaction Force

When a structural component is subjected to an external force, *internal forces* are created inside the material to hold the material and the components together. And *reaction forces* are developed at the supporting boundaries to keep the object held in position as planned. Reaction force will be explained later in details when we discuss bending of beams.

Stress

Stress provides a measure of the intensity of internal forces acting over an area.

Unit of stress is $\frac{N}{m^2}$ which is also known as **Pascal** or *Pa*.

Consider the part of a structure shown in Figure 111 where the force is applied at an angle. The tendency of the horizontal component of the force is to shear the plate, and the tendency of the vertical component is to compress the plate.

The ratio of the normal (vertical) component of the force to the area is called the **normal stress**, and the ratio of the horizontal component of the force (the component of the force that is parallel to the plate surface) to the area is called the **shear stress**.

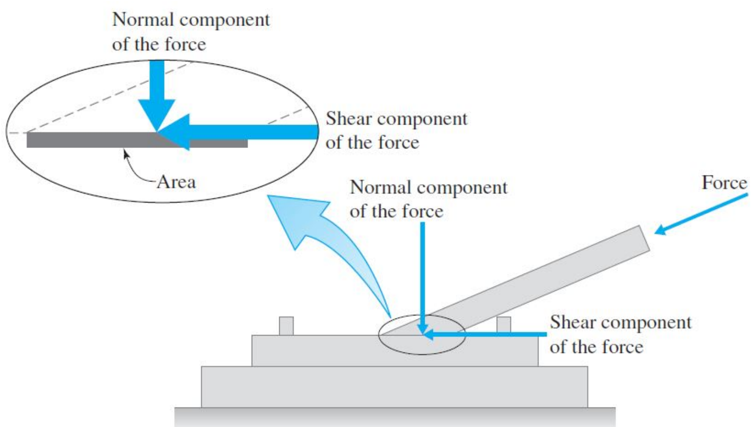
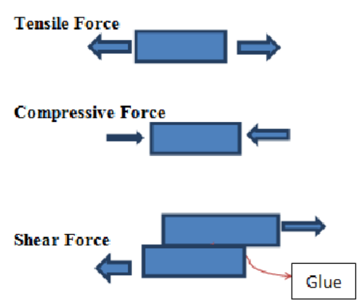


Figure 111: Normal stress and shear stress (Source: Saeed Moaveni, *Engineering Fundamentals - An Introduction to Engineering*)

Normal stress can be

- **Compressive** as shown in Figure 111 where the tendency of the force compresses the object, or
- **Tensile** where the tendency of the force is to pull the component in the axial direction. The rod in Figure 109 is under tensile stress.



When the applied force has a tendency to bend the object, **bending stress** is developed inside the object. Bending stress is the result of simultaneous presence of both tensile and compressive stresses.

When an object is being squeezed from all sides, like a submarine in the depths of an ocean, the stress developed is called **bulk stress** (or volume stress).

Strain

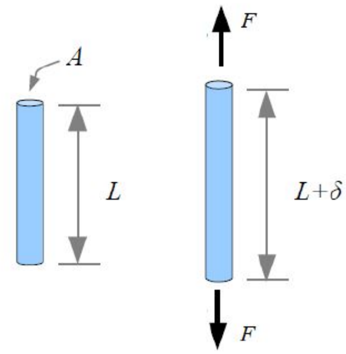
An object under stress becomes deformed. The quantity that describes this deformation is called **strain**.

Strain is given as a fractional change in either length (under tensile stress) or volume (under bulk stress) or geometry (under shear stress). Therefore, strain is a dimensionless number. Strain under a tensile stress is called tensile strain, strain under bulk stress is called bulk strain (or volume strain), and that caused by shear stress is called shear strain.

Let a rod of length L $[m]$ and area of cross-section A $[m^2]$ be pulled by force F $[N]$ from both ends. As a result, the rod will be elongated by δ $[m]$. The normal stress (σ) and normal strain (ϵ) are

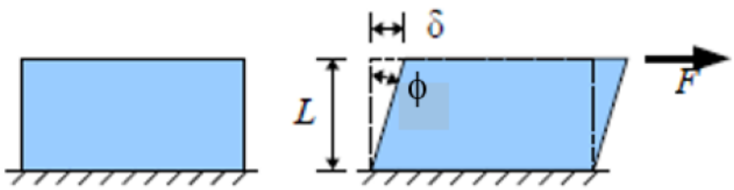
With a compressive stress, the strain is negative

$$\sigma = \frac{F}{A}[\frac{N}{m^2}], \quad \epsilon = \frac{\delta}{L}.$$



Consider a rectangular block loaded in shear (shown below). The top end is moved by an amount δ due to the shear force. Shear strain is the ratio of the distortion δ to the length L perpendicular to the distortion, i.e.,

$$\gamma = \frac{\delta}{L}.$$



Stress-Strain Graph

The greater the stress, the greater is the strain. However, the relation between them is not linear. The strain is proportional to the stress only when the stress is sufficiently low,

$$shear \propto strain \quad \Longrightarrow \quad shear = constant \times strain.$$

For normal stress, the proportionality constant is known as the **Young's Modulus** (E):

Young modulus tells us the stiffness of the material

$$\sigma = E\epsilon.$$

For shear stress, the constant is called the **Shear Modulus** (G). If τ and γ are shear stress and shear strain, respectively,

$$\tau = E\gamma.$$

The stress-versus-strain graph of a ductile material *e.g.*, iron, is shown in Figure 112.

- For stress smaller than the point marked as **Yield Strength**, the deformation is elastic. If the stress is removed, the object would go back to its original size.
- Once the yield point is passed, some fraction of the deformation will be permanent and non-reversible.
- Ultimate strength corresponds to the maximum stress that can be sustained by the structure in tension. The ultimate tensile strength is often shortened as *tensile strength*. At this point, a ductile material experiences *necking* where the cross-sectional area reduces locally.

- The **fracture point** is the point where the material physically breaks.

