ELECTRONICS

Ali M. Niknejad

PUBLISHED BY ALI NIKNEJAD

RFIC.EECS.BERKELEY.EDU/EDOGS.HTML

Third printing, September 25, 2015

Contents

1. Current and Voltage

1.1 Charge and Current

Current is charge in motion. A conductor is a material where chargers are free to move about. Even in “rest", the charge carriers are in rapid motion due to the thermal energy. Typical carriers include electrons, ions, and “holes" (in semiconductors)[1]

$$I = \frac{\text{Net charge crossing surface in time } \Delta\text{t}}{\Delta t} \tag{1.1}$$

When positive charges move in the positive direction (say right), we say the current is positive. If negative charges move in the same direction, we say the current is negative. Likewise, if a positive charge moves in the negative direction (say left), we also say the current is negative.

When both positive and negative charge are moving (Fig. 1.1), the net charge motion determines the overall current.

1.1.1 Counting Charges

Suppose that the charge carriers each have a charge of q. Let’s count the number of charges (n) crossing a surface in time Δt and multiply by the electrical charge (Fig. 1.2)

$$I = q\frac{n}{\Delta t} \tag{1.2}$$

To find n, let’s make the simple assumption that all the charges are moving at a speed of v to the right. Then the distance traversed by the charges in time Δt is simply $v\Delta t$, or in other words if we move back from the surface this distance, all the charges in the volume V formed by the cross-sectional surface A and the distance $v\Delta t$ will cross the surface in time Δt. This means that

$$V = v\Delta t A \tag{1.3}$$

[1] Holes are vacancies in a crystal, or the absence of an electron (positive charge), which can move around like bubbles, but they have mass and act like charged particles. A full understanding of holes requires knowledge of quantum mechanics.

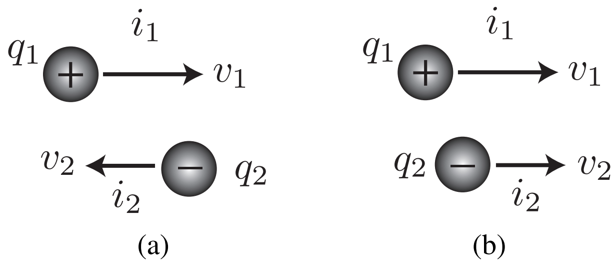


Figure 1.1: Net current is determined by the net positive charge moving to the right. (a) Since the charge q_2 is negative and moving to the left, it contributes to current positively, $Q_{net} = q_1 - (-|q_2|) = q_1 + q_2$. (b) Now q_2 is also moving to the right, and due to its negative charge, it subtracts, $Q_{net} = q_1 - |q_2|$.

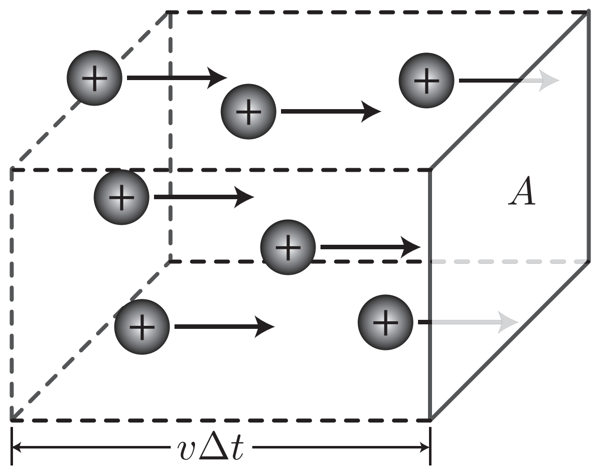
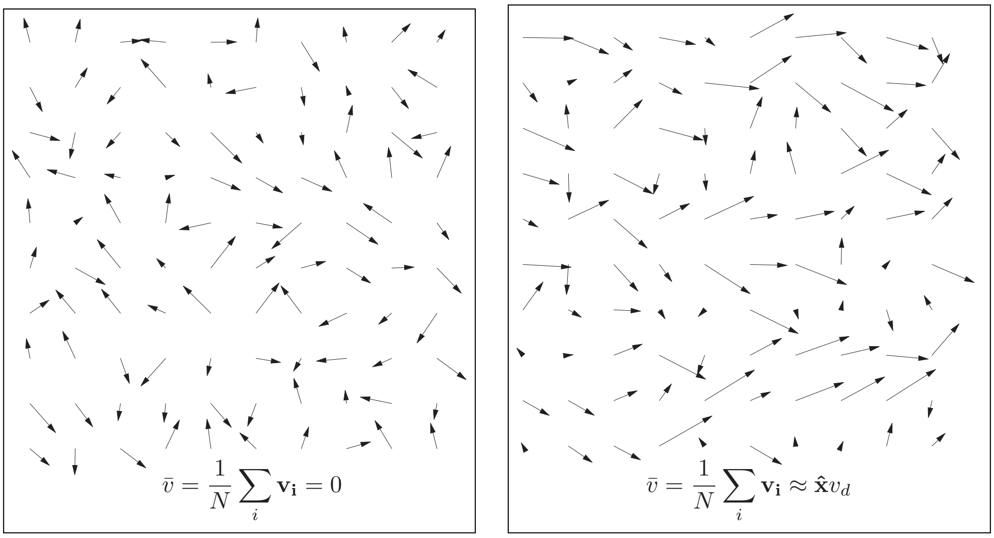


Figure 1.2: The number of charges crossing the cross sectional area A is given by the number of charges defined by the volume of A multiplied by $v\Delta t$, or the distance traveled by the charges in a time internal of Δt.



