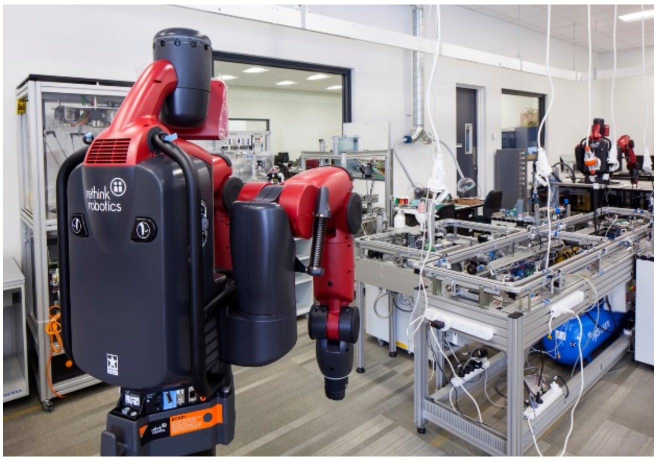
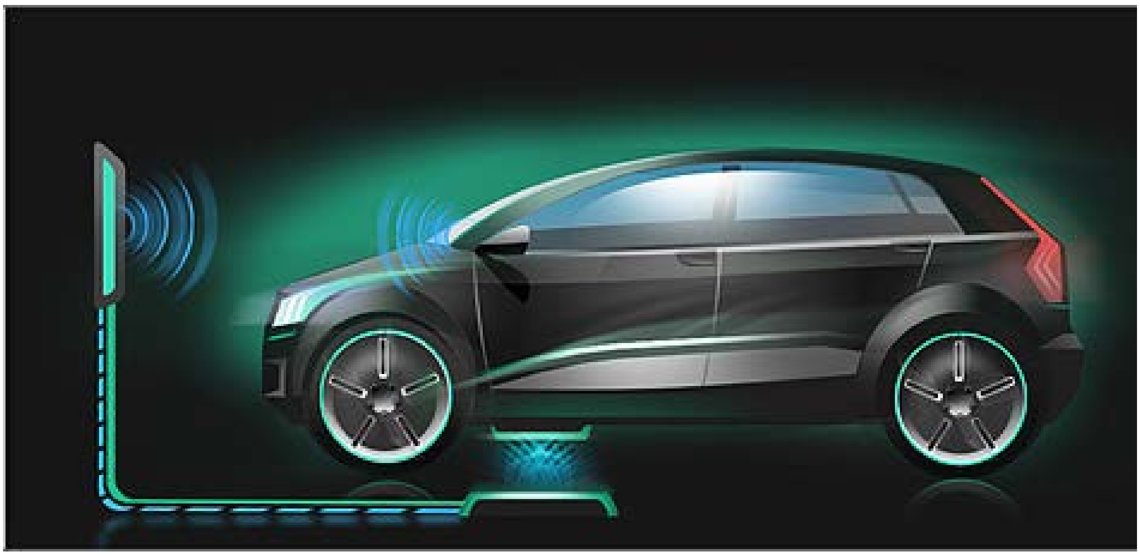
ELECTENG 101

Electrical and Digital Systems

2019





Department of Electrical, Computer, and Software Engineering

The Department of Electrical, Computer, and Software Engineering currently offers three of the nine specialisations in the Bachelor of Engineering (Hons) degree:

Electrical and Electronic Engineering

Computer Systems Engineering

Software Engineering

Cover Photos:
Graduates of the Department are making significant contributions towards the society of the future.

Top Left
There is considerable pressure for countries to move away from their dependence on fossil fuels to a variety of alternative energy sources. Solar and wind are two of these that are being extensively developed both in New Zealand and worldwide.

Top Right
Robots are increasingly being used in our modern world. The photo shows one of the Department's robots being used as part of a research project into the next generation of automated manufacturing.

Bottom
Driverless electric vehicles that sense their environment will soon be a reality for all of us. These will be charged inductively using technology developed in the Department over several decades.

Learning Resources

Online resources (available through the Canvas Course page)

In addition to the weekly lectures, these online resources have been developed to assist students in their learning:

- **Fundamental Circuit Concept Tutorials –** This is a set of online interactive questions that focuses on school level electric circuit concepts. Students are strongly encouraged to use this resource as a way of familiarising the necessary foundational understandings which the new knowledge in the course are built upon.

- **GECKO –** GECKO (Growing Epistemic Circuit Knowledge Outcomes) is a web-based platform; it was developed out of research into student learning by Bill Collis with the aim of integrating knowledge about models of electric circuits with the concepts that underpin them. The online assignments use GECKO. The questions are mostly numeric in nature, and different values are given each time you practice the question. There is feedback in many of the questions and you have the opportunity to repeat most questions until you are satisfied you have integrated the concept with the mathematical model that describes it.

- **Multiple-choice Questions (MCQs) –** A large bank of MCQs has been developed, covering various aspects of the course material. These will be made available progressively in line with the course delivery as a way for self-assessments of the relevant concepts covered throughout the course.

- **Piazza –** Piazza is an online forum where students can freely engage in anonymous discussions, debates, and collaborations over the course content. The forum will be monitored closely to provide guidance, and to be in tune with common misconceptions that arise throughout the course.

Recommended texts

Some of the course notes are based on the following three sources:

- Allan R. Hambley, *Electrical Engineering – principles and applications;* Sixth edition, Pearson Education, 2018

- James Diefenderfer, *Principles of Electronic Instrumentation;* Second edition, Holt-Saunders International Editions, 1979

- Tony R. Kuphaldt, *Lessons in Electric Circuits, Volume 1 - DC*
  This is a free series of textbooks on the subjects of electricity and electronics: http://www.ibiblio.org/kuphaldt/electricCircuits/index.htm

I have long held an opinion, almost amounting to conviction, in common I believe with many other lovers of natural knowledge, that the various forms under which the forces of matter are made manifest have one common origin; or, in other words, are so directly related and mutually dependent, that they are convertible, as it were, one into another, and possess equivalents of power in their action.

Michael Faraday (1791-1867)

Note:

Michael Faraday was an English scientist who contributed to the study of electromagnetism and electrochemistry. Among his major discoveries are the principles underlying electromagnetic induction.

Acknowledgements

A number of past and present colleagues, in particular,

Dr. Mark Andrews,

Prof. Gerard Rowe

Dr. Chris Smaill, and

Prof. Udaya Madawala

have influenced and contributed to the notes presented herein. The work and effort of all these individuals is here gratefully acknowledged.

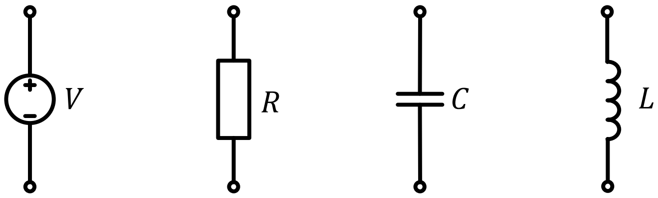
Table of Contents

1 Introduction

The study of electrical and digital systems is important because electrical technology is utterly pervasive and permeates everyday life to the extent that we are rarely conscious of its effects unless a device or system fails. Electrical technology has transformed human civilisation, improved quality of life, extended lifetimes, and has progressively reduced the burden of human and animal labour throughout the world. The revolution that started over a century ago with the telegraph and electric light bulb – and now gives us autonomous robots on other planets – shows no sign of slowing down and seems to be limited only by the imagination of engineers.

Physicists in the 19th century greatly expanded our understanding of electrical phenomena, such as the origin of electrical charge, electric and magnetic fields and how they interact. Consequently, we have very precise ways to measure quantities such as current and voltage. We also understand how to move electrons efficiently from one location to another, how to accumulate charge and regulate the flow of electrons – all through the use of ***electrical circuits***.

The simplest electrical circuits are constructed from components you *may* already be familiar with: batteries (to supply current and voltage), wires (metal conductors to deliver the current or voltage), resistors (to impede the flow of current and generate desired voltage), inductors (coiled conductors to intensify the magnetic field and store energy in the magnetic field), capacitors (flattened conductors that concentrate an electric field and store energy in the electric field).



Because we are so very adept at controlling the voltages and currents in a circuit, it turns out to be extremely useful to use electrical quantities as stand-ins (or ***analogues***) for other physical quantities (such as temperature, mechanical strain, velocity, etc.) and manipulate the analogue quantity (that is, voltage or current) instead of the physical quantity itself.

Note that this means we must have a mechanism for converting or ***transducing*** the physical quantity into an electrical analogue (and possibly vice versa). A simple example might be a "smart" heating system: A thermostat is a very simple temperature sensor that outputs a voltage to indicate "too cold", "about right", and "too hot". This output voltage can then be used to control the voltage applied to resistive heating elements producing heat to achieve the desired temperature automatically ***without*** manual intervention.

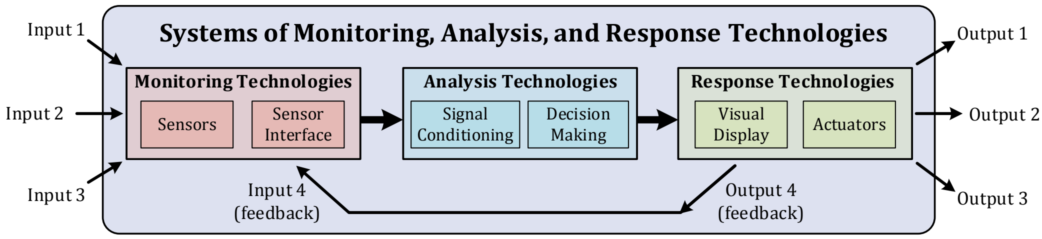
This simple self-regulating heating system reveals the internal workings of almost *all* modern electrical and digital systems that appears to have some perception of *automation* and *intelligence* – they consist of technologies that

- ***monitor*** the environment and convert the physical quantities of interest into electrical analogues;
- ***analyse*** the electrical analogues for contextual interpretations to make informed decisions;
- ***respond*** appropriately to the situation based on the gathered information to achieve pre-determined purposes.

The practice of modern electrical, computer systems, and software engineering therefore revolves around the study, design and engineering of these electrical and digital **S**ystems of **M**onitoring, **A**nalysis, and **R**esponse **T**echnologies.

1.1 'SMART' Engineering

While electrical systems can be extremely varied in nature, but a useful block diagram that applies to a wide range of realistic systems is shown below.



Monitoring Technologies

The monitoring technologies sense (or measure or sample) the environment in which the circuits exist. For certain inputs this may require transduction (converting a physical quantity into an analogue electrical quantity) or it may take direct electrical input (for example, from another circuit).

Analysis Technologies

The analysis technologies take the electrical output(s) from the monitoring technologies and process them in some way. For example, an audio amplifier may take the millivolt-level output from a microphone and amplify the voltage to several tens of volts required by the audio speaker.

In more elaborate systems (typically digital systems) the circuitry may respond differently depending on the state of one or more inputs ***and*** the current state of the processing circuitry. The system is therefore making decisions based on the current and past inputs. This is an extremely important concept and underpins all computer-systems thinking.

Response Technologies

The response technologies generate the action or response (if any) to the inputs as determined by the processing circuitry. In its simplest form, the action might be to turn on a light or display text on an LCD panel. More sophisticated actions might entail detailed behaviour such as when a mobile robotic device has plotted a trajectory through a cluttered environment.

Frequently, the action itself is sensed so that the system can determine that the action has actually taken place or to the desired precision. The general notion of an output being measured as another input to the system is called ***feedback*** and it turns out to be a very powerful and important tool in electrical engineering.

For example, when a car's cruise control is set to 80 km/h the cruise control computer needs to know if the target speed of 80 km/h has been achieved so that the car's acceleration can be changed accordingly. Accelerating from 50 km/h to 80 km/h requires different amount of engine torque depending on whether the car is travelling on the flat or up an incline.

1.2 Signals and Systems

In the previous discussions we have casually referred to "circuits", "amplifiers", and "computer systems" without being very precise about the differences. There are many ways to classify different types of circuits, together with their inputs and outputs, but the most important distinction is between continuous-time and discrete-time.

General Signals

A signal is the generic term applied to any quantity that changes with time. Voltages and currents are therefore signals and it is often handy to refer to "the signal" without necessarily specifying if it is a current or voltage. A physical quantity prior to being sensed is also a signal. For example, we may refer to "the temperature signal".

Continuous-time Signals

A continuous-time signal is one in which the signal is defined at all possible times[1]. For example, the voltage signal

$$v(t) = 100\cos(2\pi 50t)\ \text{V}$$

is a 50 Hz sinusoid with a peak value of 100 V. We can substitute any value of t into the expression for $v(t)$ and get a sensible value. A continuous-signal is also commonly referred to as an ***analog signal***.

[1] Note that this doesn't mean that the *signal* is continuous. Rather, the signal is defined at a contiuum of times and we could, in principle, measure or know the value of the signal at any instant of time.

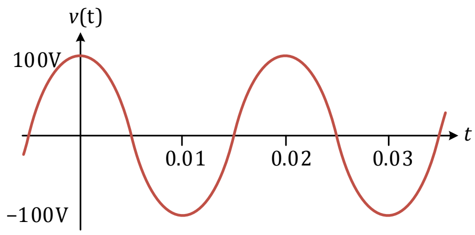
Discrete-time Signals

A discrete-time signal is one in which the signal is defined only at certain instants of time. For example, the voltage signal

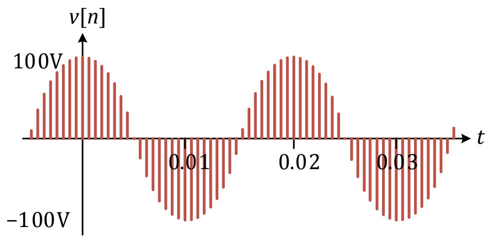
$$v[n] = 100\cos\left(2\pi 50 \cdot \frac{n}{100}\right)\ \text{V}$$

(where n is an integer) traces out the shape of a 50 Hz sinusoid with a peak value of 100 V. However, because n is an integer, we only "know" the signal at the times $t_n = n/100 = 0.01n$ (that is, every 10 milliseconds). In a minor abuse of terminology, we will refer to a discrete-time signal as a ***digital signal***.

Digital signals arise in practice because we measure or *sample* an analog signal at discrete instants in time to process them in a digital circuit[2]. This is referred to as *analog to digital conversion* or ADC[3].



Analog Signal



Digital Signal

General Systems

A system is the generic term applied to the mechanism that is acting on the inputs to produce an output. A simple circuit containing two resistors connected as a voltage divider is therefore a system. The term system can also be used quite broadly. For example, a high-level computer program that scans a text file looking for a search phrase is also a system in the sense that we have defined it.

- An *analog system* or analog circuit is any system that processes analog signals.
- A *digital system* or digital circuit is any system that processes digital signals.

[2] What is the "original" signal doing between, say, $t_1 = 0.01$ ms and $t_2 = 0.02$ ms? The answer is: we don't know! However, it is an amazing fact that provided we take "enough" measurements we don't need to know the signal value between sampling instants.

[3] We can also go back the other way. Starting with a stream of samples we can generate an analog voltage. Typically, the analog voltage is held constant until the next sample is available, producing a type of staircase effect.

1.3 Dealing with Complex Systems

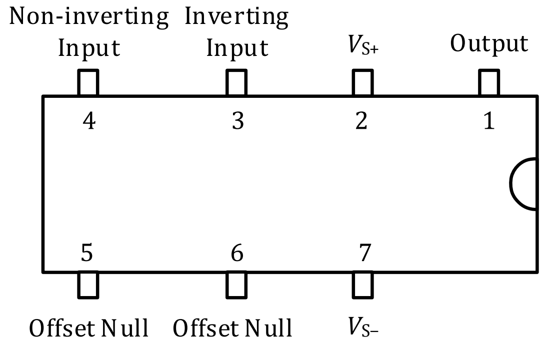
The design of electrical and digital systems is a well-understood process and as time has progressed the complexity of these systems has increased immensely. The microprocessor is one example where the complexity (defined in this case to be the number of transistors, or switches, on a single chip) is now enormous. The original Intel 4004 microprocessor released in 1971 contained 2,300 transistors. At the time this was a massive achievement and this level of device density (or *integration*) was unprecedented.

Since 2010 the leading-edge designs have never had fewer than a billion switches. For example, the Intel 10-core Xeon released in 2012 has approximately 2,600,000,000 switches and even this isn't the most highly integrated device on the market. Transistor counts alone are not a very good measure of complexity since a microprocessor has a very regular architecture with many duplicated structures. Nevertheless, how are engineers supposed to deal with this level of detail?

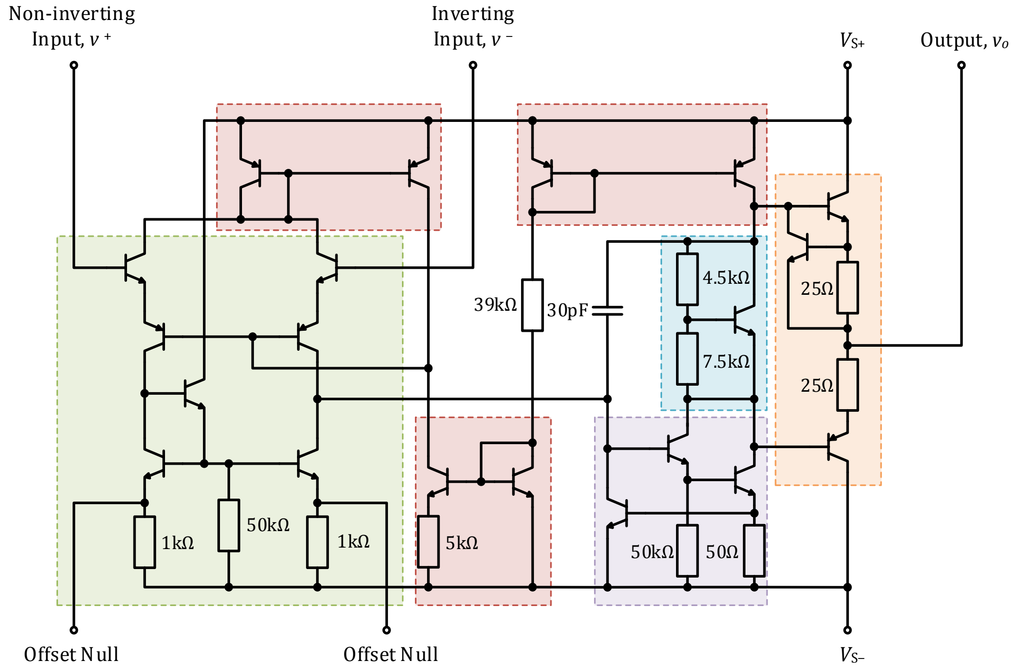
The key to dealing with complex systems is to *compartmentalise* a complex system into many smaller sub-systems (indeed, the overall system is designed by combining many smaller systems in the first place). Each sub-system is simple enough to be understood by an average engineer and at no time is it necessary to grapple with all details of the complete system. This keeps everyone sane.

The circuit diagram below shows the internal structure of the LM741 *operational amplifier* – one of the workhorses of analog circuit design. It looks rather complicated and a bit intimidating, but the dotted colour boxes delineate sub-systems that are actually easy to understand. By thinking about how the sub-systems interact (and not their internal components) it's a much easier task to figure out how the entire circuit works. As you will soon begin to see throughout this course that once we have figured out *how* a system functions, we can *abstract* away the complexity, reducing it down to an *equivalent* circuit modelling its external electrical behaviour that is easy to understand – we no longer need to worry about *why* it behaves the way it does – and focus our attention purely on its *engineering* implications

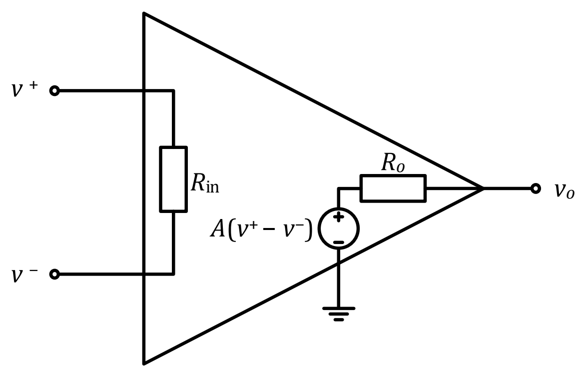
What we see physically:



What is inside internally:



What we use to model its electrical behaviour externally:



2 Fundamentals of Electrical & Digital Systems

The study of electrical and digital systems revolves around the *manipulation* and *interpretation* of the electrical quantities *voltage* and *current*. We do so by cleverly altering, suppressing, and amplifying these quantities through electrical circuits containing electrical elements such as resistors, capacitors, inductors, etc. While the manipulation is standardised (via circuits), the meaning and purpose of these signals, however, depends completely on what *we* want them to be.

For example, the output voltage of a thermistor *may* be fed into a microcontroller to drive a seven-segment display as a direct indication of the temperature, but it could also be used as the input of a regulatory system to control the current through its resistive heating elements thereby creating a self-regulated heater[4]. In both cases, the voltage signal (the electrical analog of temperature) is processed universally via (different) circuits, but the former decides the on/off state of the display LEDs, while the latter controls the magnitude of the heating current.

On the other end of the spectrum, the lowest level of the Internet transports vast quantities of numbers from one location to another in the form of voltages and currents. The transmission and reception of these signals are uniformly achieved via wired(less) transceivers (which are just electrical circuits), but the meaning of those numbers is up to us – they could represent an image, an audio, an e-mail, or an actual number, etc.

Knowledge of electrical technology is vital for all disciplines of engineering because it is simply not possible to do *any* engineering without encountering electrical principles. But in order for us to control, utilise, and ultimately *engineer* electrical and digital systems, we must first develop an appreciation of (1) what these underlying analogue quantities mean, (2) how they behave in an electrical circuit, and (3) the tools necessary to analyse them.

2.1 Circuit Concepts

At the heart of every electrical circuit is a collection of electrons. These electrons are harnessed and energised through various engineering methods (i.e., *electrical sources*). By providing a conducting path (e.g., *wires*) toward other electrical devices (e.g., *resistors*, *capacitors*, *inductors*, etc.), the resulting flow of electrons (i.e., *current*) enables the delivery of energy to these devices over time. Through careful monitoring of the energy changes in the electrons (i.e., *voltage*) across these devices, we are able exploit, and take advantage of, these devices' electrical behaviours to store energy, accumulate *charges*, control the current, regulate *power*, and in doing so, produce useful circuits that perform meaningful tasks.

It is important that we understand, and become fluent with, the terminologies and nomenclatures we use in describing the various constituents, quantities, and

[4] The idea of *transducing* a physical quantity into an electrical analogue (and vice versa) forms the basis behind *all* disciplines of engineering due to the pervasiveness of electrical technology.

phenomena pertaining to an electrical circuit for they form the *language* of *everything* we practice and do in electrical engineering – it is simply **not** possible to do *any* engineering without a concise way of articulating what we wish to convey.

2.1.1. Charge & Current

Intended Learning Outcomes

- Be able to recognise, and make use of, the relationship between charge and current.
- Be able to recognise the difference between conventional current and electron-flow current pertaining to the modelling of electric circuit behaviour.

Electrostatic force is a fundamental force in nature which we have attributed its origin to *electric charge* (symbol q or Q), a physical property of matter that causes it to experience such a force when placed next to another. Through experiments, we have discovered that there are two types of charges – like ones repel, and opposite attracts – this allowed us to *define* the notion of *positive* and *negative* charge. While we now have developed a model that describes, quite accurately, the *behaviour* of these charges in terms of the electrostatic force they exert upon each other, namely[5],

$$F \propto \frac{q_1 q_2}{r^2},$$

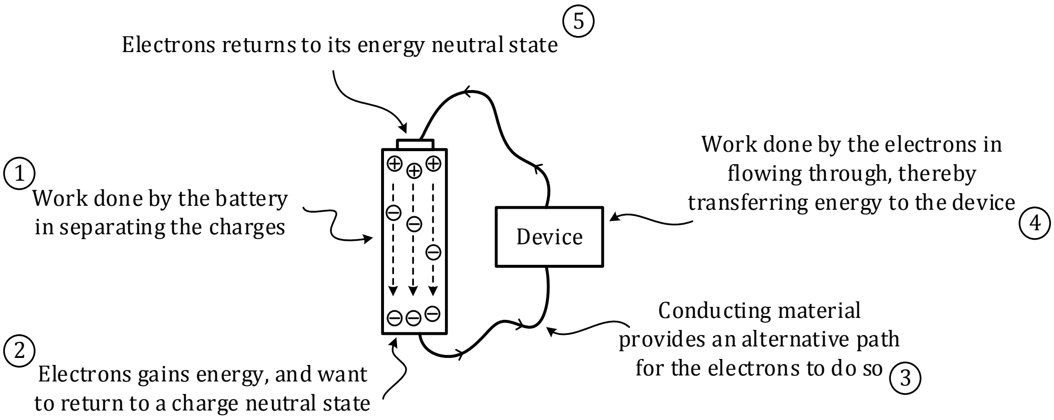
what *exactly* are they is still a mystery to some extent today. We do know that the source of negative charge is the *electron*, but we have not completely narrowed down the origin of positive charge[6]. Thus, based on what *we* can observe, we have associated the 'flow of charges' to the *movement of electrons*.

Surprisingly, no one knows what an electron *is* to date, but we only know how it *behaves*. Its behaviour has led us to link what we do know in terms of the conservation of energy to how and why an electrical circuit work the way it does – by clever means of working *against* the electrostatic forces, physicists discovered ways to separate electrons from positive charges[7], and in doing so the electrons *gain* energy which can be transported to electrical devices through conducting materials for them to make use of.

[5] This is known as the Coulomb's law which states that the force between two charges is proportional to the magnitudes of the charges (q_1, and q_2) and decreases with distance r.

[6] We do know that *protons* have positive charge, but these are not the only subatomic particles to possess it. *Positrons* which is similar to electrons in all characteristics except its behaviour is opposite is known to exist but hard to observe.

[7] E.g., through electrochemical reactions such as that inside a battery, or through rotating a wire in a magnetic field such as that of a generator.



The amount of charge a matter has is measured in Coulombs (symbol C), and the smallest free-standing particle known to date that possess charge is the electron which has approximately $(-)1.6 \times 10^{-19}\ \mathrm{C}$ of charge. In other words, $1\ \mathrm{C}$ of charge is equivalent to approximately 6.3×10^{18} electrons.

Current

A current (symbol i or I) is the number of charges flowing through a particular point in one second. In other words, it measures the *rate* at which electric charges pass through a section of an electric circuit at any given time, and this is denoted mathematically by the *derivative* of the total number of charges q passing the given section over time t,

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

Currents thus have the units *Coulombs per second* (Cs^{-1}), but commonly we use *Ampere* (symbol A).

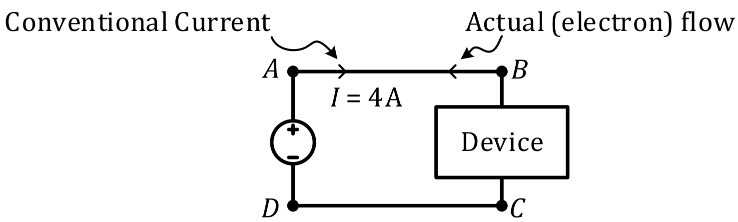
We are generally more interested in the rate at which charges flow (i.e., current) through an electrical device at a given instance in time rather than the total number of charges that have gone by the device up to that time because the rate gives us an indication of the electrical 'movement' that is happening in the device and therefore tells us if the device is 'working'. This in turn reveals how *fast* energy is being absorbed (or delivered) by the device, which we are usually more interested in, as opposed to how *much* energy has the device received (or supplied)[8], this is because all electrical devices requires a *continuous* supply of energy to function.

[8] The obvious exception is when we have limited total energy capacity such as when considering that of a practical battery. Most of the time, however, when we are interested in the behaviour of electrical devices we would assume it can be powered indefinitely by an unlimited source of energy (e.g., from the mains supply).

Conventional Current

Historically, it was once believed that the flow of charges was attributed to *positive* charge 'carriers', and consequently the *conventional* current actually described the rate at which these positive charge carriers flow. Much of the inventions developed in electrical engineering went on following such a notion, and regrettably we ended up with theories and notations all based on the conventional current[9].

While we now know this is *incorrect* – it is the *electron flow* that occurs in the conducting material – fortunately since the movement of an electron in one direction can be viewed relatively to be a displacement of a 'lack of an electron' in the opposite, other than the direction of charge flow, adhering to the conventional current had **no** impact on the theories developed in terms of how electric devices behave. Therefore, given the pervasiveness of our innocent error, the conventional current remains in use today in describing *all* electrical operations.



We denote (conventional) current in circuits by using arrows along the wires in the direction of which *positive* charge flows. Sometimes we may also use a double-subscript notation to indicate the direction the charges are flowing in the order of **from** and **to**. Thus, for the circuit shown above, we can also denote the 4A clockwise current by

$$I_{ab} = I_{bc} = I_{cd} = I_{da} = 4\text{ A}.$$

Example 1. A current of 0.5A flows through the light bulb in a torch for 8 minutes. Find the total number of charges that passed through the light bulb in this time. How many electrons are needed to carry this charge?

[9] For example, the symbol for the *diode*, an electrical device that allows current in one direction but blocks in the other, takes on an arrowhead that points in the direction of conventional current. Similarly, the *transistor*, an integral building block of all digital logics, has also inherited such a notation in its symbolic representation.

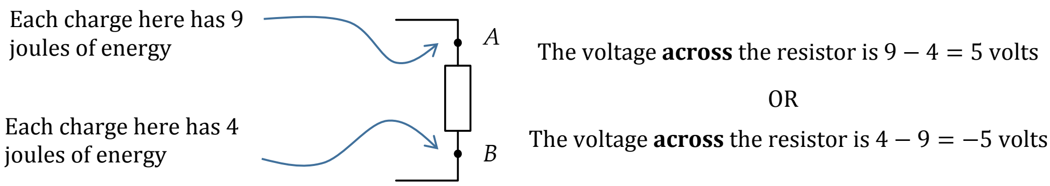
2.1.2. Voltage & Common Reference

Intended Learning Outcomes

- Be able to relate and interpret the meaning of voltage in an electric circuit to the energy characteristic of the underlying charges in the circuit.
- Be able to represent the changes in the energy of the charges across electrical components using the notion of voltage and voltage arrows.

A voltage (symbol v or V) is a mathematical *difference*, it is the difference in electrical potential energy of each unit of charge *between two points* in a circuit. As such, a voltage is also commonly known as a **potential difference** and has units *Joules per Coulomb* ($\mathrm{JC^{-1}}$) but more commonly we use *Volt* (symbol V).

Because it is a difference, for a voltage value to make sense we **must** know which point (in the circuit) that the difference is calculated with respect to. The figure below shows two ways in which we can describe the voltage *across* a resistor:



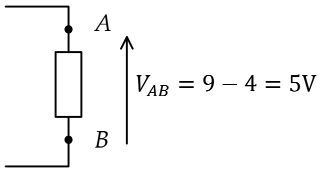
- When we say the voltage across the resistor is 5 volts, we mean that one end (in this case, A) of the resistor has 5 joules of energy **more than** the other (B)

- Likewise, we can also say the voltage across the resistor is -5 volts, and this implies that one end (in this case, B) of the resistor has 5 joules of energy **less than** the other (A)

Both statements above are correct, and such an ambiguity arises because we do **not** know, from the two voltage values, which end of the resistor is at the higher/lower potential. In other words, we do not know whether the difference is calculated with respect to A or B in the circuit.

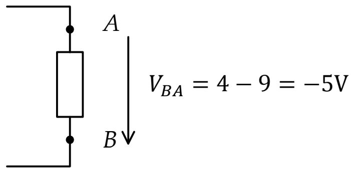
To rid of this ambiguity in defining voltages, every voltage value **must** be accompanied by a *voltage arrow* [10], pointing in the direction of a *potential rise*, or in other words, the tail indicates the end the difference is taken with respect to:

[10] An alternative is the use of positive/negative sign pair as used in defining voltage sources – negative denotes the end the difference is taken with respect to.



(a)

We say the potential at A is 5 volts **higher** than the potential at B.



(b)

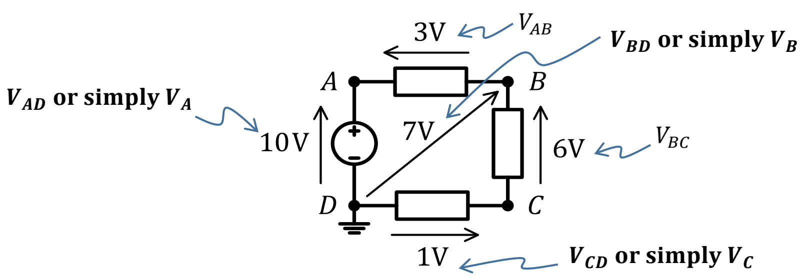
We say the potential at B is -5 volts "higher" than the potential at A. Equally, this means the potential at B is 5 volts **lower** than that at A, in agreement with (a).

Both cases now identify the voltage across the resistor unanimously – a **negative** voltage simply means the true polarity is opposite to what was assumed. We say "the voltage at A, with respect to B, is V_{AB}.", and clearly $V_{AB} = -V_{BA}$. Notice voltage values are **relative** and not absolute – we no longer know the exact potential at A or B given the voltage across the resistor V_{AB} or V_{BA}.

Common Reference or Ground

Often, voltages are defined or measured in a circuit with respect to a common reference point (may be across multiple components) which we denote as "*ground*" in the circuit and is represented by the symbols[11]:

The ground is simply a symbolic indicator for a point in the circuit where the potential is *taken* or *defined* to be zero, providing a convenient reference in which voltages at various points in the circuit can be measured with respect to. This reference is often implied or "understood" and so we omit its subscript when indicating voltages with respect to that point:



[11] The "triangle ground" is used as the ground for digital circuits while the "line segment ground" is used as the ground for analog circuits. We won't make such a distinction and will assume they mean the same.

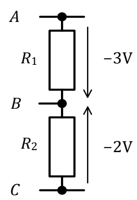
Thus, we commonly abuse the notation and say that the voltage "at" A is 10V, at B is 7V, and at C is 1V, and we can say so because **those voltages are understood to be relative to ground**:

$$V_A = V_{AD} = 10 \text{ V}, \qquad V_B = V_{BD} = 7 \text{ V}, \qquad V_C = V_{CD} = 1 \text{ V}, \qquad \text{and } V_D = V_{DD} = 0 \text{ V}.$$

Notice the placement of ground is purely arbitrary and does **not** affect the actual voltage across (nor the current through) the components – the circuit has not changed, only the reference at which the point voltages are measured relative to has. If the ground in the circuit above is placed at C instead, then we have simply *defined* the potential at that point to be zero, and this simply means that the voltage at each labelled points (with respect to this new reference) are now

$$v_A = v_{AC} = 9 \text{ V}, \qquad v_B = v_{BC} = 6 \text{ V}, \qquad v_C = v_{CC} = 0 \text{ V}, \qquad \text{and } v_D = v_{DC} = -1 \text{ V}.$$

Example 2. The voltage across each of the two electrical components R_1 and R_2 connected end-to-end is shown. The voltage across the two components R_1 and R_2, i.e., V_{AC} is



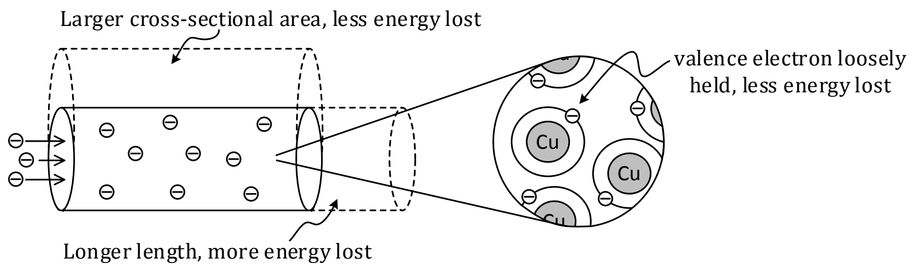
(a) -5 V (b) -1 V (c) 1 V (d) 5 V

2.1.3. Resistance & Ohm's law

Intended Learning Outcomes

- Be able to interpret and relate Ohm's law to the energy changes in the charges through a resistive conductor in terms of the voltage across, current through, and the resistance of the material.
- Be able to recognise the characteristics of an ideal conductor and infer the implications of these characteristics on its behaviour in an electric circuit.
- Be able to determine relevant electrical quantities pertaining to a resistor in an electric circuit in terms of its voltage, current, and resistance.

When an electrical charge flow through a conducting material (e.g., metals), it must necessarily *lose* energy as it repels and collides with other charges (of the material) in between. The amount of energy lost is subject to how 'difficult' it is for the charges to flow through the material which, intuitively, must depend on both the chemical (e.g., stability of valence electrons, temperature) and physical (e.g., length and cross-sectional area) properties of the conductor itself.



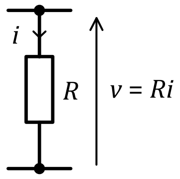
From empirical evidence it was observed that the amount of energy *lost* through a conductor is directly proportional to the *rate* at which the charges are flowing through it – the larger the amount of charges are forced through a conductor at a given instance, the larger the amount of energy is lost by each charge as they propagate through, and vice versa. Given our understanding of the meaning of *voltage* and *current* we therefore have a very concise mathematical way of describing this observation via the well-known **Ohm's law**:

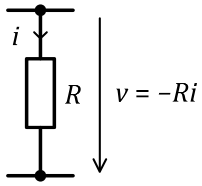
$$\underbrace{v}_{=\text{ difference in energy across the conductor}} = \underbrace{R}_{=\text{ constant of proportionality}} \times \underbrace{i}_{=\text{ rate of charge flow}}$$

The constant of proportionality, R, in Ohm's law is known as the **resistance** of the material whose value provides a measure of its ability to impede or *resist* the flow of charges (i.e., current), *summarising* its intrinsic and extrinsic properties.

From our understanding of voltage and current, a voltage only make sense if it is accompanied by a voltage arrow, likewise a current only make sense with an indicated direction. Notice this means there must be an inherently assumed voltage polarity and current direction associated with Ohm's law (which the symbols themselves cannot convey) – **the current is taken to be in the direction of a voltage drop**. Such a convention makes intuitive sense since charges can

only lose energy through a resistor[12]. Note that this also means **if** the current is taken to be in the direction of a voltage rise (or vice versa), then Ohm's law still applies, but we must include a negative sign in order to reflect reality:





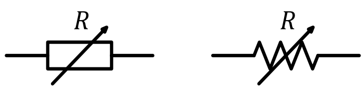
From its origin, the unit of resistance is thus *Volts per Ampere* (VA^{-1}) but more commonly we simply use *Ohm* (symbol Ω). Ohm's law makes intuitive sense because

- For a fixed amount of charges passing through the material at a given instance (current), a higher resistance in the material means a larger amount of energy must be lost (voltage) by the charges as they traverse through, and vice versa
- For a fixed amount of energy lost in the charges passing through the material (voltage), a higher resistance in the material means a smaller amount of charge must be flowing through at that instance (current), and vice versa

Resistor

While the resistance of a material might seem like a non-ideal consequence of the underpinning physics governing the movement of charges, the ingenuity in electrical engineering lies exactly in *exploiting* such an undesirable characteristic of the real world and using it to our advantage in achieving useful purposes. Indeed, through careful manufacturing methods, we can (and do) *intentionally* produce conductors of specific resistances with the intention of using them to do meaningful tasks. These 'resistive' conductors are also known as **resistors**, and we represent them symbolically by:

Fixed Resistors



Variable Resistors

It is important that we understand how the voltage across and current through a resistor behave for they are one of the fundamental building blocks of electrical circuits, and familiarity with its behaviour will allow us to develop an appreciation for how they can be used to design and engineer useful electrical circuits.

[12] This is also known as the *passive-sign convention* and in fact is used by electrical engineers to define the voltage-current behaviours (i.e., terminal characteristics) of *all* electrical elements (e.g., capacitors, inductors, diodes, etc.) – they all, by default, taken to *receive* energy (from the charges).

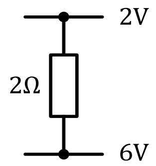
Example 3. The direction and magnitude of the current through the resistor in portion of the circuit shown is

(a) 1A downwards

(b) 2A upwards

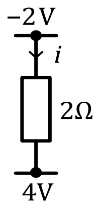
(c) 2A downwards

(d) 3A upwards



Example 4. In the circuit shown, the current i through the 2Ω resistor in the indicated direction is

(a) -3 A (b) -1 A (c) 1 A (d) 2 A



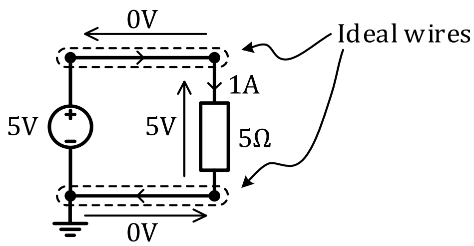
Ideal Wire

An **ideal** wire is a *perfect* conductor that has **zero** resistance, the most desirable property that we would *like* an electrical wire to have – charges can freely flow through an ideal wire **without** any loss in energy as Ohm's law would suggest:

$$\underbrace{R}_{=0\,\Omega} \times \underbrace{I}_{\text{irrespective of the size of current}} = \underbrace{0\text{ V}}_{\text{no difference in } E\ \textbf{across} \text{ the ideal wire}}.$$

Note the meaning implied by Ohm's law here: it says the difference in energy **across** an ideal wire is zero regardless of how many charges are flowing through it at a given time. Said differently, the charges anywhere along an ideal wire must have the **same** amount of energy so that the difference (i.e., voltage) between any two points along the ideal wire is zero.

This also means that the current through an ideal wire **depends** on the circuit that results from the ideal wire connections, it **cannot** be determined by simply examining the ideal wire on its own. This make sense because irrespective of the current through the ideal wire, the voltage across it is always zero (since $R = 0\ \Omega$) and by looking at such an isolated view there is thus insufficient information to characterise its behaviour[13] . This contrasts with that of a resistor – if the voltage across a resistor is zero, then the current through the resistor **must** be zero (since $I = 0/R = 0$ if $R \neq 0\ \Omega$).



It should be clear that the current through the ideal wires is 1A, determined by the circuit completed by the ideal wires which cannot be found by looking at the wires themselves: The 5V source ensures that each unit of charge at the positive terminal carries 5 joules of energy **more** than the negative (which is referenced to have zero joules). The ideal wires enables the charges to travel to and from the 5Ω resistor without losing any energy. This ensures that there is a potential difference across the resistor of 5V, and by Ohm's law the current through the 5Ω must therefore be 1A which must have come through the ideal wires.

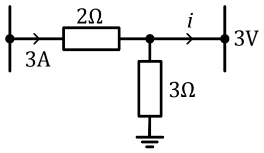
While all *real* wires have some non-zero resistances, in practice these are negligibly small[14] in comparison to the resistances of other circuit components, and we can therefore approximate them to be ideal in order to facilitate rapid analysis. Of course, in situations where it does matter, we may simply *model* the real wire as a resistor.

---

[13] Attempting to apply Ohm's law to an ideal wire will result in $I = V/R = 0/0$ A an indeterminate form indicating there is insufficient information about the current by just examining the piece of ideal wire.

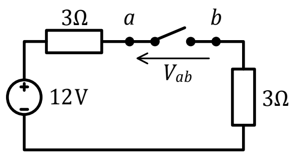
[14] For example, a typical copper wire has a resistance of approximately 5 mΩ per metre at room temperature.

Example 5. The magnitude of the current i through the ideal wire in the portion of the circuit shown is



(a) 0 A
(b) 1 A
(c) 2 A
(d) 3 A

Example 6. The voltage, V_{ab}, across the open switch in the circuit shown below is



(a) 0 A
(b) 1 A
(c) 2 A
(d) 3 A

2.1.4. Voltage & Current Sources

Intended Learning Outcomes

- Be able to recognise the characteristics of an ideal voltage and current source and infer the implications of these characteristics on their behaviour in a circuit.
- Be able to determine voltage across and current through an ideal source given knowledge of its surrounding electrical behaviour.

An *electrical source* is a device that *usually*[15] powers, or provides energy to, an electrical circuit. The most ideal characteristic that we would *like* a source to have is the ability to maintain its intended operating behaviour with zero variability and be able to do so *indefinitely*. In other words, *ideal sources* are those that can supply (or receive) infinite amount of energy without any deviation from their specifications regardless of what they are connected to. Of course, such a behaviour is purely fictional in practice – all *real* sources have finite energy capacities and will necessarily deviate from their rated output over time. Not only so, depending on what these sources are connected to, their output may also vary due to internal imperfections, and electrical noise. We are nonetheless interested in these fictitious ideal sources for several reasons:

- In most well-designed engineering applications, the sources are designed to operate within their rated limits, we are thus less concerned by their finite limitations within the scale in which they operate[16] and may therefore be assumed ideal
- Ideal sources provide very good approximations to most practical behaviours and facilitate rapid analysis that would otherwise result in analytical inconvenience with minimal gain, and if a design requires a more accurate description, we may simply refine our initial estimates through more in-depth models should the need arise
- One of the most powerful theorems in electrical engineering (see **Sect. 2.2.5** later) allows us to reduce almost *all* circuits to simply an ideal source with a resistor, abstracting the complexities away from the analysis when they are not of the focus

There are two types of electrical sources that are commonly used to model, and describe almost all phenomena that supply energy: the *ideal voltage source* and the *ideal current source*.