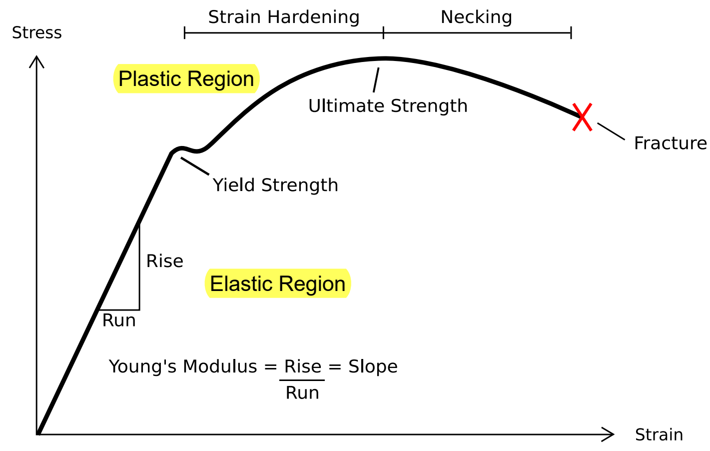


Figure 112: Stress-strain curve of a ductile material (Source: commons.wikimedia.org). The yield strengths vary from 35 MPa for a low-strength aluminum to greater than 1400 MPa for very high-strength steels. Ultimate tensile strengths vary from 50 MPa for an aluminum to as high as 3000 MPa for very high-strength steels.

This curve is for ductile materials

For brittle materials, the ultimate strength is very close tot the yield strength.

For a brittle material, necking does not occur. The breaking occurs at ultimate strength.

Note that the relation between stress and strain is an observed relation, measured in the laboratory.

Example 1:

Suppose an iron rod with circular cross-section is used to hang an object from the ceiling. The area of cross-section is $0.8 \times 10^{-4}\ m^2$. The rod is made of cast iron (Yield strength $\sigma_Y = 70\ MPa$ and Ultimate strength $\sigma_U = 100\ MPa$).

1. What is the maximum mass that would not cause permanent deformation of the rod?

2. What is the maximum mass that would ensure that the stress is smaller than σ_U?

Solution:

If the mass is $m\ kg$ then the working normal stress is

$$\sigma_W = \frac{m \times 9.8}{0.8 \times 10^{-4}} = (122{,}500 \times m)\ Pa.$$

(1)

$$122{,}500 \times m < \sigma_U \implies m < 571\ kg.$$

(1)

$$122{,}500 \times m < \sigma_Y \implies m < 816\ kg.$$

Failure due to Transverse Loading

In the example of the hanging bulb, the force of gravity acts in the direction along the length of the rod. This is called **axial loading**.

But, in case of a quad-copter arm, the force exerted by the propeller acts perpendicular to the direction along the length of the arm. This is called **transverse loading**.

The quad-copter arm is a type of structural element known as **beam**, which is long, slender object with length (L), breadth (b) and thickness or width (h) such that

$$L >> b, \quad L >> h.$$

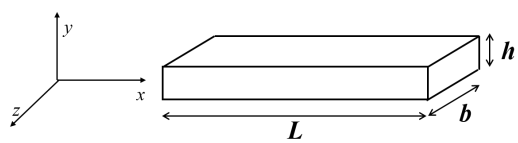


Figure 113 shows all possible internal forces and moments at a cross-section of a beam subjected to transverse load oriented in the direction of y-axis. All of them need not be present in any particular case.

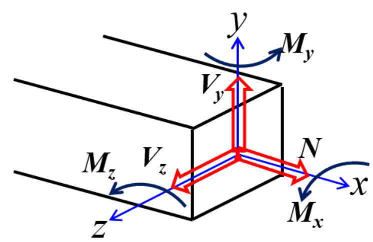


Figure 113: Possible internal forces and moments at any cross-section of a beam

- Force (N) acting in the direction of the x-axis is perpendicular to the cross-section. This is the **Normal force** (also seen under axial load).
- Froce (V_y & V_z) oriented in the direction of y-axis and z-axis, respectively, are parallel to the plane of the cross-section. These forces are the **Shear forces**. If the forces is applied vertically, V_y will be present and not V_z.
- Three moments (M_x, M_y and M_z) are also shown. M_x is present when the element is twisted. If it is bent vertically (upward or downward), then M_z is present. When bent forward/backward, then M_y is present.

Why is the normal force produced under transverse loading? The transverse force tends to bend the element which makes one side elongated (under tension) and the other side shortened (under compression) as shown in Figure 114. So, compression and tension (normal forces) occur concurrently. Length is unaffected at the center of the beam; the corresponding surface is known as the *neutral surface*.

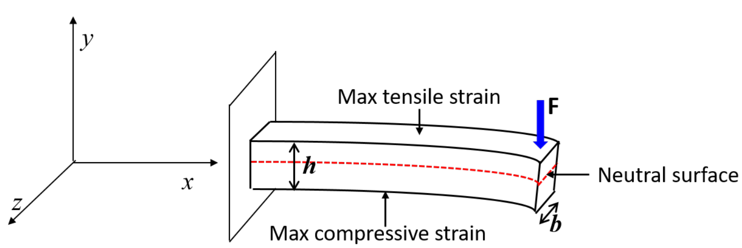


Figure 114: This is a cantilever beam - one end is fixed to a rigid structure and the other end is free. A downward force acting at the free end tends to bend the beam downward. For this case, the top half will experience tensile strain and the bottom half will experience compressive strain, with zero strain at the neutral surface. Internal forces V_y and N (both compressive and tensile) will be produced, but not V_x. Internal moment will be M_z only which, effectively, is the result of the normal forces.

If the beam is homogeneous, the strain varies linearly as function of y. It is 0 at the neutral surface ($y = 0$), $+\epsilon_{max}$ at $y = \frac{h}{2}$ and $-\epsilon_{max}$ at $y = -\frac{h}{2}$, where h is the height of the beam. That means the normal force too will vary linearly as function of y.

At neutral surface, the 2 forces cancel out and largest force experienced occurs at the the 2 surfaces of the material

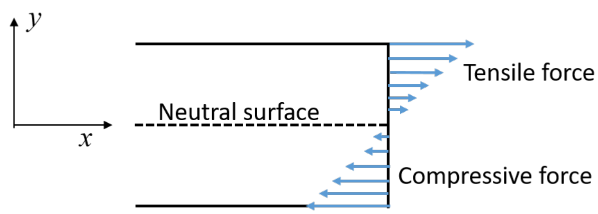


Figure 115: Distribution of tensile and compressive forces along a cross-section of the beam. For simplicity, a two-dimesional drawing is used. At the top end, the tensile strain is maximum ($+\epsilon_{max}$) and, therefore, the tensile force is maximum. Tensile strain and the tensile force decrease with decreasing value of y and becomes zero at the neutral surface. Similarly, at the bottom surface, the compressive strain is maximum ($-\epsilon_{max}$) and, hence, the compressive force is also maximum.