(a) (b)

Figure 1.3: (a) A random distribution of velocities leads to a net zero velocity (on average). (b) A small drift velocity is now added to the random distribution, which shows that on average charges move to the right.

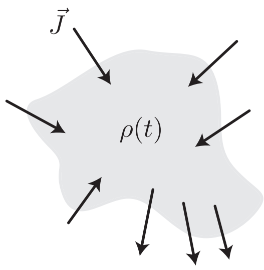


Figure 1.4: Conservation of charge is a fundamental physical fact. If the charge in a region is changing, it must be due to current flow into or out of that region.

If the volume density of charged carriers in this volume is given by N, then

$$I = q\frac{V \cdot N}{\Delta t} = q\frac{v\Delta t A \cdot N}{\Delta t} = q(NA)v \tag{1.4}$$

Aside: Conservation of Charge We know from fundamental physics that charge is conserved. That means that if in a given region the charge is changing in time, it must be due to the net flow of current into that region (Fig. 1.4). This is expressed by the current continuity relation in physics (which can be derived from Maxwell's equations)

$$\nabla \cdot \mathbf{J} = -\frac{\partial \rho}{\partial t}$$

The divergence is an expression of spatial variation of current density whereas the right-hand-side is the change in charge density at a given point.

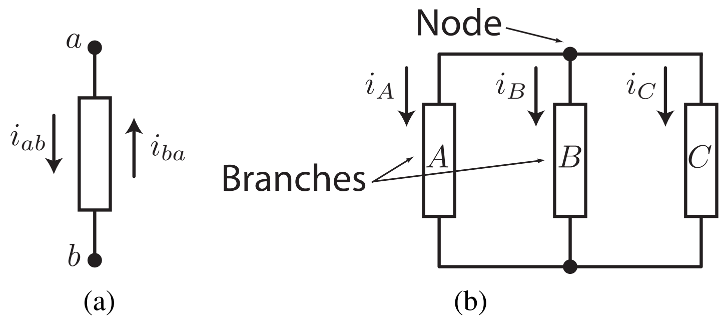


Figure 1.5: (a) The definition of the current through a component. (b) Branch currents are defined when components are connected in parallel through common nodes.

1.1.2 Motion of Real Charges

The above result emphasizes that current is associated with motion. In our simple example, we assumed all carriers move at a velocity v. In reality, as you may know, electrons move very rapidly in random directions due to thermal motion ($mv^2 \sim kT$) and v is the net *drift velocity*. This is shown in Fig. 1.3, where we have actually exaggerated the amount of drift velocity in most typical situations. Electrons actually move really fast due to thermal energy. The random motion gives rise to “noise" (think of static on a radio or snow on a TV screen), and signals are just perturbations on this random jiggling about.

1.1.3 Current Flow Through a Component

When current flows into a component (resistor, lamp, motor) from node a to b, as shown in Fig. 1.5a, we call this current i_{ab}. Note that the current i_{ab} is the same as $-i_{ba}$. When several components are connected in a circuit, we call the components branches and associate a current with each branch, as shown in Fig. 1.5b.

Current into a Component

Suppose that we now consider the current flow into a component. If we count the amount of charge Δq flowing into the component in a time interval Δt, then in the limit as $\Delta t \to 0$, the ratio is exactly the current flowing into the component

$$I = \lim_{\Delta t \to 0} \frac{\Delta q}{\Delta t} = \frac{dq}{dt}$$

1.1.4 DC versus AC Currents

Various example current waveforms are shown in Fig. 1.6. A constant current is called a “Direct Current" (DC). Otherwise it’s AC. Some AC typical waveforms are shown above. Sine waves are the waveforms coming out of an electric outlet and are very common. A square wave is the clock signal in a digital circuit, which has a well defined edge which can be used for timing. Any time-varying current is known as an AC, or alternating current. Note that the sign of the current does not necessarily have to change (the current does not have to alter direction), as the name implies.

1.2 Voltage

Due to both the attractive and repulsive nature of charges, moving them around may take energy, or it may return energy. For example, if we design a circuit that moves a lot of charges into a region

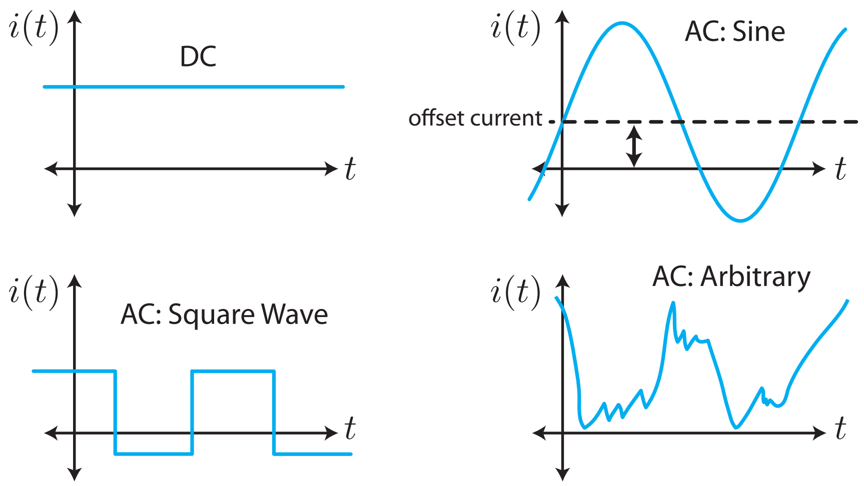


Figure 1.6: Typical current waveforms, such as a constant or DC current, a sinusoidal current, a square waveform with a DC offset, and an arbitrary waveform.