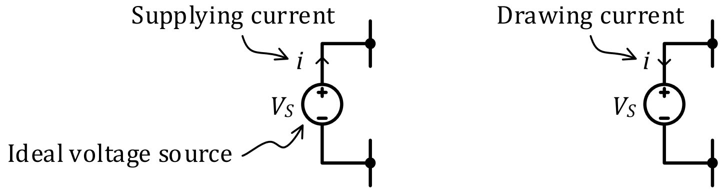
---

[15] In circuits containing multiple sources, they could also receive energy from other source(s) depending on the circuit configuration and arrangement. For example, the battery in your mobile phone provides energy to the underlying circuitry but it also receives energy from the mains supply when ‘plugged in’.

[16] For example, the bench power supplies in the laboratory are practically ideal for the sort of small-scale electronics that we would connect to it. But, of course, their energy capacity is limited by the reserve of hydro-electricity from the mains supply which, despite being ‘enormous’, is finite.

Ideal Voltage Source

An ideal voltage source maintains a constant voltage across its terminals regardless of how much current is developed through it due to its surrounding environment. In other words, an ideal voltage source will **supply** or **draw** whatever amount of current necessary in order to maintain the specified voltage across its terminals.



Such a behaviour of an ideal voltage source also means that the current through it *depends* on the circuit in which it is placed in. Without looking at the circuit an ideal voltage source is in, it is not possible to determine the current through it. This behaviour is in contrast to that of a resistor – no matter where a resistor is placed in, if the voltage across the resistor is known then its current is known – and is analogous to that of an ideal wire – an ideal wire is essentially an ideal voltage source of zero volt!

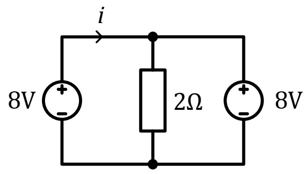
The idealised behaviour of an ideal voltage source does entail some practical considerations:

- If the terminals of an ideal voltage source are connected by an ideal wire (i.e., a "short-circuit"[17]), an infinite amount of current will result (a constant voltage over zero resistance), in practice the source will simply break

- Similarly, if ideal voltage sources of different values are connected **in parallel**, an infinite amount of current will result, in practice a large amount of current will result between the sources and energy is wasted in the sources until they reach a common voltage value (if they do not break first)

Example 7. The circuit shown contains two ideal sources. The value of the current i is

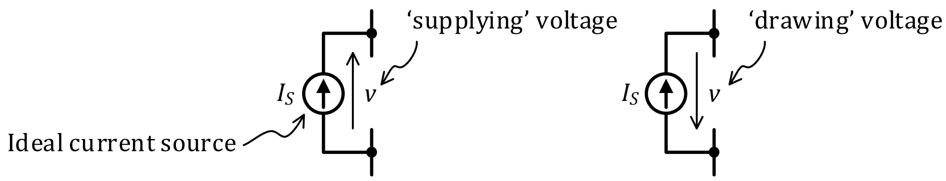
(a) 0 A
(b) 2 A
(c) 4 A
(d) 8 A



[17] If a pair of terminals is connected together with zero (or very low) resistance in between, we say the terminals are *shorted*, and we call such a connection across the terminals a *short-circuit*. Similarly, if there is no electrical connection (infinite resistance) across a pair of terminals, we say the terminals are *open*, and we call such a lack of connection across the terminals an *open-circuit*.

Ideal Current Source

An ideal current source maintains a constant current through its terminals regardless of how much voltage is developed across it due to its surrounding environment. In other words, an ideal current source will **supply** or **draw** whatever amount of energy to or from the charges (i.e., voltage) necessary in order to maintain the specified current through its terminals.



Such a behaviour of an ideal current source also means that the voltage across it *depends* on the circuit in which it is placed in. Without looking at the circuit an ideal current source is in, it is not possible to determine the voltage across it. This behaviour is in contrast to that of a resistor – no matter where a resistor is placed in, if the current through the resistor is known then its voltage is known.

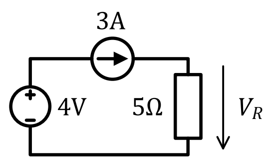
The idealised behaviour of an ideal current source also entails some practical considerations:

- If an ideal current source is not connected to anything (i.e., an "open-circuit"), an infinite amount of voltage will result (a constant current through infinite resistance), in practice the source will no longer function as intended

- Similarly, if ideal current sources of different values are connected **in series**, an infinite amount of voltage will result, in practice the sources will stop functioning as intended

Example 8. The circuit shown contains two ideal sources. The value of the voltage V_R across the 5Ω resistor is:

(a) 19 V
(b) 15 V
(c) −11 V
(d) −15 V



2.1.5. Power

Intended Learning Outcomes

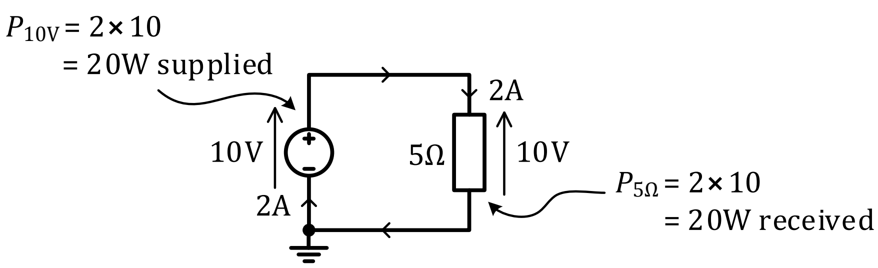
- Be able to interpret and relate the power of a circuit component to the underlying energy characteristic of the charges in an electric circuit.
- Be able to determine, and infer, the power of an electrical device based on knowledge of its voltage and current characteristics

All electrical devices function in the same way – by their ability to do *work* to the electrical charges flowing through them. The only difference between the different devices lies in the nature of the *work done* (i.e., energy conversion) that they are designed to do. For example, a battery does work by separating the charges through chemical reactions – work is done by the battery (on the charges) in isolating opposite charges, and in doing so, chemical potential energy is converted into electric potential energy. Likewise, a resistor does work by resisting the flow of charges – work is done by the charges (on the resistor) in moving through, and in doing so, electric potential energy is converted into heat.

Since all electrical devices require a continuous source of energy in order for them to operate, we often use the *rate* at which electric charges gain (from the device) or lose (to the device) electric potential energy as a measure for comparing between devices of similar purpose. Given our knowledge of the meaning of *voltage* across and *current* through a device, we therefore have a very intuitively expression for quantifying this rate:

$$\underbrace{p}_{=E\text{ supplied/received per sec.}} = \underbrace{v}_{=\Delta E\text{ across device per }Q} \times \underbrace{i}_{=\text{no. of }Q\text{ through device per sec.}}$$

The product thus measures the amount of energy transferred per second to or from the device. Dimensionally, the product also make sense since the unit is $\mathrm{JC^{-1}} \times \mathrm{Cs^{-1}} = \mathrm{Js^{-1}}$, *Joules per second*. This quantity is commonly known to you as the *power* (symbol: p or P) of an electrical device, and we usually use the unit *Watt* (symbol W) instead.



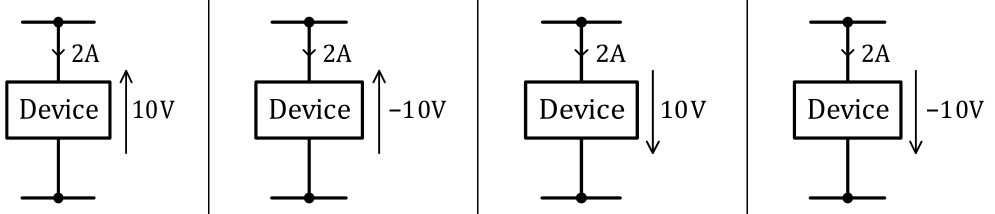
In the circuit above, it should be obvious that the power developed in both the 10V source and the 5Ω resistor is 20 Watts ($10 \cdot 2 = 20\text{W}$). However, clearly one of these components is *supplying* 20W while the other is *receiving* 20W of power:

- Ohm's law dictates there must be a 2A current through the 5Ω resistor, in the direction of the 10V voltage *drop* – each Coulomb of charge *loses* 10 joules of energy to the resistor as they flow through – the resistor receives 20W of power from the charges

- All of the 2A current must have come from the 10V source, in the direction of the 10V voltage *rise* – each Coulomb of charge *gains* 10 joules of energy from the source as they flow through – the source supplies 20W of power to the charges

There is thus a clear ambiguity in the power quantity of a device – it does **not** provide details as to whether the power calculated is supplied or received by the device, which can only be deduced by examining the underlying energy distribution of the charges across the device. This ambiguity occurs because we have failed to define the polarity and direction of the voltage and current as used in the power quantity (which the symbols themselves cannot convey).

To address this, electrical engineers have adopted an arbitrary standard convention[18] that takes the **current to be in the direction of a voltage drop** when defining the voltage-current product in the power expression. Said differently, the power quantity, by default, conveys the amount of power *received/absorbed* by the device of interest.



$$\begin{aligned} p &= +\ v \cdot i \\ &= +(10)\cdot(2) \\ &= +20\text{W} \\ &= 20\text{W received} \end{aligned}$$

$$\begin{aligned} p &= +\ v \cdot i \\ &= +(-10)\cdot(2) \\ &= -20\text{W} \\ &= 20\text{W supplied} \end{aligned}$$

$$\begin{aligned} p &= -\ v \cdot i \\ &= -(10)\cdot(2) \\ &= -20\text{W} \\ &= 20\text{W supplied} \end{aligned}$$

$$\begin{aligned} p &= -\ v \cdot i \\ &= -(-10)\cdot(2) \\ &= +20\text{W} \\ &= 20\text{W received} \end{aligned}$$

In this way we therefore have a unanimous way of measuring and expressing power of device – by measuring the voltage to be *against* the direction in which the current is measured, the (standard) product of the measured quantities (whether they are positive or negative) will tell you the amount of power *received* by the device, and a negative result simply indicate power is being *supplied* by the device.

---

[18] The *passive-sign convention*, the same one that electrical engineers have used to define the terminal characteristics of all devices. E.g., recall Ohm's law takes the voltage polarity across a resistor to be against the current direction through it (see **Sect. 2.1.3**).

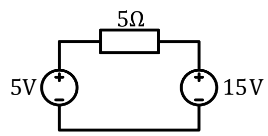
Example 9. The 5 V source in the circuit shown is

(a) Supplying 5W of power

(b) Supplying 10W of power

(c) Receiving 10W of power

(d) Receiving 15W of power



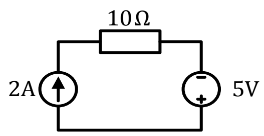
Example 10. In the circuit shown, the 2 A source is

(a) Receiving 10 W of power

(b) Receiving 30 W of power

(c) Supplying 30 W of power

(d) Supplying 50 W of power



Think about how your answer would differ in **Example 10** if the 10Ω resistor is replaced by a 1Ω resistor.

2.1.6. Kirchhoff's Current and Voltage Laws

Intended Learning Outcomes

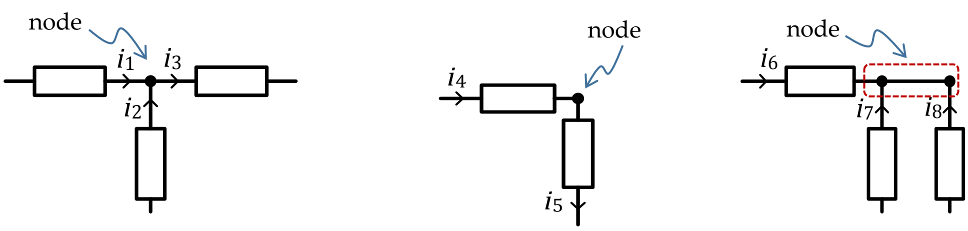
- Be able to apply, and make use of, Kirchhoff's laws to relate currents and voltages in an electric circuit
- Be able to incorporate, and integrate, Kirchhoff's laws into problems of relating unknown electrical quantities to known ones in an electric circuit.

Kirchhoff's Current Law (KCL)

A current, by definition, is the number of charges passing through a given point in a circuit in one second. Intuitively, the amount of charge arriving at any point in one second *must*[19] equal the amount of charge leaving that point in one second – this is the underpinning idea behind *Kirchhoff's Current Law (KCL)*.

KCL: *The algebraic sum of the currents leaving any point in a circuit equals zero.*

Such a statement make sense because a positive current leaving a *node*[20] in a circuit is equivalent to the *negative* of the current *entering* the node, hence by assuming all currents to be leaving[21], the algebraic signs will take care of those entering the node.



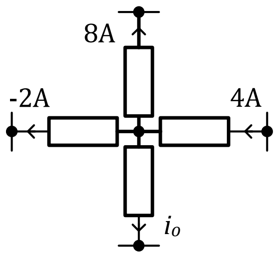
$i_1 + i_2 = i_3$ **or** $-i_1 - i_2 + i_3 = 0$

$i_4 = i_5$ **or** $-i_4 + i_5 = 0$

$i_6 + i_7 + i_8 = 0$ **or** $-i_6 - i_7 - i_8 = 0$

Example 11. For the partial circuit shown, determine the indicated current i_o.

(a) 2 A (b) -10 A (c) -6 A (d) -2 A

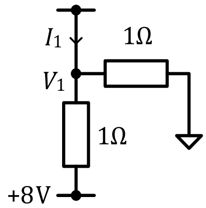


[19] If they are not equal then there would be, for example, an accumulation of charges at some point along the conducting path, and consequently a depletion of charges at some other point in the circuit by the same amount (law of conservation of charges). This separation of charges implies a force of attraction (*Coulomb's law*) along the conducting path thus invalidating the premise.

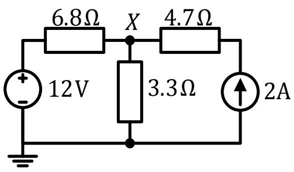
[20] A node in an electrical circuit is a point at which two or more circuit elements are connected together, either directly, or via ideal wires.

[21] You are not restricted to this arbitrary convention – you may choose to sum currents entering, or equate those entering and leaving. You will obtain the **same** KCL equations.

Example 12. Consider the part of a circuit shown. Write a KCL equation for the node of potential V_1.



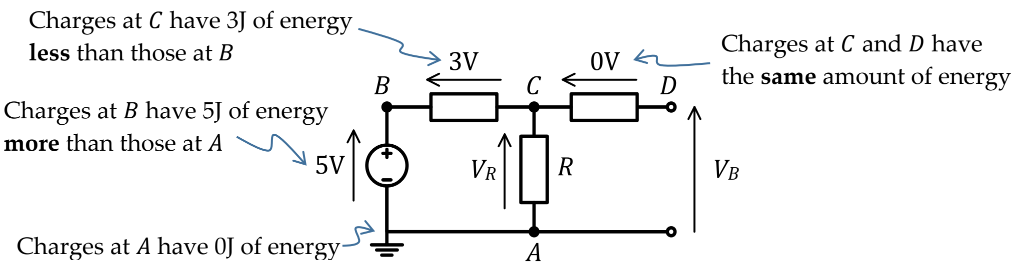
Example 13. Node X has a potential of V volts. Write a KCL equation at the node X.



Kirchhoff's Voltage Law (KVL)

A voltage, by definition, is the **difference** in energy of each unit of charge **between two points**. Such an understanding immediately suggests that the total voltage *rises* around a loop[22] in a circuit *must*[23] equal to the total voltage *drops* in the same loop – this is the underpinning idea behind *Kirchhoff's Voltage Law (KVL)*.

KVL:	*The algebraic sum of the voltages around a loop in a circuit equals zero.*



Clearly the voltage *across* the resistor R is 2V given knowledge of the energy differences (i.e., voltages) around the loop "ABC". Applying KVL we indeed have

$$5 = 3 + V_R \quad \textbf{or} \quad 5 - 3 - V_R = 0 \quad \Rightarrow \quad V_R = 2\text{V}.$$

Similarly, from the voltages around loop "ACD" we deduce that $V_B = V_R$, in agreement with KVL

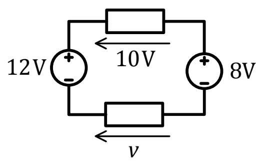
$$V_R = 0 + V_B \quad \textbf{or} \quad V_R - 0 - V_B = 0 \quad \Rightarrow \quad V_B = V_R.$$

Notice "$ABCD$" also form a loop and so we also have

$$5 = 3 + 0 + V_B \quad \textbf{or} \quad 5 - 3 - 0 - V_B = 0 \quad \Rightarrow \quad V_B = 2\text{V} \text{ as expected from above.}$$

Example 14. In the circuit shown, the voltage v as indicated is

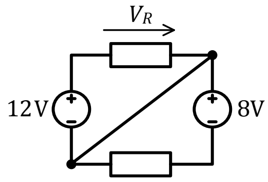
(a) 6 V (b) 10 V (c) 14 V (d) 30 V



[22] A loop in an electrical circuit is a closed path starting from a node and proceeding through the circuit, eventually returning to the same node. A loop does **not** necessarily need to be a 'complete' circuit.

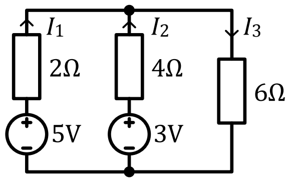
[23] Otherwise this suggests that the charges, upon returning to the original point, have, for example, gained energy, and over time more energy would be absorbed (by the charges) than is supplied, violating the law of conservation of energy.

Example 15. In the circuit shown, the voltage V_R as indicated is



(a) -12 V (b) -4 V (c) 4 V (d) 12 V

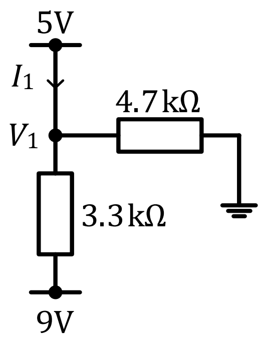
Example 16. Write a KVL equation for each of the three loops in the circuit shown[24].



[24] Note that the resulting equations are *dependent* – each one is a linear combination of the other two. We thus need more information if we were to solve for the currents, and in this case, the KCL equation $I_1 + I_2 = I_3$.

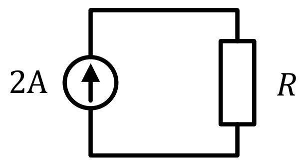
2.1.7. Problems

1. Find the voltage V_1 and current I_1 in the circuit shown.

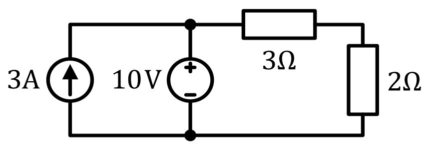


2. In your house, two 230 V light bulbs are turned on. Bulb A has a thinner tungsten wire filament than Bulb B. Which one is likely to be brighter? Explain your answer.

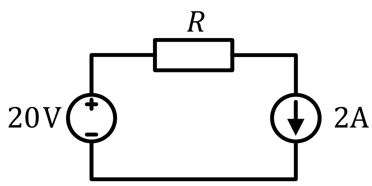
3. A resistor is connected in series with a 2A current source. If the resistor receives 20 W of electrical power. Find the voltage v across the current source, and the resistance of the resistor.



4. Find the current (magnitude and direction) through the voltage source and through the 2Ω resistance. Find the powers developed for each of the two sources. State whether each source is delivering power or receiving it.



5. For the circuit shown below



(a) If the current source receives 20 W of power, determine the value of the resistance R and the power of the voltage source. State whether the voltage source receives or supplies power.

(b) Investigate the power characteristic of the current source when the resistor is replaced with a $R = 10\Omega$ resistor.

(c) Repeat Question 6(c) when $R = 15\Omega$.

Solutions to Problems

1. $V_1 = 5$ V, $I_1 = -0.15$ mA.
2. Bulb B.
3. $v_{2\text{A}} = 10$ V polarity upwards, $R = 5\ \Omega$.
4. $i_{10\text{V}} = 1$ A downwards, $i_{2\Omega} = 2$ A downwards,
voltage source absorbs 10W, current source delivers 30W.

5(a). $R = 5\ \Omega$, the voltage source is delivering 40 W of power.

5(b). The power of the current source is 0 W, it is neither supplying nor receiving power.

5(c). The current source is supplying 20 W of power.

2.2 Methods & Theorems of Circuit Analysis

When we wish to analyse a circuit, what we mean is to determine everything there is to know about the circuit. In other words, the general goal of circuit analysis is to determine sufficient information about the circuit such that the **current** through, and **voltage** across *every* elements in a circuit can be easily deduced. All other characteristics (e.g., power) can be inferred from the knowledge of these two parameters.

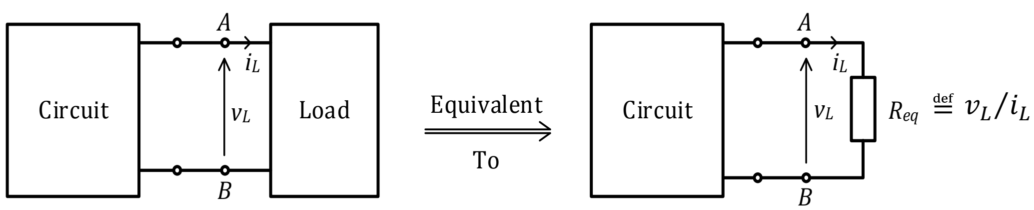
Surprisingly, as it turns out, *Kirchhoff's laws*, along with the *terminal behaviour* of the circuit elements (e.g., resistors, capacitors, and inductors) are all that is needed to characterise **all** electrical circuits containing them. Nonetheless, useful results and 'shortcuts' have been developed over time to simplify and/or speed up the analysis, thereby reducing the complexity of the task. Here, we present *some* of the common tools and results.

2.2.1. Equivalent Resistance

Intended Learning Outcomes

- Be able to recognise and identify resistive networks that are connected in series-, and parallel-connection.
- Be able to re-draw, and heuristically simplify two-terminal resistive networks consisting of series- and parallel-connections into their equivalent resistance by inspection.
- Be able to determine, interpret, and apply the notion of equivalent resistance for two-terminal resistive networks in finding relevant electrical quantities in an electric circuit.

When a load[25] is connected to a circuit, it will draw a certain amount of current and voltage from the circuit depending on the internal constituents of the load. A different load will therefore result in a different current or voltage being drawn from the same circuit. If we are only interested in the *terminal behaviour* of the load[26], and not its internal operations, then for all intents and purposes, we can replace the load as *perceived* by the circuit with a resistance according to Ohm's law – **the ratio of the voltage across the load to the current it draws from the circuit –** without affecting the circuit behaviour. This resistance is commonly referred to as the *equivalent resistance*[27] of the load seen across its terminals.



[25] A *load* is a device or circuit connected to the output of some other circuit.

[26] This may be when we are interested in optimising and fine-tuning a circuit for which a fixed load is connected to, or purely to simplify the subsequent analyses of relevant electrical quantities *external* to the load.

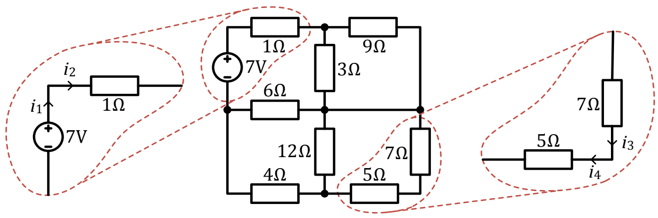
[27] Depending on the context, this is also known as the *input resistance* of the load.

You may have come across this notion of equivalent resistance when 'combining' two resistances in *series,* or in *parallel* into a single resistance. It is important that we understand the *meaning* of these 'total' resistances in relation to the original circuit they result from, as we often exploit these as means to deduce, or investigate, behaviours about the *original* circuit and **not** just as a way of 'simplifying' it.

Equivalent Series Resistance

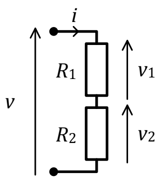
Components are said to be *in series* if and only if there is only *one* path for the charges to flow through them. In other words, we say two components are in series if they share an *exclusive* node in which nothing else is connected to, and this should make it obvious why the current through components in series *must* be the same – every charge that leaves one component has no choice but to enter the other, a direct application of *Kirchhoff's current law*.

$$-i_1 + i_2 = 0$$



$$-i_3 + i_4 = 0$$

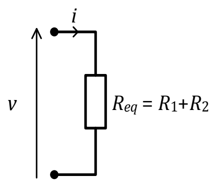
When two resistors are connected in series, the *total* voltage across, and the current through the resistors is *equivalent* to a *single* resistor whose resistance is equal to the *arithmetic* sum of the two resistances.



Applying KVL around the loop, with Ohm's law we have

$$v = \underbrace{iR_1}_{=v_1} + \underbrace{iR_2}_{=v_2} \quad \Rightarrow \quad v = i\,\underbrace{(R_1 + R_2)}_{=R_{eq}},$$

which shows the series resistor has an equivalent resistance of



$$R_{eq} = \frac{v}{i} = R_1 + R_2.$$

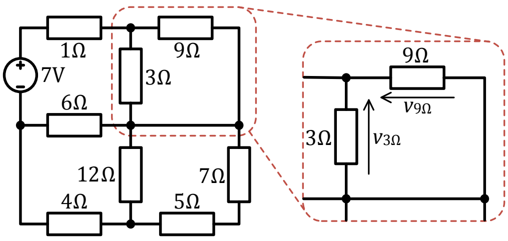
A direct generalisation of the result to N resistors in series is thus

$$R_{eq} = R_1 + R_2 + \cdots + R_N = \sum_{k=1}^{N} R_k.$$

Notice this also reveals the fact that a series equivalent resistance is **larger** than any of the resistors contributing to the equivalent, and in particular, R_{eq} is larger than the *largest* resistor in $R_1, R_2, \ldots, R_N$.

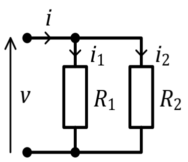
Equivalent Parallel Resistance

Components are said to be *in parallel* if and only if they are connected together on *both* ends. In other words, we say two components are in parallel if they share both nodes, and this should make it obvious why the voltage across components in parallel *must* be the same – the potential difference across each component is measured across the *same* two nodes that they share, the energy difference in the charges between the two ends of the components must therefore be the same, a direct application of *Kirchhoff's voltage law*.



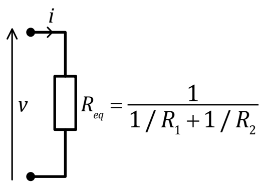
$$v_{3\Omega} - v_{9\Omega} = 0$$

When two resistors are connected in parallel, the voltage across, and the *total* current through the resistors is *equivalent* to a *single* resistor whose resistance is equal to the *harmonic* sum of the two resistances.



Applying KCL at the top node, with Ohm's law we have

$$i = \underbrace{\frac{v}{R_1}}_{=i_1} + \underbrace{\frac{v}{R_2}}_{=i_2} \quad \Rightarrow \quad v = i\underbrace{\left(\frac{1}{1/R_1 + 1/R_2}\right)}_{=R_{eq}},$$



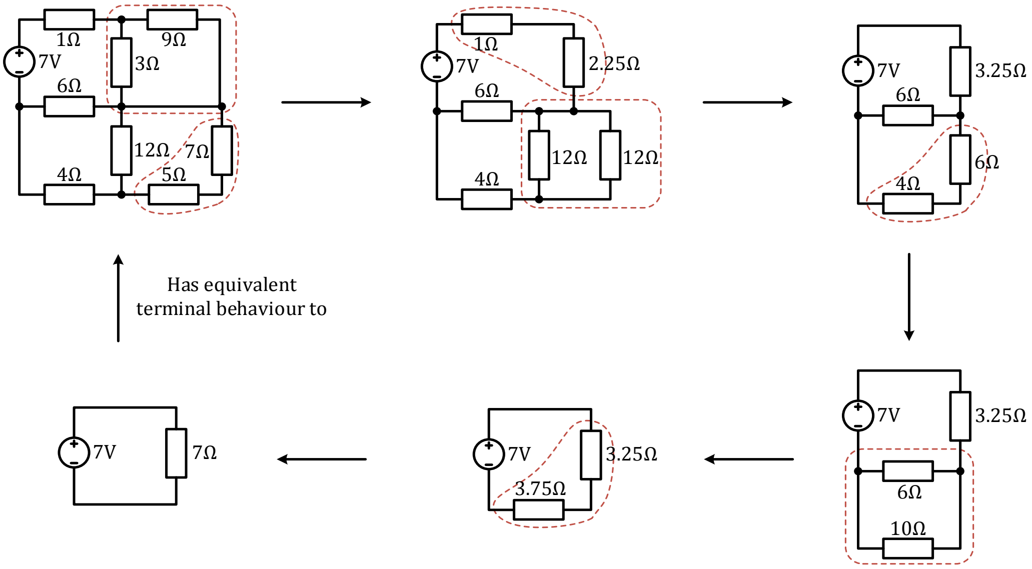
which shows the parallel resistors have an equivalent resistance of

$$R_{eq} = \frac{v}{i} = \frac{1}{1/R_1 + 1/R_2}.$$

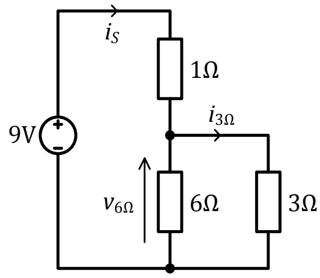
A direct generalisation of the result to N resistors in series is thus

$$R_{eq} = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \cdots + \frac{1}{R_N}} = \frac{1}{\sum_{k=1}^{N} \frac{1}{R_k}}.$$

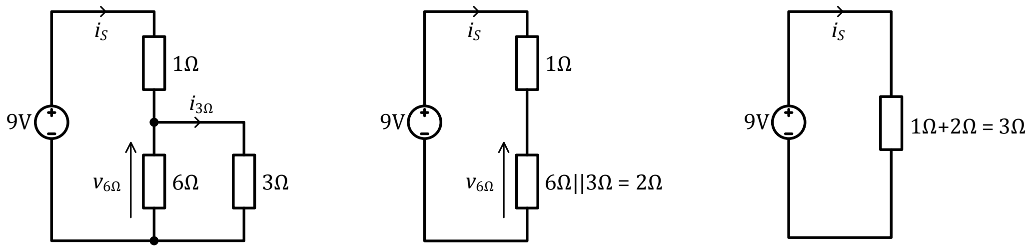
Notice this also reveals the interesting fact that a parallel equivalent resistance is **smaller** than any of the resistors contributing to the equivalent, and in particular, R_{eq} is smaller than the *smallest* resistor in R_1, R_2, ..., R_N.



It should be clear from its origins that the resulting equivalent resistance of a circuit is equivalent to the original in terms of the behaviour across the terminals in which the equivalent resistance is considered, but **not** the internal characteristics – in the above example, clearly the equivalent resistance of 7Ω is equivalent to the origin in terms of what the 7V voltage source 'sees', but the current through, say, the 3Ω in the original circuit no longer exists in the equivalent circuit[28]. We will explore this point in the following circuit:



Suppose we are interested in finding the indicated parameters i_S, $v_{6\Omega}$, and $i_{3\Omega}$ – we will do so by *interpreting* the intermediate or eventual equivalent resistances that result in relation to the original circuit. Consider the following sequence of equivalent circuits that arise from combining the relevant resistances into their equivalent:



Original Circuit — Equivalent Circuit 1 — Equivalent Circuit 2

[28] The final equivalent circuit in this case simply reveals that the total current supplied by the source (drawn by the circuit) is 1A. Notice this will corresponds to the current through the 1Ω resistor in the original, but nothing else.

It is not immediately obvious from the original circuit what any of the interested quantities would be. However, notice if our interest is solely on finding i_S then we can combine the two parallel resistances into a single equivalent resistance of 2Ω – the equivalence is in terms of the voltage across, and the current through its terminals in place of the two parallel resistors in the original circuit. This make sense since the *total* current through the parallel connection *is* i_S and this is reflected through the series connection of the 2Ω equivalent resistance to the 1Ω in Equivalent Circuit 1. Notice $i_{3\Omega}$ no longer exists, one cannot therefore deduce this current directly from Equivalent Circuit 1.

Similarly, when we are only interested in finding i_S, we can combine the two series resistances in Equivalent Circuit 1 into a single equivalent resistance of 3Ω – the equivalence is in terms of the voltage across, and the current through its terminals in place of the two series resistor in Equivalent Circuit 1. This make sense since the *total* voltage across the series connection *is* 9V and this is reflected through the single 3Ω equivalent connected directly across the 9V source in Equivalent Circuit 2. Notice now $v_{6\Omega}$ no longer exists, one cannot therefore deduce this voltage directly from Equivalent Circuit 2.

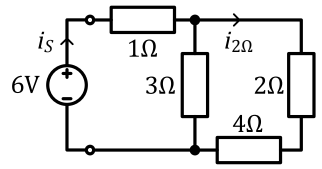
From Equivalent Circuit 2, it is now obvious that i_S is 3A (Ohm's law). This newly deduced information thus enable us to make use of in Equivalent Circuit 1 and deduce that $v_{6\Omega} = 6\text{V}$. Finally, reverting back to the original circuit with these new details reveals that $i_{3\Omega} = 2\text{A}$.

From this example, we see that the necessity in 'simplifying' a circuit, or a section of it, into its equivalent resistance is contingent on what *we* are going to use it for in the subsequent analysis, and this relies on our ability to correctly *interpret* what the 'equivalence' is in relation to the original circuit – Sometimes replacing a circuit by its equivalent resistance may directly trivialise the analysis, while sometimes it may not directly lead to the quantities of interest *but* allows quick deduction of new details about the original circuit for us to *revert* to and make use of. Other times it may simply get us further away from what we wish to achieve if we arbitrarily replace circuits by their equivalent resistances without careful planning.

It is worth noting that **not** all electrical connections are either in series or parallel[29], and so one should take care in simplifying connections into their equivalent resistances – sometimes a circuit cannot be reduced into a single equivalent resistance through series and parallel reductions.

[29] Other connections exist, such as the wye (or "Y") and delta (or "Δ") connections. The simplification of these networks are beyond the scope of this course.

Example 17. Determine the equivalent resistance of the circuit seen by the voltage source. Use the equivalent resistance to find the current i_S drawn from the source (by the circuit), and hence the current $i_{2\Omega}$ through the 2Ω resistor.



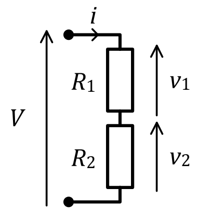
2.2.2. Voltage and Current Division Principles

Intended Learning Outcomes

- Be able to recognise the applicability of voltage and current division principles in an electrical circuit.
- Be able to incorporate, and integrate, the voltage and current division principles into problems requiring finding electrical quantities in a circuit

Voltage Division Principle

The voltage division principle or *voltage divider* is a convenient method for determining the voltage across resistive components when they are connected **in series**. It is so-called because the result suggests that the voltage across a single resistor equals the applied voltage (across all resistors) divided down by some factor. **The voltage divider is a direct consequence of KVL**.



Applying KVL around the loop, together with Ohm's law we have

$$V = \underbrace{iR_1}_{=v_1} + \underbrace{iR_2}_{=v_2} \quad \Rightarrow \quad i = \frac{1}{R_1 + R_2}V.$$

and so

$$v_1 = iR_1 = \left(\frac{R_1}{R_1 + R_2}\right)V \quad \text{and} \quad v_2 = iR_2 = \left(\frac{R_2}{R_1 + R_2}\right)V.$$

The voltage across a resistor **in series** with other resistors is directly proportional to the total voltage applied across these resistors by the fractional contribution of its resistance to the total sum of resistances.

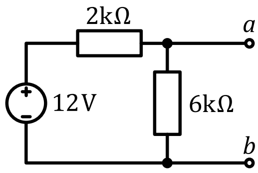
A direct generalisation of the result to N resistors in series is thus

$$v_p = \left(\frac{R_p}{R_1 + R_2 + \cdots + R_N}\right)V \qquad \text{for } p = 1,2,\ldots,N.$$

Notice the result only holds when the resistors of interest are **in series**, since it relies on the fact that the current through the resistors are equal. We emphasise this in the following example.

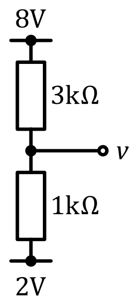
Example 18. Find the output voltage, V_{ab}, for the circuit shown when

(a) terminals a - b is an open-circuit, and
(b) a 3kΩ resistor is placed across a-b.



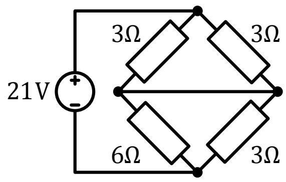
Example 19. The voltage v at the centre node in the partial circuit shown is

(a) 0.5 V (b) 1.5 V (c) 3.5 V (d) 4.5 V



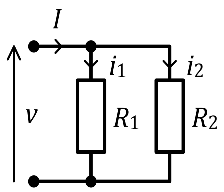
Example 20. For the circuit shown, the voltage across the 6Ω resistor is

(a) 0 V (b) 10.5 V (c) 12 V (d) 14 V



Current Division Principle

The current division principle or *current divider* is a convenient method for determining the current through resistive components when they are connected **in parallel**. It is so-called because the result suggests that the current through a single resistor equals the total current divided down by some factor. **The current divider is a direct consequence of KCL**.



Applying KCL at the top node, together with Ohm's law we have

$$I = \underbrace{\frac{v}{R_1}}_{= i_1} + \underbrace{\frac{v}{R_2}}_{= i_2} \qquad \Rightarrow \qquad v = \frac{1}{1/R_1 + 1/R_2} I.$$

and so

$$i_1 = \frac{v}{R_1} = \left(\frac{1/R_1}{1/R_1 + 1/R_2} \right) I \;\; \text{and} \;\; i_2 = \frac{v}{R_2} = \left(\frac{1/R_2}{1/R_1 + 1/R_2} \right) I.$$

> The current through a resistor **in parallel** to other resistors is directly proportional to the current supplied to these resistors by the fractional contribution of its conductance[30] to the total sum of conductances.

A direct generalisation of the result to N resistors in parallel is thus[31]

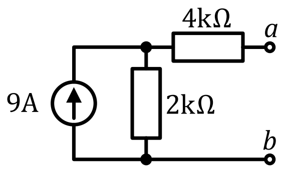
$$i_p = \left(\frac{1/R_p}{1/R_1 + 1/R_2 + \cdots + 1/R_N} \right) I \qquad \text{for } p = 1,2, \dots, N.$$

Notice the result only holds when the resistors of interest are **in parallel**, since it relies on the fact that the voltage across the resistors are equal. We emphasise this in the following example.

[30] The reciprocal of resistance is also known as the *conductance*, G, with a unit of Siemens, S.

[31] For the special case of $n = 2$, you may have seen the equivalent expression $i_1 = (R_2/(R_1 + R_2))I$ and $i_2 = (R_1/(R_1 + R_2))I$ which simply falls out of algebraic manipulation of the expression shown. However, such an expression does **not** generalise to $n > 2$, while the expression shown does and preserves similarity to that of the voltage divider.

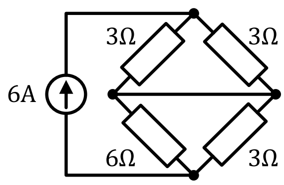
Example 21. Find the current through terminals a-b for the circuit shown when



(a) the terminals a-b is short-circuited,

(b) a 3kΩ resistor is placed across a-b.

Example 22. For the circuit shown, the current through the 6Ω resistor is



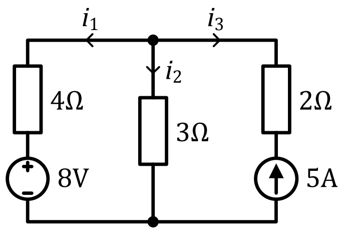
(a) 0 A (b) 1 A (c) 2 A (d) 3 A

2.2.3. Node-Voltage Analysis

Intended Learning Outcomes

- Be able to recognise and differentiate between nodes and essential nodes in a circuit.
- Be able to write down the governing node-voltage equations necessary to completely characterise a circuit containing one non-reference essential node.
- Be able to incorporate and integrate node-voltage analysis into problems requiring finding electrical quantities in a circuit.

Not all circuits can be simplified into a series/parallel equivalent[32], consequently methods such as the voltage or current dividers cannot be used in these cases, and one must fall back to Kirchhoff's laws. Consider the following circuit:



It should be clear that there are no resistors in series or in parallel, and the circuit cannot be reduced any further[33]. To analyse the circuit, we *could* apply KVL and KCL to get

KVL: $8 + 4i_1 - 3i_2 = 0,$ (1) KCL: $i_1 + i_2 + i_3 = 0,$ (2)

$i_3 = -5\text{ A},$ (3)

solving which completes the characterisation of the circuit. While feasible, such a general approach is clearly non-systematic, requiring careful planning of which laws to use and when[34].

On the other hand, **if** we can identify the voltages at the *essential* nodes[35] (i.e., the "node-voltages") of the circuit, then *all* other voltages and currents in the circuit can be trivially deduced thereby completing the analysis – this is the rationale behind the *node-voltage analysis* method.

Node-voltage analysis is the use of KCL to find the voltages at nodes by expressing the currents through those nodes in terms of their voltages.

---

[32] Recall **not** all connections are either in series or parallel, connections such as the "Y" or "Δ" exist.

[33] At least not without using additional circuit tools such as *Thevenin's theorem*, see **Section 2.2.6**.

[34] For example, why didn't we write out a KVL equation for the right loop instead? Because we do **not** know the voltage across the current source, $V_{5\text{A}}$, and would therefore end up with an additional unknown if we chose to do so: $3i_2 - 2i_3 - V_{5\text{A}} = 0$, where the polarity of $V_{5\text{A}}$ is taken to be upwards.

[35] An *essential* node is a point where **three** or more elements join.

Node-voltage analysis is carried out in the following way:

1. Select one of the essential nodes as the reference (ground) node, in which all other node voltages will be calculated with respect to.
2. Write a KCL equation for each of the remaining essential nodes in terms of the unknown essential node voltages.
3. Solve the node-voltage equations simultaneously.

Note that the reference node can be arbitrary in general, however, selecting an appropriate one may speed up the analysis.

Suppose we wish to analyse the circuit above via node-voltage analysis. We will do so by selecting the reference node to be at various points in the circuit and examine its implications.

Case 1: Selecting Bottom Essential Node as Ground

There is only one unknown node-voltage (at the top essential node), and so we only need one equation to solve the circuit. Applying KCL at this essential node we have

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

Expressing each of these currents in terms of this node-voltage (say, V_1) using Ohm's law we have

$$i_1 = \frac{V_1 - 8}{4}, \qquad i_2 = \frac{V_1 - 0}{3}, \qquad \text{and} \qquad i_3 = -5\text{ A},$$

giving the node-voltage equation $$\frac{V_1 - 8}{4} + \frac{V_1}{3} - 5 = 0,$$

solving which we deduce $$V_1 = \frac{5 + 8/4}{1/4 + 1/3} = 12\text{ V}.$$

The result says the voltage at the top node is **12V higher than the bottom node.**

Case 2: Selecting Top Essential Node as Ground

There is only one unknown node-voltage (at the bottom essential node), and so we only need one equation to solve the circuit. Applying KCL at this essential node we have

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

Expressing each of these currents in terms of this node-voltage (say, V_2) using Ohm's law we have[36]

$$i_1 = \frac{0-(V_2+8)}{4}, \quad i_2 = \frac{0-V_2}{3}, \quad \text{and} \quad i_3 = -5\text{ A},$$

giving the node-voltage equation: $$-\left(\frac{-V_2-8}{4}\right) - \left(\frac{-V_2}{3}\right) - (-5) = 0,$$

solving which we deduce $$V_2 = \frac{-5-8/4}{1/4+1/3} = -12\text{ V}.$$

The result is consistent with before – the voltage at the bottom node is **12V lower than the top node.**

Case 3: Selecting an Arbitrary Node as Ground

If we select, for example, the non-essential node between the 5A source and the 2Ω resistor as our reference, we now have two unknown essential node voltages at the top (say, V_3) and bottom (say, V_4) since it is not immediately obvious what the voltages at these nodes are with respect to the chosen reference.

Applying KCL at the top and bottom essential nodes respectively, we have

$$\frac{V_3-(V_4+8)}{4} + \frac{V_3-V_4}{3} + \frac{V_3}{2} = 0 \qquad \Rightarrow \qquad 13V_3 - 7V_4 = 24$$

and

$$-\left(\frac{V_3-(V_4+8)}{4}\right) - \left(\frac{V_3-V_4}{3}\right) + 5 = 0 \qquad \Rightarrow \qquad -V_3 + V_4 = -12,$$

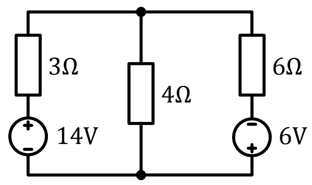
from which $V_3 = -10$V and $V_4 = -22$V.