The resultant tensile force and the resultant compressive force form a couple and its moment is the **bending moment**. It is an internal moment (as opposed to an applied moment). This is shown as M_z in Figure 113. As the force is applied vertically, it will not produce internal force V_z and internal moments M_x and M_y. We can use a two dimensional drawing for the analysis.

In summary, the effect of transverse loading can be analyzed using

1. the shear force (internal force), and
2. the bending moment (internal moment).

Steps to determine shear force and bending moment

1. Draw the free body diagram (FBD) of the beam showing all applied forces & moments, and reaction forces & moments at the support.
2. Use equilibrium condition to determine reaction forces and moments.
3. Take an imaginary cut at the cross-section of the beam where you want to find the shear force and the bending moment. This will split the beam into two imaginary halves. Use equilibrium in one of the halves to determine the shear force and the bending moment.

The step #3 above can be simplified as follows:

- Shear forces are developed in the material to balance externally applied forces in order to secure equilibrium of all parts of the beam.
  - **The shear force at any cross-section of a beam may be found by summing all the vertical forces to the left or to the right of the section under consideration.** Sum of these vertical forces is balanced by the shear force
- Bending moments are developed in the material to balance the rotating tendency of external forces.

- **The bending moment at any section of a beam may be found by adding the moments from the left or from the right of the section considered. The moment's pivot point is the location under consideration.**

These steps will be illustrated using an example. But before that, reaction at the support needs to be explained. It is done using two common types of support.

There are other types of support but only these two types are considered here.

(a) Reaction forces at support

External forces or moments will cause deformation in a structural element only if the movement of the element is restrained. (*If the element is free-to-move then the applied force will make it move without deformation.*)

1. Think of a cantilever structure, *e.g.*, the shop sign shown here. You cannot move it in any direction (resists movement along all three axes). It also resists rotation. Such a structural joint is called **fixed support**. A fixed support restrain movement of the element by producing reaction forces and moments in all directions.

2. Now think of a stapler - it's two parts are connected in a way that you can rotate one with respect to another about one particular axis. The support doesn't offer any resistance to this rotation; technically speaking, there is no reaction moment. If you make one part stationary, the other part cannot be moved in any direction. That is to say, the support produces reaction forces along all three axes. This type of structural connection is called **pin support**

Example 2:

A setup for bend testing of PCB is shown in Figure 116. This test is known as *three-point bending test*.

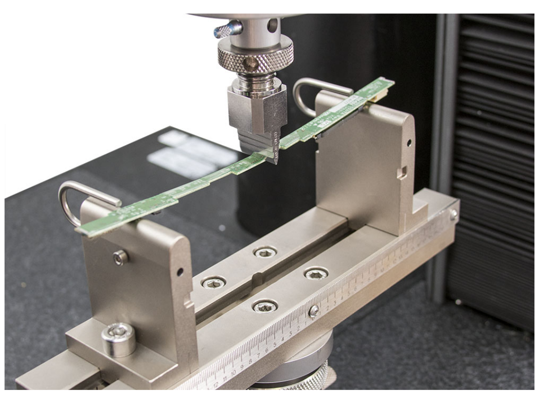


Figure 116: Setup for bend testing of PCB. The test sample (PCB) is simply supported to two ends and force is applied in the middle. Source of image: instron.us.

Assuming that the weight of the PCB is negligibly small compared to the force applied, the free-body diagram of the beam (PCB) is shown below.

Note that the beam's movement is not restrained in the horizontal direction (x and z axes), and it is free to rotate at the supports. So reaction forces are created in the y-axis direction only to counteract the applied force. No reaction moment is produced.
Considering equilibrium of forces and moments,

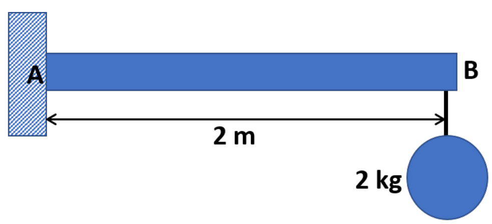
$$\Sigma F_y = 0 \quad \Rightarrow \quad R_1 + R_2 = F_{app},$$
$$\Sigma M_{at\ A} = 0 \quad \Rightarrow \quad R_1 \frac{L}{2} = R_2 \frac{L}{2}.$$

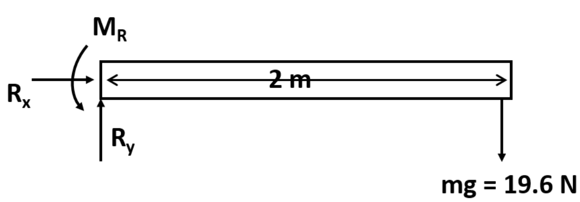
Therefore,

$$R_1 = R_2 = \frac{F_{app}}{2}.$$

Example 3:

A 2 kg mass is hung from the free end of a 2 m long cantilever beam. The beam and its free body diagram are shown below using two-dimensional drawings (it is assumed that no force is acting along the z-axis). Since the cantilever beam uses fixed support, two reaction forces and one reaction moment are generated at the support. The support resists movement in both x and y directions, and also resists rotation.





Considering equilibrium of forces and moments,

$$\Sigma F_x = 0 \quad \Rightarrow \quad R_x = 0,$$
$$\Sigma F_y = 0 \quad \Rightarrow \quad R_y = mg = 2 \times 9.8 = 19.6\ N,$$
$$\Sigma M_{at\ A} = 0 \quad \Rightarrow \quad M_R = 19.6 \times 2 = 39.2\ Nm.$$

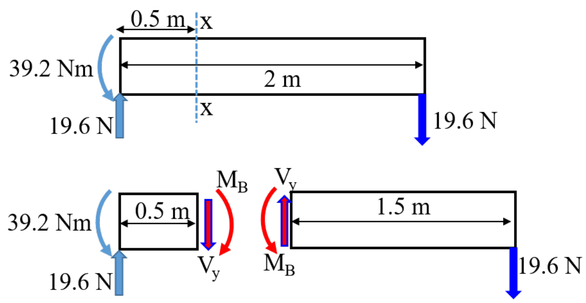
In the example above, the directions are shown using arrow for forces and moments and magnitudes are calculated. Sign convention of forces and moments will be introduced later in the note.