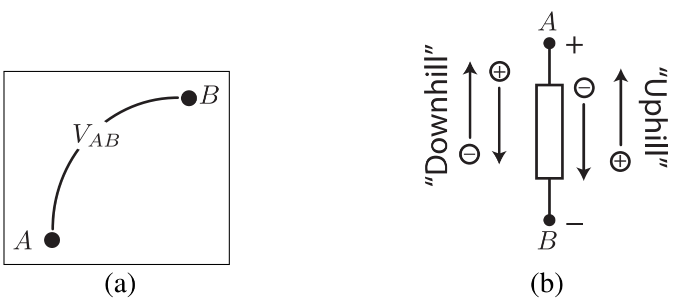


Figure 1.7: (a) The voltage V_{AB} is the energy required to move a unit of charge from A to B. (b) While the definition of voltage applies between any two points in space, we can also define the voltage drop across a component as V_{AB}. By definition, a positive voltage drop means that positive charges will “roll downhill" through the component.

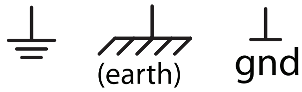


Figure 1.8: Common notation to label a node of an electrical circuit as the ground node. By convention, the ground node is arbitrarily assigned a potential of 0V.

where we store them, bunched up together on a conductor, then we have to do work because like charges repel and putting like charges onto a conductor takes a lot of energy. Note that this energy is stored since if we then allow the charges to leave the conductor, it actually takes no work at all, in fact, the circuit is willing to do work to move out of that state. In this case we say it takes negative energy.

The voltage difference V_{AB} between A and B is the amount of energy gained or lost per unit of charge in moving between the two points, as shown in Fig. 1.7a. Voltage is a relative quantity defined between two points. An absolute voltage is meaningless and usually is implicitly referenced to a known point in the circuit (ground) or in some cases a point at infinity.

If a total charge of Δq is moved from $A \to B$, the energy required is

$$E = \Delta q V_{AB} \tag{1.5}$$

If the energy is positive, then by definition energy is gained by the charges as they move “downhill". If the energy is negative, then energy must be supplied externally to move the charges “uphill". The units of voltage are Volts (after the Italian physicist Alessandro Volta), or Joules/Coulomb.

R An analogy for electrical potential is the gravitational potential for a mass (charge). While it takes energy to move mass uphill, energy is released when a mass moves downhill. The situation with electrical charge is more complicated, though, since like charges repel.

1.2.1 Voltage Across a Component

In electrical circuits, the path of motion is well defined by wires/circuit components (also known as elements). We usually label the terminals of a component as positive and negative to denote the voltage drop across the component (see Fig. 1.7).

Sometimes we don’t know the actual polarity of the voltage but we just define a reference direction. In our subsequent calculations, we may discover that we were wrong and the voltage will turn out to be negative. This is easy to detect since $V_{AB} = -V_{BA}$. By convention, when current flows into the positive terminal of a component, we say the current is positive. Otherwise the current is negative.

1.2.2 The Concept of Ground

It is common to use the ground symbol, shown above, to simplify electrical circuits. All voltages are implicitly referenced to the ground terminal. In reality, this “ground" may have a physical form, such as the earth ground, or chassis on an automobile, or a large conductor plane in an electric circuit. The requirement is that all points connected to ground should be at the same voltage, in other words ground is an equipotential surface.

This concept is of course an idealization, since no matter how conductive the ground is made, if enough current flows through the ground, then different points can be at different potentials. But usually this potential difference is smaller than the voltage drops in the circuit elements.

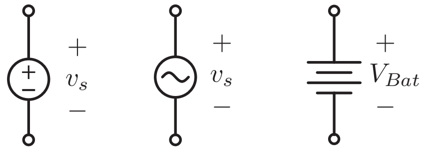


Figure 1.9: (a) The schematic representation of a voltage source, (b) an AC voltage source, and (c) a battery.

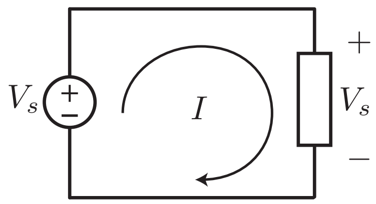


Figure 1.10: Note that in the illustration, the direction of the current flow is negative through the voltage source, indicating that charge carriers gain energy when moving through the source. On the other hand, current flowing into the component is positive.

1.3 Voltage Sources and Batteries

1.3.1 An Ideal Voltage Source

An ideal voltage source is a component, shown in various guises in Fig. 1.9, which supplies a voltage across its terminals, potentially supplying power to other circuit elements. INote that a general voltage source is shown first, and then an AC voltage source is shown next. When we say “AC" we may mean a source that outputs a constant sinusoidal voltage, such as the voltage coming out of your wall socket, or more generally a time varying source which takes on a given functional form. Finally, a battery is shown last, which is always a DC source, but generally implies a physical battery rather than an ideal source.

The current through the source component can take on any value, either positive or negative, and the voltage source maintains the same voltage regardless of the current draw. As such, we say it has no compliance, or in other words it’s an infinitely stiff source. We label the terminals of the source as positive (high voltage) and negative (low voltage). Note that the voltage at each terminal is undefined (and not necessarily negative at the negative terminal). The voltage source only enforces the *voltage difference* between its terminals to take on a certain value.