The result is, again, consistent to previous analyses – the voltage V_3 at the top node is **12V higher than the bottom node**.

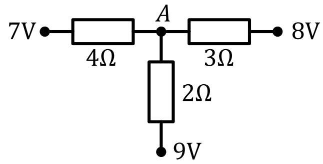
[36] Remember there is an inherent polarity and direction assumed in Ohm's law $v = iR$: the polarity of v and direction of i are **opposite**. This means for example, given the direction of i_1, the polarity across the 4Ω resistor must be **upwards**, and so the voltage across it, with the top node *assumed* to be at a higher potential, is $\left(0-(V_2+8)\right)$ volts.

Notice in all three cases, the calculated node voltages are necessarily different since they are **with respect to different points** in the circuit. However, the voltage across and current through all elements in the circuit are **identical** in all three cases – they must, since the circuit has not changed, only the way we have chosen to analyse it.

Example 23. Determine the current through the 4Ω resistor by means of node-voltage analysis



Example 24. Find the voltage at node A.



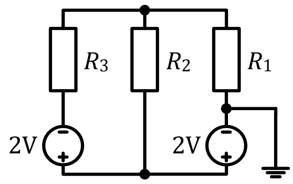
Example 25. A correct node-voltage equation for the node V is:

(a) $\dfrac{V}{R_1}+\dfrac{V}{R_2}+\dfrac{V+2}{R_3}=0$

(b) $\dfrac{V}{R_1}+\dfrac{V-2}{R_2}+\dfrac{V+2}{R_3}=0$

(c) $\dfrac{V+2}{R_1}+\dfrac{V}{R_2}+\dfrac{V+2}{R_3}=0$

(d) $\dfrac{V}{R_1}+\dfrac{V-2}{R_2}+\dfrac{V}{R_3}=0$



2.2.4. Superposition Theorem

Intended Learning Outcomes

- Be able to recognise and re-draw the equivalent of a circuit containing independent sources when they are set to zero.
- Be able to incorporate the notion of superposition in finding electrical quantities in a circuit containing multiple independent sources.

When tackling complex engineering problems, a common strategy is to reduce, or break, the problem down into manageable sub-problems. By solving these 'simpler' problems separately, the solution to the original problem is found upon integrating the results together. The *superposition theorem* is a powerful tool in circuit analysis that demonstrates such a philosophy for solving complex circuits in an intuitive way. The theorem says:

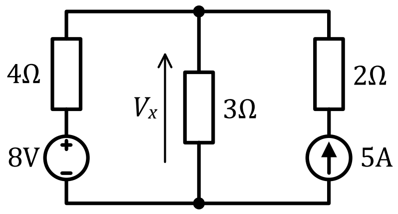
Any linear[37] response (i.e., voltage or current) in a circuit containing multiple sources is equal to the algebraic sum of the contributions to that response due to each sources acting individually with all others sources set to zero.

To set a voltage source to zero we replace it with a ***short-circuit***, to set a current source to zero we replace it with an ***open-circuit***[38].

[37] The proof of the theorem is beyond the scope of the course, but it relies on the so-called *linearity* property of the circuit response of interest. Roughly, a response $y = f(x)$ due to some input x is *linear* if it satisfies the property $f(ax_1 + bx_2) = af(x_1) + bf(x_2)$. One can easily check that the voltage and current behaviour of a resistor ($v = Ri$), a capacitor ($i = C\,dv/dt$), and an inductor ($v = L\,di/dt$) are all linear. The *power* quantity $p = vi$ in those components on the other hand, is **not**, and so superposition cannot be used to find the power developed.

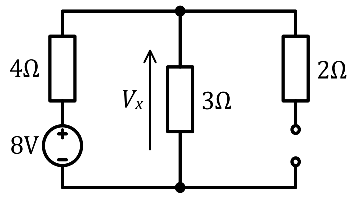
[38] Why? Because an ideal wire (i.e., short-circuit) has zero voltage across it (but potentially non-zero current through it). Likewise, a lack of electrical path (i.e., open-circuit) has zero current through it (but potentially non-zero voltage across it).

Consider the circuit shown below, we will determine the voltage V_x across the 3Ω resistor by superposition.



Response due to 8V Source

First, we consider the contribution of the 8V source to the response V_x. Setting the current source to zero, the circuit looks like:

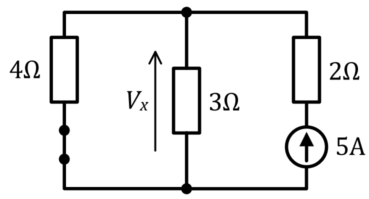


The voltage across the 3Ω resistor can be found via, for example, the voltage divider

$$V_{x(8\text{V})} = \frac{3}{3+4} \cdot 8 = \frac{24}{7}\ \text{V, polarity } \textbf{upwards}.$$

Response due to 5A Source

Second, we consider the contribution of the 5A source to the response V_x. Setting the voltage source to zero, the circuit looks like:



The voltage across the 3Ω resistor can be found via, for example, the current divider and Ohm's law:

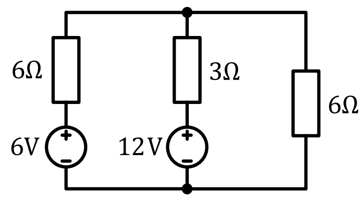
$$i_{3\Omega} = \frac{1/3}{1/3+1/4} \cdot 5 = \frac{20}{7}\ \text{A} \quad \Rightarrow \quad V_{x(5\text{A})} = 3i_{3\Omega} = \frac{60}{7}\ \text{V, polarity } \textbf{upwards}$$

By the superposition theorem, the actual voltage V_x across the 3Ω resistor with both sources in effect is therefore the algebraic sum of these two responses

$$V_x = V_{x(8\text{V})} + V_{x(5\text{A})} = \frac{24}{7} + \frac{60}{7} = 12\ \text{V polarity upwards.}$$

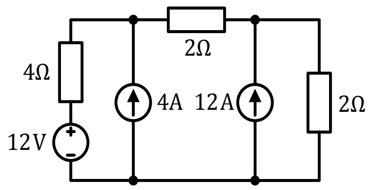
Notice the result is consistent with what we have deduced via node-voltage analysis from the previous section.

Example 26. By means of superposition[39], find the current (magnitude and direction) through the 3Ω resistor.



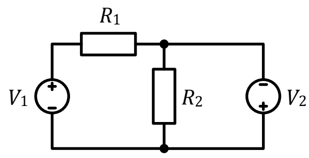
[39] As further practice try determining the current by using node-voltage analysis, and convince yourself that you must obtain the same answer.

Example 27. By means of superposition[40], find the voltage across the 4Ω resistor, make sure to indicate its polarity.



Example 28. [Optional] In the circuit shown, R_1 receives 3W of power from each of the sources when they are considered separately (by setting the other to zero). The power absorbed by R_1 when both sources are in effect is:

(a) 3W (b) 6W (c) 9W (d) 12W



[40] As before, try finding the current of interest via node-voltage analysis to convince yourself of the equivalence between the methods. Note that in this case, you will need to solve a minimum of two node-voltage equations simultaneously as there are at least two unknown essential node voltages.

2.2.5. Thevenin's Theorem

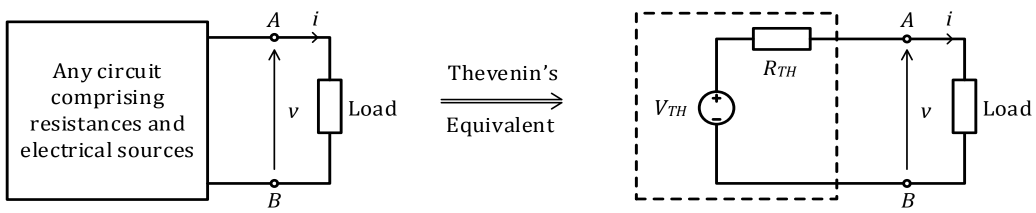
Intended Learning Outcomes

- Be able to identify and distinguish between the terms 'circuit' and 'load' as relevant in the context of defining the terminal characteristics across a pair of terminals in an electrical system.
- Be able to apply, and provide relevant interpretation to, Thevenin's theorem in the context of representing the terminal behaviour of an electric circuit.
- Be able to incorporate, and make use of, Thevenin's theorem in analysing resistive electrical systems.

Often, when we connect a load (e.g., mobile phone, speaker) to a circuit (e.g., power supply, amplifier), we are more interested in *what* the voltage and current supplied (by the circuit) to the load are, as opposed to *how* the circuit is able to produce them. In other words, we are usually more concerned with the *terminal characteristics* of the circuit (its output voltage, and output current) as seen by the load over the details of the circuit's internal operations that give rise to the observed responses.

When we are just interested in the terminal behaviour of a circuit, the *Thevenin's theorem* is a convenient tool that 'hides' the complexity of the circuit behind its terminals, thereby considerably simplifies the analysis of the *external surroundings* connected to the circuit. The theorem says:

The external behaviour of any linear[41] two-terminal circuit at its terminals can be modelled by an equivalent circuit consisting of a voltage source V_{TH} in series with a resistance[42] R_{TH}, where V_{TH} is the open-circuit voltage across the terminals, and R_{TH} is the equivalent resistance looking into the circuit from the terminals when independent sources are set to zero[43].



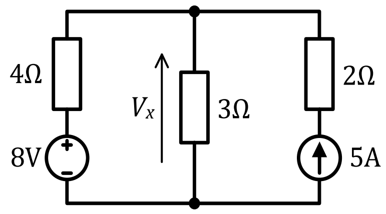
Thevenin's theorem thus allows us to effectively analyse the behaviour of a circuit from its terminals subject to different external environments without needing to re-analyse the entire system internally due to the external change.

[41] The proof of Thevenin's theorem relies on the *superposition* theorem which in itself requires the *linearity* property of the responses in the circuit.

[42] Thevenin's theorem also applies to steady-state responses due to AC excitations in circuits containing *reactive* components (inductors or capacitors). In these cases, the resistance is generalised to an *impedance* which consists of a series resistance and *reactance*. This is, of course, outside the scope of this course.

[43] Remember to set a voltage source to zero we replace it with a short-circuit, while to set a current source to zero we replace it with an open-circuit.

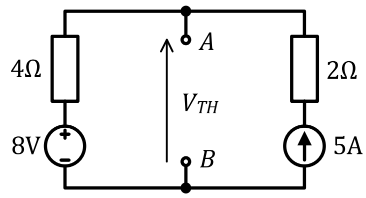
Consider the circuit shown below, we will determine the voltage V_x across the 3Ω resistor via Thevenin's theorem.



By Thevenin's theorem, we can replace the circuit seen by the 3Ω resistor by a series voltage source V_{TH} and resistance R_{TH} without affecting the behaviour experienced by the 3Ω. To do so, we *disconnect* the 3Ω from the circuit, and determine (1) the open-circuit voltage across the pair of terminals the 3Ω was connected to, and (2) the equivalent resistance of the circuit as seen from the open terminals.

Thevenin Voltage

To find the open-circuit voltage V_{AB} it should be clear that, with the 3Ω disconnected, the current through the 4Ω must be 5A, and so by KVL we have[44]

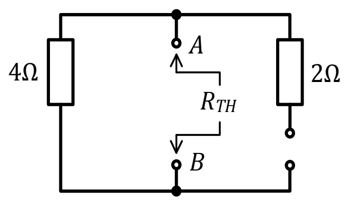


$$V_{AB} = 8 + 4 \cdot 5 = 28\text{ V},$$

this is the Thevenin voltage V_{TH}, i.e., $V_{TH} = V_{AB} = 28$ V.

Thevenin Resistance

To find the equivalent resistance seen from the terminals A-B we ask the question[45] – if a charge is to flow from terminal A, through the circuit, to terminal B (or vice versa), what resistive network (and hence the equivalent resistance) will the charge see?



Clearly, in this case since there is no conducting path along the 2Ω branch, the charge has to flow through the 4Ω in order to go from terminal A to terminal B and so immediately,

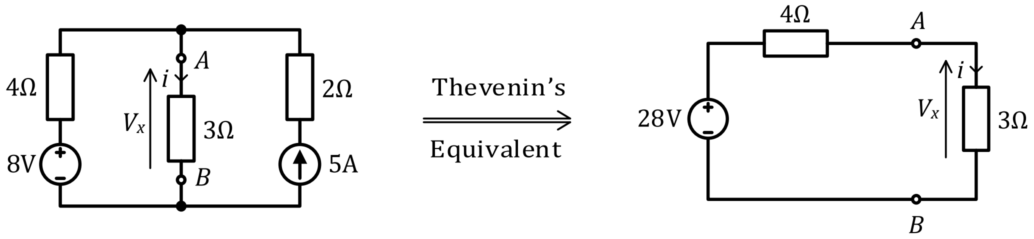
$$R_{TH} = 4\Omega,$$

this is the Thevenin resistance.

[44] Alternatively, apply node-voltage analysis at node A with node B selected as ground to obtain the node-voltage equation $-5 + (v - 8)/4 = 0 \Rightarrow v = 28$ V as expected.

[45] Recall the formal approach in calculating the equivalent resistance across a pair of terminals is to apply a *notional* voltage (current) across it, and determine the current (voltage) drawn from the source (by the circuit). The ratio of the two (Ohm's law) then gives the equivalent resistance of the circuit. In this particular case, applying, arbitrarily, an 8V across the terminals A-B, the current drawn is 2A and thus $8/2 = 4\Omega$ as expected.

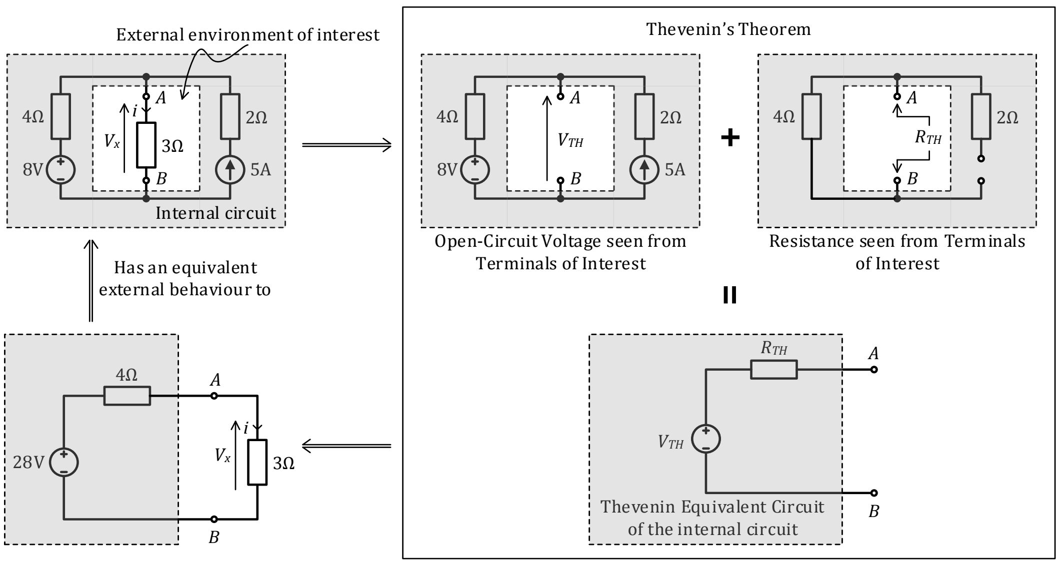
The Thevenin equivalent of the circuit seen by the 3Ω resistor thus consists of a voltage source of 28V in series with a resistance of 4Ω. This means we can replace the circuit that the 3Ω is connected to by this equivalent *without* affecting the voltage across and current through the 3Ω



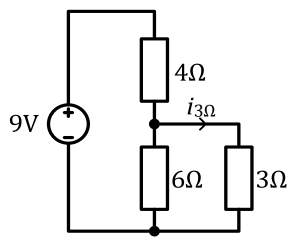
and this trivially reduces the problem of finding V_x to that of a voltage divider

$$V_x = \frac{3}{3+4} \cdot 28 = 12 \text{ V}$$

which is consistent with the prior analyses via *node-voltage analysis* and *superposition theorem*.

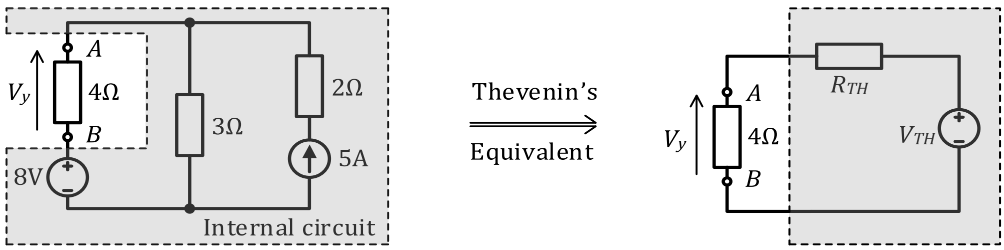


Example 29. Determine the Thevenin equivalent of the circuit seen by the 3Ω resistor, and use it to find the current $i_{3\Omega}$ through it.



Notice when applying Thevenin's theorem, *we* have designated where the 'load' is which consequently dictated what the 'internal circuit' (to be replaced by its Thevenin equivalent) is. The choice of where the external environment is in a circuit is purely decided by what *we* are interested in.

For example, if we are interested in determining the voltage across the 4Ω resistor instead, we would designate the load to be the 4Ω and replace the circuit it 'sees' by its Thevenin equivalent:

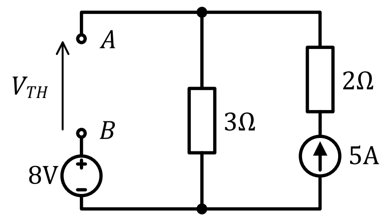


Thevenin Voltage

To find the open-circuit voltage V_{AB} across the pair of terminals the 4Ω was connected to, we note that the current through the 3Ω must be 5A, and so by KVL we have

$$V_{AB} = 3 \cdot 5 - 8 = 7 \text{ V},$$

this is the Thevenin voltage V_{TH}, i.e., $V_{TH} = V_{AB} = 7$ V.

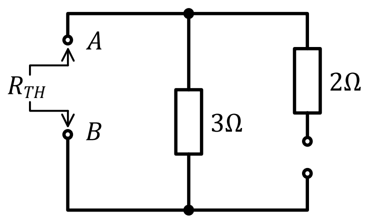


Thevenin Resistance

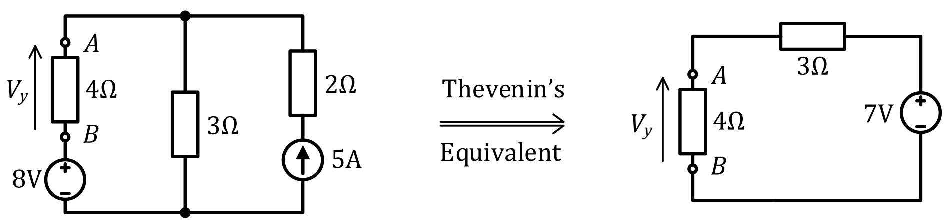
Since there is no conducting path through the 2Ω, it should be clear that the equivalent resistance seen by a charge flowing from terminal A to B is

$$R_{TH} = 3\Omega$$

which is the Thevenin resistance as seen from terminal A-B.



Replacing the circuit seen by the 4Ω by its Thevenin equivalent

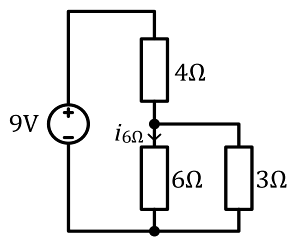


trivially reduces the problem down to, once again, a voltage divider

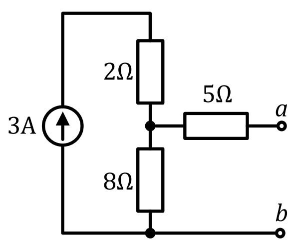
$$V_y = \frac{4}{3+4} \cdot 7 = 4 \text{ V},$$

and is consistent with the prior analysis since by KVL $V_y = V_x - 8 = 12 - 8 = 4$ V as expected.

Example 30. Determine the Thevenin equivalent of the circuit seen by the 6Ω resistor, and use it to find the current $i_{6\Omega}$ through it.



Example 31. Find the Thevenin equivalent of the circuit as seen across the terminals *a*-*b*, and use this equivalent circuit to determine how much current would pass through a 3Ω resistor if connected across the terminals *a* and *b*.



It is important to emphasise that, as seen from all of the examples, the Thevenin equivalent circuit only preserves the external behaviour of the original circuit at the terminals interest, but **not** its internal behaviour – with no load connected, there is no current in the Thevenin equivalent circuits, yet a non-zero current clearly exists in all of the replaced circuits[46].

Finally, it is worth noting that while *any* linear two-terminal circuit has a Thevenin equivalent, **not** all Thevenin equivalents have a *determinant* form. For instance, attempting to find the Thevenin equivalent circuit seen by the 2Ω resistor in the circuit presented on **Page 49** will result in $V_{TH} = \infty$, and $R_{TH} = \infty$ for the Thevenin parameters, and consequently any attempts to characterise the voltage across, and current through the 2Ω using this Thevenin equivalent will yield an *indeterminant* value of ∞/∞! Note that this does **not** mean that the electrical behaviour across the 2Ω is undefined (e.g., there is clearly a 5A current through it) but rather it says that the particular method of analysis has failed to incorporate all of the available information (in the circuit) which would otherwise uniquely characterise the behaviour of interest[47].

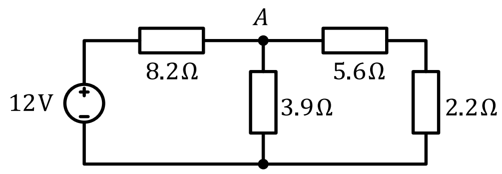
---

[46] This notion of preserving the external equivalence is no different to, for example, the conversion of two series resistances into an equivalent 'total' resistance – the equivalence of this total resistance to the original two is in terms of the external voltage across the two resistors, and current through them. One can no longer identify, say, the internal voltage across one of the series resistances from this total resistance.

[47] This is because in this particular situation, irrespective of what the current source value is, we end up with the same Thevenin equivalent circuit of $V_{TH} = R_{TH} = \infty$. The resulting equivalent circuit therefore fails to incorporate the current source value pertaining to this circuit that would otherwise uniquely determine the effect on the 2Ω.

2.2.6. Problems

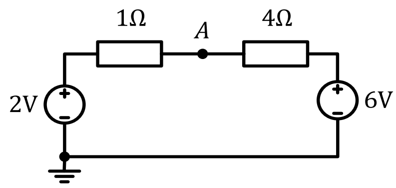
1. In the circuit shown, four resistances are connected to a voltage source.



Find the current through the 2.2 Ω resistor by

(a) finding the supply current and making use of the current division principle,
(b) making use of node-voltage analysis to determine the voltage at node A with respect to the bottom rail, and
(c) making use of Thevenin's theorem to replace the circuit seen by the 2.2Ω by its Thevenin's equivalent.

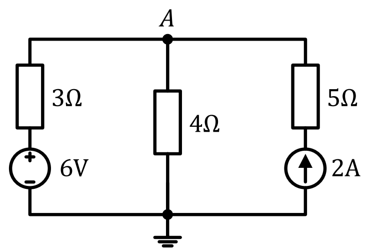
2. Consider the circuit show below.



Determine the voltage at node A by making use of

(a) the voltage division principle,
(b) node-voltage analysis, and
(c) superposition theorem.

3. Consider the circuit shown below.

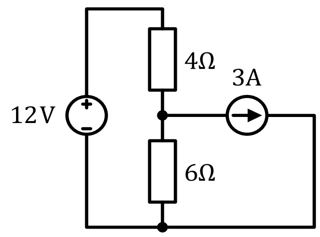


Determine the voltage at node A by making use of

(a) node-voltage analysis,
(b) superposition theorem, and
(c) Thevenin's theorem (to replace the circuit seen by the 4Ω resistor).

Hence, or otherwise, find the voltage across the 2A current source.

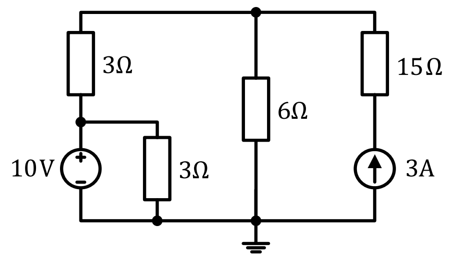
4. Consider the circuit shown below.



Determine the current through the 6Ω resistance by making use of

(a) node-voltage analysis,
(b) superposition theorem, and
(c) Thevenin's theorem (to replace the circuit seen by the 6Ω resistor).

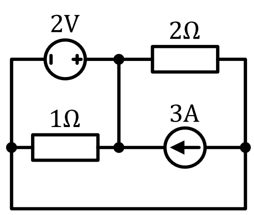
5. Consider the circuit shown below.



Determine the current through the 10V source by making use of

(a) superposition theorem, and
(b) Thevenin's theorem (to replace the circuit seen by the 10V source).

6. Determine the current through the 2V source in the circuit shown.



Solutions to Problems

1. $I_{2.2\Omega} = 10/27 \approx 0.37$ A.

1(a). $I_{\text{supply}} = 10/9$ A to be divided between 3.9 Ω and (5.6 + 2.2) Ω.

1(b). $V_A = 26/9$ V across (5.6 + 2.2) Ω, then use Ohm's law.

1(c). The Thevenin equivalent seen by the 2.2Ω is $V_{TH} \approx 3.87$ V and $R_{TH} \approx 8.24$ Ω.

2. $V_A = 2.8$ V.

2(a). $V_{1\Omega} = 0.8$ V and $V_{4\Omega} = 3.2$V, apply KVL.

2(b). Directly $V_A = 2.8$V.

2(c). $V_{A(2\text{V})} = 1.6$ V, $V_{A(6\text{V})} = 1.2$ V.

3. $V_A = 48/7 \approx 6.857$ V, $V_{2\text{A}} = 118/7 \approx 16.857$ V polarity upwards.

3(a). Directly $V_A \approx 6.857$ V, apply KVL to find $V_{2\text{A}}$.

3(b). $V_{A(6\text{V})} = 24/7$ V, $V_{A(2\text{A})} = 24/7$ V.
$V_{2A(6\text{V})} = 24/7$ V, $V_{2A(2A)} = 94/7$ V.

3(c). The Thevenin equivalent seen by the 4Ω is $V_{TH} = 12$ V and $R_{TH} = 3$ Ω.
The Thevenin equivalent seen by the 2A is $V_{TH} = 24/7$ V and $R_{TH} = 47/7$ Ω.

4. $I_{6\Omega} = 0$ A.

4(a). $V_{6\Omega} = 0$ V, then use Ohm's law.

4(b). $I_{6\Omega(12\text{V})} = 1.2$ A upwards, $I_{6\Omega(3\text{A})} = 1.2$ A downards.

4(c). The Thevenin equivalent seen by the 6Ω is $V_{TH} = 0$ V and $R_{TH} = 4$ Ω.

5. $I_{10\text{V}} = 22/9 \approx 2.44$ A upwards.

5(a). $I_{10\text{V}(10\text{V})} = 40/9$ A upwards, $I_{10\text{V}(3\text{A})} = 2$ A downwards.

5(b). The Thevenin equivalent seen by the 10V is $V_{TH} = 4.5$ V and $R_{TH} = 2.25$ Ω.

6. $I_{2\text{V}} = 0$ A.

2.3 Equivalent Circuit Model

Learning Objectives

Understand how equivalent circuit models are used to simplify analysis and guide design processes for electronic circuits.

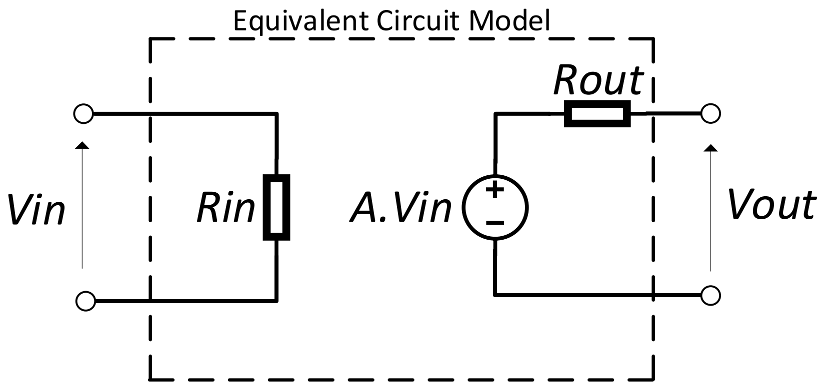
Success Crit**eria**

- Describe the equivalent circuit model for a dc circuit block
- Explain the importance of using terminal characteristics
- Explain loading effects
- Analyse the effect of loading on a circuit's input or output
- Carry out design using the parameters of the equivalent circuit model

2.3.1. Introduction

The concept of equivalence is used widely in electric circuits to model complex circuits, making them easier to analyse and design. For example multiple resistors can be modelled by a single equivalent value; and a complex circuit can be modelled by an equivalent Thevenin circuit to capture only the most important aspects of the relationship between the circuit and a load connected to it. In each case the details inside the circuit are hidden and abstracted or modelled by a set of terminal characteristics.

The general equivalent model for the input and output of a dc circuit is shown below. It has an input represented by a resistance and an output represented by a Thevenin equivalent circuit. These two pieces of information allow the details of a complex circuit to be hidden (or black-boxed). The voltage source of the Thevenin output is a fixed amplification factor *A* of the input voltage, i.e. multiplied by the input voltage.

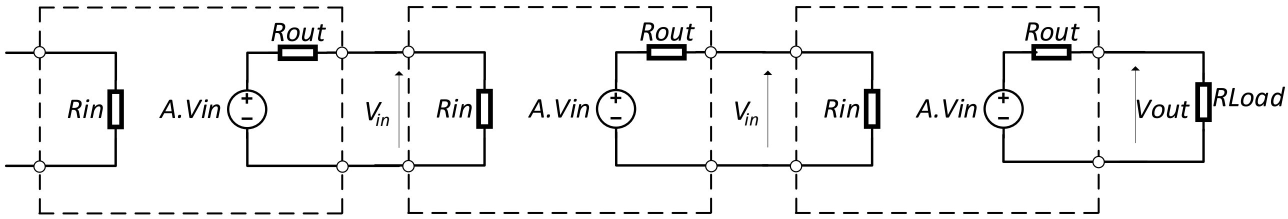


The terminal characteristics are:

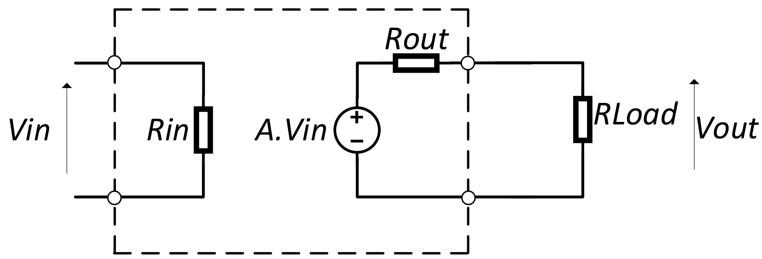
- Input resistance (*Rin*)– ideally very high to reduce the load on the prior circuit
- Amplification factor (*A*)
- Output resistance (*Rout*) – ideally very low to avoid loss of output voltage when a load is connected to the circuit.

While this model is used to specify the characteristics of a single circuit block, it only has meaning when used in conjunction with other blocks and/or circuits.

Using equivalent circuit models allows complex circuits to be systematically analysed, as relationships between smaller blocks.



Example 32.



<u>Write a mathematical model for Vout</u>

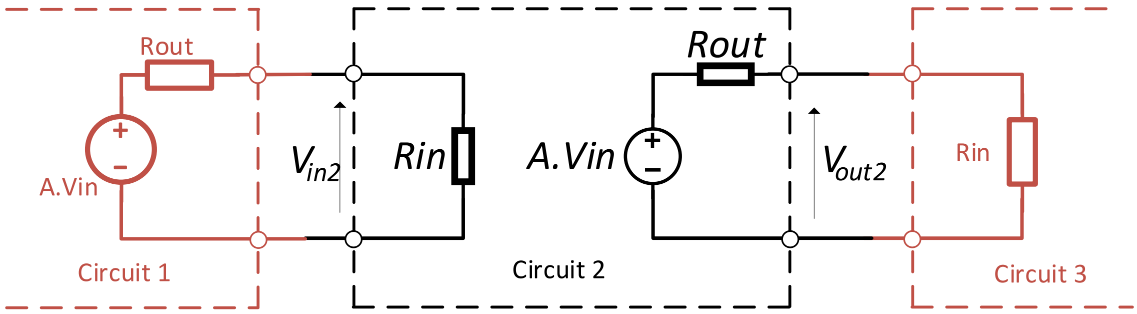
If $A = 13.4,\ R_{in} = 10M\Omega, R_{out} = 1R7\Omega, R_{Load} = 250\Omega, V_{in} = 33mV$
$calculate\ V_{out}$

When working on designs, engineers often have to work within a set of specific constraints provided to them. In your engineering practice, it is essential that you adhere to constraints provided to you, and understand the implications of not doing so.

Example 33.

In the model below A.V_{in1}, R_{out1} and R_{Load} may be known parameters and given to the engineer as constraints to design within.

Develop a mathematical model to find A_2 in terms of these parameters (and any other characteristics from circuit 2).

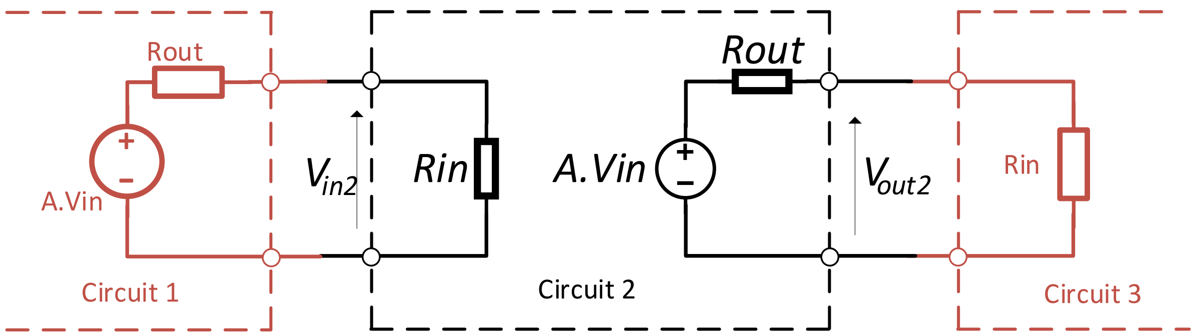


Example 34.

Given the following constraints: $A_1.\ V_{in1} = 0.7V$, $R_{out1} = 120\Omega$, $R_{in3} = 2k\Omega, and\ V_{in3} = 32V$

What is the value for A when $R_{out2} = 112\Omega\ and\ R_{in2} = 580\Omega$

Often problems are not at all like the one above, they are 'ill-defined', and engineers often have to work with 'what-if' or 'could-we' type questions, rather than putting numbers into a formula.



Example 35. Extension exercise

The value from the previous exercise is quite high for a single amplifier; can it be lowered by 15% ?

Remember: the three parameters as the design engineer you have control over are A_2, $R_{in2}\ and\ R_{out2}$, in circuit 2.

2.4 Alternating Current (AC) Concepts

Our studies of circuit behaviours thus far have been largely limited, in scope, to a specific "class" of inputs – *constant*, or *direct current* (DC), signals. These are signals in which the polarity of the voltage, and direction of the current do *not* change (e.g., a battery), and consequently, the responses of a circuit to these signals must also be time-invariant. In practice however, all interesting phenomena vary with time to convey some sort of intelligible information (e.g., a musical piece, a video recording, or the voltage output of an ECG implant, etc.), and naturally, in order for us to design and create electronics to interact with these phenomena, we need to extend our understanding of circuits to how they would behave when excited by *time-varying*, or *alternating current* (AC), signals.

While it may appear insurmountable at first – signals come in practically *infinite* number of forms and realisations – one of the great advances of the 19th century was the work of Joseph Fourier who discovered that *any* practical signal can be expressed as a *sum of sinusoids*, and in doing so, paved the path for an unifying way of representing arbitrary signals of all shapes and forms using just a set of "basic" building-blocks – the humble sinusoids.

The ability to represent arbitrary signals as a combination of sinusoids means that if we are able to understand how a circuit would respond to a *single* sinusoidal input, then we have a way to analyse the circuit subject to *any* arbitrary AC inputs – all analyses can be reduced to a sum of individual sinusoidal analysis via the *superposition theorem* (see **Sect. 2.2.4**) – we can therefore restrict our study of circuit behaviours to just *sinusoidal* sources[48] without loss of generality!

[48] This study is also important since the generation, transmission, distribution, and consumption of electrical energy from our mains supply are all in sinusoidal forms due to the ease at which they can be produced.

2.4.1. Sinusoidal Representation

Intended Learning Outcomes

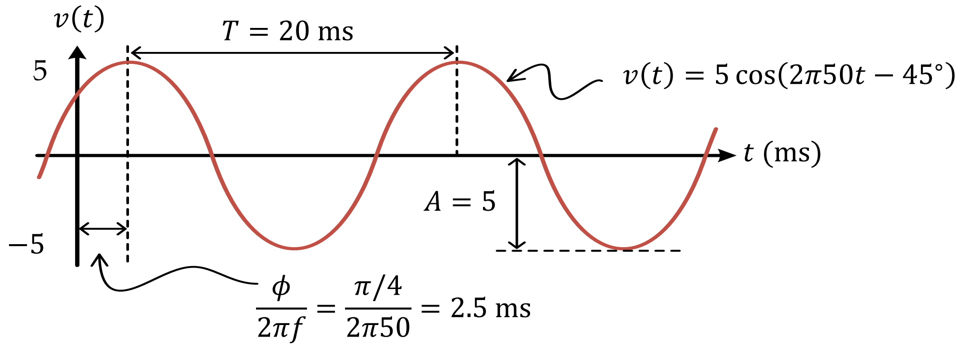
- Be able to represent and model signals as a sinusoidal function in time and make relevant interpretation about the underpinning signals they represent.

A sinusoid is a *periodic* signal[49] that varies sinusoidally with time. Mathematically, we can express a sinusoidal signal using either the *sine* or *cosine* function. Here, we will arbitrarily take the cosine convention, and define a sinusoid to be of the form

$$x(t) = A\cos(\omega t + \phi) = A\cos(2\pi f t + \phi),$$

where

A (*amplitude*)	is the magnitude (≥ 0) of the maximum deviation from the average value of $x(t)$.
ω (*angular frequency*)	is the rate (in rad s^{-1}) at which the signal 'travels' and is related to the *frequency*, f (in Hz), and *period*, T (in second), by $\omega = 2\pi f = 2\pi/T$.
ϕ (*phase*)	is the angle (in $^\circ$ or rad) that the waveform has shifted from a standard cosine, this corresponds to a time shift of ϕ/ω along the time-axis.



Note that the phase ϕ is often expressed in degrees, however, the unit of ωt is in radians[50]. Thus, in order to evaluate, say $v(5\text{ms}) = 5\cos(0.5\pi - 45^\circ)$, the argument of the cosine must be converted to the same unit via the unit conversions:

$$1^\circ = \frac{\pi}{180}\text{ rad}, \quad \text{and} \quad 1\text{ rad} = \frac{180^\circ}{\pi}.$$

[49] A *periodic* signal is one that repeats its value in regular intervals. Specifically, a signal $x(t)$ is said to be periodic if there is a non-zero constant T such that $x(t+T) = x(t)$ for all time t. The number T is known as the *period* of the signal, and the smallest of such a number is known as the *fundamental period*.

[50] Simply because we are preconditioned to be able to interpret and visualise the meaning of degrees in comparison to the more mathematically convenient counterpart, radians.

Thus, we can evaluate $v(5\text{ms})$ as

$$v(5\text{ms}) = 5\cos\left(\underbrace{0.5\pi}_{=90^\circ} - 45^\circ\right) = 5\cos(45^\circ) = \frac{5}{\sqrt{2}}\ \text{V}$$

or equivalently,

$$= 5\cos\left(0.5\pi - \underbrace{45^\circ}_{=0.25\pi}\right) = 5\cos(0.25\pi) = \frac{5}{\sqrt{2}}\ \text{V},$$

depending on your (calculator's!) preference.

Example 36. A sinusoidal current $i(t)$ has a maximum measured amplitude of 10A. The current passes through one complete cycle in 1 ms and the magnitude of the current at time $t = 0$ was measured to be 5A and decreasing. Determine

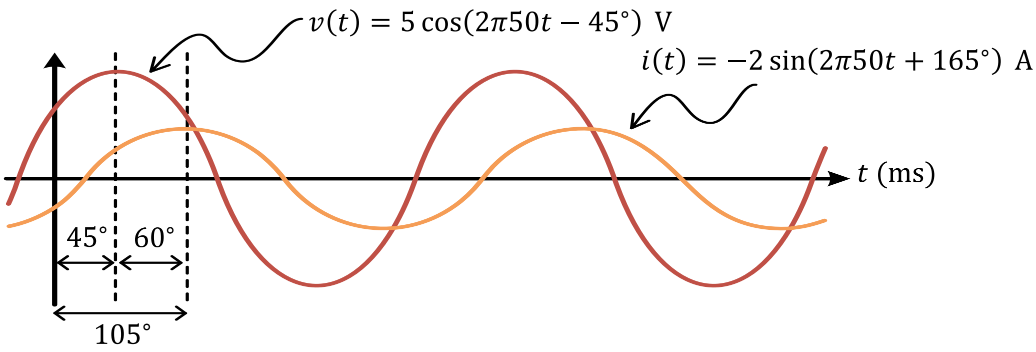
(a) the frequency of the current in Hz and rads^{-1},
(b) an expression for $i(t)$ using the cosine function,
(c) the instantaneous value at $t = 2/3$ ms, and
(d) the first time after $t = 0$ that $i(t) = -5$ A

Since trigonometric functions are periodic over 360° (2π), their representations are therefore *non-unique* and we have *infinite* ways of representing the same signal, e.g.,

$$v(t) = 5\cos(2\pi 50t - 45^\circ) = 5\cos(2\pi 50t - 45^\circ \pm 360^\circ n) \quad \text{for } n = 0,1,2, \ldots.$$

In order to manage this rather chaotic situation, the agreed convention is to always express the phase between $-180^\circ < \phi \leq 180^\circ$, the so-called *principal values*. In doing so, *all* sinusoids can thus be uniquely represented a corresponding one whose phase is within the principal values.

It should also be clear that in order for us to make meaningful comparisons between sinusoids (of the same frequency), they must be expressed in the same form. For example, if the sinusoidal voltage across, and current through a component is $v(t) = 5\cos(2\pi 50t - 45°)$ and $i(t) = -2\sin(2\pi 50t + 165°)$



it is not immediately obvious in its existing form how far 'behind' or 'ahead' (in phase, or time) the current $i(t)$ is relative to the voltage $v(t)$. But when expressed in the *same* form, this comparison becomes immediately obvious since the phase are now expressed with reference to the same trigonometric function (cosine):

$$i(t) = -2\sin(2\pi 50t + 165°) = 2\cos(2\pi 50t - 105°) \text{ A}$$

revealing that the voltage is ahead (or '*leads*') of the current through the load by $-45° - (-105°) = 60°$. Equivalently, we can say that the current is ahead of the voltage by $-105 - (-45) = -60°$ which is identical to the former statement as the negative implies that it is behind (or '*lags*') the voltage by 60°.

Convince yourself of the following equivalence:

$$\sin(x) = \cos(x - 90°), \quad \text{and} \quad -\cos(x) = \cos(x \pm 180°).$$

Example 37. Calculate the phase difference between the two voltages $v_1(t) = -10\cos(2\pi 50t + 50°)$ and $v_2(t) = 12\sin(2\pi 50t - 10°)$, and state which sinusoid is leading in time.

2.4.2. Root Mean Square (RMS) Value

Intended Learning Outcomes

- Be able to characterise common AC periodic signals in terms of their root-mean-square values

While sinusoidal signals can be described and modelled precisely through the use of trigonometric functions, unlike DC signals these functions are rather awkward to work with and compare between due to their time-varying nature. Often, we need not know the specific value of the AC signal at any given point in time, but are usually content with a descriptive measure that captures its central behaviour, allowing a quick way to discern between the various AC signals – some sort of measure of *average*.

While the standard *arithmetic* average is one of such measures, it is **not** very informative when it comes to quantifying AC, and in particular, sinusoidal signals. This is because the standard average when applied to an arbitrary sinusoidal signal of the form $x(t) = A\cos(2\pi f t + \phi)$ will *always* be **zero** and therefore does not provide any sort of differentiation between them. A commonly used measure in place of the normal average is the *root mean square* (rms or RMS) value. Much like the arithmetic average, the RMS value is just another type of average that characterises signals.