Example 4:

For the same cantilever beam problem, let's determine the shear force and the bending moment at a distance of 0.5 m from the fixed end.

Solution:

Make an imaginary cut (shown as x-x in the figure below) at a distance of 0.5 m from the fixed end.



There is only one vertical force, the support reaction (19.6 N) to the left of the cut x-x. The shear force at the surface is

$$V_y = 19.6\ N$$

pointing downward to balance the support reaction. You can get the same result considering the half to the right of the cut, but the shear force will be pointing upward there.

Similarly, the moment due to the forces and moment to the left of the cut is

$$39.2 - 19.6 \times 0.5 = 29.4\ Nm$$

in the counterclockwise direction which is resisted by the bending moment. So the bending moment at this cut will be 29.4 Nm clockwise.

You can also use the other section (1.5 m long) to find the shear force and bending moment at the imaginary cut.

Points to Note:

1. In engineering, we are concerned about the maximum values of the shear force and the bending moment because, if the maximum shear force (or bending moment) does not lead to failure, smaller values will definitely be safe. The maximum bending moment and the maximum shear force can be determined by sketching the **bending moment diagram** (BMD) or **shear force diagram** (SFD), which will be illustrated soon.

2. In the example above, logical argument is used to decide the direction for shear force and bending moment. For sketching SFD and BMD, we need to use +ve and -ve sign to distinguish between opposite directions. The sign convention for these forces and moments are explained in the next sub-section.

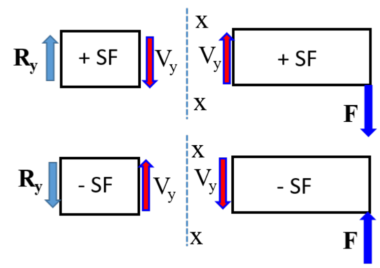
Sign convention for internal forces and moments

- Normal force: Tension is taken as positive and compression as negative.

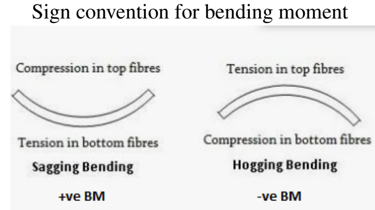
- Shear force: For the coordinate system used, V_y pointing downward on the surface left to the imaginary cut is positive and V_y pointing upward on the surface right of the cut is positive. According to this, the V_y shown in Example 3 is positive.

- Bending moment is positive if it tends to bend the beam with upward curvature (smiley face) - it is called sagging beam. If the beam bends with downward curvature (hogging beam), then the bending moment is taken as negative. Bending moment in Example 3 is negative.

Sign convention for shear force



Sign convention for bending moment



Shear Force Diagram (SFD) and Bending Moment Diagram (BMD)

The shear force diagram (SFD) is a graph showing the variation of shear force ($V_y(x)$) along the length of a beam. And the bending moment diagram (BMD) is a graph showing the variation of bending moment ($M_B(x)$) along the length of the beam.

You can find $V_y(x)$ and $M_B(x)$ as function of x by varying the distance x of the imaginary cut from $x = 0$ to $x = L$, where L is the length of the beam.

Example 5:

Sketch the SFD and BMD for the beam in Example 2.

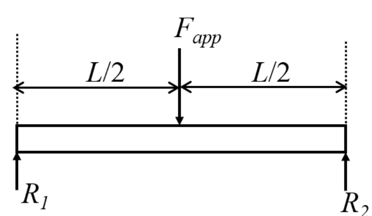
- Downward force at the mid-point between the supports: F_{app} newton

- Length of the beam, L meter.

Solution:

- Determine the reaction forces and moments. As found in Example 2, there is no reaction moment and reaction forces are

$$R_1 = R_2 = \frac{F_{app}}{2}.$$

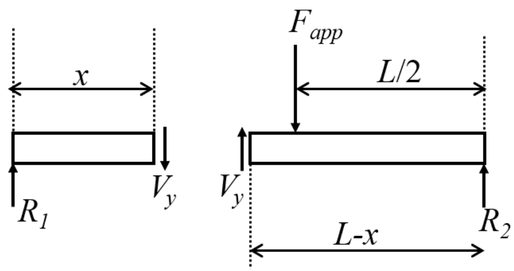


- **Shear Force Diagram (SFD):**

1. Make an imaginary cut at a distance x from the left end between the support and the applied point force.

$$\Sigma F_y = 0 \quad \Rightarrow \quad V_y(x) = R_1 = \frac{F_{app}}{2}.$$

According to the sign convention, this shear force is positive.



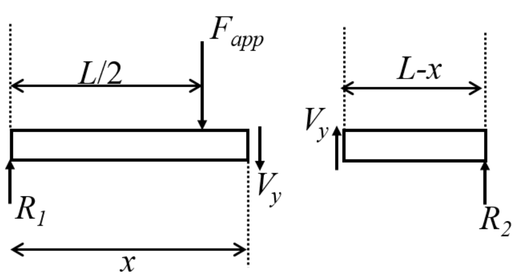
You should get the same result if you consider the right side section to determine the shear force.

This FBD is incomplete as it does not satisfy the equilibrium of moments. The internal moment, known as the bending moment, is not shown as the shear force can be determined by considering the equilibrium of forces only.

2. As there is no vertical force in the region $0 < x < \frac{L}{2}$, the above equilibrium holds for this range of x, and the shear force is

$$V(x) = +\frac{F_{app}}{2}, \quad 0 < x < \frac{L}{2}.$$

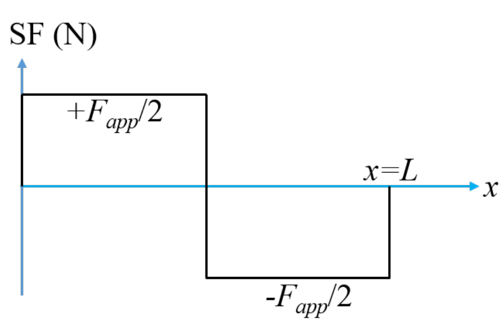
3. Make another imaginary cut in the range of $\frac{L}{2} < x < L$ and sketch the FBD showing the all forces including the shear force.