Note that the voltage source can be left open-circuited and as a result no current flows into or out of the source. When the voltage source is connected to a load (Fig. 1.10), a circuit is formed and current can flow from the source into the load. When current flows out of the positive terminal, we say the current is negative. In this illustration, current is flowing out of the source, or it’s negative for the source, and positive for the component.

In this scenario, charge moves from the source at a high energy and gives up its energy to the load, converting to other forms such as heat, motion, or potential chemical energy, and the charges return to the source in a low energy state. The voltage source then boosts the energy of these charges back up to the higher energy state and the cycle continues. An ideal source never depletes of energy, and maintains the terminal voltage difference for all time, past, present, and future.

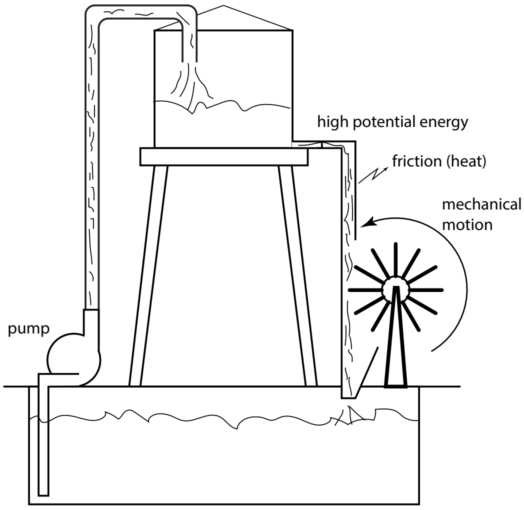


Figure 1.11: A water pump analog for an electrical circuit containing a voltage source (pump) and passive loads (pipes and water wheel).

1.3.2 Batteries

An ideal battery, like a voltage source, can supply energy to a circuit by converting stored chemical energy into electrical energy. It has a fixed voltage across its terminals and can support any current. Real batteries, though, have only a fixed amount of “fuel", or fixed energy, so their capacity to supply power to an external passive load will run out after some time. Unlike an ideal source, real batteries have a maximum current that they can supply (when short circuited), and heat up or can potentially explode when delivering this large current. Moreover, the output voltage of a battery will drop as we try to draw more current from the battery. Unlike an ideal stiff voltage source, a real battery has some compliance.

1.3.3 Water Pump Analogy

Imagine a closed circuit containing a pump, some pipes, and a water wheel, as shown in Fig. 1.11. In this example, energy is supplied by the pump, which lifts the water to a height thereby increasing the potential energy of the water. The water then flows downhill through pipes, transferring its energy to heat (friction in the pipes) and to mechanical energy, in turning the water wheel. When the water reaches the ground level, the cycle continues. In this analogy, the pump is perfectly analogous to the voltage source we have been discussing, and the water wheel and pipes represent the passive load.

1.4 Current Source

A current source is the dual of a voltage source, as shown in Fig. 1.12. It supplies a constant current regardless of the external voltage applied to its terminals. This property means that it has infinite compliance, or infinite output resistance. In contrast, a voltage source has a constant voltage regardless of the current drawn, or equivalently, it has zero output resistance.

Unlike a voltage source, there are no everyday examples (such as batteries) that can be used to illustrate the principal of a current source. Transistors in the “forward active" region do behave like current sources, but for that you’ll have to wait until later. Any real current source has finite output resistance, but it’s typically very large. Contrast this with a real voltage source.

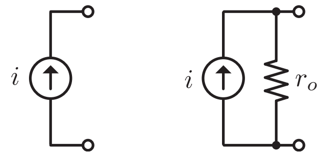


Figure 1.12: An ideal (left) current source supplies a given current (fixed or a waveform) independent of the load connected. For this reason, it's known as an independent source. A real current source (right) has non-zero output conductivity, modeled by a shunt resistance.

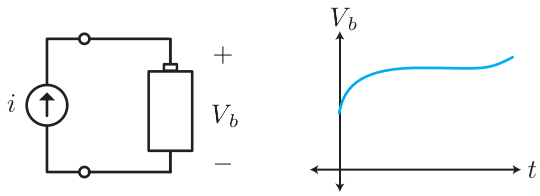


Figure 1.13: A current source is an ideal battery charger. It supplies a fixed charging current regardless of the battery terminal voltage (left). The battery cell voltage increases in a non-linear fashion as current enters the cell (right).

1.4.1 Battery Charger

A simple battery charger can be made with a current source, as shown in Fig. 1.13a. A real battery has a discharge curve, in other words as you draw current from the battery, the output voltage drops. A current source can be used to charge the battery, and it will deliver current regardless of the battery voltage. Of course, we need provisions to turn off the current source when the battery is full. In the old days, a timer was used to control how long the current was on. In more modern units, the charger monitors the battery voltage and detects when the battery is full (Fig. 1.13b).

1.5 Dependent Sources

Current and voltage sources are *independent* sources because their output is independent of any other currents or voltages in the circuit. In contrast, dependent sources depend directly on the the voltage or current in other parts of the circuit. Four flavors are possible (Fig. 1.14): Voltage Controlled Voltage Source (VCVS), Voltage Controlled Current Source (VCCS), Current Controlled Voltage Source (CCVS), and Current Controlled Current Source (CCCS). A VCCS is an ideal voltage amplifier or attenuator (gain < 1), while a CCCS is an ideal current amplifier or attenuator. Sometimes the signal of interest is a current while the desired output is a voltage. A CCVS is

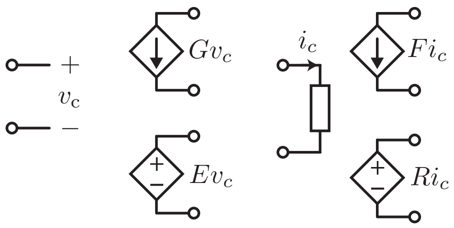


Figure 1.14: Family of dependent sources. In these examples, the dependent outputs depend either on the control voltage v_c or the control current i_c.

an ideal trans-resistance amplifier, because the gain has units of resistance. A trans-conductance amplifier does the inverse.