The RMS value of a signal $x(t)$ is the square-**root** of the **mean** of the **square** of the signal over **one period** T. Mathematically, this is captured concisely as

$$X_{\text{RMS}} = \sqrt{\frac{1}{T}\int_{t=t_0}^{t_0+T} \big(x(t)\big)^2 \, dt}.$$

As we will see later on, the RMS value[51] conveniently allows us to do *power* calculations due to AC excitations in a familiar and meaningful way.

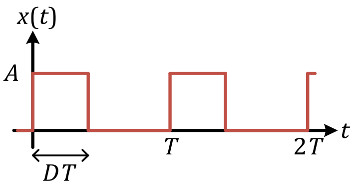
[51] The RMS value is also sometimes referred to as the *effective* value because it turns out to be the value of a DC current (or voltage) that effectively delivers the same amount of *average power* (more on what this means later on) to a resistor as the underpinning AC current that has this RMS value.

Example 38. Show that the RMS value of a general sinusoidal signal of the form $x(t) = A\cos(2\pi f t + \phi)$ is $X_{\text{RMS}} = A/\sqrt{2}$.

Hint: you *may* find the identity $\cos(2A) = 2\cos^2(A) - 1$ *useful.*

From the above example we see that the RMS value of a sinusoid does **not** depend on either its frequency or the phase shift. It should also be obvious that the RMS value of a signal necessarily *depends* on the shape and size of the signal itself, and is therefore *different* depending on the underlying signal in which the RMS value is to be calculated for. For example, we have seen that the RMS value of a standard sinusoid relates to its amplitude by a factor of $\sqrt{2}$, however, this is **not** necessarily true for a square wave as demonstrated in the following example.

Example 39. Determine the RMS value of the square-wave signal of period T that is 'ON' for a fraction D of the period, and 'OFF' for the remainder of the period.



When in doubt, apply the *definition* of RMS value and see where the result leads!

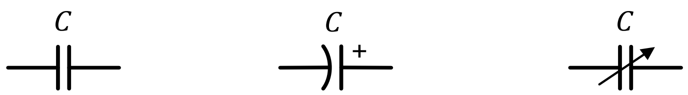
2.4.3. Capacitor & Inductor

Intended Learning Outcomes

- Be able to recognise the time-domain terminal characteristic of capacitors and inductors, and their implications towards their behaviours under DC, and under sinusoidal excitations.

Capacitor

A capacitor is an electrical component that consists of two conductive 'plates' separated by a non-conductive (i.e., insulating) region (e.g., air, ceramic, plastic, etc.) – there is **no** direct conducting path through a capacitor. Capacitors are presented symbolically by



Fixed Capacitor   Polarised Capacitor   Variable Capacitor

When a voltage is applied across a capacitor, positive charges are accumulated onto one of the plates which in turn *attract* (repel) *equal* number of negative (positive) charges from the other plate. This displacement of equal but opposite charges on the two separated plate establishes a force of attraction (i.e., Coulomb's law, see **Sect. 2.1.1**) between them. As more charges are displaced, the attraction force increases until it balances out with the force underpinning the voltage applied across the capacitor – equilibrium is reached with no more movement of charges.

From the description of its operation, it should become intuitive that the number of positive (or negative) charges accumulated on the plate is directly proportional to the voltage applied across the capacitor:

$$q = Cv,$$

where the constant of proportionality is known as the *capacitance* of the capacitor characterising its "capacity" to "store" charge (energy).

In order to relate the current (*rate* of flow of charge) to the capacitor to the voltage across it, we can simply take the derivative of the governing behaviour above to get

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

the terminal characteristic of the capacitor. The terminal characteristic says in order for a current to be present "through" the capacitor, the voltage across the capacitor must be *changing*, as otherwise the capacitor will simply be at its equilibrium state.

Capacitors are therefore not particularly interesting at DC (although there are certainly uses for them!) since it will simply "charge" to its balanced state at which $dv/dt = 0 \Rightarrow i = 0$, and ***behaves* like an open-circuit**. On the other hand, when an AC, sinusoidal, voltage $v(t) = A\cos(2\pi ft + \phi_v)$ is applied across a capacitor we see that[52]

$$i(t) = C\frac{dv(t)}{dt} = C \cdot (-2\pi f \cdot A\sin(2\pi ft + \phi_v))$$

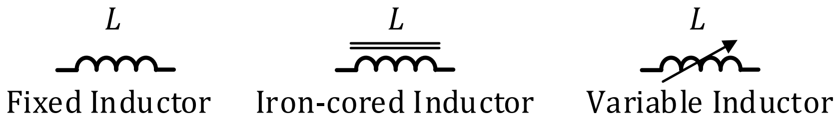
$$= \underbrace{2\pi fC \cdot A}_{\text{amplitude of } i(t)} \cos\left(2\pi ft + \underbrace{(\phi_v + 90°)}_{\text{phase of } i(t)}\right)$$

which says that the resulting **current "through" the capacitor *leads* the voltage across it by exactly 90**°:

$$(\phi_v + 90°) - \phi_v = 90°.$$

Inductor

An inductor (or a *coil*) is another electrical component that has the ability to store energy. It consists of a conductive wire wound in a spiral shape around a magnetic core (ceramic, ferrite, iron, etc.) – there thus a direct conducting path through an inductor. Inductors are presented symbolically by



A current through a wire creates a continuous circular magnetic field around it – the larger the current, the stronger this magnetic field becomes. When the wire is in the shape of a coil, the particular geometry allows the induced magnetic field within the coil to 'link' together, creating a resultant flux λ inside the coil along one direction[53]. This resultant flux is therefore directly proportional to the current i through the coil:

$$\lambda(t) = Li(t)$$

where the constant of proportionality L is known as the *inductance*[54] of the inductor characterising its ability to 'induce' magnetic field.

---

[52] Recall the derivative of $y = \cos(ax)$ is $dy/dt = -a\sin(ax)$.

[53] This direction is commonly determined via the so-called *right-hand rule*.

[54] The inductance depends on the number of coil windings, the shape of the inductor, but also the magnetic core the coil is on, etc. These are factors that determines the 'strength' of the resultant magnetic field inside the inductor.

In order to relate the voltage across the inductor to the current through it, we make use of *Faraday's law*[55] and take the derivative of the above to get

$$\frac{d\lambda(t)}{dt} = L\frac{di(t)}{dt} \qquad \Rightarrow \qquad v(t) = L\frac{di(t)}{dt}$$

the terminal characteristic of the inductor. The terminal characteristic says in order for a voltage to be present across an inductor, the current through the inductor must be *changing*, as otherwise the flux through the inductor will be constant.

Like the capacitor, the inductors are therefore not particularly interesting at DC, since they will simply have uniform magnetic flux through them (due to constant current) at which $di/dt = 0 \Rightarrow v = 0$, and ***behaves* like a short-circuit**. On the other hand, when an AC, sinusoidal current $i(t) = A\cos(2\pi ft + \phi_i)$ is applied across an inductor we see that

$$v(t) = L\frac{di(t)}{dt} = L \cdot (-2\pi f \cdot A\sin(2\pi ft + \phi_i))$$

$$= \underbrace{2\pi fL \cdot A}_{\text{amplitude of } v(\text{t})} \cos\left(2\pi ft + \underbrace{(\phi_i + 90^\circ)}_{\text{phase of } v(t)}\right)$$

which says that the result voltage across the inductor *leads* the voltage across it by exactly 90°. Equivalently, this says the **current through the inductor *lags* the voltage across it by exactly 90°:**

$$\phi_i - (\phi_i + 90^\circ) = -90^\circ.$$

[55] Faraday's law states that any change in the magnetic flux in a coil will induce a voltage proportional to the rate of change of the flux to *oppose* the current creating the flux.

2.4.4. Power Characterisation

Intended Learning Outcomes

- Be able to differentiate, relate, and interpret between average power, reactive power, apparent power, and instantaneous power of an underlying device driven by sinusoidal excitations.
- Be able to relate, and make connections, between the power factor of a device to its power characteristics, and provide relevant interpretation towards its equivalent electrical component composition.
- Be able to determine the relevant power characteristics of a device subject to sinusoidal excitations.

Power is perhaps the most important quantity in all electric utilities, electronics, and communication systems as the underlying operating principles of these systems all revolve around the transmission, conversion, and utilisation of power. As such, every industrial and household electrical devices has a power rating – the nominal power required by the device.

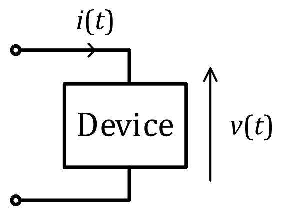
Conveniently, sinusoidal sources are also the predominant means of providing electric power due to the ease at which they are produced, it is therefore important that we understand, and be able to interpret, the power characteristics of a device under sinusoidal excitations.

Instantaneous Power

The power of an electrical device is, by definition, the product of the voltage across, and the current through the device. Since the voltage $v(t)$ and current $i(t)$ of the device in an AC circuit varies over time, the power must therefore also be a function of time

$$p(t) = v(t)i(t).$$

This is the *instantaneous* power expression (or simply *the* power) that completely describes the power **absorbed**[56] by the device *at any instance in time* t (compare with the DC case where $p(t) = VI$, a constant for all time t).

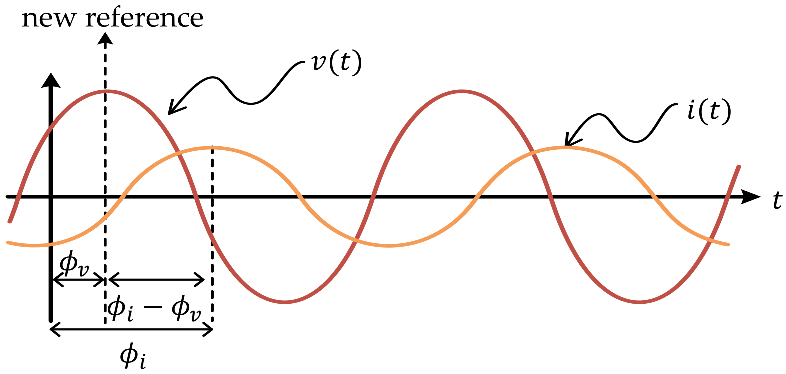


[56] Recall the passive-sign convention takes the voltage and current to be in opposing directions, which means their product gives the amount of power *absorbed* or *received*. If the product is negative (at certain instances in time) then it means the device is *delivering* or *supplying* power at those instances (see **Sect. 2.1.5**).

Our interest in instantaneous power revolves around sinusoidal voltage and current of the general form

$$v(t) = V_p \cos(\omega t + \phi_v)\,, \quad \text{and} \quad i(t) = I_p \cos(\omega t + \phi_i),$$

where ϕ_v and ϕ_i are the phase angle of the voltage and current through the device respectively. Since the signals are periodic, the zero-time reference is arbitrary, and so we lose nothing by choosing the zero-time instance to be when the voltage is at its positive peak for convenience.



This means, we can equivalently consider

$$v(t) = V_p \cos(\omega t)\,, \quad \text{and} \quad i(t) = I_p \cos\big(\omega t + (\phi_i - \phi_v)\big)$$

to describes the exact same situation[57]. The power $p(t)$ through the device due to sinusoidal excitation is thus

$$\begin{aligned} p(t) &= V_p I_p \cos(\omega t) \cos\big(\omega t + (\phi_i - \phi_v)\big) \\ &= \underbrace{\frac{1}{2} V_p I_p \cos(\phi_v - \phi_i)}_{=\text{constant}} + \underbrace{\frac{1}{2} V_p I_p \cos\big(2\omega t + (\phi_i - \phi_v)\big)}_{=\text{time-varying}}. \end{aligned}$$

Notice we have deliberately expressed[58] $p(t)$ in a much more meaningful form from which we can deduce that the instantaneous power that results from sinusoidal excitations generally consists of

- a constant component whose value *depends* on the phase difference between $v(t)$ and $i(t)$,
- a sinusoidal time-varying component that varies at **twice** the frequency of the underpinning voltage and current, and
- may be negative for a portion of each cycle – power is delivered by the device (to the circuit it is connected in) due to capacitors and/or inductors.

[57] The only reason we have done so is merely to simplify the mathematics in the discussions to come. Of course, all of the results still follow if we did not make this shift in time reference but the mathematics become unnecessary.
[58] Use the trigonometric identity $2\cos(A)\cos(B) = \cos(A - B) + \cos(A + B)$.

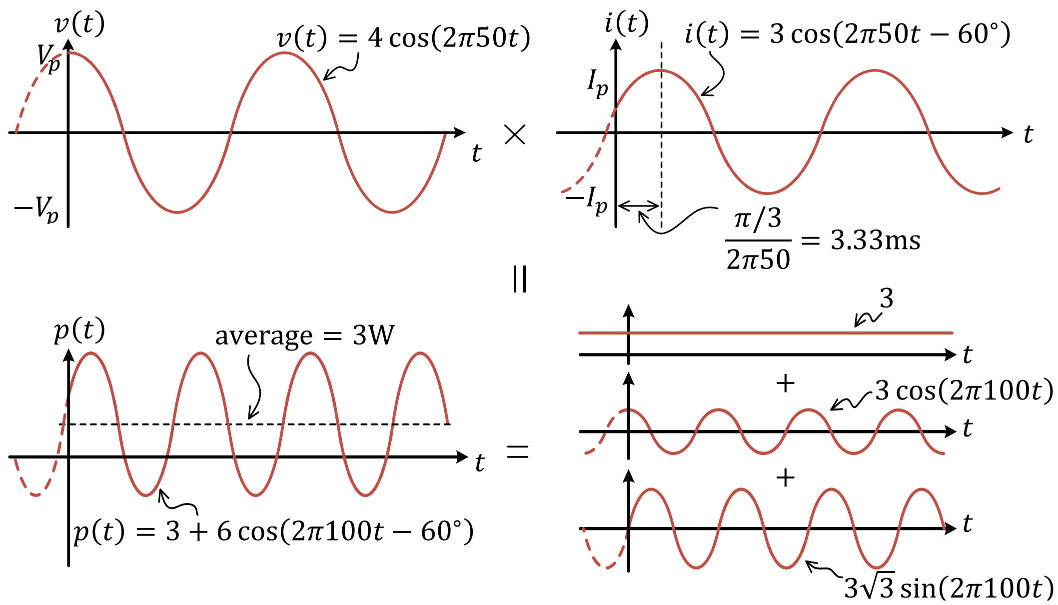
A closer examination of $p(t)$ above shows we can further decompose the time-varying component[59] and re-write $p(t)$ equivalently as

$$p(t) = \underbrace{\underbrace{\frac{1}{2}V_pI_p\cos(\phi_v - \phi_i)}_{=P} + \underbrace{\left[\frac{1}{2}V_pI_p\cos(\phi_v - \phi_i)\right]}_{=P}\cos(2\omega t)}_{\text{due to resistive elements}} + \underbrace{\underbrace{\left[\frac{1}{2}V_pI_p\sin(\phi_v - \phi_i)\right]}_{=Q}\sin(2\omega t)}_{\text{due to non-resistive elements}}$$

The reason for putting $p(t)$ in this seemingly convoluted form is this: while $p(t)$ completely describes the power characteristic of an element at any given time, its time-varying nature makes it a difficult quantity to convey useful information about the device. Knowledge of the instantaneous evolution of $p(t)$ is also, for the most part, unnecessary – its average and salient features are sufficient in describing the overall behaviour in a coherent way.

As seen from the decomposed expression above, the instantaneous power is characterised by essentially four features, or metrics:

Average Power:	$P = \frac{1}{2}V_pI_p\cos(\phi_v - \phi_i)$	**Reactive Power:**	$Q = \frac{1}{2}V_pI_p\sin(\phi_v - \phi_i)$
Apparent Power:	$S = \frac{1}{2}V_pI_p$	**Power Factor:**	$\text{p.f.} = \cos(\phi_v - \phi_i)$



[59] Use the trigonometric identity $\cos(A + B) = \cos(A)\cos(B) - \sin(A)\sin(B)$.

Average Power

The *average* power, as its name suggests, is the average of $p(t)$ over one of its period T. If $v(t)$ and $i(t)$ of a device are sinusoidal, it should be obvious from the instantaneous power expression that the average must be

$$P = \frac{1}{2} V_p I_p \cos(\phi_v - \phi_i).$$

The average power is also referred to as the ***real*** or ***active power*** because this is the measure of power converted from electrical to non-electrical form (e.g., heat for a resistor) – power that does work.

The RMS value makes it useful appearance here, as it allows us to re-write the average power expression in a more nostalgic form in comparison to the DC case:

$$P = \frac{1}{2} V_p I_p \cos(\phi_v - \phi_i) = \underbrace{\frac{V_p}{\sqrt{2}}}_{=V_{\text{RMS}}} \cdot \underbrace{\frac{I_p}{\sqrt{2}}}_{=I_{\text{RMS}}} \cdot \cos(\phi_v - \phi_i) = V_{\text{RMS}} I_{\text{RMS}} \cos(\phi_v - \phi_i).$$

RMS values are so useful that, unless specified otherwise, values quoted in AC context are the RMS values not maximum values. The New Zealand mains voltage is 230 V(rms).

The unit of average power is the usual **Watt (W)**.

Example 40. A 150Ω resistive lamp is connected to a voltage source of 230 V(rms). Determine an expression for the power absorbed by the lamp at any instant in time, and the power absorbed by the lamp on average.

Example 41. A resistive heater draws, on average, 1.5 kW of power from the NZ mains supply – 230 V(rms). Determine the
(a) peak value of the current drawn by the heater, and
(b) the equivalent resistance of the heater.

Reactive Power

In a purely resistive load, one would intuitively expect $p(t)$ to be non-negative at all time (why?), and this is clear from the expression for the power that results as seen in **Example 29**, where

$$p(t) = \frac{v^2(t)}{R} = i^2(t)R \geq 0.$$

But it is also apparent from examining the general power expression when both $v(t)$ and $i(t)$ through the resistive load are sinusoidal: since $v(t)$ and $i(t)$ must be *in phase*, i.e., $\phi_v = \phi_i$ (because they are directly related by a cosntant – Ohm's law), we see that $Q = 0$, and $p(t) = P(1 + \cos(2\omega t)) \geq 0$ as expected.

On the contrary, a pure capacitive or inductive load has a phase difference of exactly $\phi_v - \phi_i = \pm 90°$ between the sinusoidal voltage and current through them. This entails that

- It does **not** dissipate power as non-electrical form since $P = 0$, and

- $p(t) = Q\sin(2\omega t)$, the load still draws current and voltage such that **half** of the time power is being absorbed (stored), while the same amount of power is being delivered (extracted) from the capacitor or inductor during the other half of the cycle.

There is thus **no** net transfer of energy, but energy is continuously exchanged between the energy storage elements and the circuit driving them.

To quantify this phenomenon we thus use the amplitude of the resulting $p(t)$ as a measure of the 'size' of the exchange:

$$Q = \frac{1}{2} V_p I_p \sin(\phi_v - \phi_i) = V_{\text{RMS}} I_{\text{RMS}} \sin(\phi_v - \phi_i)$$

and denote it by the term ***reactive power***[60] as this is the power that results from the elements that *react* to the change (i.e., capacitors and inductors). Notice as an artefact of the mathematics, for a inductive load the reactive power is *positive* while for a capacitive load the reactive power is *negative*. This should **not** be confused with the passive sign convention – all reactive loads absorb and deliver power continuously.

Reactive power does **not** do anything useful, it is continuously exchanged between the energy storage elements and the circuit. However, the fact that they drop voltage, and draw current gives the deceptive impression that they do dissipate power. The current drawn wastes energy across other resistive elements in the system, and so in general, reactive power is undesirable and should be minimised in practice.

The unit of reactive power is **volt-ampere reactive (VAR or var)**.

Example 42. A voltage of $v(t) = 100\cos(5t + 15°)$ V is applied across a load and the current was measured to be $(t) = -4\sin(5t - 15°)$ A (in the direction of a voltage drop). Determine

(a) the instantaneous power of the load,
(b) the average power absorbed/delivered by the load,
(c) the reactive power of the load, and
(d) state whether the load is inductive or capacitive.

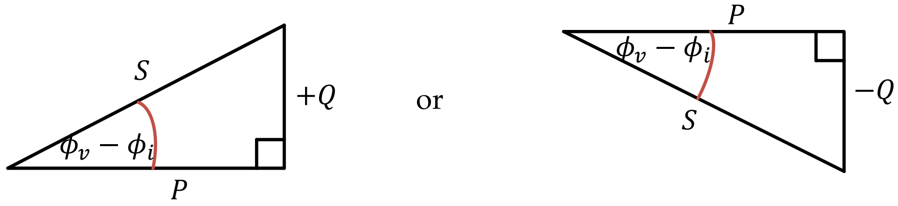
[60] This is sometimes also referred to as 'Wattless' power or 'phantom' power because this power never leaves the electrical system.

Apparent Power

A closer look at the average and reactive power expressions shows that they appear to be just a 'portion' of some underpinning quantity

$$S = \frac{1}{2} V_p I_p = V_{\text{RMS}} I_{\text{RMS}} \,,$$

that is, $P = S \cdot \cos(\phi_v - \phi_i)$, and $Q = S \cdot \sin(\phi_v - \phi_i)$. Graphically, this can be visualised by the so-called *power triangle*:



depending on whether the load is inductive (positive Q) or capacitive (negative Q). It should also be obvious from the power triangle that $S = \sqrt{P^2 + Q^2}$.

The quantity S is called the ***apparent power***[61] and gives an overall measure of the size[62] of the load: a load could do very little useful work but still draws an appreciable amount of current due to reactive elements (i.e., $P \ll Q$), the apparent power therefore gives an indicative measure of the total power 'capacity' required by the load irrespective of whether power is being drawn to do anything useful.

The unit of apparent power is **volt-ampere (VA)**.

[61] It is so called because it seems apparent that the power should be the voltage-current product, by analogy with DC resistive circuits.
[62] We say load A is 'bigger' than load B if load A draws more current, or draws more power, than load B.

Power Factor

The average and reactive power of a load for a fixed magnitude of sinusoidal current and voltage depends heavily on their phase difference, $\phi_v - \phi_i$. This phase, or angle, difference is known as the ***power factor angle***, a reminiscence of the power triangle. The *cosine* of this angle, as used in the real power expression, is known as the ***power factor*** or ***p.f.***[63]

$$\mathrm{p.f.} = \cos(\phi_v - \phi_i)$$

For a load it is always true that

$$\underbrace{-90°}_{\text{capacitive}} \leq \phi_v - \phi_i \leq \underbrace{90°}_{\text{inductive}}$$

and so

$$\underbrace{0}_{\text{capacitive } \textbf{or} \text{ inductive}} \leq \underbrace{\cos(\phi_v - \phi_i)}_{\text{p.f.}} \leq \underbrace{1}_{\text{resistive}} .$$

The nature of the load therefore cannot be determined uniquely from the power factor as we no longer know whether the current leads or lags the voltage. As such, power factors **must** be specified with the accompanying terms ***lagging*** or ***leading*** power factor[64]. These terms define the disposition of the phase of the current phase with respect to the phase of the voltage:

- If $\phi_i < \phi_v$ (current ***lags*** voltage), the load is **inductive**, or has a **lagging p.f.**

- If $\phi_i > \phi_v$ (current ***leads*** voltage), the load is **capacitive** or has a **leading p.f.**

Notice given any **pair** of the power metrics P, Q, S, and p.f., all others can be inferred and uniquely determined.

Example 43. An electrical load operates at 240 V(rms). The load absorbs an average power of 8 kW at a leading power factor of 0.8. Determine the apparent and reactive power of the load.

[63] The sine of the power factor angle, as used in the reactive power expression, is known as the *reactive factor*, or *r.f.*.

[64] This can be mitigated if one uses the *reactive factor* instead. However, this is seldom used as the real power is usually the quantity of interest.

Example 44. A single-phase fan draws 1 kW of active power when plugged into the NZ mains supply. The fan current was measured to lag the voltage by 45°. Determine

(a) the expression for the current $i(t)$ through the fan,

(b) the power factor, and the apparent power of the fan.

(c) By connecting cleverly designed electronics *in parallel* to the fan, it is possible to alter the current through the fan to be *in phase* with the voltage supplied. Determine the expression for this new current $i(t)$, and by comparing with (a), explain why it is desirable to do so.

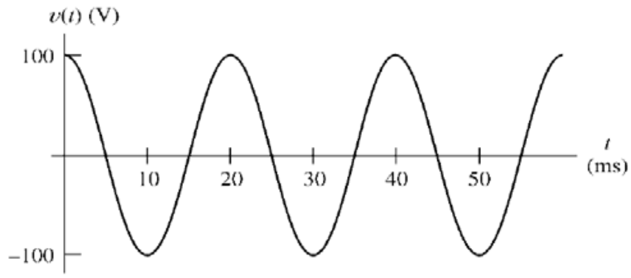
2.4.5. Problems

1. A sinusoidal voltage signal is described by $v(t) = 50\cos(30t + 10°)$ V, determine its amplitude, phase, period, frequency (in Hz), and its instantaneous value at $t = 10$ ms.

2. Express the following sinusoidal signals using the standard cosine representation:

   (a) $v(t) = 10\sin(2\pi 50t + 30°)$ V,
   (b) $i(t) = -9\sin(8t + 100°)$ A, and
   (c) $p(t) = -20\cos(2\pi 100t + 45°)$ W.

3. Describe the following sinusoidal signal using the standard cosine representation:

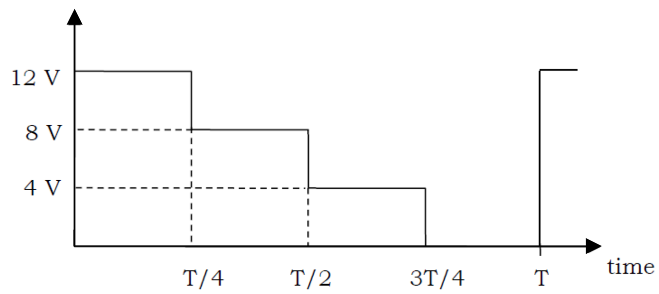


4. Determine the phase angle difference between the two current signals

$$i_1 = -4\sin(377t + 120°)\text{ A} \quad \text{and} \quad i_2 = 5\cos(377t - 65°)\text{ A},$$

   and state whether i_1 lead or lag i_2.

5. Determine the RMS value of the voltage waveform shown below.



6. The voltage across a 10Ω resistor is given by $v(t) = 50\cos(200t)$ V. Determine the *peak* and *average* power absorbed by the resistor.

7. The voltage across, and current through a load is $v(t) = 160\cos(50t)$ V and $i(t) = -33\sin(50t - 30°)$ A (measured in the direction of a voltage drop). Determine (a) the instantaneous power of the load, (b) the average power absorbed by the load, and (c) whether the load is capacitive or inductive.

8. A domestic appliance is supplied by an AC 230 V(rms) supply. If the appliance draws 150 W of average power at a lagging power factor of 0.7, determine the peak value of the current drawn by the appliance.

9. An electric heater draws 1.5 kW of average power from a 230 V(rms), 50 Hz electricity supply. The heater can be considered as a purely resistive appliance. Determine
   (a) the resistance of the heater,
   (b) the peak and RMS value of the current drawn by the heater,
   (c) the power factor and the phase angle difference between the voltage across and current through the heater, and
   (d) the reactive and apparent power supplied by the electricity supply.

Solutions to Problems

1. $A = 50\text{ V},\ \phi = 10°,\ T \approx 0.21\text{ s},\ f \approx 4.77\text{ Hz}, v(10\text{m}) \approx 44.48\text{ V}.$
2(a). $v(t) = 10\cos(2\pi 50t - 60°)\ \text{V}.$
2(b). $i(t) = 9\cos(8t - 170°)\ \text{A}.$
2(c). $p(t) = 20\cos(2\pi 100t - 135°)\ \text{W}.$
3. $v(t) = 100\cos(100\pi t)\ \text{V}.$
4. i_1 lags i_2 by $85°$.
5. $V_{\text{RMS}} \approx 7.48\text{ V}.$
6. peak power $= 250\text{ W}$, average power $P = 125\text{ W}.$
7(a). $p(t) = -5280\cos(50t)\sin(50t - 30°)\ \text{W}.$
7(b). $P = 2640\text{ W}$
7(c). Capacitive.
8. $i_{\text{peak}} \approx 1.32\text{ A}.$
9(a). $R \approx 35.27\ \Omega$
9(b). $i_{\text{peak}} \approx 9.22\text{ A}.$
9(c). $\text{p.f.} = 1,\ \phi_v - \phi_i = 0°.$
9(d). $Q = 0\text{ vars},\ S = 1500\text{ VA}.$

2.5 Boolean Algebra

Learning Objectives

Recognize the abstract or general nature of Boolean axioms and theorems (laws)

Success Crit**eria**

- distinguish the different ways that software, electrical and computer systems engineers use the axioms and theorems of Boolean algebra

Boolean algebra is a mathematical system used to describe operations on variables based upon two values, True and False. It was invented by George Boole, an English mathematician. While he developed it to express formal logical arguments, it has become highly influential in digital electronics and computer programming. Boolean algebra has several fundamental operations:

Fundamental operation	Symbol
AND (Conjunction)	$\wedge$
OR (Disjunction)	$\vee$
NOT (Negation)	$\neg$
XOR (Exclusive Disjunction)	$\oplus$

Using two variables (X and Y), each with two values (TRUE or FALSE) gives rise to the following statements (or propositions), and a set of truth tables for each.

X AND Y: Both X must be True and Y must be True for the result to be True, otherwise the result is False.

$\boldsymbol{X \wedge Y}$	***Result***
$F \wedge F$	
$F \wedge T$	
$T \wedge F$	
$T \wedge T$	

X OR Y: Either X must be True or Y must be True for the result to be True, otherwise the result is False

$\boldsymbol{X \vee Y}$	***Result***
$F \vee F$	
$F \vee T$	
$T \vee F$	
$T \vee T$	

X XOR Y: Only X or Y must be True, otherwise the result is False

$\boldsymbol{X \oplus Y}$	***Result***
$F \oplus F$	
$F \oplus T$	
$T \oplus F$	
$T \oplus T$	

NOT X: if X is True the result is False, if X is False the result is True

$\boldsymbol{\neg X}$	***Result***
$\neg F$	
$\neg T$	

Boolean Laws

Boolean Laws are expressions that are built up from variables and the constants (**T** and **F**) and used to solve logic problems.

Name	Law
Commutativity of OR	$\boldsymbol{X \vee Y = Y \vee X}$
Commutativity of AND	$\boldsymbol{X \wedge Y = Y \wedge X}$
Associativity of OR	$\boldsymbol{X \vee (Y \vee Z) = (X \vee Y) \vee Z}$
Associativity of AND	$\boldsymbol{X \wedge (Y \wedge Z) = (X \wedge Y) \wedge Z}$
Distributivity of AND over OR	$\boldsymbol{X \wedge (Y \vee Z) = (X \wedge Y) \vee (X \wedge Z)}$
Distributivity of OR over AND	$\boldsymbol{X \vee (Y \wedge Z) = (X \vee Y) \wedge (X \vee Z)}$
Identity	$\boldsymbol{X \vee F = X}$
Identity	$\boldsymbol{X \wedge T = X}$
Annihilator	$\boldsymbol{X \wedge F = F}$
Annihilator	$\boldsymbol{X \vee T = T}$
Idempotence	$\boldsymbol{X \vee X = X}$
Idempotence	$\boldsymbol{X \wedge X = X}$
Absorption 1	$\boldsymbol{X \wedge (X \vee Y) = X}$
Absorption 2	$\boldsymbol{X \vee (X \wedge Y) = X}$
Complementation 1	$\boldsymbol{X \wedge \neg X = 0}$
Complementation 2	$\boldsymbol{X \vee \neg X = 1}$
Double Negation	$\boldsymbol{\neg(\neg X) = X}$
De Morgan 1	$\boldsymbol{\neg X \vee \neg Y = \neg(X \wedge Y)}$
De Morgan 2	$\boldsymbol{\neg X \wedge \neg Y = \neg(X \vee Y)}$

Abstract Boolean operators of the form ∨∧ ¬ will not be used in this course. They are not here for you to remember but for you to know that in the context of this course and other courses that you take, the operations and Laws themselves do not change, just the way the operators are written. This will ultimately make them easier for you to understand and therefore use across all the various contexts you will encounter them in.

Boolean algebra in electrical, computer, and software engineering

In the context of Electrical, Computer, and Software Engineering they are written in the following ways.

		Boolean Algebra variables and operations			
		Contextualised use			
Boolean logic		Electronic Digital Logic	C Programming Logic	Matlab logic	Binary bitwise programming logic
True	T	**1, High, On**	**Non 0 value**	**true**	**1**
False	F	**0, Low, Off**	**0**	**false**	**0**
AND (Conjunction)	∧	+	**&&**	**and**	**&**
OR (Disjunction)	∨	·	||	**or**	|
NOT (Negation)	¬	~X $\overline{X}$	**!**	**not**	~
XOR (Exclusive Or)	⊕	⊕ ^	^	**xor**	^

2.6 Number Systems

Learning Outcome

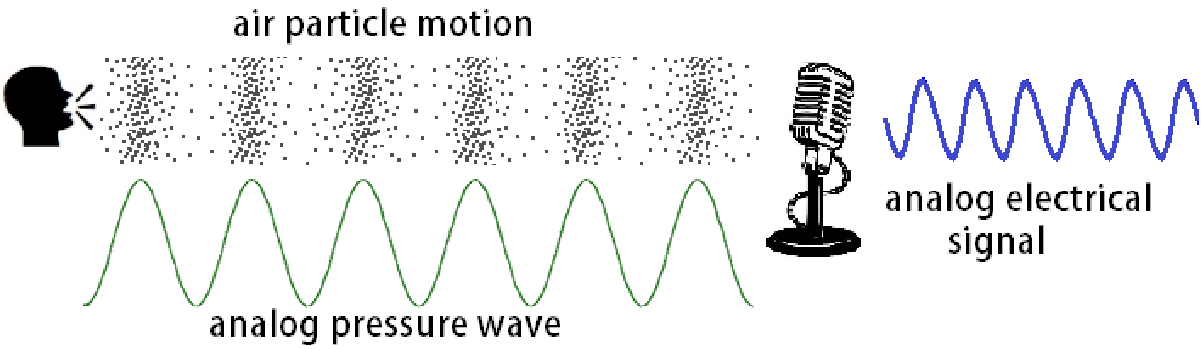
Understand the common number representation systems used in engineering and their limitations, and why they are so important.

Success Criteria

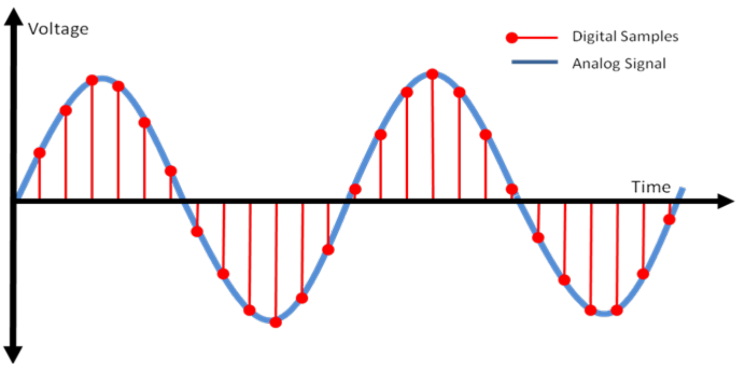
- Convert between decimal (base 10) and binary (base 2)
- Convert between decimal and hexadecimal (base 16)
- Convert between binary and hexadecimal
- Recognise the limitations of integer, signed integer and floating point number systems in electronic systems

2.6.1. Introduction

In engineering, the measurement of physical quantities (e.g. temperature, pressure, force or acceleration) is in continuous quantities. When these quantities are represented by a graph we say they are analog (from the word analogy) because they represent the actual quantity and vary proportionally with it. For example a graph of pressure or voltage is an analog to a the sound particle movement in the air.



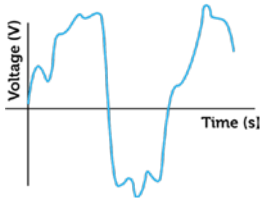
For every value of time, the analog signal below (in blue) can take one value amongst any of the values within a given range.



Source: www.mixrevu.com

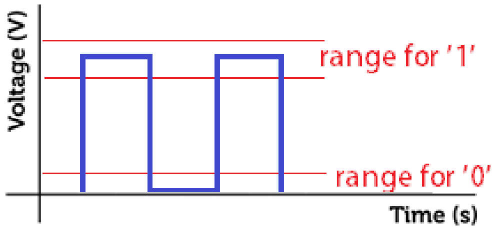
By contrast a digital signal can take only a finite number of values, as illustrated in red in the figure above.

While analog signals vary continuously (they are a direct analog of the actual input)



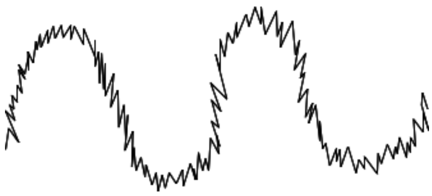
Source: ck12.org

With digital signal only a few restricted ranges of amplitude are allowed. Binary signals are the most common and these take on amplitudes in only two ranges, with the information associated with the ranges represented by the logic values 1 and 0.



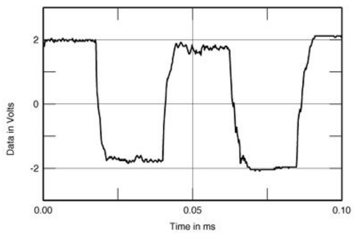
This gives digital signals an important advantage over analog signals.

In the case where noise distorts an analog signal, it is sometimes impossible to determine the precise characteristics of the original signal.



Source: we.stanford.edu

By contrast, after noise distorts a digital signal, we can still determine the logic values – provided that the noise amplitude does not take it outside its level.



Source: stereophile.com

If a logic circuit only needs to produce a voltage somewhere in the correct range, then component values in digital circuits do not need to be as precise as in analog circuits.

Another advantage of digital circuits is that modern integrated-circuit manufacturing technology, allows very complex digital logic circuits (containing millions of components) to be produced economically. Analog circuits however often require large capacitors and precise component values that are difficult, if not impossible, to manufacture into integrated circuits.

As digital systems have become so fundamental to our way of life, it is necessary to understand how to represent information in digital form, and the limitations of these systems. When several bits are grouped together they can be used to encode information. One bit can represent one of two states, or the integers from 0 to 1, or any other two-state quantity. This may be satisfactory for a single piece of yes/no logic, however more complex data requires more bits. N bits can encode 2^N pieces of information.

Binary number systems are central to a thorough understanding of digital signals and digital system (e.g. computers) so it is important to become familiar with the binary number system.

2.6.2. Number Radixes (Bases)

Base 10 (Decimal)

Our number system is positional. The numeral at each position is called a digit, and the weight or strength of its contribution to the entire number depends on the digit's position in the number. The allowable digits for any number system depend upon the radix or base. For instance in the decimal number system 10 digits (0 to 9) are used.

The general form for this is

$$\boldsymbol{d_1 b^n + d_{n-1} b^{n-1} + d_{n-2} b^{n-2} + \cdots\ d_0 b^0\ +\ d_{-1} b^{-1} +\ d_{-2} b^{-2} + \cdots\ d_{-m} b^{-m}}$$

Example 45.
Integer decimal number

$428_{10} =$

Example 46.
Decimal number with fractional part

$25.23_{10} =$

Base 2 (Binary)

The numeral at each position is called a binary digit or bit. The allowable bits for binary numbers are 0 and 1.

Example 47.
Integer binary number
$101_2 =$

We call the left-most bit of a binary number the Most Significant Bit (MSB) and the right-most bit the Least Significant Bit (LSB).

Just as numbers in decimal can have fractional pats, binary numbers can as well.

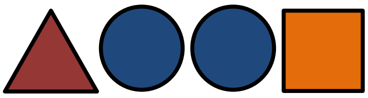
Example 48.
Binary number with fractional part

$1.101_2 =$

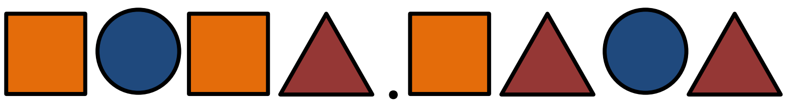
Base 3

The ternary (radix = 3) number system is not used a lot but makes an excellent exercise to check your general or abstract knowledge of number systems.

Example 49.
If the number

 $_3 = 29_{10}$

Example 50.
What is this number in base 10?

 $_3$

Base 16 (Hexadecimal)

The numeral at each position is called a hex digit. The 16 hex digits are 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F, where $A_{16} = 10_{10}$ … $C_{16} = 12_{10}$ … $F_{16} = 15_{10}$

Example 51.
Integer hexadecimal number

$BE7_{16} =$

Example 52.
Hexadecimal number with fractional part

$A.10B_{16} =$

2.6.3. Converting Between Bases

Decimal to Binary Conversion

Algorithm for integer Conversion

- Divide I_n by 2 to give a quotient I_{n+1} and a remainder R_{n+1}
- The remainder R_{n+1} is the next bit of the final binary number (from LSB to MSB).
- If I_{n+1}=0 then STOP, else go to step 1.

Example 53.
Convert 14_{10} to binary

Integer Conversion

Calculation	Bit

The final result is

Example 54.
Convert37_{10} to binary.

Integer Conversion

Calculation	Bit

The final result is

Conversions between Binary and Hexadecimal

Converting between base-16 and base-2 is particularly easy because each hex digit requires exactly four bits to represent it.

To convert from hexadecimal to binary, simply concatenate the binary representation of each hex digit.

To convert from binary to hexadecimal, simply group the bits in lots of 4 on either side of the binary point.

Hex	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
Bin	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111

Example 55.
Hex integer number to binary
$BE7_{16} =$

Example 56.
Binary integer number to hexadecimal

111010010_2 =

Other Bases

In the early days of computer development it was common to use octal numbers (base 8) where each octal digit was in the range 0 to 7. (In binary octal numbers required 3 bits to represent each digit.) Nowadays octal numbers are only used in a small number of special situations.

Other number bases are also possible but unlikely to catch on. For example, 32-bit data paths in modern microprocessors would make base-32 numbers potentially attractive. However, the catch is trying to devise unique symbols for the "digits" from 10_{10} to 31_{10}, not to mention the problem of memorising them.

Nibbles, Bytes and Words

An 8-bit number pattern is called a byte. Half a byte (4 bits) is called a nibble. An 'int' is at least 16 bits in size, but this is dependent upon the programming language, compiler and the microprocessor. So make no assumptions about variable size, check the documentation.

Express the binary numbers 10111_2 and 11101_2 in decimal form.

Example 57. $10111_2 =$

Example 58. $11101_2 =$

Express 38 in binary and hexadecimal.

Example 59. Conversion to binary

Example 60. Conversion to hexadecimal

Convert the numbers $FA5_{16}$ and 725_{16} to decimal form.

Example 61. $FA5_{16} =$

Example 62. $725_{16} =$

How many bits are needed to represent the positive decimal integers from a) 1). 0 through to 100, 2).0 through to 1000, 3).0 through to 10^6, and 4).0 through to N?

1. $2^6 = 64$ and $2^7 = 128$ and since $64 < \mathbf{100} < 128$ we deduce that 7 bits are required.
2. $2^9 = 512$ *and* $2^{10} = 1{,}024$ and since $512 < \mathbf{1000} < 1{,}024$ we deduce that 10 bits are required.
3. 2^{19} 524,288 *and* $2^{20} = 1{,}048{,}576$ and since $524{,}288 < \mathbf{10^6} < 1{,}048{,}576$ we deduce that 20 bits are required.
4. An n-bit binary number can "count" from $0\ to\ 2^n - 1$
   Where N can be represented as $N \leq 2^n - 1$, or $n \geq log_2(N+1)$ but as n must be an integer, we use the ceiling function $n = \lceil \cdot \rceil$, so the number of bits is $n = \lceil log_2(N+1) \rceil$.

Example 63. $N = 100$

Example 64. $N = 256$

Computer processors and number limits

Computer processors are designed with a fixed numbers of bits; and in electrical, computer, and software engineering the following systems are very common: 8, 16, 32 and 64 bit - although a computer systems engineer could design a 93 bit computer if needed for a special job, and they do design special computers for unique purposes.

Example 65. In an 8 bit computer positive or unsigned integers range from

The number $0101\ 1110_2$ =

Bin	0	1	0	1	1	1	1	0	$0101\ 1110_2$
Value	2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0	
Dec									

In a 16 bit computer positive integers range from $0000\ 0000\ 0000\ 0000_2$ to $1111\ 1111\ 1111\ 1111_2$ $(FFFF_{16})(1023_{10})$. This range is

It becomes easier to express 32-bit and 64-bit ranges in hexadecimal form

In a 32 bit computer, positive integers range from $0\ to\ FFFF\ FFFF_{16}$. In base 10 the range is $0\ to\ 2^{32}-1 =$

In a 64 bit computer positive integers range from $0\ to$

It is critical to understand the limits of binary numbers in a computer, as if the limit is exceeded the computer system may turn out a very strange result.

Example 66.

	Decimal	8 bit unsigned int.	Hex
	100	0110 0100	64
+	200	1100 1000	$C8$
=			

Example 67.