The shear force in this half of the beam is

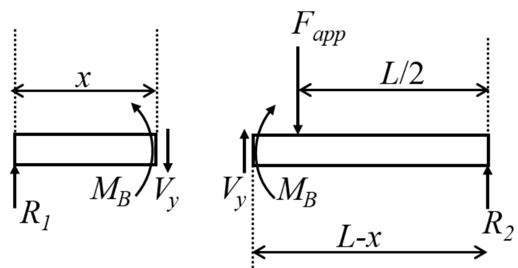
$$V(x) = -\frac{F_{app}}{2}, \quad \frac{L}{2} < x < L.$$

4. The shear force is changed abruptly at at two points, $x = \frac{L}{2}$ and $x = L$. The SFD is shown below.



- **Bending Moment Diagram (BMD):**

1. Consider an imaginary cut anywhere in $0 < x < \frac{L}{2}$ but, this time, include the internal moment is shown that prevents the two sections from rotating.



Bending moment in the middle and therefore the stress in the middle is the largest

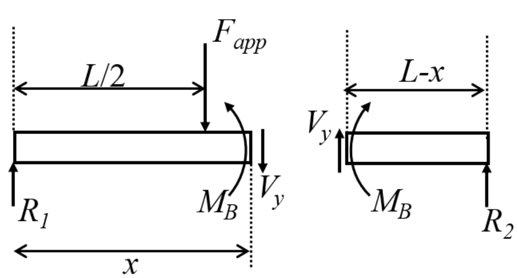
2. The forces R_1 and V_y form a couple and the resulting clockwise moment is balanced by the counter-clockwise *bending moment* M_B in the section left of the imaginary cut:

$$M_B = R_1 x = \frac{F_{app}}{2} x, \quad 0 < x < \frac{L}{2}.$$

According to the sign convention, this bending moment is positive. So for $0 < x < \frac{L}{2}$, bending moment is positive and increases with increasing value of x. At $x = L/2$,

$$M_B(L/2) = \frac{F_{app} L}{4}.$$

3. Now take the imaginary cut anywhere between $x = L/2$ and $x = L$.



Considering the equilibrium of moment in the section to the left of the cut

$$M_B = V_y x + F_{app} \frac{L}{2}.$$

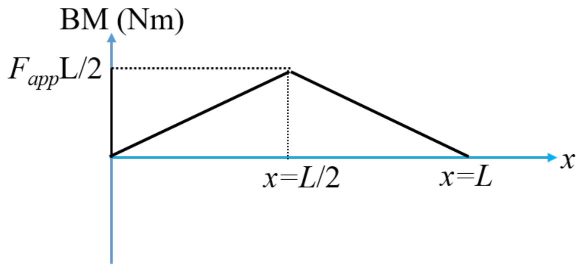
But

$$V_y = R_1 - F_{app} = -\frac{F_{app}}{2}.$$

So,

$$M_B = \frac{F_{app}}{2}(L - x), \quad \frac{L}{2} < x < L.$$

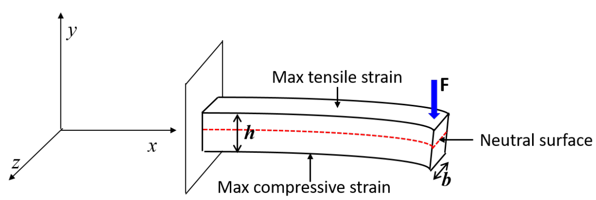
The resulting BMD diagram is shown below.



Relation between bending moment and bending stress

Shear force or bending moment is the internal reaction of the structural element in response to external forces and moments. However, failure occurs when the stress in the element exceeds the ultimate strength (σ_U) of the material.

- Once the shear force at any cross-section of the beam is found, shear stress can be calculated by dividing shear force by the area of cross-section.
- Relation between bending moment and bending stress is somehow more complicated, which is explained in this subsection.



Consider the beam shown above. The beam bends downward, *i.e.*, in the direction of $-y$. The upper half is under tension and the lower half is under compression. Assuming the strain to vary as a linear function of y:

$$\epsilon_x(y) = \frac{\epsilon_{max} y}{h/2}.$$

$\epsilon_x(+h/2) = +\epsilon_{max}$ is maximum tensile strain and $\epsilon_x(+h/2) = -\epsilon_{max}$ is maximum compressive strain.

If the Young's modulus of the material is E then,

$$\sigma_x(y) = E\frac{\epsilon_{max} y}{h/2} = \sigma_{max}\frac{y}{h/2}. \tag{39}$$

The bending moment is the result of the couple formed by the tensile and the compressive forces. However, these forces vary as function of y. So we find the force (dF) on a infinitesimally small area (dA) at a distance y from the neutral axis, and then determine the corresponding infinitesimally small moment. The resultant moment is sum of all such infinitesimally small moments, which is equivalent to the integration

$$\int y \times dF(y) = \int y(\sigma_x(y) \times dA).$$

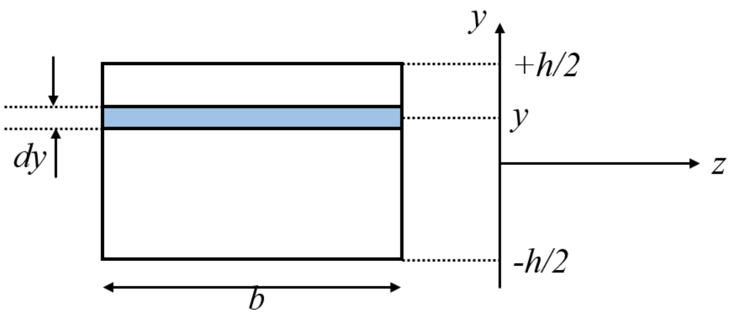


Figure 117: To find the moment of the couple formed by the tensile and the compressive forces (which vary as function of y), we start with the force on an infinitesimally small area $dA = bdy$ at a distance y from the neutral axis.

This is illustrated for a rectangular cross-section with b and h as its width and height, respectively (Figure 117).