2. KCL/KVL and Energy and Power

2.1 KVL and KCL

KVL and KCL are Kirchhoff's famous Voltage (V) and Current (C) Laws, formulated by Gustav Kirchhoff in 1845 as his first and second laws. As we shall see, they are extremely useful for analyzing circuits containing components connected in arbitrary ways.

2.1.1 KCL: Kirchhoff's Current Law

KCL states that the net charge flowing into any node of a circuit is identically zero. In the example shown in Fig. 2.1a

$$i_1 + i_2 - i_3 = 0$$

The reason for this is clear from the flow nature of current. Some of the currents flow in, some flow out, but in the net all must balance out. A direct implication of KCL is that series elements have equal currents, $i_A = i_B$, as shown in Fig. 2.1b.

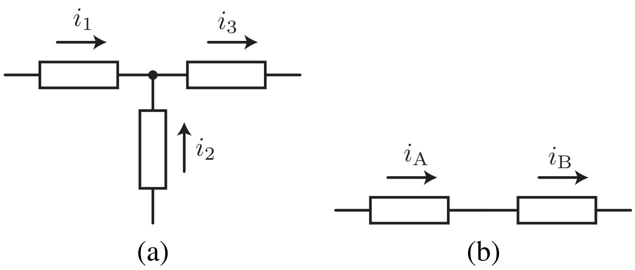


Figure 2.1: (a) KCL states that the net charge into a node of a circuit is zero, or $i_1 + i_2 - i_3 = 0$. Note that i_3 is leaving the node, hence the negative sign in the KCL equation. (b) For elements connected in series, the currents must equal, or $i_A = i_B$.

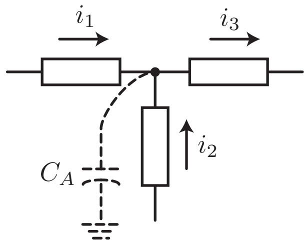


Figure 2.2: The connection between KCL and charge conservation is captured by defining capacitance at a node, a concept we will introduce later.

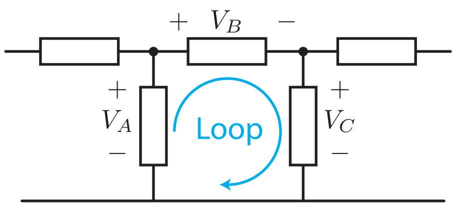


Figure 2.3: KVL applied to an example circuit loop (or mesh).

Aside: Origin of KCL KCL is related to charge conservation. Note that this is just a restatement of current continuity $\nabla \cdot \mathbf{J} = -\frac{\partial \rho}{\partial t}$. If the net current is not zero, then somehow charge must be accumulating at a node. This can only happen if the node in question has some capacitance (say to ground). But such a capacitance can always be included as a separate component (come back and re-read this after we define capacitance). So if we say that the node as zero capacitance, or $C_A \equiv 0$, then KCL makes sense.

2.1.2 KVL: Kirchhoff's Voltage Law

KVL states that the net potential around any loop in a circuit is zero

$$\sum_{Loop} V_k = 0 \tag{2.1}$$

Or more explicitly, for the example shown in Fig. 2.3, adding the voltage "drops" around the loop we get

$$-V_A + V_B + V_C = 0$$

In other words, the net energy in going around a loop is zero. Notice that if we had defined the loop in the opposite direction, then we have:

$$V_A - V_C - V_B = 0$$

or by multiplying the equation by -1, the same relation.

Connection to Electromagnetics: KVL makes sense if the voltage arises from electrostatic sources, which leads to a conservative field. Then if we calculate the net energy in traversing any closed path, including any loop in a circuit, it must be zero since we return to the same point. You may be wondering about a non-electrostatic situation in which the field is not conservative. In this situation, if there is a changing magnetic field crossing the loop (such as in a motor), then energy can be transferred into or out of the circuit (see Fig. 2.4). This can be represented as coupled inductors linking the loop to other circuitry, a detail which we will ignore for now.

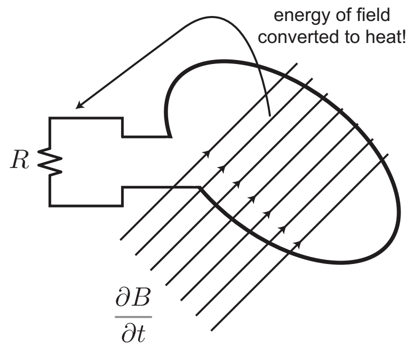


Figure 2.4: A time-varying magnetic field cuts an electric circuit, transferring energy into and out of the circuit.

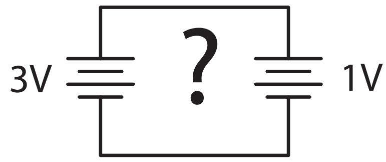


Figure 2.5: Two batteries with a differential voltage should not be connected in parallel.