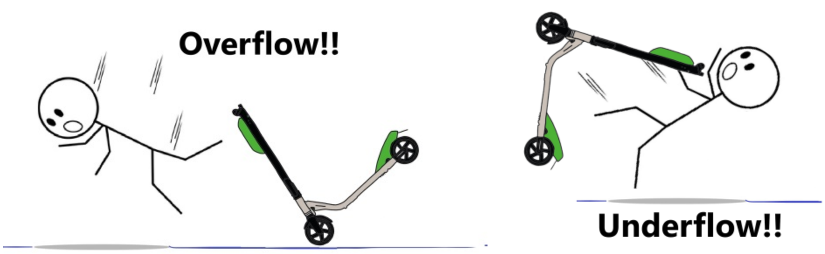
	Decimal	8 bit unsigned int.	Hex
	100	0110 0100	64
-	200	1100 1000	$C8$
=			

This can be very confusing, however it makes sense if you view these fixed number systems as circular rather than infinite number lines.

Check out Bill's 8-bit circular number line at
https://www.coursebuilder.cad.auckland.ac.nz/flexicourses/3721/publish/1/3_1.html

From the number line can you see how overflow and underflow happen.

When number systems issues happen in computer programs they are not always recognised until something goes wrong!



2.6.4. Negative integers

The binary numbers considered so far are positive integers. Computers, however, must represent both positive and negative numbers. As computers work on fixed bit sizes, they cannot be prefixed by a minus sign (-), so it is necessary to store the negative as part of the 8/16/32/64-bit number. This is done with the MSB of the number, where a 1 represents a negative integer and 0 represents a positive integer. This bit is called the sign bit and these numbers are called signed integers. This system is called two's complement.

For example a 4 bit signed integer, has numbers that range from $-8 \; to \; +7$

$$0111_2 = 7_{10}$$
$$0110_2 = 6_{10}$$
$$\ldots$$
$$010_2 = 2_{10}$$
$$001_2 = 1_{10}$$
$$000_2 = 0_{10}$$
$$111_2 = -1_{10}$$
$$110_2 = -2_{10}$$
$$101_2 = -3_{10}$$
$$\ldots$$
$$1010_2 = -7_{10}$$
$$1000_2 = -8_{10}$$

The range of the numbers is $-2^{n-1} \; to \; 2^{n-1} - 1$

While the pattern may not seem intuitive, the two's complement representation is used because it easy to manipulate binary numbers. For example all fundamental arithmetic ($+ - \div \times$) work as normal, and it is straight forward to negate a (positive or negative) number, you only need to invert all the bits (take the one's complement) and then add 1 (this is what makes it twos-complement).

Example 68. In an 8 bit computer
$-5 + 20 = 15$

0000 0101 (5)

1111 1011 (-5) Two's complement of 5

<u>0001 0100 (20)</u>

0000 1111 (15)

Note that any carry to the non-existent $9th$ bit is ignored.

The limits of 32 bit signed integers are:

$$-2^{32-1} \; to \; 2^{32-1} - 1 = -2{,}147{,}483{,}648 \; to \; 2{,}147{,}483{,}647$$

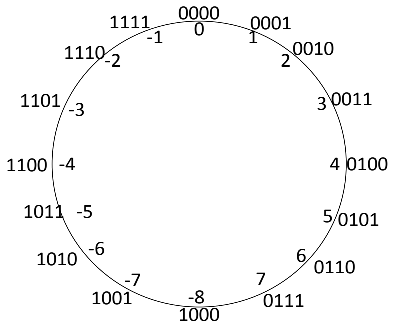
Example 69. $-17 + 34 =$

Example 70. $12 - 22 =$

Again think of this as a circular number line, but note the transition near the middle.

Take note that the MSB now represents -2^{n-1} so the 8-bit number $1101\ 1110_2$ =

Signed 4-bit integer circular number line



Bin	0	1	0	1	1	1	1	0	
Value	-2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0	
Dec	-128	64		16	8	4	2	0	94

Example 71. The limits of this system are important to understand, e.g. using the number line:

$$2 - 2 = \qquad 2 - 5 = \qquad 2 - 12 =$$

2.6.5. Fractional numbers

Computers also need to represent fractional numbers, and the IEE-754 floating-point number system is most often used. Numbers in this system are expressed in scientific notation, using different bits to encode the significant digits and the exponent separately. Numbers such as $1.07x10^{-3}$ or $-2.5476\ x\ 10^{23}$ can be expressed in the general form as $x = (-1)^s \times m \times b^e$ where s is called the sign, m is called the mantissa, b is called the exponent base and e is called the exponent. For example, for the number -2.5476×10^{23}, $s = 1,\ m = 2.5476, b = 10\ and\ e = 23$.

When programming computers ‘floats’ are numbers which have 32-bit single precision and ‘doubles’ have 64-bit double precision. However in computer systems engineering, you need to double ☺ check this as on small systems, a double may only have 32-bit precision.

The IEEE-754 32-bit single-precision floating-point representation is

$(-1)^s \times x1.m_2 \times 2^{e-127}$, where s occupies 1 bit, e is an unsigned 8-bit number $1 \leq e \leq 254$, and binary number m occupies 23 bits (giving a total of 32 bits). Note that the mantissa $1.m$ includes an implied 1 as the MSB (which is why it is more properly called the significand).

Example 72. Suppose that $s = 0, m = 101_2\ and\ e = 128_{10}$ a float encoded by these values is

$x = (-1)^0 \times 1.101_2 \times 2^{128-127}$

$= +1 \times 1.625_{10} \times 2$

$= 3.25_{10}$

Example 73.
Find the decimal floating-point number corresponding to $s = 1$, $m = 111_2$ and $e = 126_{10}$

The different length of the significand between float and double has a major effect on precision

A 32-bit float has a 23 bit significand

S EEEEEEEE MMMMMMMMMMMMMMMMMMMMMMM
0 1 8 9 31

A 64-bit double has a 51 bit significand

S EEEEEEEEEEE MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM
0 1 11 12 63

For example, π is an irrational number (cannot be expressed as a ratio of two integers). π = 3.1415926535897932384626433832795028841971693993751 ...

Here is π to 1 million digits ☺ https://www.angio.net/pi/digits/pi1000000.txt

However as a float in a computer system $\pi =$

And as a double $\pi =$

These examples indicate the resolution of floats ( ) and doubles ( ).

Example 74. When would it matter which you used?

	Float precision	Double precision
Circumference of earth	2.4 meters	10^{-9} (nanometer)
Distance to the sun	10 km	50 micrometers
Length of a day	5 mS	1 pS
Length of a century	3 minutes	$1\mu S$
Time since Big Bang	1000 years	1 minute

While the precision of a float in one calculation may be sufficient, issues using floats arise when doing many calculations e.g. in a loop, where a float was added 1 million times.

Example 75. Would it be best to use a float or a double to measure GPS coordinates?

Example 76. What is the largest positive number that can be represented by a single precision floating-point number?
https://en.wikipedia.org/wiki/Single-precision_floating-point_format

Example 77. What is the smallest positive non-zero number that can be represented by a single precision floating-point number?

Example 78. So why not just always use doubles, and never use floats?

Example 79. When would you need more precision than a double?

Where does engineering come into design and use of number systems? By knowing, understanding and respecting the limits of any system that uses numbers

2.7 Analog to Digital Conversion

Learning Outcomes

Understand the relationship between the real (analog) and the digital world.

Success Criteria

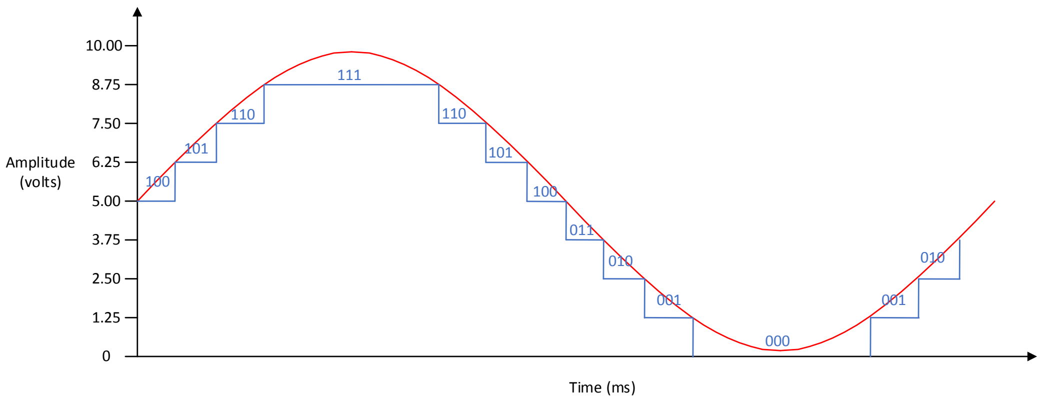
- Convert analog information to digital data
- Identify the limitation of quantization when information is converted to digital data.
- Work with the ideas of resolution

Conversion of signals from analog to digital form

Signal conditioning introduced in the previous section is used to change the output range of advice such as a sensor to a range that the following circuit can use most effectively. Often that following circuit is a digital control circuit, which makes decisions based upon the data or it could be a data logging system.

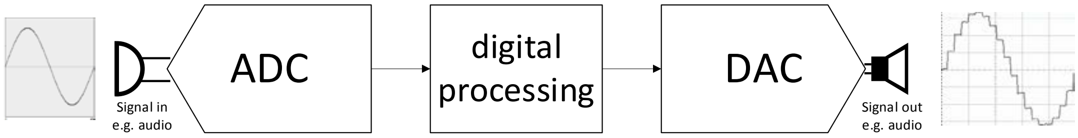
Quantization

In order to represent an analog signal in a digital format, the signal must be quantized. Quantization is the process of mapping a continuous or large range of input values to a smaller set of output values. This process is illustrated in the diagram below, in which each sample value is quantized into a three-bit number (the blue line), corresponding to the amplitude level into which it belongs.



In the diagram a continuously variable cosine input $f(t) = A\cos(2\pi f t + \phi) + B$ is quantized to 8 levels from 000_2 *to* 111_2. Because of quantization, exact signal amplitudes cannot be represented, as all amplitudes falling into a given zone have the same binary value. This reduced set of output levels introduces quantization error. For example in the 3-bit ADC system represented in the diagram, there is no difference between any input value $7.50 \leq input\ V < 8.75$, as this whole range will be represented by the binary number 110_2.

One implication of quantization error is that when a digital to analog converter (DAC) is used to recreate the original analog signal from the binary, it is possible to reconstruct only an approximation of the original signal.



Resolution

The term most often used to describe the ability to make useful measurements is resolution, which is

$$resolution\ (V) = {}^{A}/_{2^n}\ (where\ A = amplitude\ and\ n = numberof\ bits)$$

The resolution of the system described in the diagram is

$$Resolution\ (V) = {}^{10}/_{2} = 1.25V$$

Quantization error can be reduced (resolution increased) by using a larger number of levels; this, of course, requires more bits per sample to be used. In CD audio technology, 16-bit binary values are used. Using 16-bits makes it difficult for a listener to detect the effects of quantization error on the reconstructed audio signal.

When an engineer specifies instrumentation for taking measurements, she needs to determine the specifications for an ADC to convert the continuous input signal into binary. For example, suppose a signal which has a range of $-1V\ to\ +1V$ is to be measured. She needs to know the required resolution of the measurement (the acceptable quantization error) to determine how many bits will be required.

Input range	$-1\ to\ +1\ V$	$-1\ to\ +1\ V$	$-1\ to\ +1\ V$	$-1\ to\ +1\ V$
Amplitude	$2V$	$2V$	$2V$	$2V$
Required resolution	$\mathbf{0.2V}$	$\mathbf{100mV}$	$\mathbf{5mV}$	$\mathbf{0.5mV}$
Number of levels	10	20	400	4000
Min number of bits	4	5		
Actual resolution achieved				

Increasing →

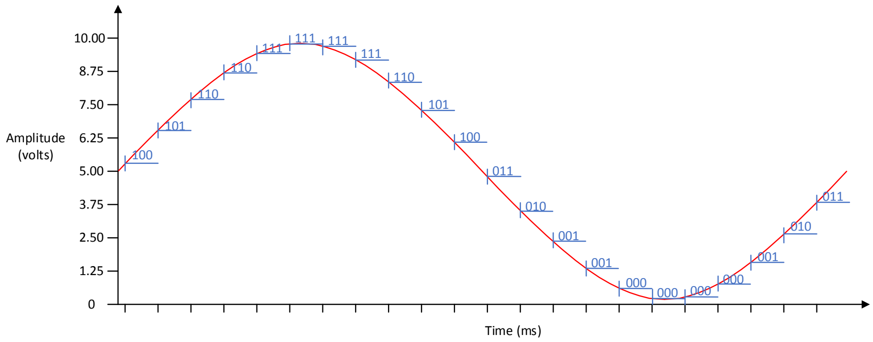
Note the way we use the term 'increasing resolution'. We say that we are increasing resolution, even though the resolution in Volts is getting smaller. What we are really saying, is that we are increasing our ability to resolve smaller and smaller changes in the input signal.

Confusingly in engineering we also use the term resolution to refer to the number of bits being used, so an ADC might have an 8, 10, 12 or 24-bit resolution.

Sampling

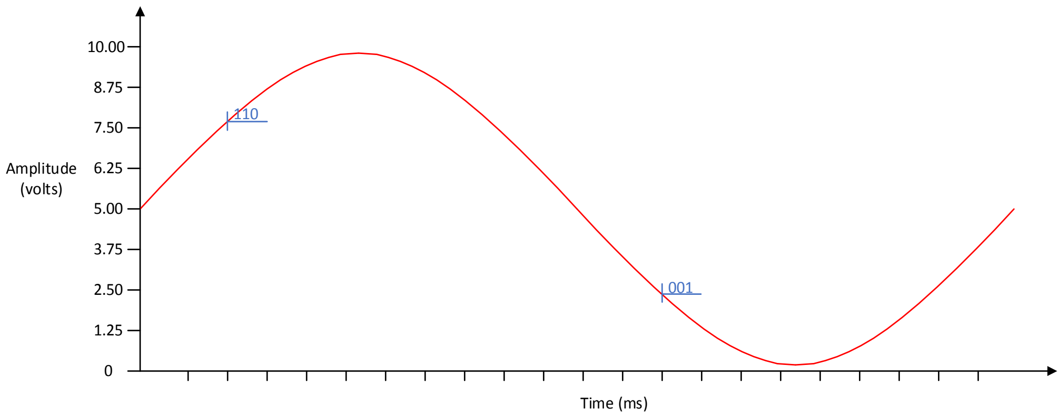
Another significant consideration of an ADC process is the frequency or rate, f_s, at which a signal is sampled (captured and digitised). This depends on the rate of change (the frequency) of the input signal. While it is desirable to use a low sampling rate (to minimize the amount of data that must be later processed or stored) there is a lower limit of sampling rate, so that we do not lose information about the input signal.

For example if a structural engineer is measuring stress on a bridge beam over a 7 day period, there is little point in sampling every millisecond, however they may miss a significant event if sampling took place every minute. If vehicles pass the point in the bridge to be measured at a maximum of 110 km/hr or 30.5 m/s; and a vehicle is 2m in length, then the rate would have to be at least 15 times per second so as not to miss a significant event.



In the diagram above the input signal (red cosine) has a total period of $22mS$, $f = 5.455Hz$

It is sampled at $1kHz$ (every 1ms). This is actually more than is required to reproduce the frequency information in the signal. Harry Nyquist determined that the sample rate only needs to be twice the highest frequency in a signal. So this signal could be sampled at $91Hz$, as in the diagram below.



The two concepts of quantization and sample rate now need to be connected together so that engineering problems can be analysed and solved.

Example 80.

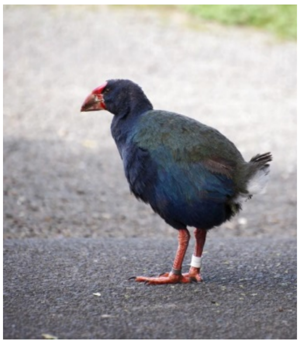
A mechanical engineer needs to record the ambient noise level next to a generator at a power station over a 24 hour period. Audio is typically sampled at 44.1kHz using 16-bit binary. What size storage device would be required?
1 second of audio data would require

Example 81.

If you put an 8GB USB drive to the noise meter, and it ran out of space, and your charge out rate is $860 per day, and your boss had to send you back to Huntly power station from Auckland (at least 4 hour return trip) to do the job again, what would be the lost profit on the job?

Example 82.

An environmental engineer from DOC is monitoring Takahē at Maungatautari, (the sanctuary mountain of 'inland island' in the Waikato). The goal is to measure the bird's temperature (to track health and mortality rates). A temperature change of 0.5 $degC$ will indicate a potential health issue. A thermistor circuit provides the range from $0\ to\ 40\ degC$, over the range of voltage $1.0\ to\ 3.2V$, what resolution would be required from the ADC to detect a 0.25 $degC$ change in temperature?



Example 83.

A monitoring device is installed subcutaneously (under the skin) of the bird (like this SIRTRACK one). If the circuit draws 3.2mA for the ½ sec it takes to record and transmit a measurement, and the required sample rate is three times per day. How often would a 30,000mAh battery have to be changed?

sirtrack.co.nz/index.php/avian/vhf/implant

Example 84.

In Brazil, near Manaus, the Amazon Tall Tower Observatory is used to capture a wide range of atmospheric data. For example CO is monitored at 8 different heights on the tower. Using a CO sensor circuit, CO in the range 1 to 100ppm is monitored over the voltage range 0.45 to 4.35V, we want a resolution of 2ppm, what resolution ADC would be required?

twitter.com/attoresearch

In general, the phenomena of interest to engineers are not inherently electrical. However, the instruments used by engineers to interpret and act upon these phenomena are electrical. This occurs because of the great ease with which electrical signals may be used, interpreted, and recorded. For instance, a Chemical engineer monitoring (and controlling) the temperature in a kiln will almost inevitably turn to electronic instrumentation to carry out this control process.

An *input transducer* (or commonly referred to as a ***sensor***) is a device which converts a non-electrical signal into an electrical signal. This signal is usually so small that it cannot be directly recorded or displayed on instrumentation or directly input to a computer-based data acquisition system. The output of the transducer is passed through a signal conditioning circuit which provides amplification and may sometimes perform other signal conditioning functions. Often, we may wish to take some action based upon our interpretation of the transducer's signal. If we wish to control the operation of some non-electrical plant, we must convert the electrical signal into a non-electrical signal. To do this we use an *output transducer* (which is more commonly referred to as an *actuator*).

These new terms are best illustrated by using the familiar example of a sound system composed of a turntable, preamplifier, amplifier, and speaker. The cartridge of the turntable is the input transducer. It converts the mechanically inscribed sound on the record into a weak electrical signal. The preamplifier amplifies (i.e., increases the amplitude of) this signal, and the amplifier provides further (higher power) amplification in order to produce a sufficiently powerful signal to drive the voice coil of the speaker. Thus, the pre-amplifier and the amplifier perform signal conditioning. The speaker is the output transducer since it converts the electrical signal into a non-electrical (sound) signal.

The kinds of measurements a Civil engineer makes are markedly different from those made by a Chemical and Materials engineer or by a Biomedical engineer. However, it is really only the *type* of transducer (input and output) that is used that differs between the various fields of engineering. Once some sort of electrical signal is obtained to represent the physical quantity being measured, it is processed by signal conditioning devices (such as amplifiers) and displayed on meters or acquired by computer-based data acquisition systems that are common to all engineering discipline. Thus, the underlying instrumentation systems in which *every* modern engineer to deal with the physical world is the same, and it is therefore crucial that we foster an appreciation for how these *systems of monitoring, analysis, and response technologies* work.

3.1 Monitoring Technologies

Many texts categorise transducers by the measurand (e.g., temperature, pressure, displacement, etc.), an alternative, electrical engineering, approach is to categorise based upon *means of operation*. In other words, instead of classifying a transducer by the type of non-electrical signal it monitors, it is more meaningful to categorise the transducer by the underlying electrical parameter whose change in magnitude relates to the non-electrical signal of interest. This way of classification provides a natural scheme for organising *all* transducers by the five common electrical analogues: *voltage*, *current*, *resistance*, *capacitance*, and *inductance*.

We will primarily focus on the operations of *resistive* transducers and examine some of the electronics necessary to interface them with subsequent analysis and response systems. The remaining types of transducer all operates via the same principle, the only difference is the electrical property that is used as the analogue to the non-electrical signal of interest and the underlying mechanisms for doing so.

3.1.1. Resistive Input Transducer

Intended Learning Outcomes

- Be able to describe, and relate, the operation of resistive transducers to the governing physical quantities.
- Be able to determine the relevant electrical and physical quantities pertaining to a resistive transducer within its nominal operating limits.

Resistive input transducers are devices in which a change in their resistances is related to the non-electrical parameter of interest. These transducers are simply resistive materials designed and constructed in clever arrangements to exploits the change in their physical properties due to variation in the non-electrical quantity being sensed. They operate by variation in length (l), area (A), or resistivity (ρ), parameters that affects the resistance R of a material by[65]

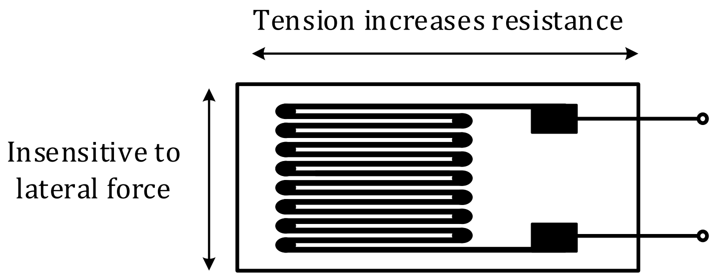
$$R = \rho \cdot \frac{l}{A} \qquad (\text{unit: } \Omega)$$

Naturally, the phenomenon of interest should only change **one** of these parameters in a resistive transducer so that a proper interpretation can be made from the corresponding change in resistance.

[65] Can you link this relationship to the intuitive description of the underlying physics of operation of a resistor as given in **Sect. 2.1.3**?

Strain Gauge

Structural engineers (a sub-discipline within the field of Civil engineering) may be concerned with monitoring of the structural health of load-bearing members in large infrastructure such as bridges and high-rise buildings. The transducer used for such monitoring is known as a *strain gauge*. The strain gauge is an example of a resistive input transducer whose change in resistance is related to changes in length. Small changes (which are indicative of strain in a member under investigation) may be accurately measured in this manner. These devices, one of which is shown below, are the size of a postage stamp or smaller.



The strain gauge is affixed to the member to be tested via an adhesive coating supplied on the back of the gauge. As the member under test (or subject to long-term monitoring) is stretched or compressed along the indicated direction, changes in the length of the long axis of the gauge produce a directly proportional change in the resistance. This is described quantitatively by

$$\frac{\Delta R}{R_o} = G \cdot \underbrace{\left(\frac{\Delta l}{l_o}\right)}_{=\text{ strain, } \epsilon}$$

where the constant of proportionality, G, is the *gauge factor* which is specific to the material and the manufacturing process of the device[66].

A complicating feature is that the resistance of the gauge also depends on its temperature (see *Resistance Thermometers*), introducing a potential confounding factor. To overcome this difficulty, a second strain gauge (of properties similar to the first and held at the same temperature, but not attached to the test member) is used commonly used as a reference in the interfacing circuitry to the first to compensate for any temperature variation during the measurement (see the *Wheatstone Bridge* later on in **Sect. 3.1.2**).

[66] For example, the commercially popular constantan gauge has a gauge factor of 2.

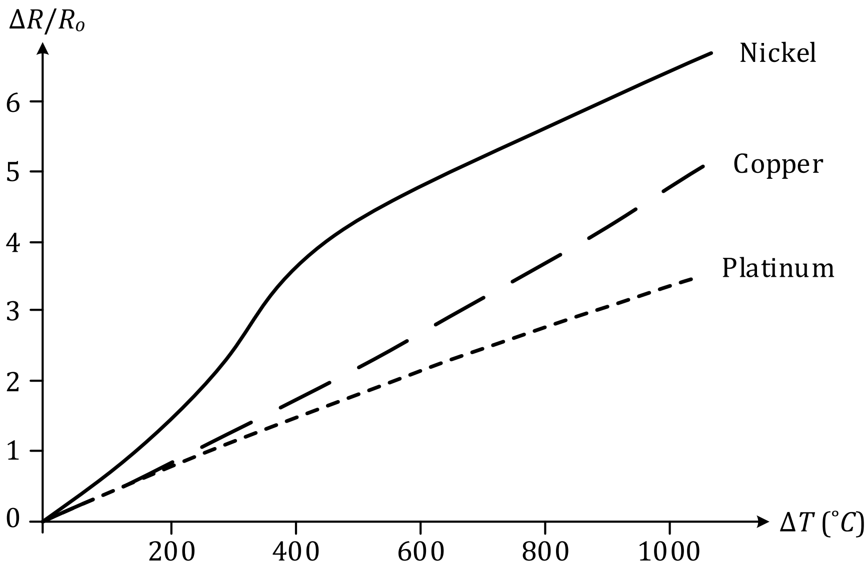
Resistance Thermometers

One of the ways in which we can measure temperature is to make use of the fact that the value of any resistance is temperature dependent – what appeared to be a nuisance for the strain gauge is now exploited to an advantage. Specifically, the *resistivity* of the material changes with temperature according to the approximate linear relationship

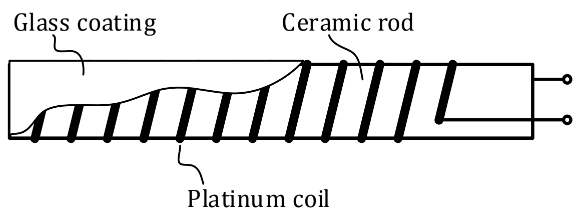
$$\rho(T) = \rho_0\big(1 + \alpha(T - T_0)\big) \qquad \Rightarrow \qquad R(T) = R_0\big(1 + \alpha(T - T_0)\big)$$

where ρ_0 is the resistivity at temperature T_0, and α is the *temperature coefficient* of the resistive material[67]. This means the change in resistance of the material can be used as a direct surrogate to measure the change in temperature from T_0

$$\frac{\Delta R}{R_0} = \alpha \cdot (\Delta T).$$



Generally, the resistance thermometer is constructed by winding platinum wire around a former, the longer the length of the wire, the more sensitive the thermometer is to changes in temperature[68]. Platinum is commonly used for its linear temperature dependency over other materials despite its modest coefficient in comparison to others, and is the standard from -190°C to 660°C.



[67] Temperature coefficients of some common materials are: nickel (0.0067), copper (0.0043), silver (0.0041), and platinum (0.0039). Note that in general α also varies as a function of temperature, but over limited range it is approximately constant.

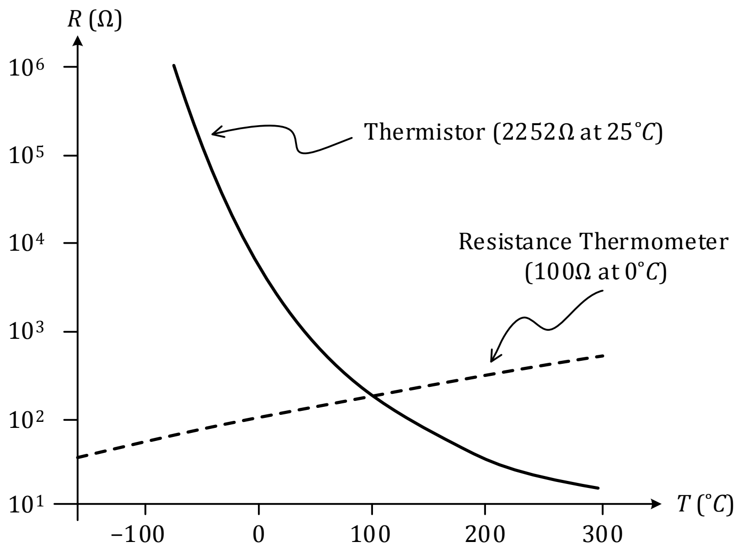
[68] A longer wire necessarily means a larger resistance, a change in temperature will thus give rise to a larger measurable difference in the resistance.

Care must be taken to avoid i^2R heating (or *Ohmic* heating), which becomes significant as the current becomes too large and would invariably confound the measurement through self-heating.

It is worth noting that the resistivity ρ of a material is not only a function of temperature, but also subject to a variety of other physical phenomena such as pressure, light intensity, magnetic forces, etc. Resistive transducers can thus be made to detect numerous other non-electrical quantities. Notice the only difference is the mechanism in which the material property is changed.

Thermistor

Another common resistive input transducer used for temperature sensing is the *thermistor*. The active material of a thermistor is a type of *semiconductor*[69] (e.g., germanium, silicon, metallic oxides, etc.) whose resistance changes with temperature. What makes thermistors particular attractive over the resistance thermometer is their relatively large temperature dependence.

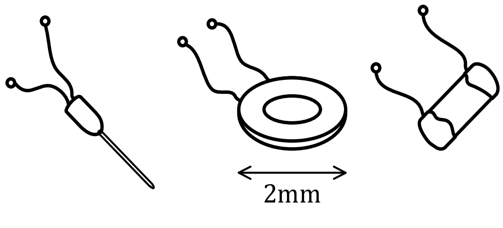


While the response is a non-linear, exponentially decaying, function of temperature

$$R(T) = R_0 e^{\beta\left(\frac{1}{T} - \frac{1}{T_0}\right)}$$

where β is a material specific constant, and R_0 is the known resistance at a given temperature T_0, over a small range of temperatures a linear approximation is valid. In general thermistors are used over the temperature range from 0°C to 100°C, although higher temperature measurements are possible.

Having a large sensitivity means the physical size of thermistors can thus be small (typically 1 to 2 mm) and be made cheap in comparison to their counterpart. Once again, care must be taken to avoid i^2R heating which is especially detrimental for the thermistors due to their highly *negative* temperature coefficient leading to *thermal runaway* – an increase in temperature due to self-heating would decrease the resistance and thereby further increase the current until catastrophic failure unless a current-controlling resistance is also in the circuit.



[69] A semiconductor is a material that acts as both an insulator and a conductor depending on the temperature and the potential difference applied across it.

3.1.2. Sensor Interface Circuitry

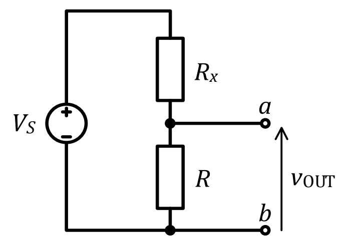
Intended Learning Outcomes

- Be able to determine, and provide specifications for, the relevant circuitry necessary to interface a resistive transducer to subsequent electrical systems.
- Be able to anticipate, and provide the specifications for, the necessary signal conditioning circuitry, given the characteristics of both the transducer circuitry and the subsequent circuitry

In order for transducers to be useful, we must have accurate means of monitoring the value, or at the very least the changes in the value, of their electrical properties as the non-electrical quantities of interest vary. For a resistive transducer, this means we need electronic circuits that can facilitate the measurement of miniscule changes in resistance, but also flexible enough to allow the sensor to be interfaced directly for use with subsequent control systems for automatic analysis and response without manual intervention.

The Voltage Divider

The most primitive but intuitive circuit for determining the resistance of a resistive transducer is just a basic voltage divider with a known resistor R in series with the transducer of unknown value R_x, supplied by a known constant voltage supply V_S:



Standard analysis shows that the output voltage v_{OUT} is related to the transducer resistance R_x by

$$v_{\text{OUT}} = v_a - v_b = \left(\frac{R}{R + R_x}\right) \cdot V_S \qquad \Rightarrow \qquad R_x = \left(\frac{V_S}{v_{\text{OUT}}} - 1\right) \cdot R.$$

Thus, by measuring the output voltage v_{OUT} over time, we can directly infer what the transducer resistance is, and hence changes in the underpinning physical quantity being monitored.

In practice it may not be necessary to know the actual resistance, instead, we may simply use the output voltage as an *analogue* to R_x as it changes (inverse) proportionally to R_x. In this way, the variation in the output tracks the changes in resistance thereby providing a useable voltage representation of measurand. This voltage can then be used by subsequent electrical or digital systems that make use of the physical quantity being sensed (e.g., to regulate the temperature, to trigger a warning indicator, or simply to display the measurand in a meaningful way).

Example 85. A Pt100 sensor[70] is a platinum thermometer with a resistance of 100Ω at $0^{\circ}\mathrm{C}$ and a temperature coefficient of $0.0039/^{\circ}\mathrm{C}$. The sensor is to be used to monitor the temperature between 0 to $95^{\circ}\mathrm{C}$. Design a voltage-divider interface with a 5V battery such that the output will vary nominally about 2.5V at $25^{\circ}\mathrm{C}$, and determine the largest output swing possible over the monitored range.

Despite its simplicity, the voltage-divider interface comes with many weaknesses and sources of error:

1. The output voltage is a function of the supply voltage, which has inherent noise, imprecisions, and will likely drift over time leading to inaccurate measurements and error in the inferred value of R_x.

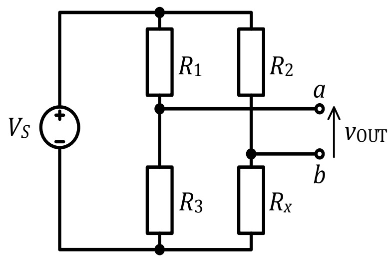
2. Any temperature-related drift in the sensor R_x will confound the measurement even if the measurand is unrelated to temperature (e.g., strain).

3. Changes in the output of the divider (due to variation in R_x) is a small variation about a large offset (DC) voltage. If this output voltage is to be used by subsequent systems, it necessarily requires a modest amount of amplification (typically 10 to 1000 times), but this means the meaningless offset will also be amplified by the same amount and likely become incompatible with the following systems, unless the offset is removed.

[70] "Pt" is the symbol for platinum, and "100" stands for the resistance in ohms at $0^{\circ}\mathrm{C}$. A Pt1000 sensor therefore has a resistance of 1000Ω at $0^{\circ}\mathrm{C}$. Pt100 is most commonly used.

The Wheatstone Bridge

The shortcomings of the simple voltage-divider interface can be mitigated by modifying it to contain an additional divider in parallel, and taking the output to be *across* the two branches instead,



giving rise to the topology commonly referred to as the *Wheatstone Bridge*.

By placing the transducer R_x in one 'arm' of the bridge, standard analysis shows that the output v_{OUT} is given by

$$v_{\mathrm{OUT}} = v_a - v_b = \left(\frac{R_3}{R_1 + R_3} - \frac{R_x}{R_2 + R_x}\right) \cdot V_S.$$

This clever design overcomes the aforementioned weaknesses (1) to (3) in the following ways:

1. To measure the unknown resistance R_x, we can adjust[71] the resistor values R_2 and R_3 with a known R_1 **until** $v_{\mathrm{OUT}} = 0$ (at which we say the bridge is *balanced*) and can deduce immediately

$$\frac{R_3}{R_1 + R_3} - \frac{R_x}{R_2 + R_x} = 0 \qquad \Rightarrow \qquad R_x = \frac{R_2 R_3}{R_1}.$$

   Notice via this method the inferred value of R_x is **independent** of the supply voltage and only depends on the resistor ratio.

2. If we wish to make the measurements *independent* of temperature, e.g., for a strain gauge, we can decouple v_{OUT} from such a dependency by replacing R_3 with an identical gauge to the active one, R_x, but mounted in a strain-free placement. Both gauges will now track in parallel with variations in temperature **cancelling** its effect, leaving only the effects of strain in v_{OUT}.

3. Since v_{OUT} is taken as a *difference* across the two branches, its changes can be made to vary about zero volt (or in practice some small offset) – we set the known resistors such that at the nominal R_x value, the bridge is balanced. The output voltage thus directly represent the physical quantity **without the** (meaningless) **offset**.

---

[71] To do so, we usually use a coarser variable resistor for R_2 and a finer variable resistor for R_3. Therefore, by varying R_2 we get a ballpark estimate for R_x while R_3 provides the fine tuning necessary to narrow down the value.

Example 86. A Wheatstone bridge has R_1 a fixed 1kΩ resistor, while R_3 can be adjusted in 1-Ω steps from 0 to 1100Ω, and R_2 can be selected to be 1kΩ, 10kΩ, 100kΩ, or 1MΩ.

(a) If the bridge is balanced with $R_3 = 732\Omega$ and $R_2 = 10\text{k}\Omega$, what is the value of the resistive transducer R_x?

(b) What is the largest value of R_x for which the bridge can be balanced?

(c) Suppose that $R_2 = 1\text{M}\Omega$. What is the increment between values of R_x for which the bridge can be precisely balanced?

Practical Considerations

While the variation in the output voltage of the sensor interface circuitry is representative of the underpinning physical quantity being monitored, in practice this variation is extremely small and cannot be used directly to drive or control any subsequent systems. For example, strain measurements rarely involve quantities larger than a few *milli-strain* (i.e., $\Delta l/l$ is in the ballpark of 10^{-3}), resulting in only a fraction of a percent change in the resistance of a strain gauge. Similarly, the temperature coefficients of resistive thermometers are typically very small, between 2×10^{-5} and 2×10^{-2} per $^{\circ}\mathrm{C}$ leading to almost unmeasurable change in output voltage. While the thermistors have a larger sensitivity (typically -4% to -7% of R per $^{\circ}\mathrm{C}$) the corresponding changes in the output voltage is still miniscule in comparison to the sort of magnitudes required by typical electronics.

Further complications arise in the form of unavoidable offset (as previously mentioned) inherent to the operation of the chosen interface circuitry, but more noticeably, the undesirable effect the subsequent systems has on the output of the sensor circuitry once connected – the *loading effect* (see **Sect. 3.2.1** later on) – causing the voltage to deviate from its expected value.

We must therefore have some sort of electronics that *amplify* the output voltage of the sensing circuit, *remove* the unwanted effects inherent to the sensing circuit where appropriate, and *minimise* the impact the subsequent systems have on the output of the sensor circuit *before* it can be interpreted and used subsequently. The *signal conditioning circuitry* used to alleviate these undesirable artefacts form the basis behind the *analysis technologies* to be discussed in **Sect. 3.2.1**.

3.1.3. Problems

1. Describe (with the aid of suitable diagrams if necessary) the operation of a resistive transducer which is based on variation in length, and temperature. Describe how a resistive *pressure* sensor might operate based on your understanding of the physical properties of resistive transducers.

2. The output of a Wheatstone bridge used to interface an input transducer produces a signal in the range 0 V to 5 mV. The subsequent system can only accept inputs in the range 0 V to 5 V. How much amplification is required in the signal conditioning circuitry in order to make the sensor signal compatible?

3. The output of a voltage divider interface used to interface an input transducer produces a signal in the range of 5 mV to 10 mV. The subsequent system can only accept inputs in the range of 0 V to 5 V. What is the voltage offset that should be removed, and the amplification required in the signal conditioning circuitry in order to make the sensor signal compatible?

Solutions to Problems

1. See Coursebook.
2. 1000.
3. 5 mV and 1000.

3.2 Analysis Technologies

The second stage in our model of SMART systems is analysis. In this section of the course, signals from sensors are conditioned for use in digital logic circuits, and for use by software in microcontroller based embedded systems.

3.2.1. Signal Conditioning – The Operational Amplifier

Electrical signals are generated by many sources and vary in amplitude and frequency. This course focusses predominantly on DC signals – those that vary in amplitude, e.g. voltage from sensor circuits. While there are different components and circuits for this, there is one key electronic device that is comonly used, the operational amplifier or OpAmp.

Learning Outcomes – OpAmps

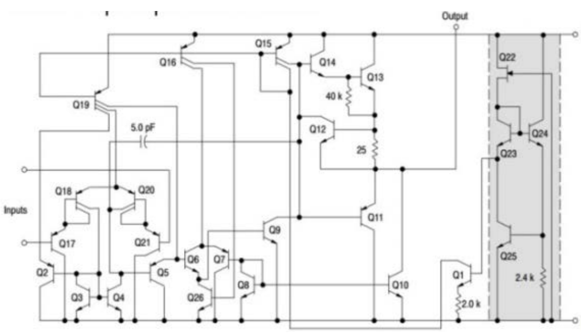
Understand how the basic principles of OpAmp devices, along with circuit analysis techniques, are used to analyse and design OpAmp circuits to condition DC signals.

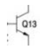
Success Criteria

- Describe the characteristics of an ideal OpAmp device.
- Describe the main characteristic of an OpAmp in a negative feedback circuit.
- Explain the difference between amplification and gain.
- Apply ideal OpAmp characteristics and circuit analysis techniques to analyse OpAmp circuits.
- Explain loading and calculate loading effects in circuits.
- Use standard OpAmp models to design sensor conditioning circuits.

Introduction to the OpAmp

An operational amplifier is an electronic circuit that can be easily modified by external components, to perform closely controlled amplification of signals. Its early uses include audio signal amplification for telephone lines, and use in analog computers. In analog computers, numeric values are represented by different levels of voltage and OpAmp circuits represent mathematical equations (including differential equations). While analog computers are not so common ([they still exist)](), applications of OpAmps have only increased and are a core component used in the design of analog circuits. In this course OpAmps are explained through using them to condition or amplify signals from sensors.



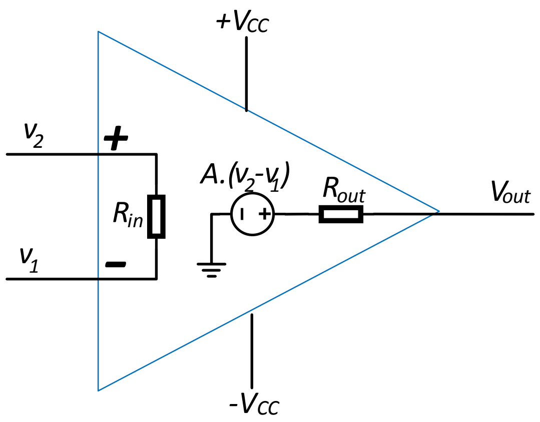
The internal circuit of an OpAmp is complex, such as the circuit of the LM318 IC in this diagram. It is made of ___  transistors. An IC is made from a single piece of silicon. Some ICs such as computer chips have more than a billion transistors in them. The transistor (invented in 1947) could be considered the most significant invention of the 20^{th} century.

While the OpAmp is highly complex, it can be represented using the equivalent circuit model, which allows all its complexity to be abstracted down to a few simple descriptions of its terminal characteristics.

<u>**The main characteristic of an OpAmp is that it amplifies the difference between its inputs.**</u>

Where *A* is the amplification factor of the OpAmp device.

<u>Equivalent circuit model for an OpAmp</u>



<u>Ideal OpAmp characteristics</u>

1.
2.
3.
4.

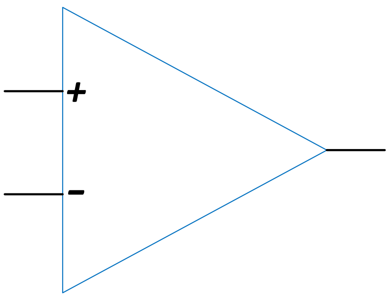
<u>Typical OpAmp characteristics</u>

$A =$

$R_{in} =$

$R_{out} =$

Usually the OpAmp is drawn like this



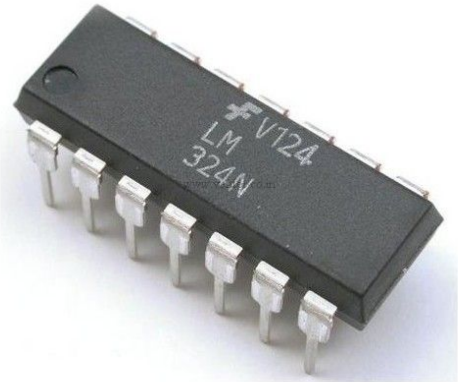
Where:

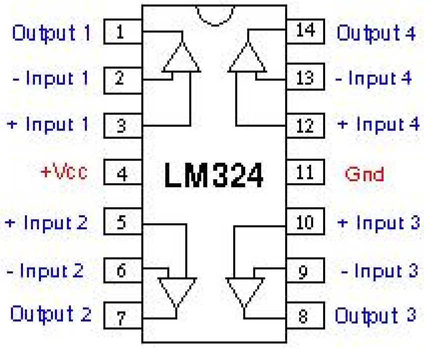
+ is called the

− is called the

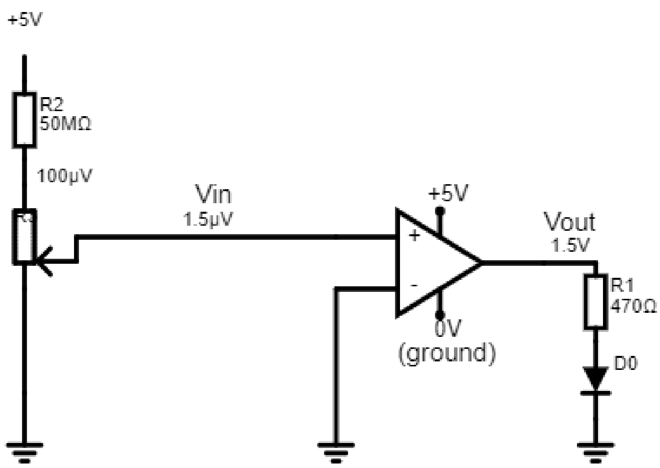
Note that the power supply connections are usually not drawn to make circuit diagrams easier to read – though they are still used - otherwise real OpAmps would not work!