For this example, the infinitesimally small area is

$$dA = b \times dy,$$

corresponding force is

$$dF = \sigma_x(y) \times bdy,$$

and hence the bending moment is

$$M_B = \int_{-h/2}^{+h/2} y(\sigma_x(y) \times bdy).$$

Substituting

$$\sigma_x = \sigma_{max} \frac{y}{h/2}$$

from equation 39,

$$M_B = \int_{-h/2}^{+h/2} y(\sigma_{max} \frac{y}{h/2} \times bdy).$$

$$M_B = \frac{\sigma_{max}}{h/2} \int_{-h/2}^{+h/2} y(y \times bdy). \tag{40}$$

The term $\int_{-h/2}^{h/2} y(y \times bdy)$ is known as the **second moment of area** and it quantifies the resistance to bending due to the geometrical shape and dimension of the beam's cross-section. Let the symbol I_z represent the second moment of area when the beam bends about the z-axis, *i.e.*,

$$I_z = \int_{-h/2}^{h/2} y(ybdy).$$

Substituting this is equation 40,

$$M_B = \frac{\sigma_{max}}{h/2} I_z \tag{41}$$

$$\sigma_{max} = \frac{M_B}{I_z}(\frac{h}{2}). \tag{42}$$

σ_{max} is the maximum stress caused by bending or the **maximum bending stress** at a cross-section where the bending moment is M_B.

The expression in equation 42 is for maximum bending stress, which occurs at the top surface and the bottom surface. You can use this to find the bending stress at any surface which is y unit away from the neutral surface by combining equations 39 and 42 :

$$\sigma(y) = \sigma_{max} \frac{y}{h/2} = \frac{M_B y}{I_z}. \tag{43}$$

The larger the value of y, the greater the stress. Therefore, the further away the from the the neutral force, the larger the bending stress.

Effect of Second Moment of Area on Bending Stress

Larger the value of I_z, smaller is the bending stress for the same amount of bending moment.

The magnitude of I_z is determined by the geometry and dimensions of the beam cross-section and is not affected by the magnitude of the bending moment and hence the magnitude of the external forces.

Different shape, different I derivations!!

For the rectangular cross-section used in the explanation, the second moment of area is

$$I_z = b \int_{-h/2}^{+h/2} y^2 dy = \frac{bh^3}{12}.$$

For a beam with rectangular cross-section, the width does not vary with y. So bdy is taken as the infinitesimally small area. However, this is not true for all types of cross-section. For example, in a triangular-shaped cross-section, width varies linearly as function of y, and $b(y)dy$ should be taken as the infinitesimally small area.

The second moment of area for different cross-sectional shapes are usually available in textbooks and on-line resources. You will not be required to find it by integration.

Effect of External Forces on Bending Stress

The bending stress is proportional to the bending moment M_B which depends on

1. External forces applied,
2. Length of the beam, and
3. Reaction forces from the support.

Strength analysis for beams under transverse loading

- Draw the BMD to determine the maximum bending moment ($M_{B,max}$) due to external loading.
  - You have seen that the bending moment M_B varies along the x-axis of the beam. We are interested in the maximum value only.
- Determine the second moment of area for the cross-section of the beam.
- Use equation 42 to determine the maximum bending stress developed

$$\sigma_{B,max} = \frac{M_{B,max}}{I_z}.$$

- Maximum bending stress should be smaller than the ultimate strength of the material to avoid failure.

Factor of Safety

The definition:

$$FoS = \frac{\sigma_U}{\sigma_W}, \tag{44}$$

where

- σ_U is the ultimate strength of the material used to make the component, and
- σ_W is the maximum working stress the component experiences.

If the factor of safety is two, it simply means that it is anticipated that the part can be subjected to effectively twice the load it was designed for before the design criteria of either yielding or fracture, whichever has been chosen, will be exceeded in a static application.

The concept of factor of safety is introduced into design for the purpose of minimizing the risk of potential part failure.

Where does the risk come from?

1. Approximate analytical techniques used to determine the load and strength level
2. Uncertainty and variability in the values used in the numerical calculations involved, for example, in
   (a) geometrical dimensions,
   (b) magnitude of loading, or
   (c) material properties.

Appendix

Second-Order Ordinary Differential Equations

Second-order differential equations play a central role in the physical sciences. They are found, for example, in laws describing mechanical systems, wave motion, electric currents and quantum phenomena.
Examples of such equations are

$$\begin{aligned} \frac{d^y}{dt^2} - 3\frac{dy}{dt} + 2y &= 4e^t, \\ 3\frac{d^2y}{dt^2} + y &= \sin t. \end{aligned} \tag{45}$$

A second-order ODE must include the second derivative, but may or may not have the first derivative.

Linearity and Superposition

The equations shown in Equation 45 are the so called *linear* second-order differential equation with *constant coefficients*. General expression of this class of equations is

$$a\frac{d^2y}{dt^2} + b\frac{dy}{dt} + cy = f(t), \tag{46}$$

where a, b and c are constant and $a \neq 0$.

A linear differential equation with constant coefficient is said to be **homogeneous** if

$$f(t) = 0$$

for all values of t and **non-homogeneous** otherwise.

The principle of superposition is a fundamental principle that is applicable to all linear systems including linear differential equations. Let $y_1(t)$ be the solution of

$$a\frac{d^2y}{dt^2} + b\frac{dy}{dt} + cy = f_1(t),$$

Two equations are identical except the **forcing function**, which is on the right hand side, giving two different solutions.

and $y_2(t)$ be the solution of

$$a\frac{d^2y}{dt^2} + b\frac{dy}{dt} + cy = f_2(t).$$

According to the principle of superposition, the function

$$y(t) = k_1 y_1(t) + k_2 y_2(t),$$

where k_1 and k_2 are constants, is a solution of the ODE if the forcing function is changed to

$$k_1 f_1(t) + k_2 f_2(t).$$

Stating the above phenomenon in a slightly different way, we can say that the solution of a linear second-order ODE

$$a\frac{d^2y}{dt^2} + b\frac{dy}{dt} + cy = f(t) \tag{47}$$

will be

$$y(t) = y_c(t) + y_p(t),$$

where, $y_c(t)$ is the solution of the **homogeneous equation**

$$a\frac{d^2y}{dt^2} + b\frac{dy}{dt} + cy = 0$$

and $y_p(t)$ is the solution of

$$a\frac{d^2y}{dt^2} + b\frac{dy}{dt} + cy = f(t).$$

The first solution $y_c(t)$ is known as the **complementary function** and the second solution $y_p(t)$ is known as the **particular integral**.

> The general solution of a linear, constant-coefficient second-order differential equation is given by
>
> $$y(t) = y_c(t) + y_p(t),$$
>
> where, $y_c(t)$ and $y_p(t)$ are the complementary function and the particular integral, respectively.