2.1.3 KVL Implications

All shunt (or parallel) components have the same voltage. This is why you should *never* connect batteries in parallel (like Fig. 2.5). Since each one has a fixed potential across its terminals, it is simply not possible to do this with ideal elements. In practice, a large current would flow from the higher voltage battery to the lower one, possibly burning the wires connecting them, or worse causing an explosion. The magnitude of the current is limited by the internal resistances of the batteries (come back and re-read this after we talk about resistance).

2.1.4 Your Taste for Opens and Shorts

Recall that voltage sources detest short circuits, because that violates KVL. Also, two voltage sources cannot be put in parallel. Voltage sources do like open circuits, because they don't draw any power (Fig. 2.6).

The current source has the dual relationship. It cannot drive an open circuit, because that violates KCL. Likewise, two current sources cannot be put in series. On the other hand, current sources love short circuits, and they happily deliver all of their current to a short circuit. This results in no power draw (why?).

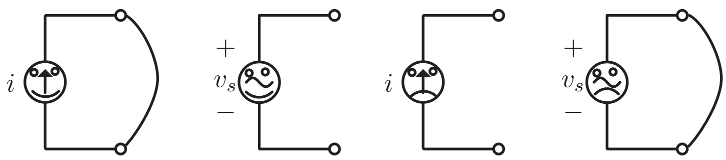


Figure 2.6: “Happy" and "sad" current / voltage sources.

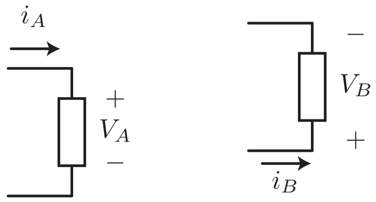


Figure 2.7: (a) A component has a current i_A flowing through it while experience a potential drop of V_A. (b) Another component has a current i_B and a voltage drop of V_B.

2.2 Power and Energy

By the definition of electrical potential voltage V, the change of energy ΔE experienced by a charge increment Δq in moving from A to B is given by $\Delta E = V\Delta q$, where V is the potential difference between points A and B ($V = V_A - V_B$). [1] Points A and B could be the nodes of a circuit element.

2.2.1 Energy Sign Convention

By convention, positive energy ΔE means the charge loses energy, whereas a negative value means the component gains energy. In Fig. 2.7a, if V_A is positive, then the charges flowing into the component due to i_A are losing energy. If V_A is negative, then they gain energy.

Then the power is simply the time rate of change of energy

$$P = \lim_{\Delta t \to 0} \frac{\Delta E}{\Delta t} = V \lim_{\Delta t \to 0} \frac{\Delta q}{\Delta t} = V \times I \tag{2.2}$$

As expected, the units of $[P] = [V][I] = [J/C][C/s] = [J/s]$. The instantaneous power $p(t) = v(t)i(t)$ is the power dissipated (positive) or supplied (negative) by a component with voltage $v(t)$ with current $i(t)$.

Example 1: Power Through a Component

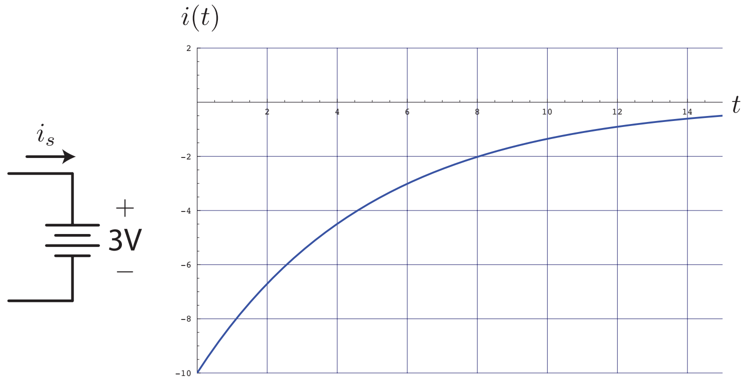
As shown in Fig. 2.7a, suppose a component has a measured voltage of $V_A = 120\text{V}$ across its terminals while supporting a current $i_A = 2\text{A}$ through it. What is the power dissipated in the component? Given that the current flows into the positive terminal of the component, then

$$P_A = V_A \cdot i_A = 240\text{W}$$

Note that the component is absorbing power. Due to conservation of energy, the power is converted into other forms (such as heat or mechanical energy). A lamp converts electrical energy into light and heat. A motor converts electrical energy into mechanical energy. A battery stores the energy.

---

[1] We like to use a small charge Δq because we don't want to disturb the system. In other words, V itself is a function of charge, so let's assume we use a very small amount of charge.



(a) (b)

Figure 2.8: A time-varying current i_s flows into a 3V battery.

Example 2: Power Through Another Component

Suppose another component has a measured voltage of $V_B = -12\text{V}$ across its terminals while supporting a current $i_B = 5\text{mA}$ through it (see Fig. Fig. 2.7b). What is the power dissipated in the component? Given that the current flows into the positive terminal of the component, then

$$P_A = V_B \cdot i_B = -60\text{mW}$$

Note that the component is absorbing negative power, or in other words it's supplying power. Due to conservation of energy, the power must be coming from another internal source, for example the chemical energy of a battery, the heat from a thermocouple, light from a solar cell, or a mechanical generator.

Example 3: Instantaneous Power

Suppose a 3V voltage source has an instantaneous current of $i_s(t) = -10e^{-t/\tau}$ mA for $t > 0$, as shown in Fig. 2.8 ($\tau = 5\text{ns}$). Find the net energy supplied by the source.