A common OpAmp integrated circuit (IC or chip), the LM324, has 4 OpAmps inside it.





To demonstrate the Amplification (A) of an OpAmp device, a circuit like this could be built.

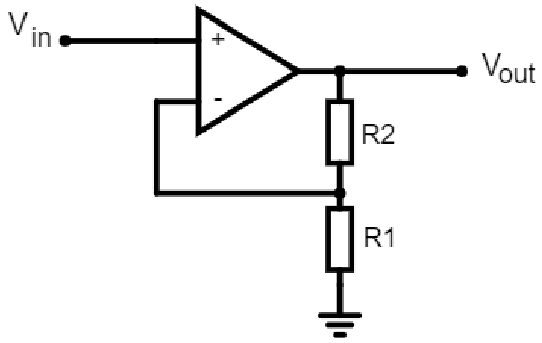


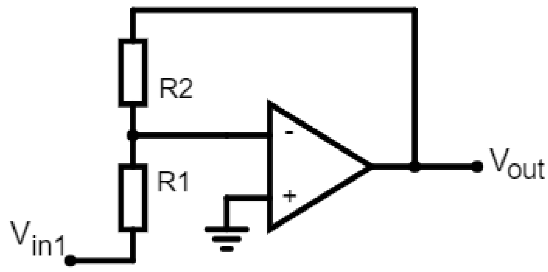
Example 87.
If the gain is 10^6

Vin(uV)	Vin(V)	Vout(V)
1		
3.5		
8		

However while this can help understanding of OpAmp amplification, it is not at all useful to have a circuit with a gain G of 10^6, circuits with much finer control are required.

From a mathematical modelling point of view, gain (G) is multiplication of an input voltage by a constant e.g. $v_{out} = G \times v_{in}$. Circuits using OpAmp devices are usually designed to amplify input signals by relatively small values of gain ($e.g. G = 5, G = 2.6,\ G = 15$). To do this a technique called negative feedback is used. Negative feedback occurs when a portion of the output is fed back to the inverting input. There are two ways OpAmp circuits are configured using negative feedback, and we describe those in terms of the input connections to the circuits.





non-inverting input

In this circuit, as the input voltage increases, the output voltage

$$v_{out} = G \times v_{in}$$

inverting input

In this circuit, as the input voltage increases, the output voltage

$$v_{out} = -G \times v_{in}$$

In each case a voltage divider is used to provide negative feedback. Negative feedback works like this: as the difference between the two OpAmp connections ($v_+$and v_-) increases, v_{out} increases. A portion of the increased v_{out} is fed back to the inverting input via the voltage divider, this reduces the voltage on the inverting input, thus reducing the difference between $v_+$and v_-. This process is fast happen almost instantaneously, so changes in V_{out} follow changes in v_- *and* v_+ rapidly.

As the gain is so high the difference between the two input signals is almost 0, which effectively means that when analysing and modelling negative feedback OpAmp circuits $v_+ = v_-$. This is another of the key properties of OpAmp circuits that can be used to analyse them.

Key OpAmp properties to know

$\boldsymbol{A \approx \infty\ 0}$ **(open loop gain is infinite)**

$$\boldsymbol{R_{in} \approx \infty\,\Omega}$$

$$\boldsymbol{\therefore i_2 = i_1 = 0A}$$

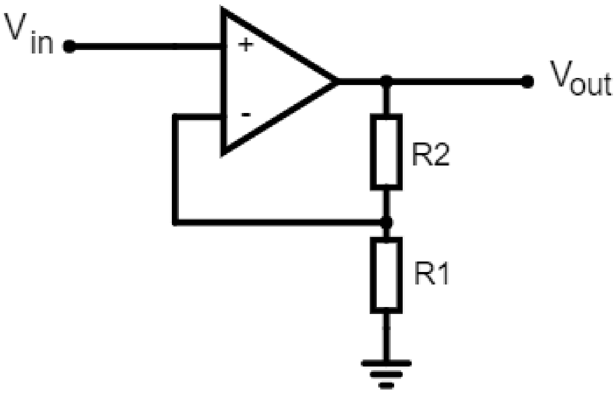
$$\boldsymbol{R_{out} \approx 0\,\Omega}$$

Negative feedback OpAmp circuit properties for analysis purposes

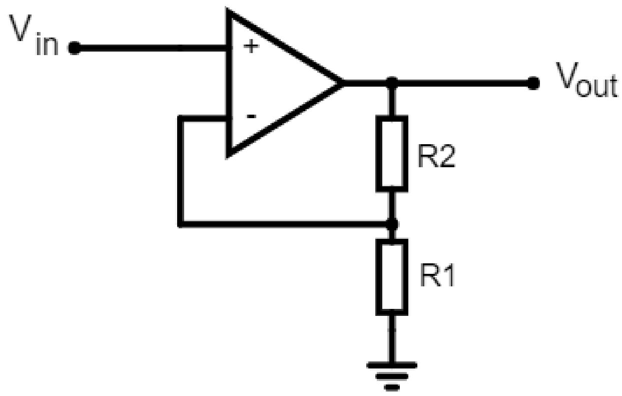
$$\boldsymbol{v_+ = v_-}$$

Be clear that **A** is the amplification factor of an OpAmp device, and **G** is the gain of any circuit (that may or may not include an OpAmp)

The non-inverting OpAmp circuit.



Example 88.



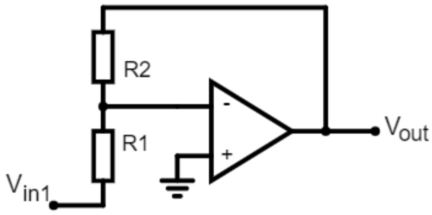
Example 89.

If $v_{in} = 0.1V, R_2 = 10k\Omega \; and \; R_1 = 1k\Omega$, work out v_{out} and the gain(G)

Example 90.

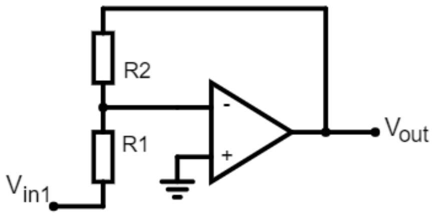
If $v_{in} = 0.3V, R_2 = 91k\Omega \; and \; R_1 = 4k7\Omega$, work out v_{out}

The inverting OpAmp circuit.



Example 91.

Apply OpAmp negative feedback circuit properties



Example 92.

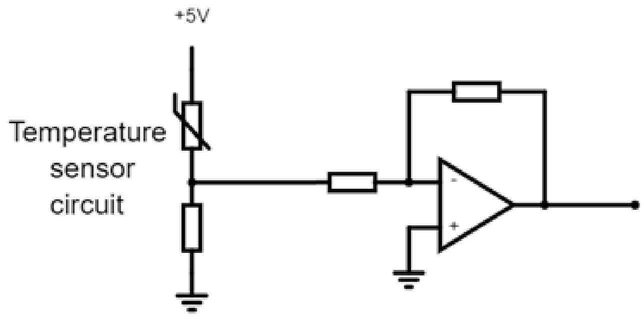
If $v_{in} = 0.1V$, $R_2 = 10k\Omega \; and \; R_1 = 1k\Omega$ work out $v_{out}$and the gain(G)

Example 93.

If $v_{out} = 81mV$, $R_2 = 27k\Omega \; and \; R_1 = 11k\Omega$ work out $v_{in}$and the gain(G)

Modelling Loading

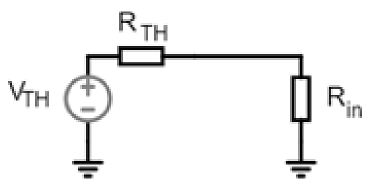
One critical aspect of circuit design is developing an understanding loading; this is when two circuits are connected together, the output characteristics of the first can have an undesirable effect on the input of the second AND/OR that the input characteristics of the second can have an undesirable effect on the output of the first. In the circuit below there is a voltage divider with thermistor in a temperature sensing circuit, connected to an inverting amplifier.



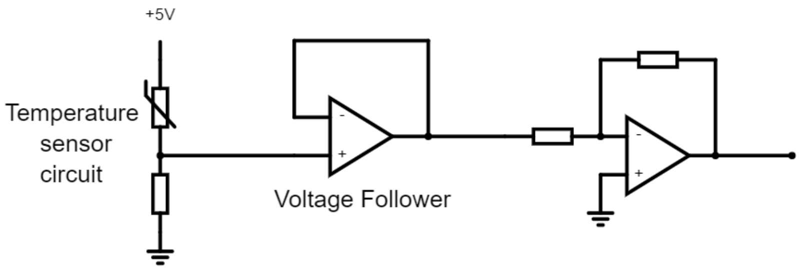
The presence of the inverting OpAmp circuit will change the output voltage of the voltage divider – this is called loading. The effects of loading can be demonstrated using the equivalent circuit model.

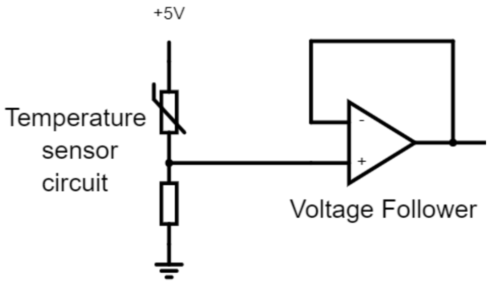
Here R_{Th} *and* V_{Th} represent the thermistor in the voltage divider and R_{in} represents the input resistance of the inverting OpAmp circuit.

Example 94.

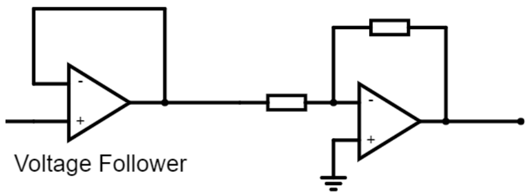
For the above circuit, at 26°C the thermistor has a value of 200$k\Omega$ and the fixed R of the voltage divider is 22$k\Omega$.  	The Thevenin equivalent is $R_{Th} = 19k8\Omega$ and $V_{Th} = 0.495V$. $R_{in} = 22k\Omega$ therefore $$V_{in} = 0.495 \times \frac{22000}{(22000 + 19800)}$$ $$V_{in} = 0.26V$$ Not only has the presence of the amplifier changed our voltage output of the sensor voltage divider but there are other problems as well. The sensor voltage divider will change the gain of the OpAmp, and when the temperature changes this Thevenin equivalent circuit will need to be updated and the calculation will become different.

The voltage follower (or buffer) is a special case of the non-inverting OpAmp circuit, where the output voltage follows (is the same as) the input voltage i.e. $V_{out} = v_{-} = v_{+}$. The critical aspects of this circuit are that $R_{in} = \infty$, so the voltage follower will not load the prior circuit, and $R_{out} = 0\ \Omega$, so the output of the voltage follower will have no impact upon the input of the next circuit. The voltage follower provides a way to connect two circuits <u>without either of them having an effect on the other</u>, solving the (bi-directional) loading problem ☺



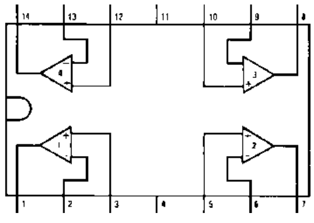


Example 95.



Example 96.

To make an engineer's life easier, OpAmp manufacturers put the output and inverting pins of OpAmps next to each other on the IC, (like this LM324) to make feedback and buffer circuits easy to make.

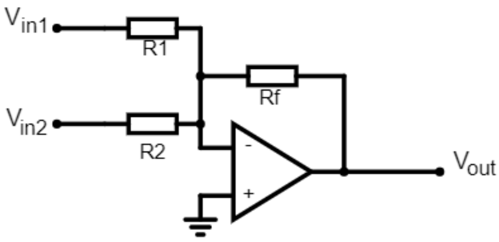


Summer and Subtractor OpAmp Circuits

So far circuits that use OpAmp's to do multiplication with a constant have been investigated. Having knowledge of two more mathematical OpAmp circuits: the summing OpAmp circuit and the subtractor OpAmp circuit, will allow some interesting problems to be solved.

<table>
<tr>
<td>

(Inverting) summing OpAmp circuit



</td>
<td>

Example 97.

The mathematical model for this is

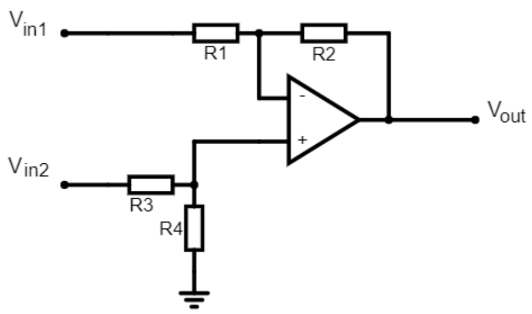
$$\boldsymbol{v_{out}} = -(\boldsymbol{G_1} \times \boldsymbol{v_{in1}} + \boldsymbol{G_2} \times \boldsymbol{v_{in2}})$$

(using super position)

</td>
</tr>
</table>

Example 98.

In an inverting summing amplifier has three input voltages: 120mV, 220mV, 512mV with $R_f = 10k$ and $R_1 = R_2 = R_3 = 4k7$. What is the output voltage?

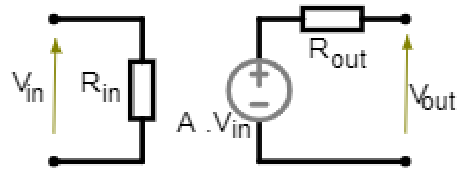
Example 99. **Subtractor (differential) OpAmp circuit**  	$\boldsymbol{v_{out} = G_2 \times v_{in2} - G_1 \times v_{in1}}$ (using super position) In this circuit G_1 is $\frac{R_2}{R_1}$ because it is an inverting OpAmp amplifier. G_2 however is more complex because it involves all 4 resistor values - the voltage divider of R3 & R4, and the non-inverting gain of R2 and R1

Example 100.

Know how to fully derive the mathematical models for the summer OpAmp circuit and the (inverting) subtractor OpAmp circuits above. Search a textbook or the internet for the non-inverting summing OpAmp and derive the model for that.

<u>Some</u> signal conditioning circuits you can now use

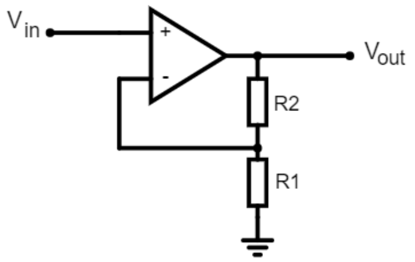
The characteristics of each circuit in terms of the general circuit model. <u>There are too many of these to remember, use your understandings wisely to become familiar with them all.</u>



For an OpAmp $\boldsymbol{A} \approx \infty$, $\boldsymbol{R_{out}} \approx \mathbf{0\ \Omega}$

$\boldsymbol{R_{in}} \approx \infty\,\boldsymbol{\Omega} \therefore \boldsymbol{i_2} = \boldsymbol{i_1} = \mathbf{0A}$

in negative feedback circuits $\boldsymbol{v_+} = \boldsymbol{v_-}$

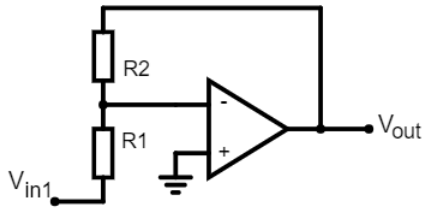


$$V_{out} = 1 + \frac{R2}{R1} \times V_{in}$$

G

R_{in}

R_{out}



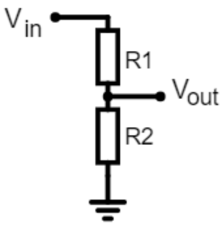
$$v_{out} = -\frac{R_2}{R_1} \times v_{in}$$

G

R_{in}

R_{out}

$(if\ R2 = R1\ then\ G$

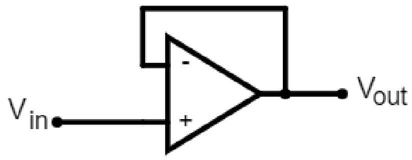


$$V_{out} = \frac{R2}{R1 + R2} \times V_{in}$$

G

R_{in}

R_{out}



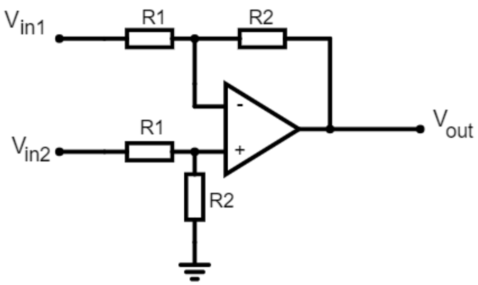
(add f/b resistor)

V_{out}

G

R_{in}

R_{out}



$$v_{out} = \frac{R_2}{R_1} \times (v_{in2} - V_{in1})$$

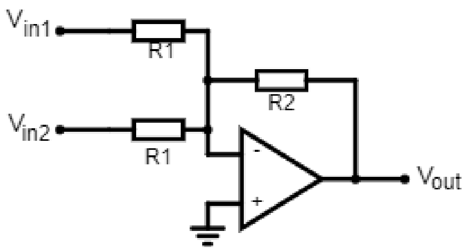
G

Rin1

R_{in2}

R_{out}

$(if\ R2 = R1\ then\ G$



$$v_{out} = -\frac{R_2}{R_1} \times (v_{in1} + V_{in2})$$

G

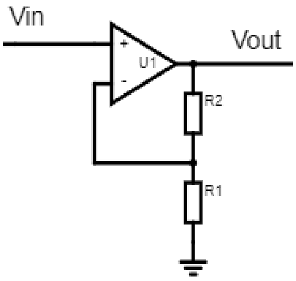
R_{in}

R_{out}

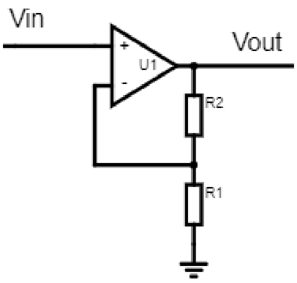
$(if\ R2 = R1\ then\ G$

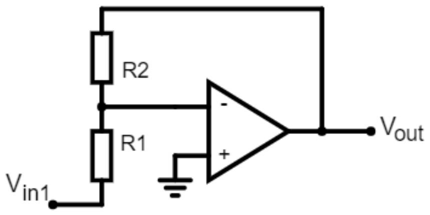
Example 101.

<table>
<tr><td colspan="2">using $\boldsymbol{A} \approx \infty$, $\boldsymbol{R_{out}} \approx \mathbf{0}\ \boldsymbol{\Omega}$, $\boldsymbol{R_{in}} \approx \infty\ \boldsymbol{\Omega}$ $\therefore \boldsymbol{i_2 = i_1 = 0A}$

and for negative feedback circuits $\boldsymbol{v_+ = v_-}$</td></tr>
<tr><td>Example 8:

</td><td>$V_{in} = 22mV, G\ is\ to\ be\ 22.6, R_1\ is\ 1k5,$

Apply characteristics of OpAmps and negative feedback OpAmp circuits, what should R_2 be? Use value(s) from the $E24$ series of resistors</td></tr>
</table>

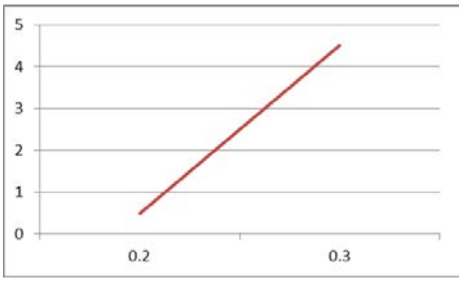
Example 102. 	$V_{in} \ = \ 0.4V, and\ V_{out} \ = \ 2.1V$ is required Choose a value for R1 and work out R2, use values from the e24 series.

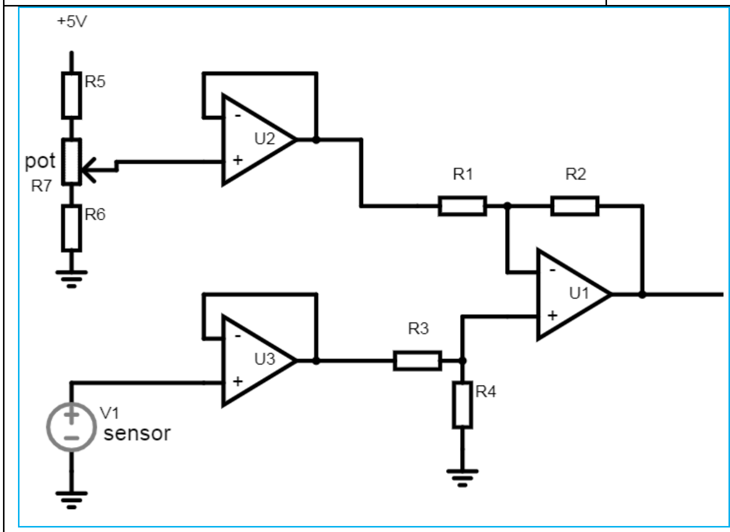
Example 103.	Vout = -0.448V, Vin = 22.4mV
 	Find two E24 values that will provide the gain required.

Example 104.

Design a circuit to follow this mathematical model: $V_{out} = 50V_{in} - 10$, ≈

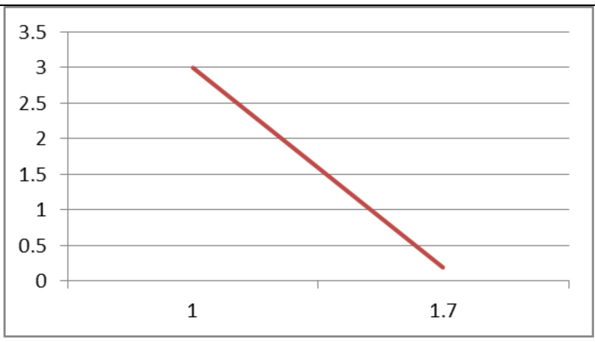
where Vout of a sensor is 0.2V to 0.3V





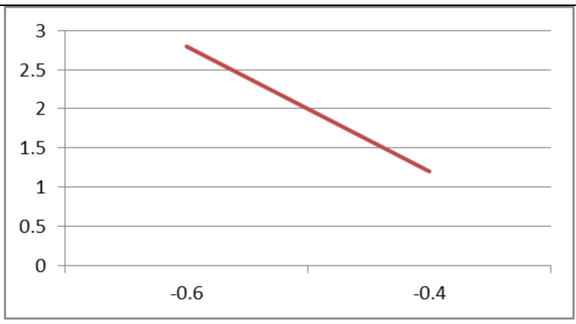
Example 105.

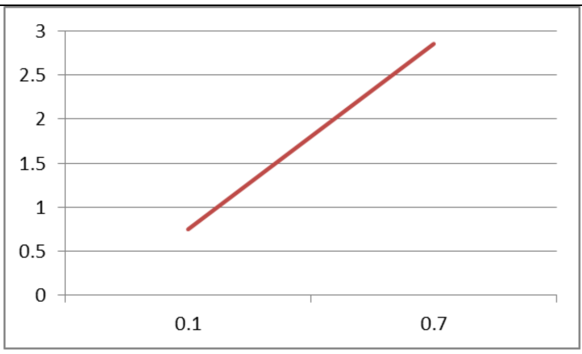
A sensor has an output of 1.7V to 1.0V, amplify this to the range 0.2 to 3.0V



Example 106.

A sensor has an output of -0.6V to -0.4V, amplify this to the range 2.8 to 1.2V



Example 107. A sensor has an output of $0.1V\ to\ 0.7V$, amplify this to the range $0.75\ to\ 2.85V$	

Signal Analysis Exercises

Huntly Power Station provides essential power to Auckland; however it is also a potential source of significant pollutants. It releases CO_2, (a greenhouse gas) and some particulate matter (fine particles that can have negative effects on lung function) into the atmosphere, and takes water for cooling and discharges it back into the Waikato River.

The Waikato is the longest river in Aotearoa New Zealand, its waters originate on Mt Ruapehu and flow $425kM$ to Port Waikato. While many people have interests (are considered stakeholders) in the Waikato River (it is used for rowing, recreational fishing, boating, electricity generation, drinking water for Auckland), to Tangata Whenua, the iwi of Waikato-Tainui, it is a core part of their identity - Ko Waikato te Awa and is endowed with the status of an ancestor.

Wikipedia.com

Waahi Marae, maorimaps.com

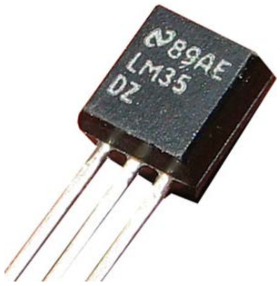
Huntly power station

Next to the power station is Waahi marae. Those whose whakapapa includes the marae, connect with the Tainui waka which arrived in the area around the year 1400. It is in the interest of Tangata Whenua - the people of te Waahi marae and Waikato-Tainui - along with all people in Auckland and Aotearoa-New Zealand that this river is cared for. Much of the specialised work falls upon teams of engineers who design and build systems to monitor and guard the river and its environment.

An important reading for this course can be found here.

Example 108.

The temperature of the cooling water from the power station must be monitored before it can be released back into the Waikato, it is not allowed to exceed 25°C. If it did it could have a detrimental effect on aquatic plant life. An LM35 temperature sensor with an output of $0.1V/°C$ is used to monitor the temperature from $0°C\ to\ 50°C$. This range needs to be amplified into the range $0.1V\ to\ 4.8V$ for the monitoring equipment it will connect to.



(a) draw a diagram of the situation using the generalised circuit model

(b) mathematically model the problem

(c) design the circuit

Example 109.

Huntly power station is powered by Gas (and Coal until Dec 2022), it emits significant levels of CO_2 (it is as high as 5000 kilotonnes p.a. when fired by coal). Whilst CO_2 is not harmful in itself (we breathe it out), it is a greenhouse gas. Emissions of CO_2 must be reduced, as it is accumulating at a higher rate than the forests in the world can break it down into C and O_2. Now that these gases have accumulated in our atmosphere, they are contributing to global warming - a slow increase in atmospheric temperature and increasing our risk of climate change.

Monitoring CO_2 levels can be carried out using an MG-811 sensor. In this case the output range of 0.45V to 0.66V is of interest and must be converted into 0.2 to 3.1v for connection to instrumentation. The sensor has an output resistance of 2k2Ω.

Digikey.com

(a) draw a diagram of the situation using the generalised circuit model
(b) mathematically model the problem
(c) design the circuit

Example 110.

SKF.com

Problems with the operation of the cooling system might cause the power station to shut down as it is not allowed to discharge warm water into the Waikato. A shut down would mean a loss of much of Auckland's power grid which would result in extensive costs to buiness and industry. Vibration sensors can be attached to the motors of water pumps to detect small issues before big problems occur. A vibration sensor has an output range of $0v\ (0mm)\ to\ 14V\ (14mm\ movement)$, an electrical engineer needs to design an amplifier to change it into the range $0\ to\ 5V$. The sensor has an output resistance of 380Ω and the circuit needs to have a low output impedance to connect to data logging equipment.

(a) draw a diagram of the situation using the generalised circuit model
(b) mathematically model the problem
(c) design the circuit

Example 111.

Huntly power station is a neighbour to Waahi Marae, several schools, kura, Waahi lake and Waahi stream. When demand for energy is high in Auckland, coal is used in conjunction with gas to fire the station, this releases a certain amount of particulate matter into the atmosphere. This sensor outputs an analog voltage from $4V\ to\ 0V$, of interest is the range $0.55\ to\ 0.5V$ which needs to be amplified into the range $0\ to\ 3.3V$

Sensiron.com

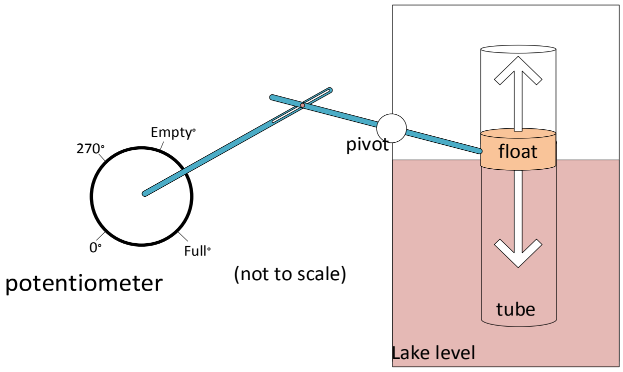
(a) draw a diagram of the situation using the generalised circuit model
(b) mathematically model the problem
(c) design the circuit

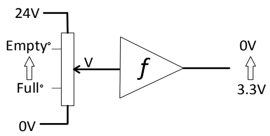
Example 112.

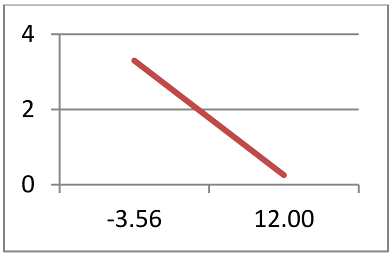
Water from [Lake Waahi]() flows via Waahi stream and then into the Waikato River. The lake's water quality is currently poor, the result of [runoff and leeching]() from farms and coal mines. The lakes outlet and flow into the Waahi stream is controlled as part of a plan by the Waikato Regional Council in conjunction with Waikato-Tainui to begin to improve the quality of the water and perhaps one day see the lake return it to its prior state.

One type of water level sensor works using a mechanical float connected to a potentiometer. The pot has a full range of motion from $0\ to\ 270\ degree$. One end of the pot is connected to $24V$ and the other to $0V$. The pot is at 135° when the lake is full and 40° when low. The signal conditioning circuit needs to translate this range to $3.32V$ when full and $0.28V$ when low.

(a) draw a diagram of the situation using the generalised circuit model
(b) mathematically model the problem
(c) design the circuit.







Example 113.

Any time that sensor circuits are left to be automatically monitored, they need to be self-monitoring in case of failure. Some extra circuits are needed to detect error conditions. Design extra circuits to output 3.3V if there is an error (hint: errors could be caused by a broken wire).

Once a signal has been measured, conditioned and digitised; digital electronic circuits can be used to process and make decisions regarding the data. This section of the course develops understandings of digital circuit (hardware) and microcontroller based (software) decision making processes.

3.2.2. Decision making - digital logic circuits

Learning Outcomes

Understand how the axioms and theorems of Boolean algebra are applied to decision making using digital logic circuits.

Success Criteria

- Draw symbols and construct truth tables for simple logic functions
- Produce a truth table for a simple logic system (<10 logic gates)
- Derive a system of gates form a truth table.
- Manipulate and simplify Boolean models for simple logic circuits

In the 1930's Claude Shannon found that the rules of Boole's algebra were useful in switching circuits, and described circuits called logic gates, using Boolean Logic statements and rules. These rules are also found in programming languages such as C. The rules are the same but written slightly differently in each context.

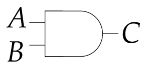
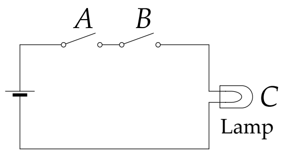
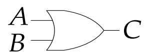
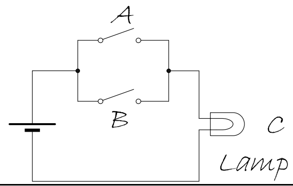
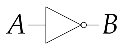
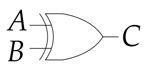
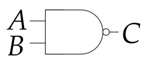
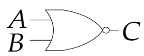
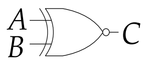
Boolean algebra

Boolean Algebra		Electronic Digital Logic variables and operations
True	T	**1, High, On**
False	F	**0, Low, Off**
AND (Conjunction)	∧	+
OR (Disjunction)	∨	·
NOT (Negation)	¬	**~X** $\overline{X}$
XOR (Exclusive Or)	⊕	⊕

Logic gates for digital decision making

Logic gates are the fundamental units of all digital systems, including microprocessors, microcontrollers, ASICs, and FPGAs, and perform the most primitive of logic functions. A thorough understanding of logic gates is necessary before complex (multidemenisonal) problems can be introduced.

A logic *value* can assume one of two values: on or off, high or low, 1 *or* 0, with the specific representation depending on the application. A logic *variable* is just a name used to identity a binary quantity. For example, a simple switch can be in one of two states: OFF (0) or ON (1) and we may choose to label the state of the switch using the variable S. You will also see X,Y,Z and A,B,C,D... often used as variable names as well, for example: $S = 0\ or\ S = 1, A = 0\ or\ A = 1$

Boolean operation	Algebraic model	Electronic gate circuit symbol	Conceptual switching circuit
AND	$\mathbf{C = A \cdot B}$ $\mathbf{C = AB}$		
OR	$\mathbf{C = A + B}$		
NOT	$\boldsymbol{B = \overline{A}}$		
XOR	$\boldsymbol{C = A \oplus B}$		
NAND	$\boldsymbol{C = \overline{A \cdot B}}$ $\boldsymbol{C = \overline{AB}}$		
NOR	$\boldsymbol{C = \overline{A + B}}$		
XNOR	$\boldsymbol{C = \overline{A \oplus B}}$		

AND:

A	B	C
0	0	
0	1	
1	0	
1	1	

OR:

A	B	C
0	0	
0	1	
1	0	
1	1	

NOT:

A	B
0	
1	

XOR:

A	B	C
0	0	
0	1	
1	0	
1	1	

NAND:

A	B	C
0	0	
0	1	
1	0	
1	1	

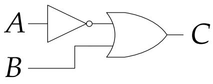
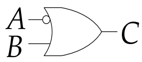
NOR:

A	B	C
0	0	
0	1	
1	0	
1	1	

XNOR:

A	B	C
0	0	
0	1	
1	0	
1	1	

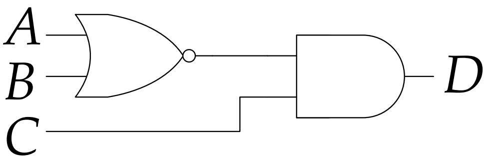
Even small logic circuits can become time consuming to draw; one way of short cutting this process is to compact the not gate as an <u>inversion bubble</u> e.g.

<table>
<tr><th>Expression</th><th>Circuit</th><th>Shortcut circuit</th><th>Truth Table</th></tr>
<tr><td>$C = \overline{A} \cdot B$</td><td>
</td><td>
</td><td>
<table>
<tr><th>A</th><th>$\overline{A}$</th><th>B</th><th>C</th></tr>
<tr><td>0</td><td></td><td>0</td><td></td></tr>
<tr><td>0</td><td></td><td>1</td><td></td></tr>
<tr><td>1</td><td></td><td>0</td><td></td></tr>
<tr><td>1</td><td></td><td>1</td><td></td></tr>
</table>
</td></tr>
</table>

Logic techniques practice

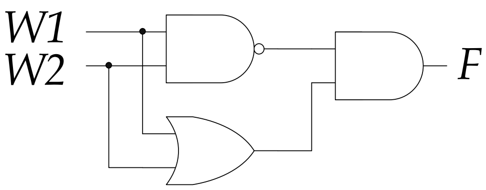
Logic functions are common in electronics and programming (and generally in life), so being able to work fluidly between logic expressions, gates and truth tables are useful techniques to know.

Example 114. Develop expressions and truth tables for the circuits below.



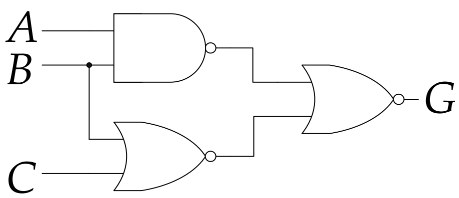
A	B	C	$\overline{A+B}$	D
0	0	0		
0	0	1		
0	1	0		
0	1	1		

Example 115.



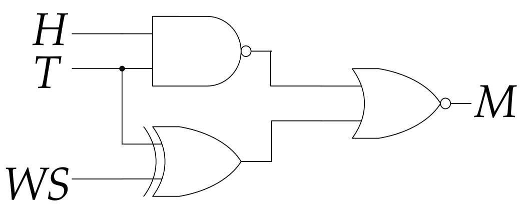
$W1$	$W2$	$W1 \quad W2$	$W1 \quad W2$	F
0	0			
0	1			
1	0			
1	1			

Example 116.



A	B	C			G

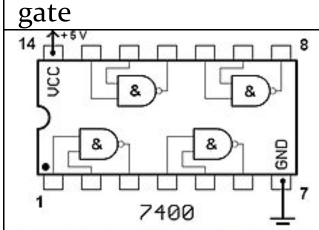
Example 117.

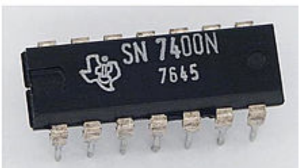
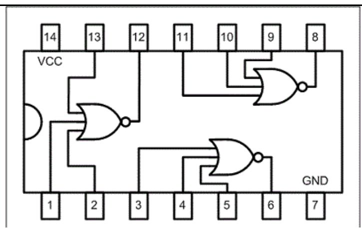
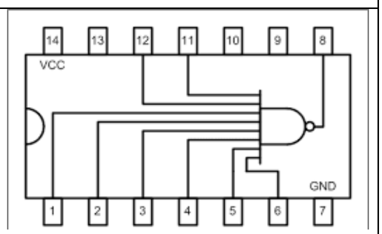


H	T	WS			M

Gates with three or more inputs

Logic gates can be bought and used in circuits, examples of three logic IC's (integrated circuits) or chips that you can buy are below.

<table>
<tr><td>7400</td><td>7427</td><td>7430</td></tr>
<tr><td>Quad 2-input NAND gate</td><td></td><td></td></tr>
<tr><td>


</td><td>
</td><td>
</td></tr>
</table>

Check out a more complete list on [wikipedia]

Boolean Laws contextualised for logic circuits

When logic problems need modelling with more than a simple algebraic function, it is often possible to manipulate the model to reduce it down to something that is easier to build, i.e. uses fewer gates.

Boolean Law	Boolean Expression	Description
Identity	$A + 0 = A$	OR with 0 leaves a variable the same
Identity	$A \cdot 1 = A$	AND with 1 leaves a variable the same
Annihilator	$A + 1 = 1$	OR with 1 results in 1
Annihilator	$A \cdot 0 = 0$	AND with 0 results in 0
Idempotence	$A + A = A$	OR with self leaves a variable the same
Idempotence	$A \cdot A = A$	AND with self leaves a variable the same
Complementation	$A + \overline{A} = 1$	OR with complement results in 1
Complementation	$A \cdot \overline{A} = 0$	AND with complement results in 0
Commutativity of OR	$A + B = B + A$	Order is unimportant
Commutativity of AND	$A \cdot B = B \cdot A$	
Associativity of OR	$A + (B + C) = (A + B) + C$	Bracket removal and regrouping variables
Associativity of AND	$A \cdot (B \cdot C) = (A \cdot B) \cdot C$	
Distributivity OR	$A(B + C) = (AB) + (AC)$	Multiplying or factoring out an expression
Distributivity of AND	$A + (BC) = (A + B)(A + C)$	
Absorption of AND	$A + (A \cdot B) = A$	
Absorption of AND	$A \cdot (A + B) = A$	
Double Negation	$\overline{\overline{A}} = A$	
De Morgan 1	$\overline{A} + \overline{B} = \overline{AB}$	
De Morgan 2	$\overline{A}\,\overline{B} = \overline{A + B}$	

Boolean algebra in action

A lift door will close ($D = 1$) when a timer times out ($T = 1$) or a button is pressed ($B = 1$) as long as the door is clear($C = 1$)

1. Identify the inputs and output from the problem description
2. Draw a truth table for the logic.
3. Write the Boolean model (function) and simplify if possible.
4. Draw the logic circuit.

T	B	C	D	
0	0	0	0	
0	0	1	0	
0	1	0	0	
0	1	1	1	$\overline{T}BS$
1	0	0	0	
1	0	1	1	$T\overline{B}S$
1	1	0	0	
1	1	1	1	TBS

It can be seen that the above problem is true in one of three conditions.

We can fully express this with the following (sum-of-products) model.

Example 118.

$$D = \overline{T}BS + T\overline{B}C + TBC\ (has\ 6\ functions\ with\ 3\ levels)$$

$D = \overline{T}BC + CT(B + \overline{B})$

$D = \overline{T}BC + CT \cdot 1$

$D = \overline{T}BC + CT\ (has\ 4\ functions\ with\ 3\ levels)$

$D = C \cdot \left(\overline{T}B + T\right)(has\ 4\ functions)$

$D = C \cdot ((\overline{T} + T) \cdot (B + T))$

$D = C \cdot \left(\overline{T}B + T\right)$

$D = C \cdot (\left[\overline{T}B\right] + T)$

$D = C \cdot (\left[\overline{T + \overline{B}}\right] + T)$

$D = C \cdot (\overline{\left[T + \overline{B}\right] \cdot \overline{T}})$

$D = C \cdot (\overline{T \cdot \overline{T} + \overline{B} \cdot \overline{T}})$

$D = C \cdot (\overline{\overline{B} \cdot \overline{T}})$

$D = C \cdot (1 \cdot (B + T))$

$D = C \cdot (B + T\ )\ has\ 2\ functions\ with\ 2\ levels$

Example 119.

What are the implications of reducing a problem from 6 to 2 functions for:

an electrical engineer?

a computer systems engineer?

a software engineer?

The binary adder

Example 120.

Design a circuit that adds two binary numbers together.

The first stage is the addition of two 1-bit numbers as shown below. (Here, the "+" symbol refers to addition, not OR.)

$$\begin{array}{r} 0 \\ +0 \\ \hline =0 \end{array} \qquad \begin{array}{r} 0 \\ +1 \\ \hline =1 \end{array} \qquad \begin{array}{r} 1 \\ +0 \\ \hline =1 \end{array} \qquad \begin{array}{r} 1 \\ +1 \\ \hline =10 \end{array}$$

If we have two 1-bit numbers *A* and *B*, then

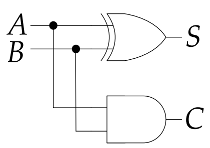
$$\begin{array}{r} A \\ +B \\ \hline =C\,S \end{array}$$

Now two outputs of the circuit are required - *S* the *sum* and *C* the *carry bit*.

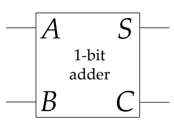
For any two values *A* and *B*, we can write down the truth table for outputs *C* and *S*:

A	*B*	*C*	*S*		
0	0	0	0	$0+0=0_2$	(0_{10})
0	1	0	1	$0+1=1_2$	(1_{10})
1	0	0	1	$1+0=1_2$	(1_{10})
1	1	1	0	$1+1=10_2$	(2_{10})

Looking at the columns for *C* and *S* we see that output $C = A \cdot B$ and output $S = A \oplus B$. Our 1-bit adder circuit therefore looks like this:



It is out practice to abstract (hide) the complexity of circuits by constructing an equivalent model, a 'block', that exposes only the essential characteristics of a highly detailed circuit. In this case the block has input terminals *A* and *B* and output terminals *S* and *C*. For the 1-bit binary adder above we might draw a block like this:



Can we somehow cascade N copies of the 1-bit adder in the previous example to create an N-bit binary adder? No. The reason is that our simple 1-bit adder cannot cope with an input carry bit from the previous stage.

Adding two binary numbers together by hand is conceptually the same as with decimal numbers: add each column and copy the "carry" to the next column. For example, to add the binary numbers $A = 1010_2$ and $B = 1110_2$ we would do this:

$$\begin{array}{cccccc} & 1 & 1 & 1 & 0 & \\ \hdashline & & 1 & 0 & 1 & 0 \\ + & & 1 & 1 & 1 & 0 \\ \hline = & 1 & 1 & 0 & 0 & 0 \end{array} \quad \leftarrow \text{Carry bits}$$

In general, adding two N-bit numbers looks like this:

$$\begin{array}{ccccccc} & & C_{N-2} & C_{N-3} & \cdots & C_1 & C_0 & \\ \hdashline & & A_{N-1} & A_{N-2} & \cdots & A_2 & A_1 & A_0 \\ + & & B_{N-1} & B_{N-2} & \cdots & B_2 & B_1 & B_0 \\ \hline = & C_{N-1} & S_{N-1} & S_{N-2} & \cdots & S_2 & S_1 & S_0 \end{array}$$

Note that sum S_n depends on A_n, B_n and the carry bit C_{n-1} from $A_{n-1} + B_{n-1} + C_{n-2}$. We need a block that accommodates an input carry bit!

An improved truth table that can cope with an input carry bit is shown below.