A. Finding the complementary function

For the ODE in Equation 47, the corresponding homogeneous equation is

$$a\frac{d^2y}{dt^2} + b\frac{dy}{dt} + cy = 0. \tag{48}$$

We start with a trial solution

$$y = e^{\lambda t}, \tag{49}$$

where λ is an unknown constant.

$$\begin{aligned} y &= e^{\lambda t}, \\ \frac{dy}{dt} &= \lambda e^{\lambda t}, \\ \frac{d^2y}{dt^2} &= \lambda^2 e^{\lambda t}. \end{aligned}$$

Substituting the solution and its derivatives into the homogeneous equation

$$(a\lambda^2 + b\lambda + c)e^{\lambda t} = 0. \tag{50}$$

Hence $y = e^{\lambda t}$ will be a solution of Equation 48 if and only if λ satisfies the condition

$$a\lambda^2 + b\lambda + c = 0. \tag{51}$$

Equation 51 is known as the **characteristics equation** or the **auxiliary equation** of the homogeneous linear constant-coefficient second-order differential equation.

The characteristics equation (51) is a quadratic equation and hence has two roots λ_1 and λ_2 giving two solutions of the homogeneous equation, *i.e.*,

$$y_1(t) = e^{\lambda_1 t}, \quad y_2(t) = e^{\lambda_2 t}.$$

If the two roots are distinct, the general solution to the homogeneous equation is

$$y_c(t) = Ce^{\lambda_1 t} + De^{\lambda_2 t}.$$

The roots of the characteristics equation are

$$\lambda_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a}, \quad , \lambda_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}.$$

Assuming that the coefficients a, b and c are real, there are three cases to consider, depending on the sign of the discriminant

$$b^2 - 4ac$$

:

- $b^2 - 4ac > 0$, the roots are distinct and real,
- $b^2 - 4ac < 0$, the roots are distinct and complex,
- $b^2 - 4ac = 0$, the roots are equal and real.

Next, we consider these cases one at a time.

(a) Distinct and real roots:

$$y_c(t) = Ce^{\lambda_1 t} + De^{\lambda_2 t}$$

In order to find the unknown coefficients C and D, we need to know the initial conditions of y and its derivative

$$\frac{dy_c}{dt} = C\lambda_1 e^{\lambda_1 t} + D\lambda_2 e^{\lambda_2 t}.$$

Let's assume that the initial conditions are

$$y(0) = Y_0, \quad \frac{dy}{dt}(0) = 0.$$

Substituting these conditions in the solution and its derivative

$$Y_0 = C + D,$$

$$0 = C\lambda_1 + D\lambda_2.$$

There are two unknowns C and D which can be obtained by solving the two simultaneous equations.

(b) Distinct and complex roots:

The roots are

$$\lambda_1 = \alpha + j\beta, \quad \lambda_2 = \alpha - j\beta.$$

Therefore,

$$y_c(t) = Ae^{(\alpha+j\beta)t} + Be^{(\alpha-j\beta)t},$$

where A and B are arbitrary constant which can be complex numbers. We usually need a real valued solution and hence it is better to express the result without complex number. To do that, we first express the solution as

$$y_c(t) = e^{\alpha t}(Ae^{j\beta t} + Be^{-j\beta t}).$$

Applying Euler's formula,

$$\begin{aligned} y_c(t) &= e^{\alpha t}(A\cos\beta t + jA\sin\beta t + B\cos\beta t - jB\sin\beta t), \\ &= e^{\alpha t}((A+B)\cos\beta t + j(A-B)\sin\beta t), \\ &= e^{\alpha t}(C\cos\beta t + D\sin\beta t), \end{aligned}$$

where $C = (A+B)$ and $D = j(A-B)$ are two arbitrary constants whose values can be determined using two initial value conditions.

(c) Equal and real roots:

We may encounter a problem if we use the solution

$$y_c(t) = Cy_1(t) + Dy_2(t)$$

if the individual solutions y_1 and y_2 are not linearly independent.

Suppose, y_2 is a constant multiple of y_1, *i.e.*,

$$y_2(t) = ky_1(t)$$

for a constant k, then the two solutions are linearly dependent. In such case

$$y_c(t) = Cy_1(t) + Dky_1(t) = (C + Dk)y_1(t),$$

which is effectively only one solution.

In the case of equal roots ($\lambda_1 = \lambda_2 = \lambda$), the general solution of the homogeneous equation is given by

$$y_c(t) = (C + Dt)e^{\lambda t}. \tag{52}$$

Two unknown constants can be determined using the initial conditions as explained for other two cases.

Consider the series RLC circuit of Figure 118. There is no voltage or current source in the circuit (undriven circuit). In order to initiate energy exchange between capacitor and inductor, the system must be perturbed initially (the capacitor voltage is V_0 before the switch is closed) The ODE

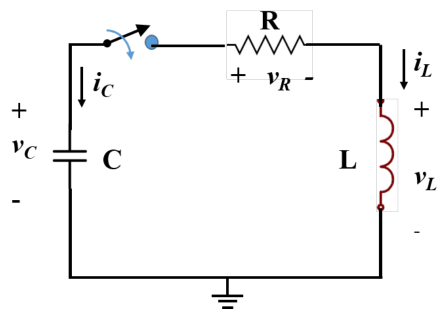


Figure 118: Undriven series RLC circuit

model for this circuit is

$$\frac{d^2 v_C}{dt^2} + \frac{R}{L}\frac{dv_C}{dt} + \frac{1}{LC} = 0,$$

with initial conditions

$$v_C = V_0, \quad \frac{dv_C}{dt} = 0, \quad at\ t = 0.$$

You can find the expression for $v_C(t)$ by solving the linear homogeneous constant-coefficient ODE for the given initial conditions.