Given that the current flows into the positive terminal of the component, then

$$p_s(t) = 3\text{V} \cdot i_s(t)$$

If we integrate the power over time, we arrive at the net power supplied in this interval (note that $i(t) < 0$ for all time so the power always flows out of the source)

$$E_s = \int_0^\infty p_s(t)dt = -30\text{mW} \int_0^\infty e^{-t/\tau} dt = \frac{+30\text{mW}}{1/\tau} \left(e^{-\infty} - e^0\right)$$

$$= -30\text{mW} \times 5\text{nsec} = -150\text{pJ}$$

3. Conductance and Resistance

3.1 Conductors

3.1.1 Ideal Conductors

An ideal conductor is simply a material that carries electrical current unimpeded. We have been implicitly using ideal conductors as “wires" to connect components. Ideal wires, which connect components in an electrical schematic, are made from ideal conductors.

R An ideal conductor is an equipotential body. It can support any current without incurring a potential drop – i.e. it has zero resistance. Between any two points on an ideal conductor, $V_{AB} \equiv 0$, as shown in Fig. 3.1.

An ideal ground would use an ideal conductor to serve as a constant reference potential for the circuit. In a PCB (printed circuit board), a “ground plane" is commonly used to provide a low resistance path to the power source. It is also convenient for wiring, since any component can connect to the ground, often made with a “via", a hole in the PCB that connects layers together, thus eliminating most wire traces to ground (Fig. 3.2).

3.1.2 Resistance

Real conductors have “resistance", which impedes the flow of current by producing a voltage drop across the path of the current. The higher the current, the higher the voltage drop. Remarkably, the

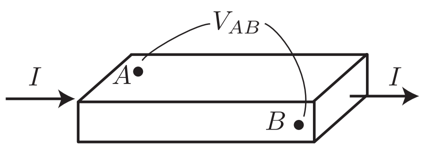


Figure 3.1: An ideal conductor is defined as a material that can pass current while maintaining zero voltage drop across its terminals.

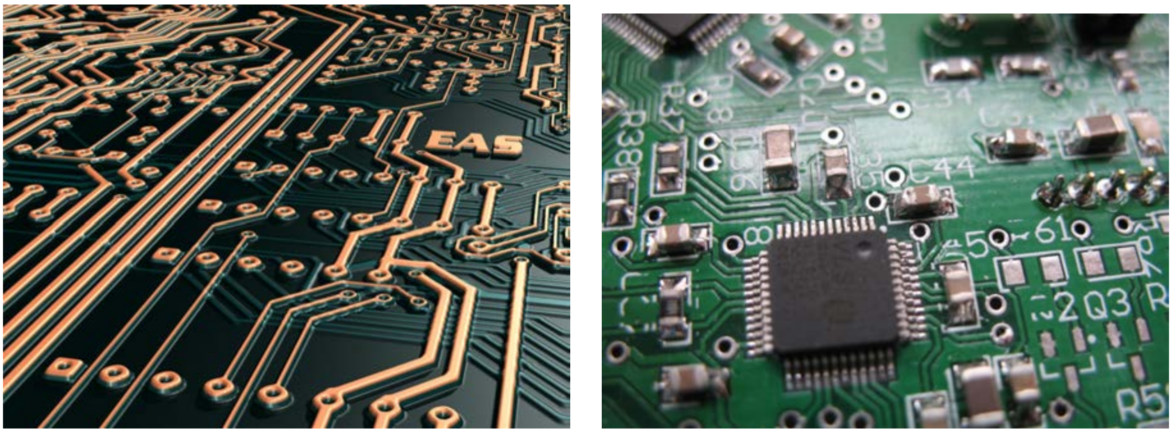
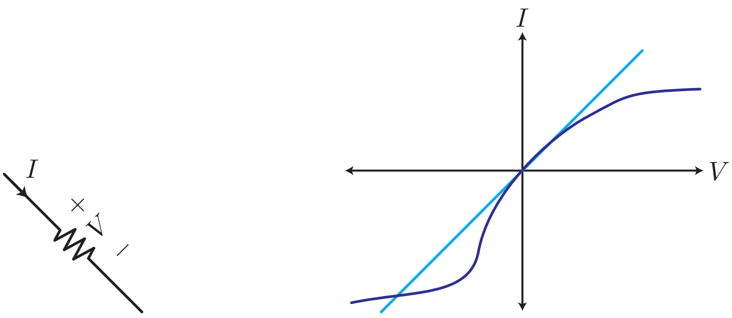


Figure 3.2: In a PCB, layers of conductors are sandwiched with insulating layers. Components are soldered on the top and bottom surfaces. One or more layers are used to approximate an equipotential surface (ground / supply) by using a solid sheet of conductor. Components that are connected to these points use a via to connect up or down to the conducting planes.



(a) (b)

Figure 3.3: (a) Current flow into a resistor results in a voltage drop proportional to the current. (b) While one would expect the current-voltage (I-V) to be a complicated function, in reality for most materials, the result is remarkably linear.

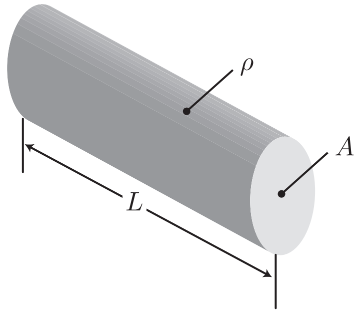


Figure 3.4: The resistance of a wire is proportional to its length ℓ and the material resistivity ρ, and inversely proportional to its cross sectional area A.

voltage drop is exactly proportional to the current. This is a statement of Ohm's Law

$$V = I \cdot R \tag{3.1}$$

The proportionality constant is known as the resistance of the conductor. The unit of R is Voltage/Ampere, or more commonly "Ohms" (Ω). Such a conductor is known as a resistor. The value of resistance can be altered by using different materials in the construction of the component, or by modifying the physical dimensions of the conductor. The schematic representation for a resistor is shown in Fig. 3.3a, which hints that the current is impeded (zig-zag path).