	C_{n-1}	A_n	B_n	C_n	S_n		
1	0	0	0	0	0	$0+0+0=0_2$	(0_{10})
2	0	0	1	0	1	$0+0+1=1_2$	(1_{10})
3	0	1	0	0	1	$0+1+0=1_2$	(1_{10})
4	0	1	1	1	0	$0+1+1=10_2$	(2_{10})
5	1	0	0	0	1	$1+0+0=1_2$	(1_{10})
6	1	0	1	1	0	$1+0+1=10_2$	(2_{10})
7	1	1	0	1	0	$1+1+0=10_2$	(2_{10})
8	1	1	1	1	1	$1+1+1=11_2$	(3_{10})

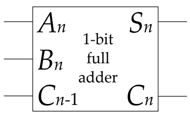
Close examination of the truth table shows that[72]

$$S_n = \overline{C_{n-1}} \cdot (A_n \oplus B_n) + C_{n-1} \cdot \overline{A_n \oplus B_n}$$
$$C_n = \overline{C_{n-1}} \cdot A_n \cdot B_n + C_{n-1} \cdot (A_n + B_n)$$

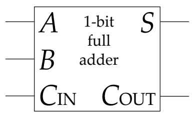
[72] Are these the most compact representations? You should be able to show that these equations can be simplified slightly to $S_n = (A_n \oplus B_n) \oplus C_{n-1}$ and $C_n = C_{n-1}(A_n + B_n) + A_n B_n$.

We can now construct a circuit to compute S_n and C_n using simple gates (AND, OR, XOR, etc).

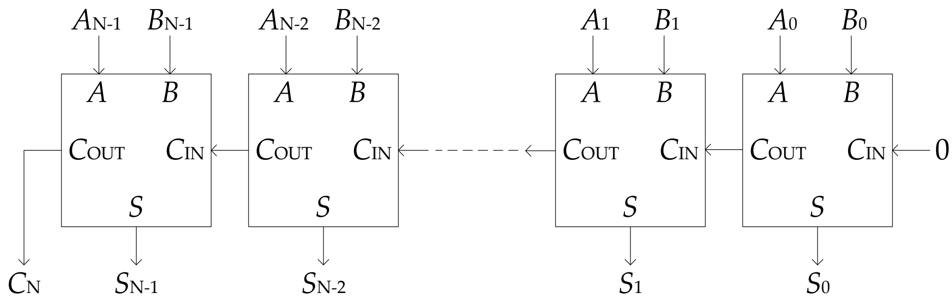
This circuit is called a *full adder* because it correctly handles the input carry bit[73]. We can also draw the circuit in block form



or

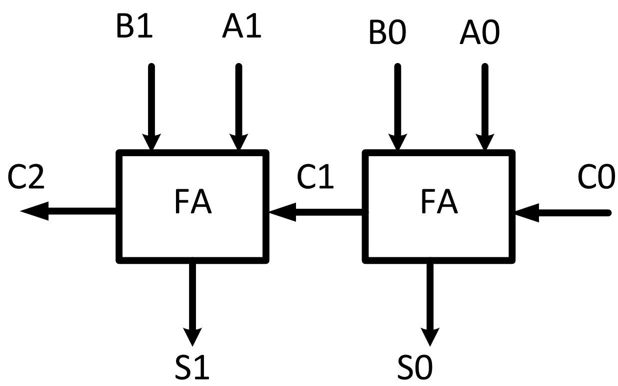


To create an *N*-bit binary adder we simply cascade the full-adder blocks like this:



Example 121.

For the 2-bit binary full adder shown, if the input numbers $A = A_1A_0 = 11$ and $B = B_1B_0 = 10$ are added and the carry bit C0=1. What would the values of the carry bits C1, and C2 be respectively?



(a) 11
(b) 00
(c) 01
(d) 10
(e) None of the above

[73] Similarly, the 1-bit adder in the previous example is called a *half-adder* because it does not handle an input carry bit.

3.2.3. Decision making - embedded systems.

Learning Outcomes

Understand how embedded systems are used to capture information about the real (analog) world, store and manipulate in the digital world in order to control the real world.

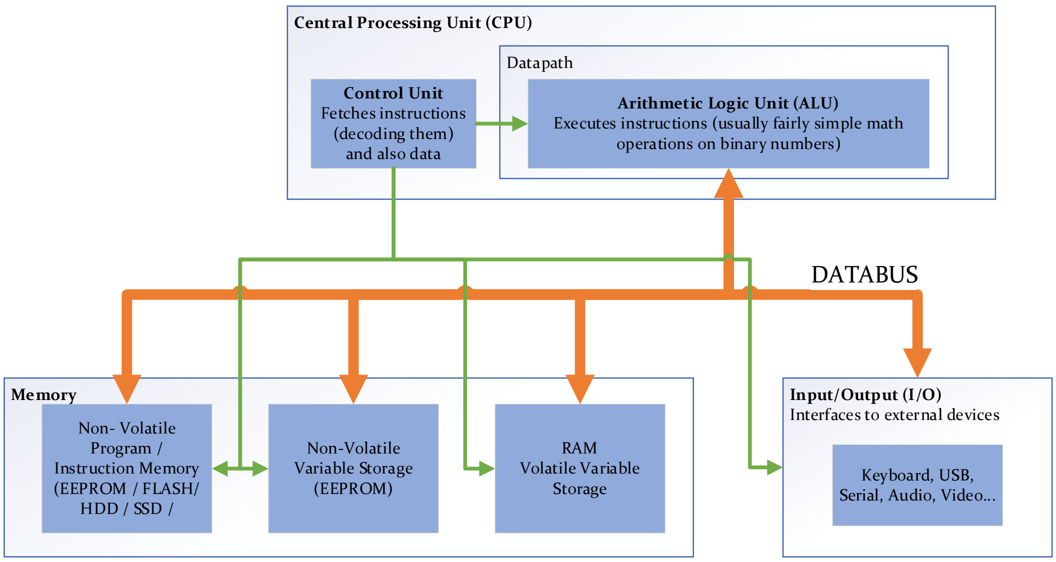
Success Criteria

- Recognise the role of microcontroller technologies in our modern world
- Convert analog information to digital data using a microcontroller,
- Apply the axioms and theorems of Boolean algebra to decision making using software in a microcontroller.

Almost every modern mechanical and electronic product today has a small computer-based electronic device inside it to control it. These are called embedded systems, and they are most often constructed using an IC called a microcontroller. A microcontroller has the same general components as a desktop computer or laptop, but on a small, sometimes miniature, scale, making them ideal for embedding inside all sorts of equipment.

The digital computer

The concept of a computer is the same across all digital computing devices. Computers systematically follow instructions to process and manipulate data. Both instructions and data are in binary number form and stored in memory. Usually different types of memory are used to store program instructions and data. The CPU has a control unit that sequentially fetches instructions from program memory. These instructions control which I/O device to get data from (or send it to) and how these data values are to be manipulated by the arithmetic logic unit (ALU).



In a PC, laptop, phone or tablet, the CPU is within a microprocessor. The word micro and processor hint at the type of instructions a computer does. They are micro (small and simplistic) such as adding or subtracting numbers in binary form. Instructions and data are transferred between the internal parts of a computer along parallel electrical connections called a data bus. These may be 8, 16 or 32 'bits' wide. The process of writing a computer program is to break a large process down into a number of very simple mathematical processes and carry them out at great speed.

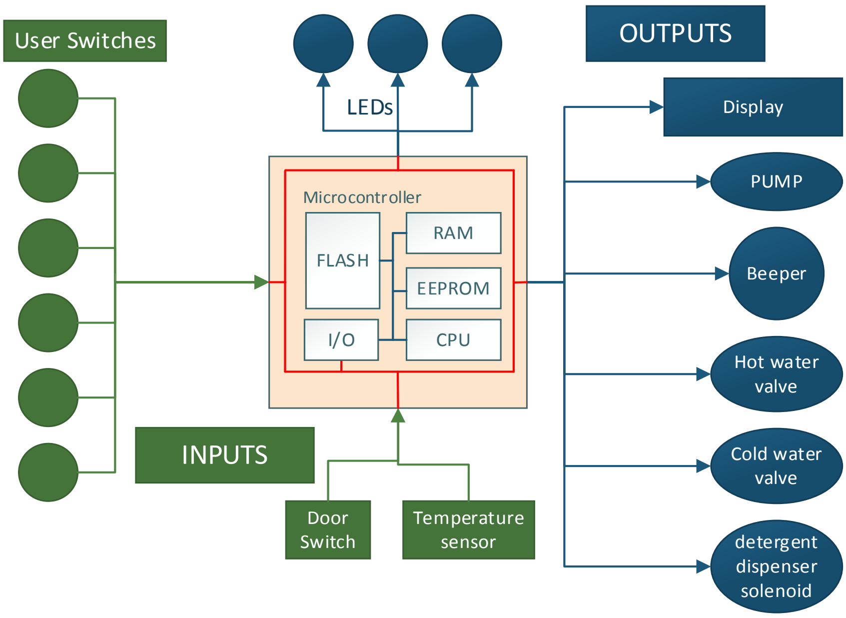
The microcontroller (uC)

Effectively a microcontroller is different to a larger computer only in term of its size. Inside a microcontroller IC is: the CPU (control unit and ALU), and a small amount of: SRAM (static random access memory) for temporary storage of data, FLASH for long term storage of programs, EEPROM for long term storage of data. The microcontroller has I/O (input and output) pins to connect input transducers (sensors) and output transducers (actuators). In a desktop PC or laptop, these same parts exist as separate components.

Embedded Systems

There are billions of microcontrollers embedded inside equipment within our world: cars can have dozens of them, a digital watch has one, and every modern washing machine and dishwasher has one. While the size is a key difference between an embedded system and a general computer system, the main difference is effectively that the embedded system usually has a single fixed function.

Inside a dishwasher, a microcontroller based embedded system will check the state of various input transducers or sensors, such as user buttons, a temperature sensor and a door switch; and control output transducers or actuators, such as a display, LEDs, solenoid valves to allow hot and cold water into the machine, a pump to pump water around the machine, a beeper, and a solenoid to open the detergent dispenser.

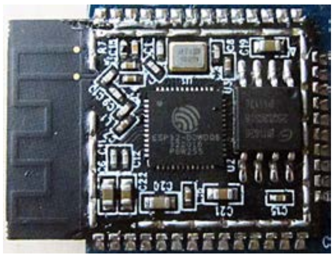
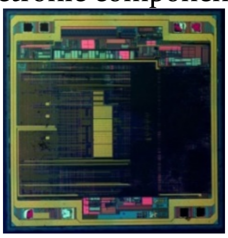


System block diagram for a dish washer

Washing machine PCB (printed circuit board) with microcontroller

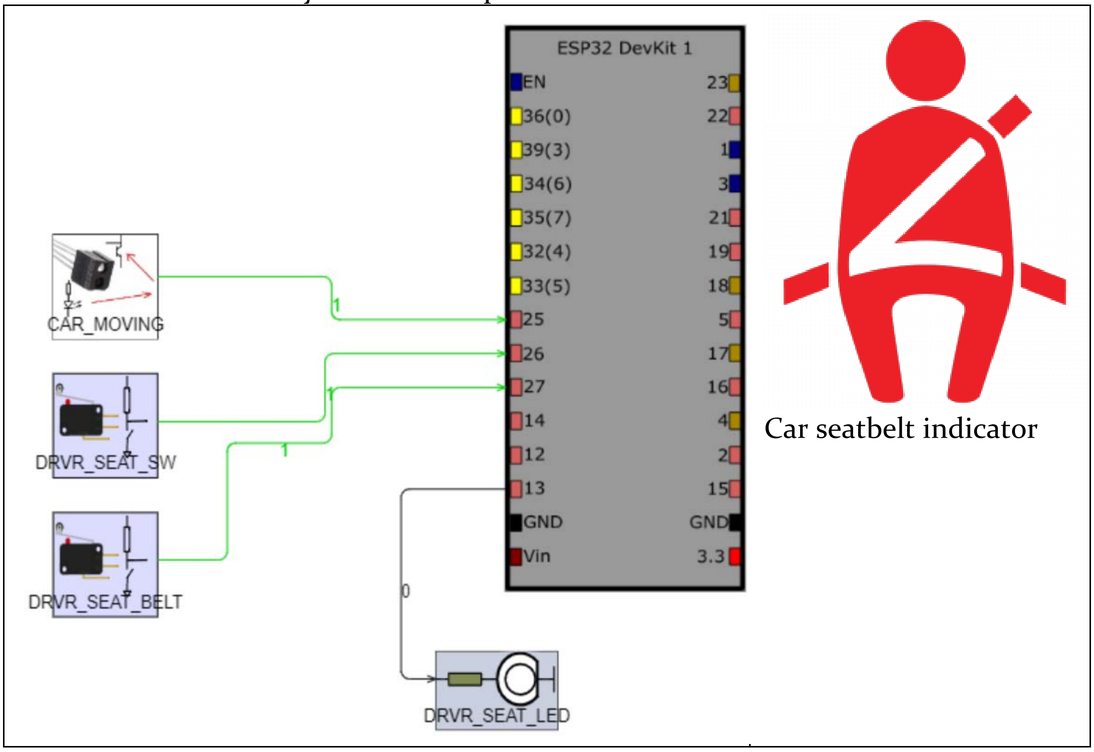
Embedded system development boards/kits

To enter into the world of programing embedded systems, manufacturers build small development boards with interfaces for connecting to a PC. On the PC, a software application called an integrated development environment (IDE) is used to write program code and upload it to the microcontroller. There are a great many introductory microcontroller 'development boards' or kits for learners of embedded systems. A very common one is the Arduino Uno, another one growing in popularity is the $ESP32$.

There are several different types of $ESP32$ development boards. This is the $ESP32$ DOIT $V1$ development board. These boards can be bought for about $NZ\$7$ dollars.  www.aliexpress.com	Under the metal 'can' on the $ESP32$ dev board is the microcontroller IC or chip and some SRAM,  www.wikimedia.org	Magnified up close the silicon inside a microcontroller IC looks like a small city of electronic components.  www.zeptobars.com (it's interesting looking at their hi-res pictures!)

Embedded System Software

The most common computer language for writing software for embedded systems is C. In this system block diagram are the components for a system that indicates to a car driver that they need to do up their seatbelt.



The logic for this system can be represented by a truth table. Tale special note of the use of active low logic, a '0' means something has happened.

Example 122.

DRVR_SEAT_BELT	DRVR_SEAT_SW	CAR_MOVING	DRVR_SEAT_LED
0=done up	0=sitting	0=moving	1=on
0	0	0	
0	0	1	
0	1	0	
0	1	1	
1	0	0	
1	0	1	
1	1	0	
1	1	1	

Developing a microcontroller program

In this question in the assignment, the simulator can be used to write and test your code to see if your solution will correctly control the warning light.
The simulator in GEZCKO was developed to provide only a small subset of the code that can be written for the *ESP*32. Just enough to get you started if you want to go on to learning about IoT - the internet of things.
Example 123.

<table>
<tr>
<td>

```
/***** Includes *****/
#include <Arduino.h>

/***** Hardware macros *****/
#define DRVR_SEAT_LED_PIN 13
#define CAR_MOVING_PIN 25
#define DRVR_SEAT_SW_PIN 26
#define DRVR_SEAT_BELT_PIN 27

/***** Declare & initialise global variables *****/
int car_moving_val = 0;
int drvr_seat_sw_val = 0;
int drvr_seat_belt_val = 0;
int pb_sw_1_val = 0;


/***** Run once code goes here *****/
void setup() {
  /***** IO Hardware Config *****/
  pinMode(DRVR_SEAT_LED_PIN, OUTPUT);
  pinMode(CAR_MOVING_PIN, INPUT_PULLUP);
  pinMode(DRVR_SEAT_SW_PIN, INPUT_PULLUP);
  pinMode(DRVR_SEAT_BELT_PIN, INPUT_PULLUP);
} //setup end

//***** Loop code *****/
void loop() {
  /***** Actions *****/
  // digitalWrite(DRVR_SEAT_LED_PIN, HIGH);
  // digitalWrite(DRVR_SEAT_LED_PIN, LOW);
  car_moving_val = digitalRead(CAR_MOVING_PIN);
  drvr_seat_sw_val = digitalRead(DRVR_SEAT_SW_PIN);
  drvr_seat_belt_val = digitalRead(DRVR_SEAT_BELT_PIN);
  /***** Your code *****/
  if (







} //loop end
```

</td>
<td>Important point to note:
The transparent and systematic naming of every IO device, pin and variable.</td>
</tr>
</table>

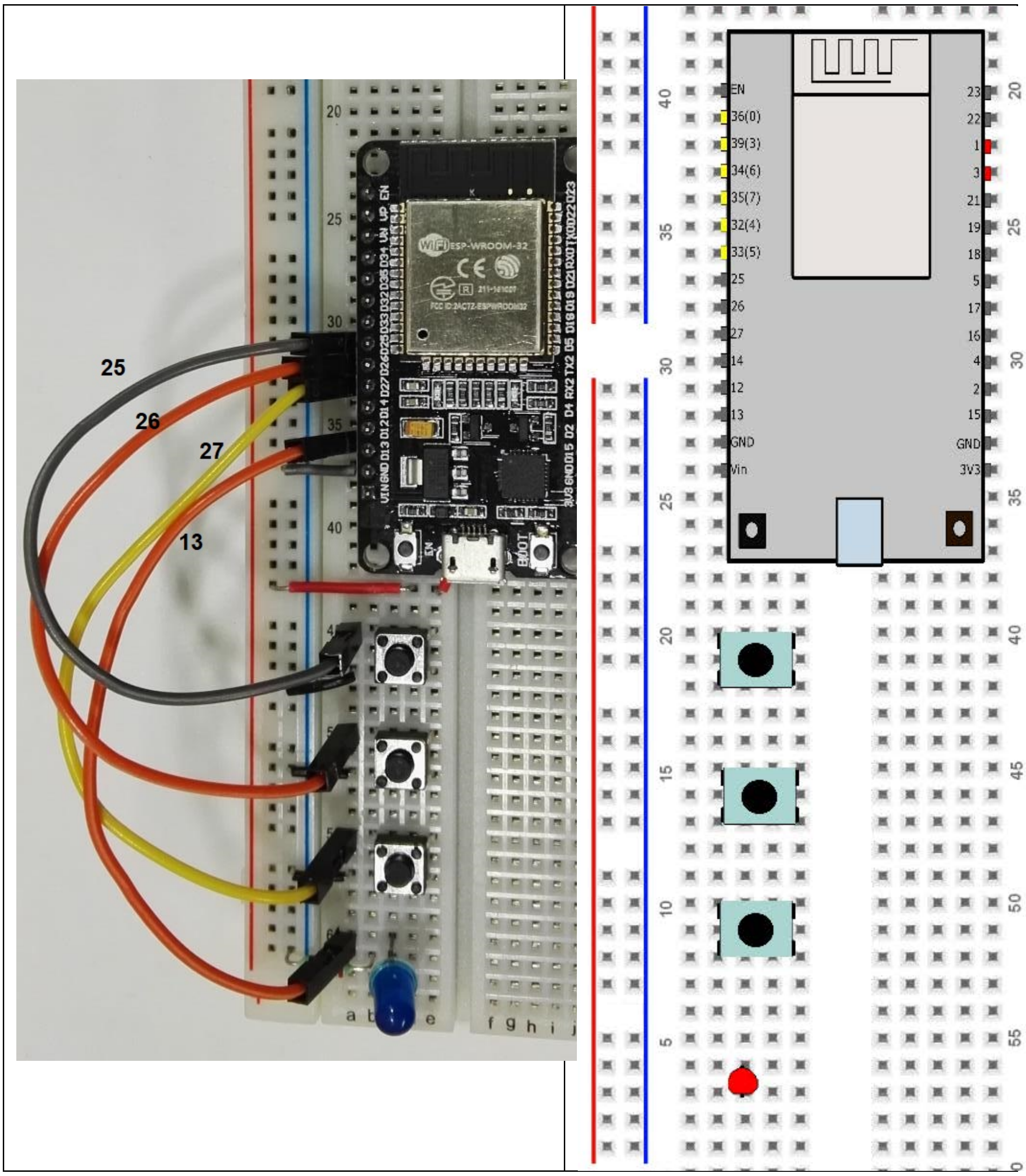
Building it your self

There is no need to build these circuits yourself; but with the huge growth in IoT (the internet of things), very few engineers will go through their careers now without being exposed to these systems in some form. So building some circuits can be useful educationally, and is of course also fun.
A list of suitable parts can be found here.

Construction points

- There are a couple of wires partially hidden under the ESP32, you need to add.
- The switches have to be oriented correctly to work.
- The LED is polarised so must be oriented correctly to work.

Example 124.



Programming the ESP32

To complete a practical exercise the ESP32 requires an IDE. The easiest introductory IDE is the Arduino software. There are many websites and several videos on YouTube that describe how to set this up. Two great sets of resources are the website and YouTube channel RandomNerdTutorials and the YouTube channel of Andreas Speirs.

The automaton

An embedded system is an example of an automata or automaton – it automatically runs immediately from power on, and stays running until power down, without human control required. A program for a PC is best described as a transactional process, one that runs once under human control and generally have single non-repetitive functions.

Variables and Memory

Computers store variables (binary data) in RAM (random access memory). In a PC this is generally large, often being 4/8/16 GigaBytes. In an embedded system a microcontroller might have as little as 512 bytes of RAM. The $ESP32$ development boards typically have $4Mbits$ = __________ bytes of RAM.

Manufacturers of equipment will choose a microcontroller that has enough memory for their specific application, and no more, so as to reduce manufacturing costs. Engineers carefully manage their use of microcontroller RAM by clearly understanding the limits of the information that needs to be stored and choosing variable types that are big enough but do not waste memory.

Variable type	Microcontroller type	Size (bytes)	Range of numbers
char	int8_t	1	$-128\ to\ +127$
unsigned char	uint8_t	1	$0\ to\ 255$
short *	int16_t	2	
	uint 16_t	2	
int *	int32_t	4	
	uint32_t	4	
float	float	4	
double **	double	$8\ **$	

* depending upon the programming language an int might be 2 or 4 bytes – as engineers want to be explicit and clear, it is useful to use the specific type.
** depending upon the microcontroller a double might only be 4 bytes

Choosing the best variable type depends upon the application

Application	Chosen type	Reasoning
Example 125. The number of minutes of sunshine in a day in a solar power installation		
Example 126. The number of minutes the bulb in a data projector has been on for.		
Example 127. The number of		

people at a 25,000 seat sports stadium are counted as they go through turnstiles.		
Example 128. Water turbidity (amount of suspended particles) in a stream are measured using a nephelometer, over the range 0 to 4000 NTU.		

Note that RAM is volatile, when power goes off the data stored in the RAM is gone. EEPROM is non-volatile, data is retained when the power goes off. Which of the above might be better stored in EEPROM rather than RAM?

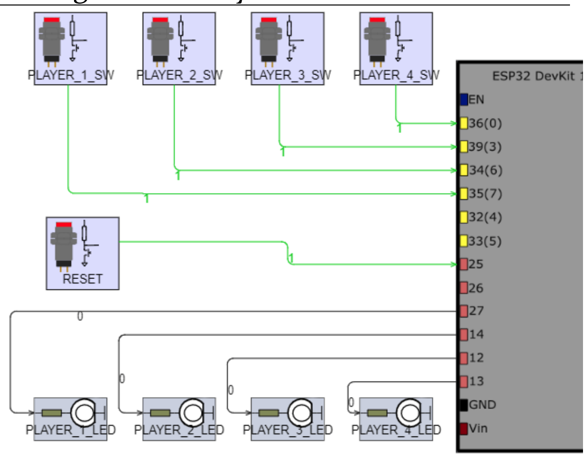
Decision making logic in embedded systems

An 'algorithm' is a process or sequence of operations, and the way a computer program works is described using algorithms. It is best to think about the algorithm before starting programming.

A quiz game controller algorithm.

When the first contestant presses their button, their light comes on, this locks out all other contestants' lights from coming on if they press their button. When the reset button is pressed the system will work again normally.

instructables.com



Example 129. While there are several ways to implement this algorithm. One common way requires some form of memory being used to keep track of which button was pressed. Here is part of a program from the question in the assignment.

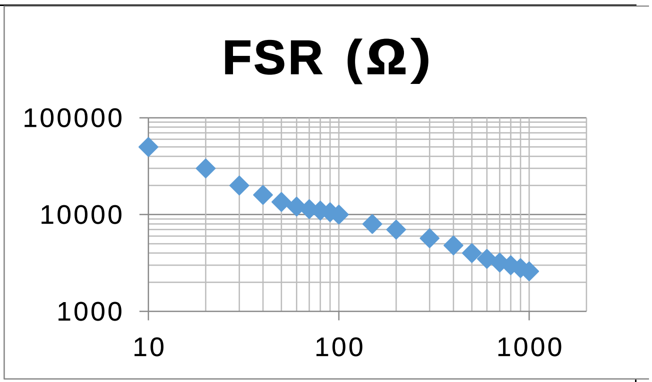
```
//Global variable
…
loop( {
    if (SW_1

    if (SW_2

    if (SW_3

    if (SW_4

    if (RESET

} //loop end
```

Making decisions with analog data

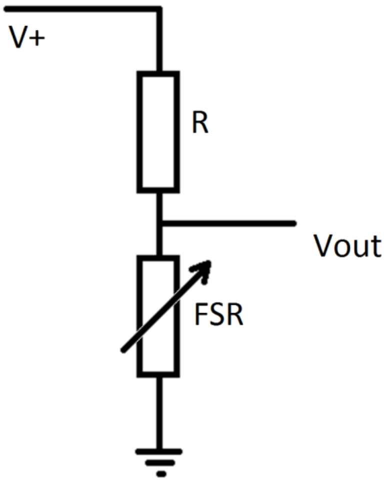
In the real world information is not binary it is analog, and to be used within an embedded system this must be converted to digital or binary numbers. The ESP32 microcontrollers have a 12-bit ADC for this purpose.

Example 130. An FSR is a force sensitive resistor, and is described by the graph below.

13mm diameter FSR
Interlink Electronics



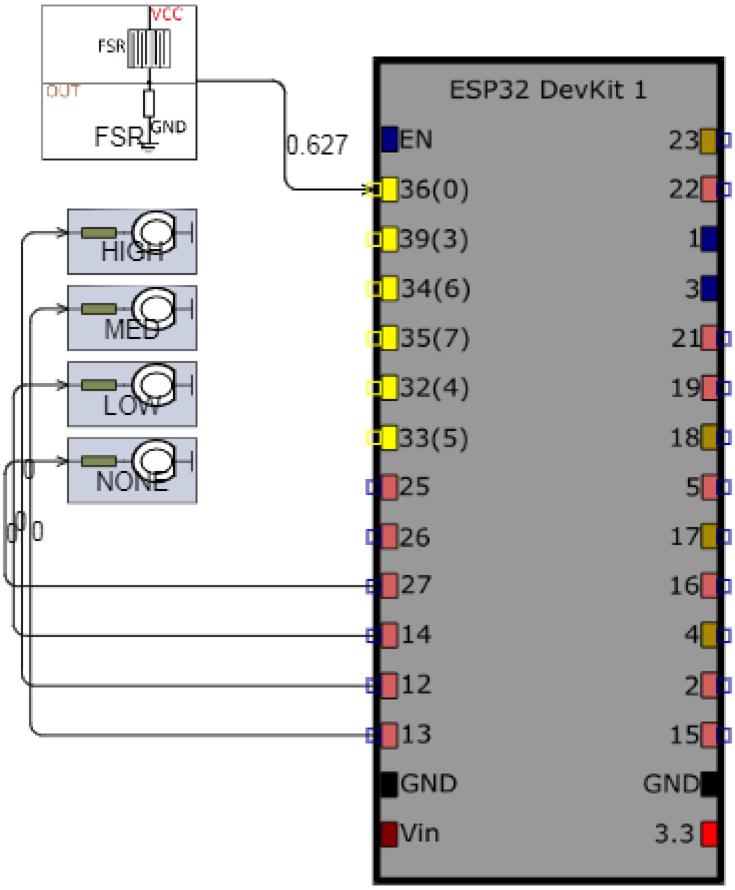
V+ = 3.3V,
Vref = 3.3V,
R= 20K,
12bit ADC



Force(gm)	FSR (Ω)	Vout	ADC
10	50000	2.750	3412
20	30000		
30	20000		
40	16000		
50	13500		
60	12100		
70	11400		
80	11000		
90	10600		
100	10000		
150	8000		
200	7000		
300	5700		
400	4800		
500	4000		
600	3500		
700	3200		
800	3000		
900	2800		502
1000	2600	0.380	471

Example 131.

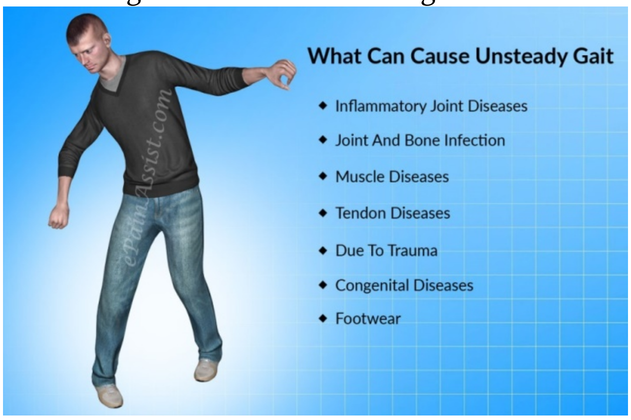
Here is an example of taking ADC readings, using a microcontroller.




```
/***** Declare & initialise global variables *****/
int fsr_val = 0;

/***** Run once code goes here *****/
void setup() {
    /***** IO Hardware Config *****/
    pinMode(MED_PIN, OUTPUT);
    pinMode(HIGH_PIN, OUTPUT);
    pinMode(LOW_PIN, OUTPUT);
    pinMode(NONE_PIN, OUTPUT);
} //setup end

//***** Loop code *****/
void loop() {

    /***** Actions *****/
    fsr_val = analogRead(FSR);
    /***** Your code *****/
```

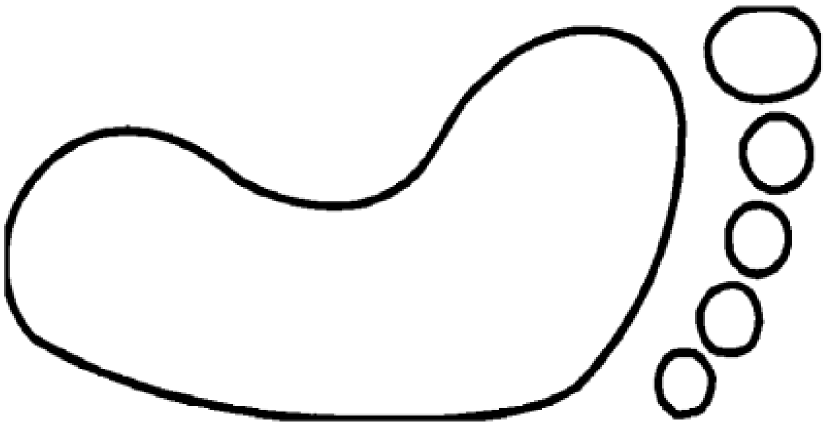
Example 132.

Corrective gait device: a person has a gait (manner of walking) issue, an engineer wants to develop a system to provide immediate feedback to them while they are walking, to help them recognise and correct their gait.



epainassist.com

FSR sensors are to be added inside the sole of a shoe to give feedback to the wearer about their gait, how many sensors would be required in the shoe and where would they be?



3.3 Response Technologies

The third stage of a 'SMART' system is to respond back to the real world, here a few simple visual and kinetic response systems are described.

Learning Outcomes

Understand how signals in the electronics world are converted to real world visual and kinetic phenomena.

Success Criteria

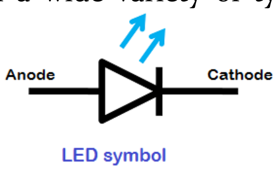
- describe how display technologies convert electronic signals to visible information
- design control circuits for LEDs based upon provided design parameters
- describe how electro-mechanical devices convert electrical to kinetic energy.
- recognise the important design parameters used to design drivers for electro-mechanical devices.

3.3.1. Visual Response Technologies

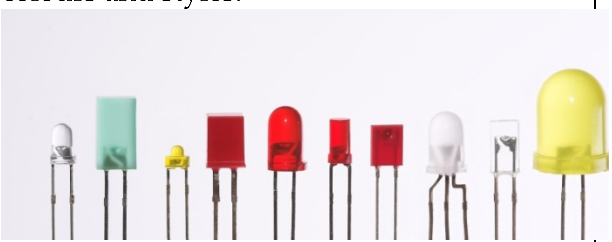
The visible output devices that are connected to SMART systems are LED's and displays.

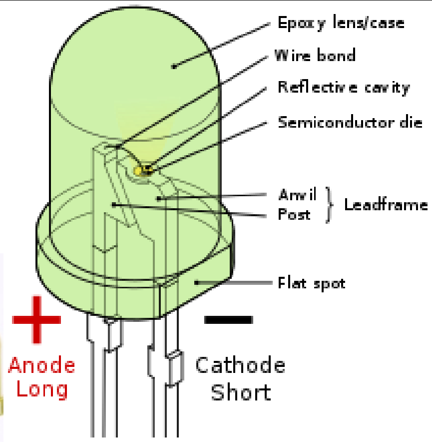
LED's

The LED is a semiconductor device made of silicon. It comes in a wide variety of types,



colours and styles.



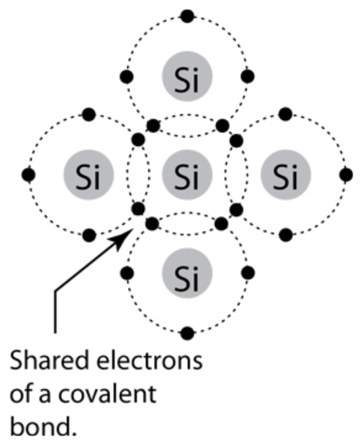
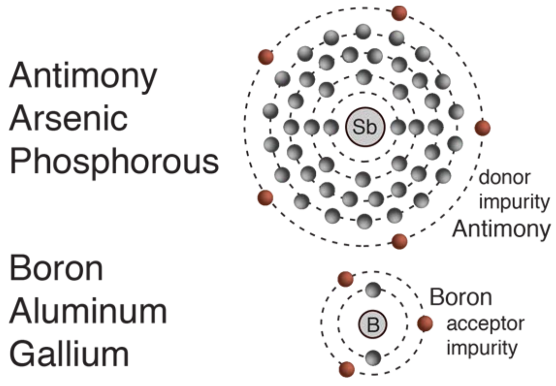


wikipedia

Silicon has 4 valence electrons and consequently forms a strong lattice structure, which because of its covalent bonding makes it a good insulator. Diodes are made by doping - adding to one end of the piece of silicon very small amounts of two other materials. One end could have Antinomy added and the other Boron. The main differences between the two doping materials are:

- one has a valence shell with 3 electrons and the other a valence shell with 5 electrons,
- each has a different number of shells or energy levels

This makes the resultant material a partial or semi-conductor.

Silicon	Added impurities
	

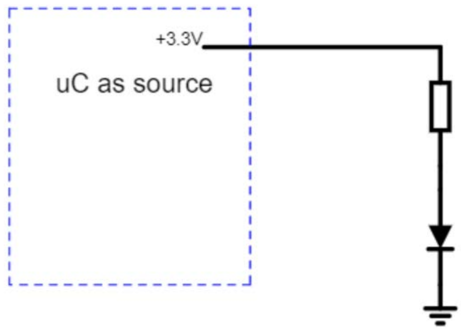
This means the energy levels between the valence shell on one impurity and the valence shell on the other impurity are different. When the silicon diode is connected to an energy source (a battery) and charges move across the diode between the impurities the charges also move between the shells/energy levels of the atom; this creates a significant change in the charge's energy level. This energy change is released as photons - light. Using different impurities, results in different frequencies of light. Interestingly this process works in the opposite direction as well, when photons are projected onto a semiconductor, electrical energy is produced.

LEDs have many advantages and some disadvantages.

https://en.wikipedia.org/wiki/Light-emitting_diode#Advantages

One disadvantage of LED's is that they are sensitive to voltage level, and different coloured and types of LED's require specific levels of electric potential to conduct, and produce the required frequency and amount of light.

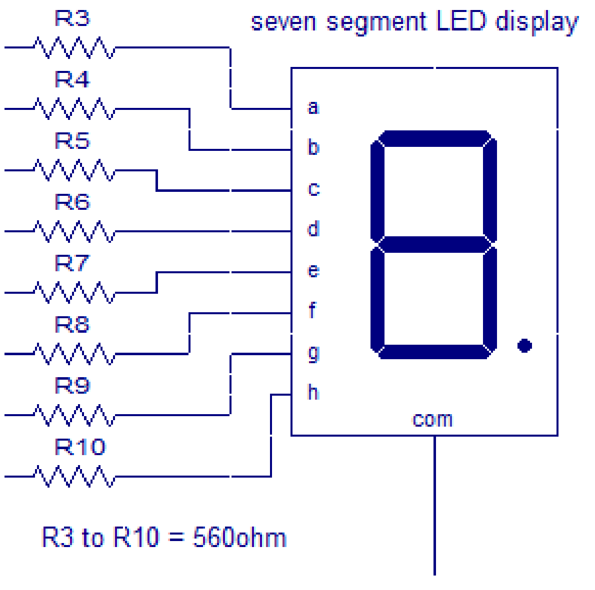
Example 133.

A typical indicator LED will need a supply of approximately $2V$, and will be brightest with a current of $20mA$. What value of 'current limit' resistor (from the $E24$ series) should be used with this LED?	
What power is dissipated in the resistor?	

Example 134.
What $E24$ value resistor should be used with a $2.2V, 35mA$ LED connected to a car battery?

What power is dissipated in the resistor?

LEDs are usually connected individually to single pins of a microcontroller, so 8 LEDs will require 8 pins. The 7-segment LED display has 8 LEDs.



The 7-segment has been a very common display used in equipment for a long time, Today it is being replaced more and more by newer display technologies.

Example 135.
To display the number 4 would require turning on (and off) which segments?

Example 136.
To display the hex number A would require turning on (and off) which segments?

Display technologies

Modern displays have progressed from Cathode Ray Tubes to LCDs to OLED's. While these displays are very different, they have one shared concept of operation; they are made up of individual pixels that must be separately electronically driven.

TFT LCD

OLED

Small displays such as these can be bought cheaply (< $5) for personal projects. The TFT (thin film transistor liquid crystal display) has an LED backlight and tiny segments of liquid that vary between opaque and transparent depending upon whether a voltage is applied to them or not. Each pixel has three segments RGB and each can act as a filter allowing more of less light through. The OLED (organic LED) has a similar pixel arrangement, but no backlight as the pixels are electroluminescent. The TFT and OLED displays above have built in controllers that are typically driven by an embedded system with serial data using a few wired connections. OLED's are more efficient and faster than LCDs.

Example 137.

A 1.8" TFT LCD (1.8 is the diagonal measurement) is 160 pixels horizontal x 128 pixels vertical. Each pixel needs 3 bytes of data (1 byte per colour for RGB). How much data must be sent to the display to do a full screen?

Example 138. :

How much data would need to be sent per second to refresh the screen at $60Hz$?

Example 139.

The 0.96" OLED display above is a single colour (1 bit per pixel). It measures 128 * 128 pixels.
How much data would need to be sent to do a screen refresh?

Example 140.

How much data would need to be sent per second to refresh the screen at $75Hz$?

3.3.2. Actuators

SMART systems need to do, as well as display. The devices that move things are usually electro-mechanical, and include both solenoids and motors. These are constructed of coils of wire and therefore inductive in nature.

Driving Actuators

Actuators require higher power to drive them than a microcontroller pin can typically provide.
The ESP32 can only supply $3.3V$ at a maximum current of $40mA$. This is too small to drive most inductive devices. We need an interface device that a microcontroller can switch that will in turn control larger devices. One common device is the logic MOSFET (metal oxide silicon field effect transistor).

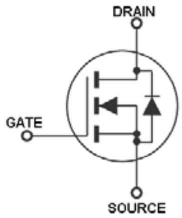
The logic MOSFET

There are several different types of transistor –the BJT and the FET are common. The only type considered in the course is the logic MOSFET, a specific type of FET that can be driven by voltages used in $3.3V$ or $5V$ logic circuits.

There are many different logic FETs made by very many different manufacturers, one common low power Logic FET is the $2N7000$.

Physical appearance	Scematic symbol	Characteristics of interest

$2N7000$



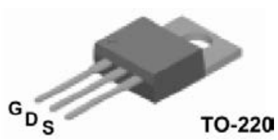
BV_{DSS} / BV_{DGS}	$R_{DS(ON)}$ (max)	$I_{D(ON)}$ (min)
60V	5.0Ω	75mA

I_D (continuous)*	I_D (pulsed)	Power Dissipation @ T_C = 25°C
200mA	500mA	1W

$2N7000$ datasheet

One higher power logic FET is the $FQP30N06L$

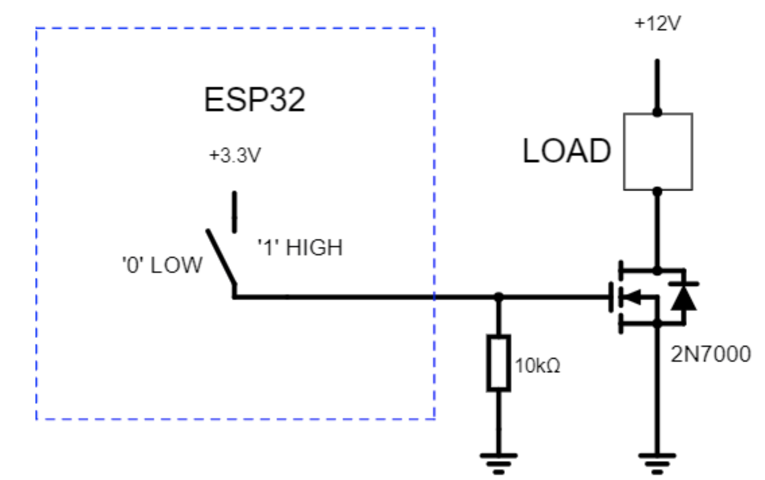
$FQP30N06L$



Symbol	Parameter		FQP30N06L	
V_{DSS}	Drain-Source Voltage		60	V
I_D	Drain Current	- Continuous (T_C = 25°C)	32	A
		- Continuous (T_C = 100°C)	22.6	A
I_{DM}	Drain Current	- Pulsed	128	A

Symbol	Parameter	Min	Typ	Max	
$V_{GS(th)}$	Gate Threshold Voltage	1.0	--	2.5	V
$R_{DS(on)}$	Static Drain-Source On-Resistance	-- --	0.027 0.035	0.035 0.045	Ω

We connect a Logic FET to a microcontroller like this:

Schematic	Equivalent circuit
	

The 10k resistor is not necessary for the circuit to work, it is there because the when the microcontroller turns on it takes a finite amount of time to set the pin as an output. During that time the input to the FET would effectively be unconnected without the resistor and the FET could turn due to stray electric fields from the circuit around it.

Example 141.

Connecting a $150mW$ load to the circuit above.
What is the current in the load?

Example 142.

What is the voltage V_{DS}?

Example 143.

What is the power dissipated by the FET?

Example 144.

Replacing the $2N7000$ with a high power FET, such as the $FQP30N06L$, and changing to a $150W$ load - what is the current in the load?

Example 145.

What is the voltage V_{DS}?

Example 146.

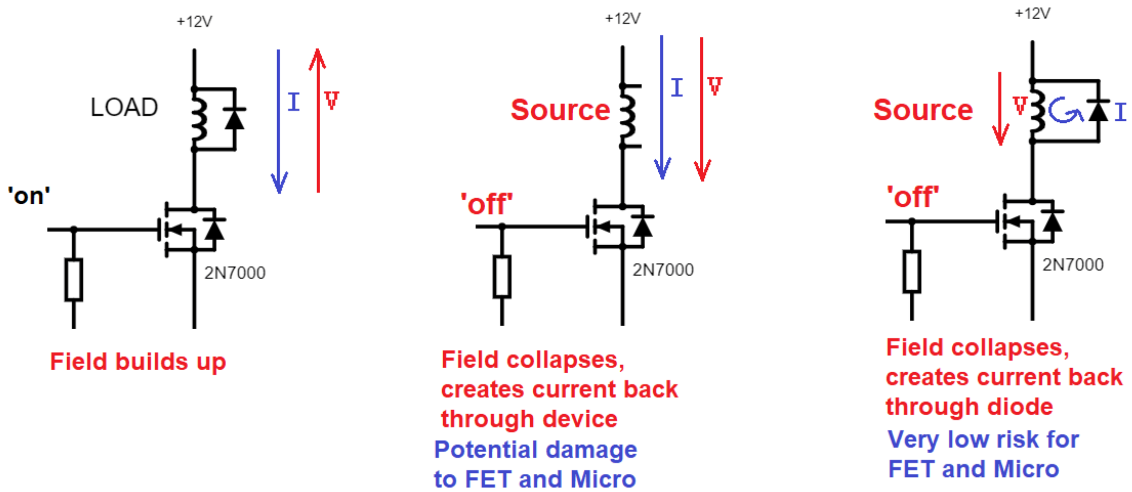
What is the power dissipated by the FET?

Inductive Loads

Faraday's Law states that when the magnetic flux linking a circuit changes, an electromotive force (EMF) is induced in the circuit proportional to the rate of change of the flux linkage (this is the principle behind how generators work).