Rewriting the equation in terms of natural frequency (ω_n) and damping ratio (ζ),

$$\frac{d^2 v_C}{dt^2} + 2\zeta\omega_n \frac{dv_C}{dt} + \omega_n^2 = 0, \tag{53}$$

where

$$\omega_n = \frac{1}{\sqrt{LC}} \quad and \quad \zeta = \frac{R}{2}\sqrt{\frac{C}{L}}.$$

The roots of the characteristics equation

$$\lambda^2 + 2\zeta\omega_n\lambda + \omega_n^2 = 0$$

are

$$\lambda = -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1}.$$

For under-damped case, *i.e.*, $\zeta < 1$,

$$\lambda = -\zeta\omega_n \pm j\omega_n\sqrt{1 - \zeta^2} = -\alpha \pm j\beta.$$

The solution for distinct complex roots of CE was given earlier. Following that solution,

$$v_C(t) = e^{-\alpha t}(C\cos\beta t + D\sin\beta t). \tag{54}$$

Evaluating the initial condition for v_C,

$$V_0 = C, \quad as \quad \cos 0 = 1, \quad \sin 0 = 0.$$

Differentiating $v_C(t)$,

$$\frac{dv_C}{dt} = e^{-\alpha t}(-C\beta \sin \beta t + D\beta \cos \beta t) - \alpha e^{-\alpha t}(C \cos \beta t + D \sin \beta t).$$

Evaluating the initial condition of $\frac{dv_C}{dt}$,

$$0 = (-0 + \beta D) - \alpha(C + 0).$$

Therefore,

$$D = \frac{\alpha}{\beta}C = \frac{\zeta\omega_n}{\omega_n\sqrt{1-\zeta^2}}C = \frac{\zeta}{\sqrt{1-\zeta^2}}V_0.$$

We have already found $C = V_0$, so $D = \frac{\zeta}{\sqrt{1-\zeta^2}}V_0$. Substituring the values of these constants in Equation 54,

$$v_C(t) = e^{-\alpha t}V_0(\cos \beta t + \frac{\zeta}{\sqrt{1-\zeta^2}} \sin \beta t). \tag{55}$$

Instead of writing it as a sum of sine function and cosine function, we can express it with a single function using the trigonometric identity

$$P\cos(\beta t - \phi) = P\cos\phi\cos\beta + P\sin\phi\sin\beta.$$

Rewriting Equation 55,

$$\begin{aligned} v_C(t) &= e^{-\alpha t}V_0\frac{1}{\sqrt{1-\zeta^2}}\cos(\beta t - \phi), && (56) \\ &= e^{-\zeta\omega_n t}V_0\frac{1}{\sqrt{1-\zeta^2}}\cos(\sqrt{1-\zeta^2}\omega_n t - \phi), && (57) \end{aligned}$$

where $\tan\phi = \frac{\zeta}{\sqrt{1-\zeta^2}}$.

B. Finding the particular integral

If the series RLC is connected across a voltage source $V_S(t)$, then the ODE model becomes

$$\frac{d^2v_C}{dt^2} + 2\zeta\omega_n\frac{dv_C}{dt} + \omega_n^2 = V_S(t), \tag{58}$$

Steps to find the solution:

1. Find the complementary function $v_{C,c}(t)$ but do not determine the constants
2. Find the particular integral $v_{C,p}(t)$

3. Overall solution is

$$v_C(t) = v_{C,c}(t) + v_{C,p}(t).$$

4. Determine the unknown constants by taking initial conditions into consideration.

In this section, only the steps required to find the particular integral are shown.

- Start with a trial solution; the trial solution is a function with coefficients that are unknown.
  - Choice of trial solution depends on the type of forcing function, *i.e.*, the right hand side of Equation 58.
- Substitute the trial solution into the differential equation and determine the unknown coefficients of the trial solution.

Choice of trial solution:

Table 12: Caption

Type	Forcing function	Trial solution
Polynomial	$f(t) = \Sigma_{i=0}^{i=n} m_i t^i$	$y(t) = \Sigma_{i=0}^{i=n} p_i t^i$ coefficients $p_n,\ p_{n-1}$, etc. are unknown. $f(t) = constant$ falls under this type with all coefficients except p_0 being 0.
Exponential	$f(t) = me^{kt}$	$y(t) = pe^{kt}$
Sinusoidal	$f(t) = m\cos kt + n\sin kt$	$y(t) = p\cos kt + q\sin kt$ p and q are unknown.

Example:

Solve the following non-homogeneous ODE

$$\frac{d^2y}{dt^2} + 2\frac{dy}{dt} + 2 = 10\sin 2t.$$

Step 1: Find the complementary function

The roots of the characteristic equation

$$\lambda^2 + 2\lambda + 2 = 0.$$

are

$$\lambda_1 = -1 + j, \quad \lambda_2 = -1 - j,$$

where $j = \sqrt{-1}$. The complementary function is

$$y_c(t) = Ce^{(-1+j)t} + De^{(-1-j)t} = e^{-t}(Ce^{jt} + De^{-jt}).$$

Step 2: Find the particular integral

The forcing function is sinusoidal and, therefore, the trial solution is

$$y_p(t) = p\cos 2t + q\sin 2t.$$

Then,

$$\frac{dy_p}{dt} = -2p\sin 2t + 2q\cos 2t,$$
$$\frac{d^2y_p}{dt^2} = -4p\cos 2t - 4q\sin 2t,$$

Substitute these values in the given ODE:

$$(-4p\cos 2t - 4q\sin 2t) + 2(-2p\sin 2t + 2q\cos 2t) + (p\cos 2t + q\sin 2t) = 10\sin 2t.$$

Reorganizing the functions,

$$(-4p + 4q + p)\cos 2t = 0,$$

$$(-4q - 4p + q - 10)\sin 2t = 0.$$

Both theese conditions will be simultaneously true if and only if

$$\begin{aligned} -3p + 4q &= 0, \\ -4p - 3q - 10 &= 0. \end{aligned}$$

Solving them we get

$$p = -\frac{8}{5}, \quad q = \frac{6}{5}.$$

And

$$y_p(t) = -\frac{8}{5}\cos 2t + \frac{6}{5}\sin 2t.$$

Step 3: Find the coefficient of the complementary function

The general solution of the ODE is

$$y(t) = y_c(t) + y_p(t) = e^{-t}(Ce^{jt} + De^{-jt}) - \frac{8}{5}\cos 2t + \frac{6}{5}\sin 2t.$$

Find the unknown parameters C and D by taking the initial conditions into consideration.

Exceptional cases: The method given above for finding the particular integral fails if the suggested trial solution is part of the complementary function. For example, the complementary function of the ODE

$$\frac{d^y}{dt^2} - 4y = 2e^{2t}$$

is

$$y_c(t) = Ce^{2t} + De^{2t}.$$

In this case, pe^{2t} as a trial solution will not work. Difficulties like this are usually overcome by using modified trial solution which is the proposed trial solution multiplied by the independent variable. This issue is beyond the scope of this handout.