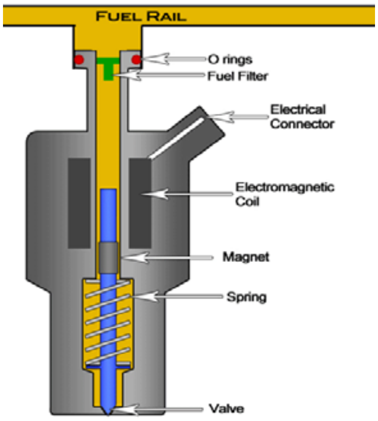
This however can have catastrophic effects in electronic circuits. If we connect a motor or solenoid (an inductor) to a microcontroller using a FET, when the FET is turned on and there is current in the device a magnetic field builds up. When the device is turned off the magnetic field collapses (rapidly) and creates an EMF. The effect is that the current tries to be maintained, this turns the inductor from a load to a source, and the voltage across the inductor reverses polarity – we use the term back-EMF to describe this reversed polarity. The inductive device becomes a small battery for a short period of time – enough time to create enough current at a high voltage that will most likely damage the FET and then the microcontroller- sometimes spectacularly!

The built in diode of the $2N7000$ protects it to a certain level; however we always connect a diode (called a flyback diode in this context) across the load. What is a diode? It is a semiconductor that conducts when the polarity across it is in one direction and does not conduct when the polarity across it is in the other direction.



Solenoids

A solenoid is simply a coil of wire around a metal core - an electromagnet. Electrical to mechanical energy conversion takes place when the magnetic field produced by the solenoid interacts with some mechanical part (lever or plunger) that is free to move.

A portable oxygen delivery system or an environmental gas sampler will use sub-miniature solenoids with an integral pneumatic valve.	 www.leeproducts.co.uk
Solenoids are used in cars to move the starter motor drive into contact with the car engine, to control the flow of fuel into car cylinders, and to control exhaust gas emission recirculation.	 cecas.clemson.edu

Solenoids are rated in terms of their Voltage, Power, travel distance (stroke).

Example 147.

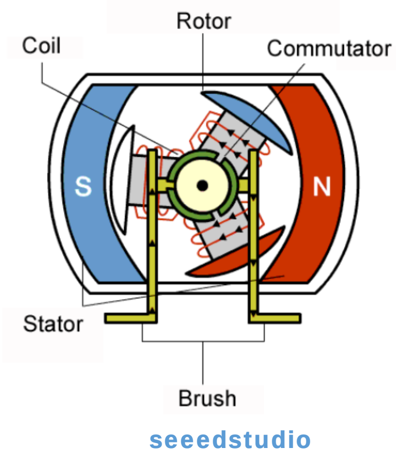
A 12V 40W solenoid is connected to the earlier FET circuit, what power is dissipated by the FET?

Motors

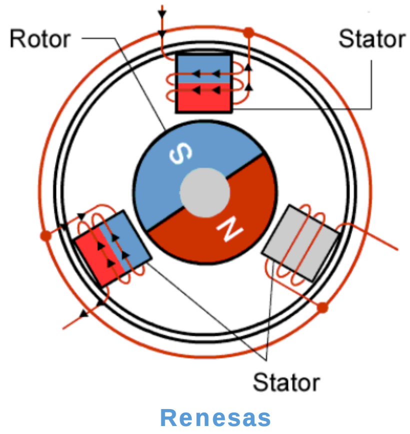
Motors are found everywhere: toys, tools, appliances, electric vehicles, and even in your phone. There are many types, although all operate using the same basic principle of operation - two magnetic fields interacting and creating a resultant attracting/repelling force between them. Think of two magnets, when similar poles are pushed toward each other, their fields want to force the magnets apart, when two opposing fields are pushed together they want to bring the magnets together.

In a motor at least one magnetic field is developed using electrical energy, the other may be created electrically or by using permanent magnets. One part of the motor must be free to move – the rotor (usually the inner part), and the other part is fixed in position - the stator (usually the outer part).

DC Brushed Motor

The DC motor is found across the world in many toys, tools and is used in cars as starter motors. They are not highly efficient (e.g. 75%) as energy is lost across the brushed that transfer the DC to the rotor coils. They have good torque at both low and high speeds, and are reversible (e.g. the battery drill). Unfortunately they are electrically noisy (you can often see the sparks from a battery drill motor), and the brushes do require maintenance.	

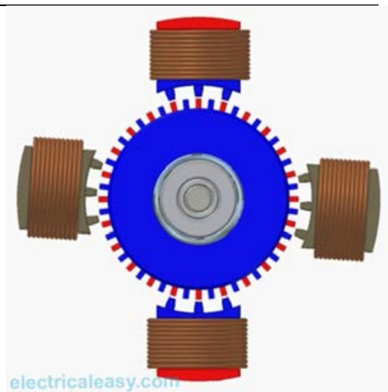
DC Brushless Motor

 	This type of motor is used in computer fans and electric vehicles – while the source of energy is DC, it has to be switched in polarity by smart electric circuits. This makes a quieter, more reliable and more efficient motor. The electronic control however makes them more expensive than DC brushed motors.

Stepper Motor

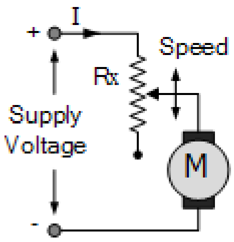
The stepper motor has many fixed magnets on the rotor, and multiple windings and multiple poles on the stator. As the windings are energised, they attract and repel the magnets of the motor. The windings can be easily controlled by a microcontroller to move the rotor a single step at a time. This gives the stepper motor very precise repeatable positional control. They are consequently used in $3D$ printers, photocopiers, laser cutters, and milling machines.

Stepper motors tend to be lower power and small motors (up to approximately $70mm$). While they have low torque at high speeds they have very high torque when stationery. The motor can also be reversed direction easily by changing the sequence of energising the windings. The number of steps in a revolution is controlled by the relationship between the number of poles on the stator and number of magnetic poles or 'teeth' on the rotor.



Control of DC Motors - PWM

The DC motor is the easiest motor to learn a little about speed control. An obvious choice may be to use a rheostat (large variable resistor) in series with the motor. This however is highly wasteful of energy.



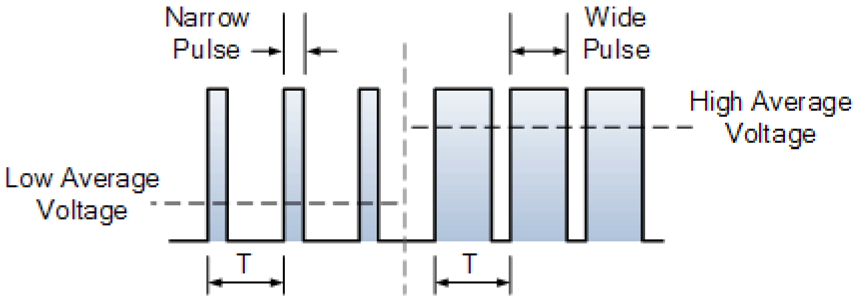
Example 148.

If the Motor is a $12V$ $6W$ type, and the rheostat is 20Ω. What is the energy wasted in the rheostat when it is set to mid position?

Example 149.

What is the power developed by the motor at this setting?

The usual process is to turn the motor on and off rapidly using a process called pulse width modulation (PWM). A narrow pulse means less energy is delivered to the motor, and a wider pulse means more energy is delivered to the motor. Keeping the voltage at its maximum (or higher) means torque is not lost at lower speeds and there is no wasteful power loss.

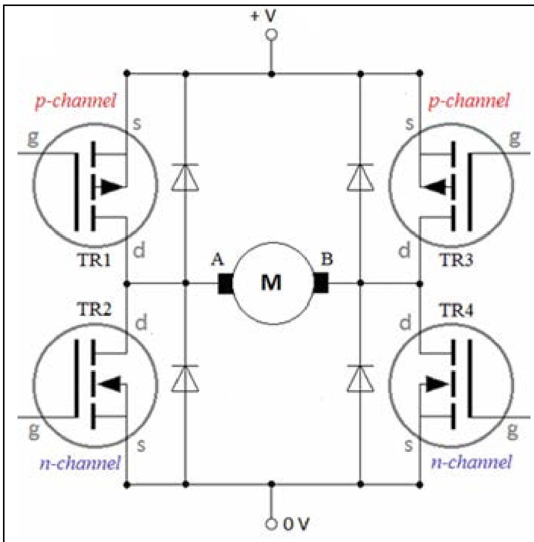


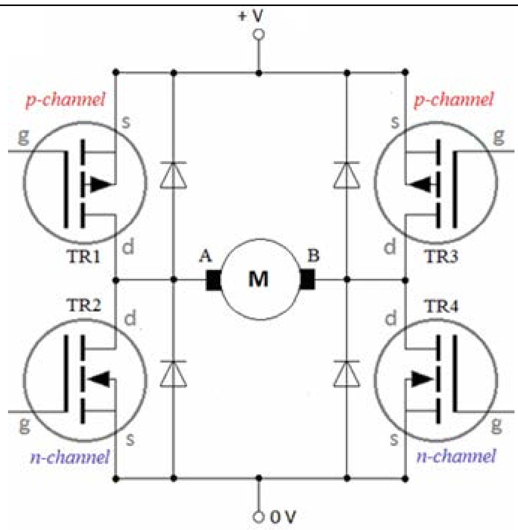
Example 150.

The technique of PWM is also used to control the brightness of LEDs.
Which device would have a higher PWM frequency – a motor or an LED?

H-Bridge

We often require the ability to drive a motor in both directions. With an electric drill this is done with a reversing switch, however a very useful circuit, one used extensively in robotics, is the H-Bridge. There are many commercial H-Bridge ICs available; here is the basic principle of operation.

<table>
<tr>
<td>


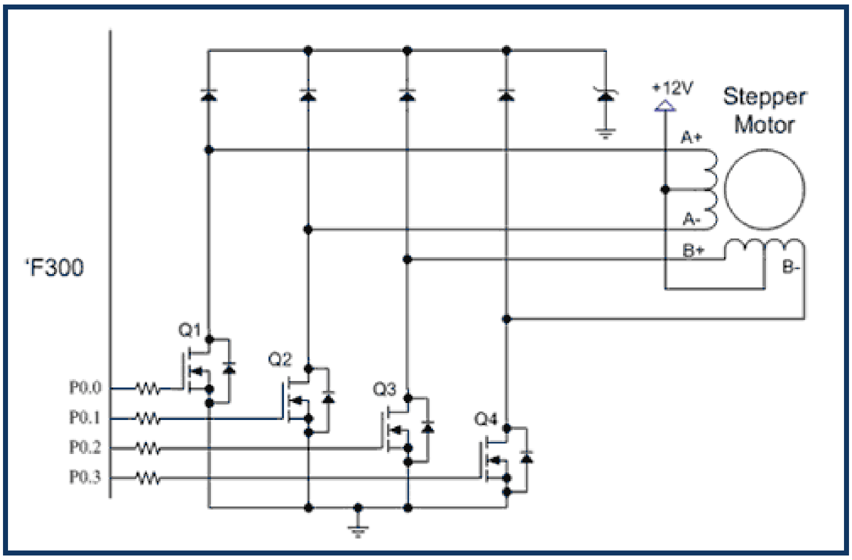
To turn the motor in one direction
</td>
<td>


To reverse direction
</td>
</tr>
</table>

Example 151.

What could happen if the sequence got messed up?

Simplified stepper motor control circuit



Using 4 outputs of a microcontroller we could write software to drive the stepper motor in small steps at varying speed in either direction. Usually we use a special IC that allows fine drive control of the motor and provides other functions as well.

4 Electricity Supply

Learning Outcomes

Develop an overview of how electrical energy is produced and delivered in Aotearoa- NZ.

Success Criteria

- explain why all engineers need some knowledge of the NZ power industry.
- describe the main features of the NZ power industry.
- explain why electrical energy is distributed as AC or DC
- describe the characteristics of 3 phase AC
- mathematically model transmission systems
- explain how the fundamentals of electric circuits influence safety.

All engineers either directly or indirectly work with electrical energy. So there is important information all engineers need so that they can do their work and interact knowledgably with clients and colleagues.

Electrical Network Safety for Emergency Services Personnel (Guide)

In the photo above, civil engineers would have been involved in engineering or determining the suitability of the ground structures for the power poles as well as the pole itself. A mechanical engineer could have designed the traffic light pole. A traffic engineer might review accident statistics and identify new traffic flows in the area and want to relocate power poles and transformers. Bio-medical engineers work with life-saving and life-sustaining equipment that would be used at the scene, and need to understand the risks associated with electricity in situations such as these. Environmental engineers investigate visual and other pollutants that relate to power generation and distribution systems.

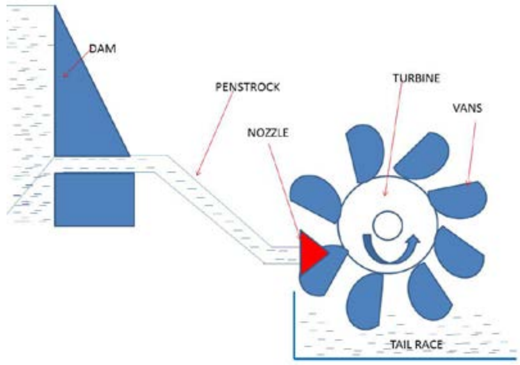
In each case some background understandings about electrical energy; and its generation, transmission and distribution systems is essential. Such as:

- Why are there so many wires? What do they all do?
- Why are they separated by so much distance?
- What are they made of? What do they weigh?
- Are they insulated? What do transformers do?
- How dangerous is electricity and why?

- Why can't we put all the power cables underground?
- Why is the voltage so high, if it is so dangerous?

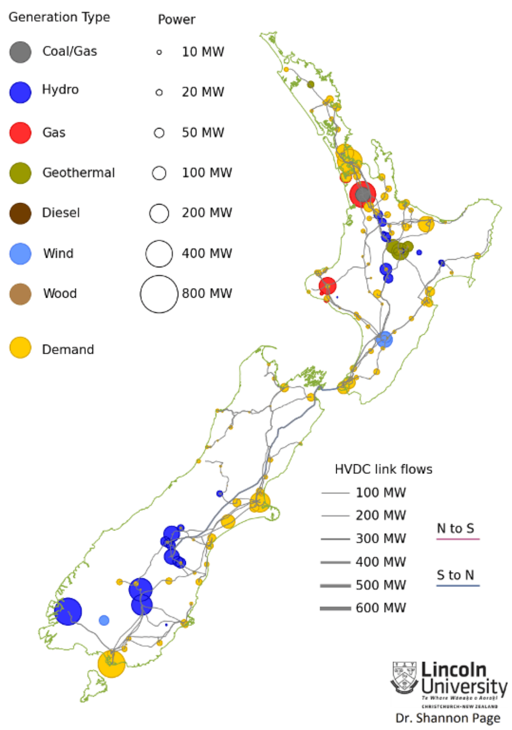
What other questions do you have?

4.1 Generation of Electricity



mech4study.com

The first stage in understanding our electrical systems is to know the process of where electrical energy comes from. The law of the conservation of energy means that electrical energy is not made but converted from some other form of energy. In Aotearoa New Zealand several different forms of energy are converted into electrical energy. The most common conversion is that some sort of fluid flow is used to turn a turbine. The turbine is mechanically connected to an AC generator. Hydro energy is the most common form of this.
Energy in other forms such as gas, geothermal, coal and oil are also used to heat steam to turn turbines. On a smaller scale wind or air flow is used to turn turbines as well, solar is different – no turbine or generator is used.

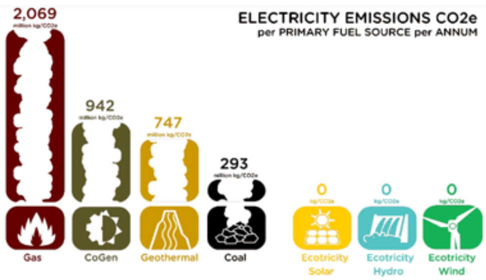


This means that the generation of electricity must take place where the source of energy is, and there are 219 electricity generation stations in NZ owned by a number of companies. Water is collected in natural or man-made lakes for hydro power generation, predominantly along the Waikato River and in the lower South Island. Hydro produces 59% of NZ Electricity. This large percentage means our power supplies are susceptible to dry winters.

In Aotearoa New Zealand, access to river systems and land was taken by the government under the Public Works Act to build dams and power stations. For instance in the period of 1929 to 1970 eight dams and nine hydro stations (owned by Mighty River Power) were built along the Waikato River. Also control gates were built to manage the flow of water from Lake Taupo into the Waikato. These dams and lakes divide the river, permanently changing its ecosystems and the lives of the iwi who live along the river. Dams prevent eels from being able to spawn in the sea, native fish species have disappeared being replaced by other species. Archaeological sites, rock paintings, hot springs, and wahi tapu (sacred places) along the river have been flooded and lost. ("Tupuna Awa" by Marama Muru-Lanning).

Geothermal energy is predominantly harvested at

Wairakei, close to Lake Taupo, and is used to produce 17% of NZ electricity. Gas, from New Plymouth, and coal, predominantly from the Waikato, are used to produce thermal energy which supplies 16% of NZ power.



ecotricity.co.nz/news/carbon-knowledge/

A number of wind turbines can be found at Te Apiti and Tararua wind farms in the Tararua ranges 11km from Palmerston North, as well as smaller farms in other places. Wind energy is used to produce 5% of NZ electricity. Tidal energy is also being explored near Raglan and in the Cook Straight.

Of critical significance some forms of energy used to produce electrical energy release large amounts of CO_2, a greenhouse gas, into the atmosphere, while others do not. Greenhouse gases contribute to global warming which has dramatic effects on our environment.

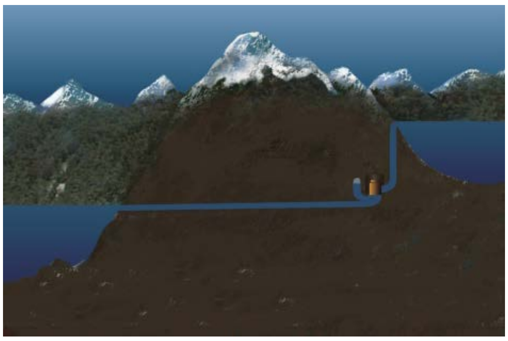

Tararua Wind farm, teara.govt.nz

4.1.1. Power Stations

Power station generators are large, for example each of the seven turbines at Lake Manapouri turns at 250rpm producing 121.5 megawatts of energy at 13,800VAC. This is three phase AC.


Lake Manapouri: witharoha.wordpress.com

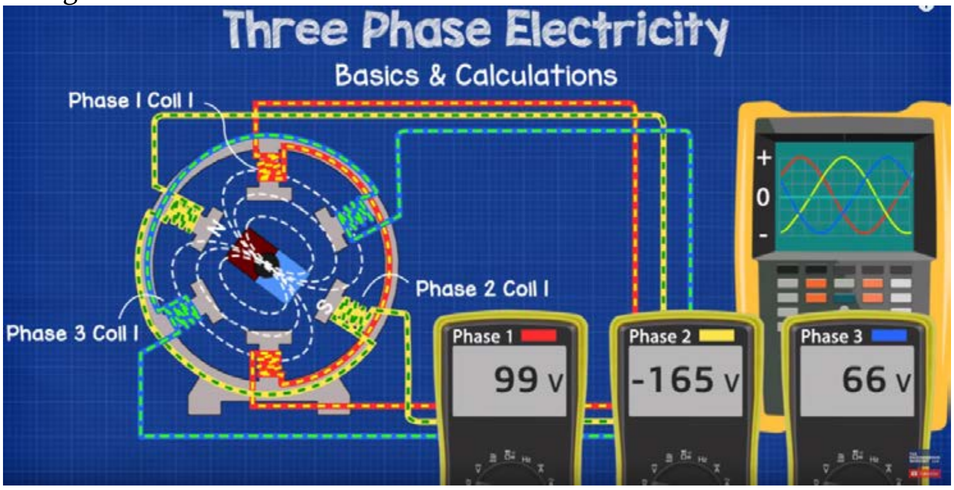
These generators require a considerable amount of energy to turn, and the site is ideal because there is a natural 178m drop from the lake to the sea at Doubtful Sound - part of Fiordland National Park. The original plan was to raise the lake level by 30m merging it with Lake Te Anau, this would have had a devastating effect on the environment, destroying local ecosystems and submersing several small islands. An environmental campaign forced the government to change plans to flood the lake, and the seven generators at Manapouri are now contained 178m below the lake in a man-made cavern. The water then flows 10km into Doubtful Sound and out in the Tasman Sea.


Lake Manapouri: www.dcintinc.com

Generation capability relats to: winter rainfall, lake levels and water travel time. For instance water takes 18 hours to flow from Aratiatia near Lake Taupo to Karapiro, the last power station on the Waikato river; water needs to be at each dam so that at peak usage times, there is sufficient capability to generate electricity.

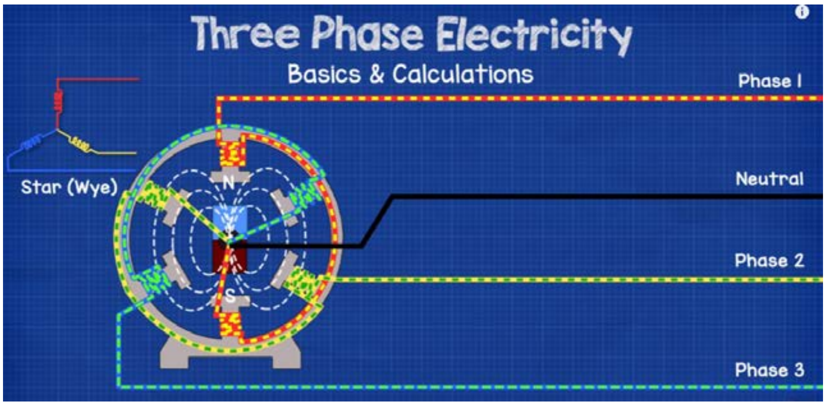
4.1.2. Three phase AC

As a generator rotates the instantaneous voltage in each phase changes, this voltage is modelled by a cosine wave. With a three phase generator there are three windings, each 120 degrees apart, the three AC voltages are not in phase, but offset by 120 degrees.



https://www.youtube.com/watch?v=qthuFLNSrlg

The wires used to carry three phase AC can be connected in two ways, either in a 3- wire delta or 4-wire star combination the $4th$ wire is a common 'neutral' or return wire.



www.youtube.com/watch?v=qthuFLNSrlg

While companies and industry will usually have all three phases connected, homes usually have only a single phase of $230 - 240V$ power supply, consisting of three wires a phase, a neutral and an earth. Along your street different houses will be connected to different phases. So if every third house loses its power there is a clue as to what might be wrong.

It is not unusual for the voltage to fluctuate several volts during the day; the frequency however, will be constant at $50Hz$. Note that in some countries, mains power distribution is at $60Hz$ and $110V$.

4.2 Electricity Transmission Systems

The demand for electrical energy is not usually where the main energy sources are. Demand is high in densely populated areas (such as Auckland), and where there are major industries (such as the aluminium smelter at Tiwai Point in Southland).

Auckland power pylons: stuff.co.nz

The transmission system used to carry electrical energy is known as the 'national grid', and there are around 11,500km of transmission lines in the grid. The company that owns and operates this system is Transpower (a state owned enterprise).

Canterbury power pylons

Transmission lines: mfe.govt.nz

Most of these transmission lines are three phase HVAC (High Voltage Alternating Current), operating at 110kV and 220kV

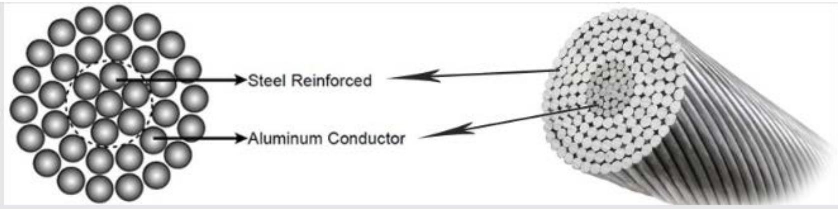
The 13.8kV produced by generators is not suitable for long distance transmission. To understand why, it is important to recognise the interrelationships between the resistivity and strength of the materials used in cable manufacture, and $P = VI$ and $I = V/R$.

Over a long distance the resistivity of the material used in cable manufacture becomes highly significant; ideally it should be as low as possible. Copper is an excellent conductor (as is silver), HVAC cables however, are usually made of aluminium, as it is lighter and less expensive. Because of its higher resistivity a cable made of aluminium might be 50% larger than a copper one, but it will be half the weight. This means a significant reduction in the engineering design requirements for power pylons and insulators.

Transmission line cables are usually of the ACSR (aluminium conductor steel reinforced) type. The steel cores inside the cable support the cable, stopping it from breaking under its own weight.
Thinking about $P = I^2R$ in terms of loses from resistance of cables, these losses can be reduced more significantly if I is reduced than if R is reduced. Another

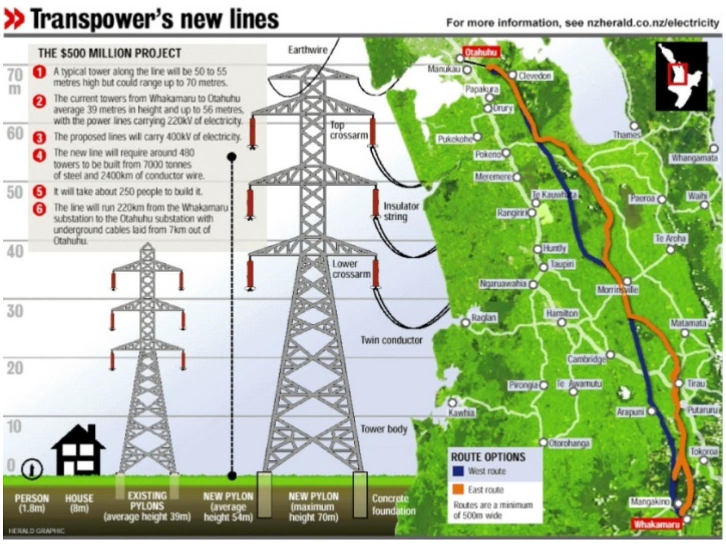
consideration is that losses as heat cause cables to expand and sag significantly; and cables generally run hot at up to $90°C$.

In terms of $P = VI$, to deliver the same amount of power to an end user, if we reduce I then we need to increase V. Consequently most longer distance transmission lines in NZ are $220kV$, while some to smaller centres are $110kV$. Zebra and Moa are common conductors used.



Conductor Code	Stranding and Wire Diameter (No/mm) Aluminium (No/mm)	Steel (No/mm)	Nominal Overall Diameter (mm)	Cross Sectional Area (mm^2)	Approx. Mass (kg/km)	Breaking Load (kN)	Modulus of Elasticity (GPa)	Coefficient of Linear Expansion (x10-6/oC	DC Resistance (Ω/km)
Zebra	54/3.18	7/3.18	28.62	484.5	1621	131.9	69	19.3	0.0674
Moa	76/3.72	7/2.89	38.4	871.9	2577	180.6	63	21.2	0.0366
Chukar	84/3.70	7/3.70	33.3	759.0	2710	203.0	66.5	20.7	0.0373

Because HV cables generate heat, having them uninsulated and putting them in the air helps dissipate this heat. Keeping the cable cool reduces sag and increases the upper current limits of the cable. Uninsulated HV cables however require significant separation of distance to avoid the air ionising and arcs forming between them. The new power pylons built by Transpower in the Waikato and South Auckland had to be 70m high rather than 40 m high because they wanted to be able to run at a higher voltage (400kV). These attracted significant environmental concerns and required considerable consultation with communities.



https://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=10010914

AC Power losses

Example 152.

Three phase electricity is delivered to a dairy factory 25km from a distribution company. The dairy factory receives a voltage of 10.9kV and uses 2100kW. Each transmission cable has a resistance of 3Ω/100kM. Calculate the power dissipated by the power cables.

Example 153. Extension exercise

Starting from the last example, the owners of the dairy factory want to increase its capacity, and require the supplier to deliver 3,900kW of power. What is the increase in loss? What is the voltage at the dairy company?

Example 154. Extension exercise

Starting from the last example, the distribution company want to upgrade the supply voltage to 22kV. Calculate the new loss in the line.

4.2.1. HVDC

Usually transmission of electrical energy is via HVAC, however in NZ there is one HVDC line. This takes power $610km$ from Benmore power station in the South Island to Haywards substation in Lower Hutt; $40km$ of this is a submarine cable under Cook Strait. HVDC has certain advantages over HVAC:

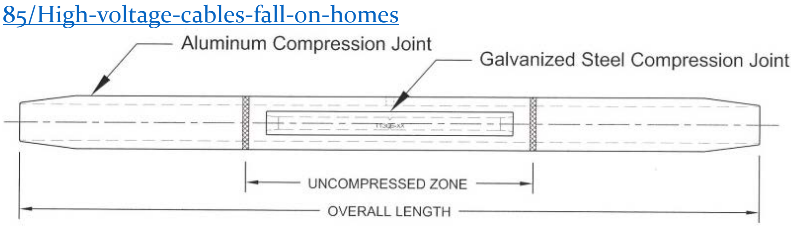
- AC does not use the whole cable, it travels around the edge of the cable, this is called skin effect, and around only 10% of the cable is used; whereas DC energy uses the full diameter of the cable, so is more efficient. When a similar cable is used for DC as AC, about twice as much power can be transmitted using DC.
- The AC peak voltage is $\sqrt{2}$ greater than an equivalent DC voltage in a cable, meaning better insulation or higher separation is required for AC transmission lines.
- AC transmission lines can produce eddy currents in neighbouring lines, which create energy lost as heat.
- AC has reactive energy losses causes by capacitance between different cables and between cables and the earth, DC does not.
- AC transmission is via three phases, requiring three conductors. DC transmission using two conductors (sometimes only one, when earth is used as the return path), which reduces costs.

However HVDC has a very significant disadvantage over HVAC; at high powers, DC cannot be easily transformed between high and low values, and the equipment is very costly. AC is much, much easier and therefore cheaper to transform.
While the costs of the equipment to transform DC means that AC is most often used for long distance transmission; there is a breakeven point for using DC. This is around $600km$ for overhead cables and $50km$ for submarine cables.

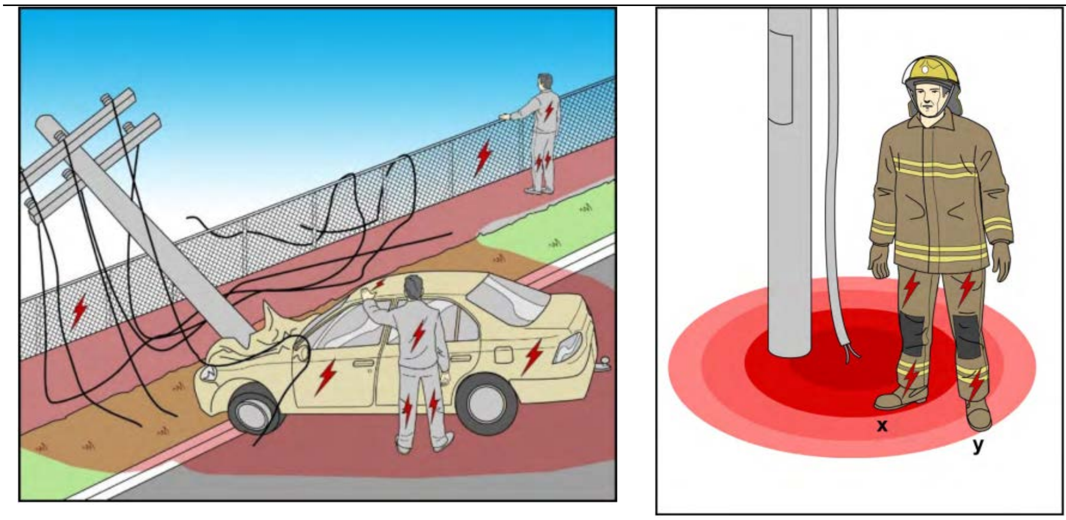
Electrical energy transmission at high voltages is complex and while it has electrical risks associated with it, the risks are more often than not more related to the work that mechanical and civil engineers do, as their roles are critical in the development of the systems and the products used within them.

http://www.stuff.co.nz/auckland/1399785/High-voltage-cables-fall-on-homes

When a $220kV$ cable dropped on 15 homes in Flat Bush in South Auckland, it was found that a join in the cable had come apart. These connections consist of an inner steel compression joint for the cable core and a $> 2m$ outer aluminium compression joint for the electrical conductors.



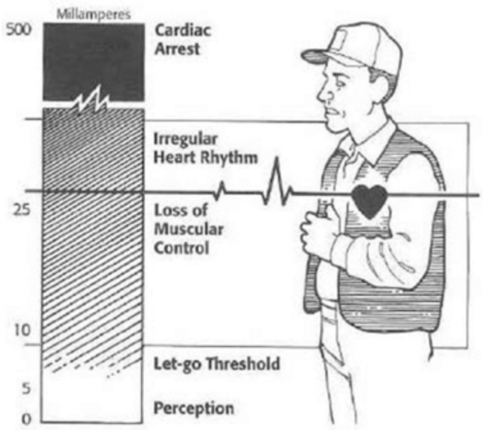
The risks created by fallen power cables are significant. This applies to urban power lines as well which run at $440V$ and $230V$. The danger is not just from the live power cable, but the area around the cable becomes 'live' as well, as there are different levels of electric potential created in the ground around a fallen cable.



Electrical Network Safety for Emergency Services Personnel (Guide)

While the above scenario presents the risks of HVAC systems under a failure condition, the risk to human life is always present with any source of electric energy – even low voltage supplies such as 12V.

An electric current occurs when an electric potential is applied to a closed circuit. When considering the human body as a circuit it is useful to recognise that we have some finite 'resistance', but that the human body is much more than some equivalent resistance. The risks from an electric shock depend upon the frequency of the supplied energy, duration, and even history of other electrical events.



http://slideplayer.com

Humans and animals are most susceptible to frequencies of 50 to 60 Hz because the frequency of the nerve signals controlling the heart is about 60 Hz. Humans can withstand 10 times more current at DC or 1000 Hz than at 50 or 60 Hz. Electro-surgical equipment operating above 100 kHz can pass high currents through the body with no effect on the heart or breathing of a patient.
The resistance of human skin is different for different people and can vary during the day. It can be as high as 100kΩ or as low as 1kΩ, and reduce to 500Ω when higher voltages are applied and it begins to break down. The presence of moisture increases the risk from electric shock.

4.3 Electricity Distribution Systems

The national grid or transmission system delivers electrical energy to 29 electricity distribution companies around the country. These companies distribute it at lower voltages to local communities and consumers, Vector is the largest of these companies.

Electricity transmission and distribution systems are very expensive, and a balance is needed between capability of the system (to reduce interruptions) and the cost of the investment.

In February 1998, electricity supplies to Auckland CBD were disrupted for 5 weeks when four $110kV$ underground cables failed, leaving only one $22kV$ cable to supply the CBD. People were trapped in lifts, traffic light systems stopped working, trains stopped, hospitals cancelled operations, food spoiled, retail shops closed, staff lost pay, more than half of employers moved their operations outside the city and some went bankrupt. Problems persisted for several months as new pylons and cables were erected.

A ministerial inquiry criticised Mercury Energy, and its predecessor the Auckland Electric Power Board, for having sub-standard industry practices. One specific issue described was that Mercury placed heavy reliance on its oil filled cables, but this was "unaccompanied by any investigation of their actual conditions".

In June 2006 during high winds, two rusted D-Shackles holding an earth wire at the Otahuhu sub-station broke, the earth wire fell across $220kV$ lines. Power was disrupted for 6 hours affecting 700,000 people in central, south Auckland and the CBD.

Power companies are mandated to provide high quality standards of delivery to consumers. In March 2019 Vector was fined $\$3.575M$ for failing to do this.

Power pylons and cables being installed in 1998.
https://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=12033654

Earth wire fallen across $220kV$ cables at Auckland substation
https://www.nzherald.co.nz

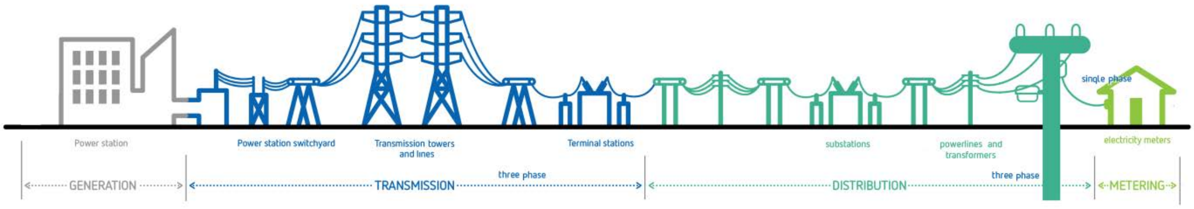
Broken D-shackle that caused the earth cable above to fail.
https://www.nzherald.co.nz

Much distribution is via overhead cables, however in newer urban development's there is a trend toward underground power distribution. Underground cables however require not just insulating, but because they are covered by a 'blanket' of soil they retain much of their heat and need to be larger and are cooled by either gas or oil. There also needs to be significant separation $(3 - 6m)$ between power cables and other underground services (gas, sewerage and water pipes).

The process of moving existing overhead lines to underground is very costly. Directional drilling machines are used to put cables underneath existing structures that cannot be disturbed, such as buildings, roads and lakes. These machines drill or push an initial pilot hole and then ream out the hole to a larger size, before pulling a conduit (pipe) through that will hold the cables.

www.youtube.com/watch?v=mdLCD6t6C-w

4.4 Transformers



Transformers are used to change AC voltage from one level to another and are used across the generation, and distribution sectors as well as within consumer organisations. Transformers at a power station such as Manapouri turn $13.8kV$ into $220kV$. At points where users need electrical energy a series of transformers in substations and along streets are used to distribute the energy at reduced voltages, such as $33kV$ and $11kV$.

The transformer works on the principal of Faraday's Law, where the rate of change of magnetic flux through a loop of conductive wire creates an electromotive force (voltage). This principle works both ways:

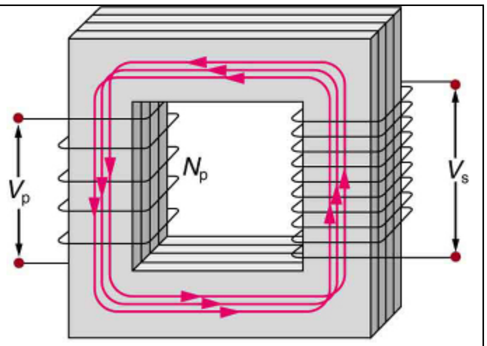
- A changing magnetic flux can produce a changing voltage.
- A changing voltage produces a changing magnetic flux.

The transformer has two windings, a primary and a secondary, both consist of many turns of wire wound around a ferro-magnetic (iron) core.
A picture such as the one below is often used to describe transformers and there are many videos online, such as https://www.youtube.com/watch?v=UchitHGF4n8

The points to observe about this diagram are:

<table>
<tr>
<td>

The magnetic flux produced by the primary is almost all fully captured by the iron core.
The core is made of multiple laminations – if it were a solid block then the magnetic flux would induce 'eddy' currents in the core, which create heating effects and energy loss.
The voltage of the primary and secondary are generally different and a simple mathematical model is used to express the relationship between number of turns and voltage of the primary and secondary windings.
For the same power transfer a higher voltage means a lower current, and therefore smaller conductors.

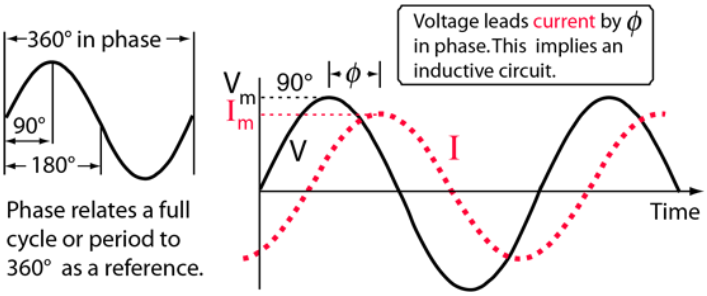
</td>
<td>


$$\frac{V_p}{V_s} = \frac{N_p}{N_s}$$

Note the relationship with current

$$\frac{I_p}{I_s} = \frac{V_s}{V_p}$$
</td>
</tr>
</table>

AC power is affected by the inductance and capacitance of the loads connected to the supply. This produces a difference between the phase of the voltage and the phase of the current in the system. One large effect is from inductive loads such as large motors. The relationship between the voltage and current is an angular measure and is called Power Factor (PF), where $PF < 1$ for inductive loads. Remember *CIVIL* or *ELI the ICE man*, V leads I in inductive loads, as the voltage changes in an inductor the current changes at a later point in time.



hyperphysics.phy-astr.gsu.edu

Transformer Specifications

All electrical equipment and appliances are designed to provide guaranteed performance when operated under specific or rated conditions. Transformers are also designed and manufactured under rated conditions, given as 'Nameplate data', and these include the rated voltage of primary and secondary, VA rating and frequency, etc.

Example 155.

Consider a single-phase transformer with the nameplate data 23 kVA, 2.3 kV/ 230 V and 50 Hz. This data implies that the transformer operates at 50 Hz and has a rated 'rms' primary voltage (V_p) of 2300 V and rated 'rms' secondary voltage (V_s) of 230 V, and both windings are capable of handling (supplying) rated apparent power (S_r) of 23,000 Volt-Amperes.

Therefore the rated currents of both primary and secondary windings and turns ratio of the transformer can be determined from;

$$\text{Rated current of primary winding}: \quad I_{r,1} = \frac{S_r}{V_{r,1}} = \qquad = \qquad A$$

$$\text{Rated current of secondary winding}: \quad I_{r,2} = \frac{S_r}{V_{r,2}} = \qquad = \qquad A$$

$$\text{turns ratio}: \quad N = \frac{V_{r,1}}{V_{r,1}} = \qquad =$$

Exercises

Example 156.

A 1.15 kVA, 230V/110V step-down transformer is connected to a 1 kW resistive load. Determine the primary and secondary currents, turns ratio and the load resistance reflected on to the primary.

Example 157.

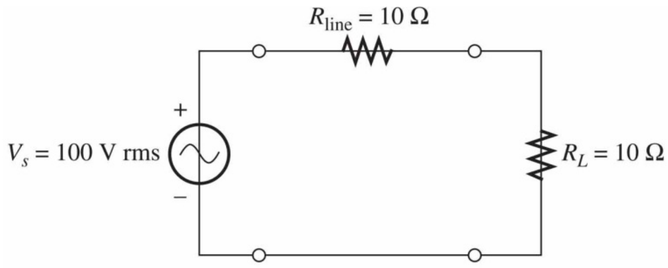
A single-phase transformer is rated 50 kVA, 33 kV / 230 V. At a certain time it is actually operating at 45 kVA with a primary voltage of 30 kV. Find

(i) the primary current

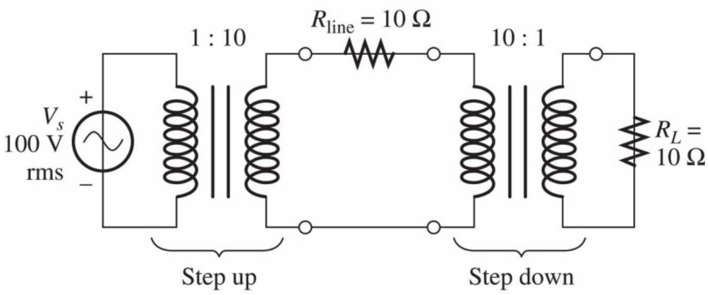
(ii) the secondary voltage.

Example 158.

A voltage source Vs is to be connected to a resistive load RL = 10 Ω by a transmission line having a resistance Rline = 10 Ω as shown below. In part (a), no transformers are used. In part (b), a step-up transformer is used at the sending end of the line, and a step-down transformer is used at the load end. For each case, calculate the power delivered by the source, the power dissipated in the line resistance, the power delivered to the load, and the efficiency, defined as the power delivered to the load as a percentage of the source power. [Ans : (a) 50 %, (b) 99.01 %]



(a)



(b)

4.5 Domestic (Consumer) Electricity Supply

The domestic or mains supply in Aotearoa-NZ is $230 - 240VAC$. The power comes into the meter (usually outside the house), then into a switchboard or distribution board (usually inside the house). The distribution board has isolators (switches), circuit breakers and sine 2003 there must be residual current devices (RCD's). For safety reasons the neutral is connected to ground (via large rods pushed into the ground) at several points, it is known as the multiple earthed neutral system (MENS).

<table>
<tr>
<td>In the switchboard there will be at least 1 RCD and multiple circuit breakers for such things as:
• Hot water cylinder
• Stove/oven
• Heat pump
• groups of lights (e.g. one for upstairs and one for downstairs)
• groups of power points</td>
<td>

RCD's and circuit breaker</td>
</tr>
</table>