3.1.3 Calculating Resistance

For a wire of uniform cross area, the resistance is calculated as follows

$$R = \frac{\rho \ell}{A} \tag{3.2}$$

It's proportional to the length ℓ of the wire, inversely proportional to the cross sectional area A, and proportional to the material resistivity ρ. The resistivity of some common conductors and insulators is shown in the Table 3.1. Note the enormous range in resistivity.

Material	Resistivity $\Omega - \mathrm{m}$
Copper (Cu)	1.72×10^{-8}
Gold (Au)	2.27×10^{-8}
Silicon (Si)	$10^{-5} \sim 1$
Quartz (SiO_2)	$> 10^{21}$
Teflon	10^{19}

Table 3.1: Resistivity of some representative materials.

3.1.4 Conductance and Conductivity

It's sometimes convenient to think in terms of conductivity rather than resistivity. We re-cast Ohm's Law as

$$I = G \cdot V \tag{3.3}$$

where G is the conductance of a resistor. Note that this is simply the inverse of resistance, $G = 1/R$. The units of G are inverse Ohms (Ω^{-1}), also called Siemens (S). Similarly, the conductance of a

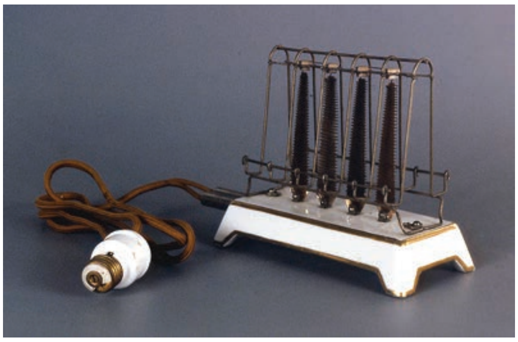


Figure 3.5: Toaster ovens come in any shapes and forms, but at their heart they are just resistors that convert electrical energy into heat.

material can be calculated from the equation

$$G = \frac{\sigma A}{L} \tag{3.4}$$

σ is the material conductivity, which is the inverse of the resistivity $\sigma = 1/\rho$.

3.2 Power Loss in Resistors

From the equation for the power loss in a component, we have

$$P = V \cdot I = (R \cdot I) \cdot I = I^2 R \tag{3.5}$$

Or in terms of conductance

$$P = V \cdot I = V \cdot (G \cdot V) = V^2 G \tag{3.6}$$

This power is lost to heat or “Joule Heating". In fact, most electric ovens use resistors to heat up the oven.

R Nichrome (nickel-chromium alloy) is often used as the heating element. It has a high melting point of 1400°C and a hight resistivity and resistance to oxidation at high temperature. It is widely used in ovens, hair dryers, and toasters.

Ideal Wires

An ideal wire is an equipotential surface, and so it cannot support a voltage from one end to the other. The power dissipated by an ideal wire is also zero. An ideal wire is also called a “short circuit", especially when placed from one point to the other, we say the nodes are “shorted".

3.3 Resistors versus Resistance

It’s important to realize that we often use a resistor in a schematic to model the equivalent resistance of a component which may not be resistor at all. Take for example a loudspeaker, which primarily converts electrical energy into sound (pressure waves). Such a component does not ideally dissipate any power as heat, and yet the power conversion into sound can be represented by an equivalent resistance ($R_{\text{speaker}} = 8\Omega$).

Other examples include a light bulb (converts electricity into heat and light), an antenna (which converts electricity into electromagnetic radiation), or an entire house which contains hundreds of individual devices dissipating energy.

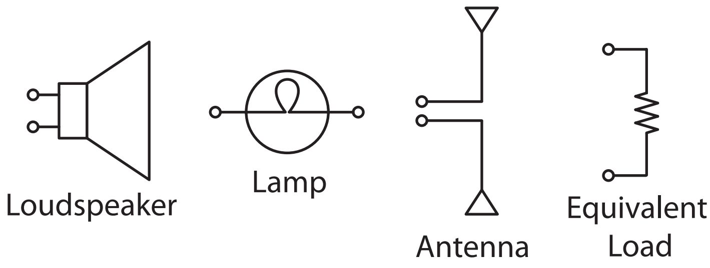


Figure 3.6: Many components that convert electrical power to other forms of energy, such as sound, light or electromagnetic radiation, can be represented as an equivalent resistor.

Example 4: Application: Strain Gauge

A strain gauge uses the change in resistance induced by the change in the dimensions of materials $L = L_0 + \Delta L$ under strain, $R = \rho \frac{L}{A}$. Define

$$G \equiv \frac{\frac{\Delta R}{R_0}}{\varepsilon}$$

For example, suppose $R_0 = 500\Omega$, $G = 3$ and a 1% strain is applied. Then $\Delta R = R_0 \cdot G \cdot \varepsilon = 15\Omega$. Precise changes in resistance can be measured with a Wheatstone Bridge (see Ch. 4).

3.3.1 Why Power Is Delivered with High Voltages

For economic and environmental reasons, power is usually generated remotely (wind farms, electric dam, nuclear power plants, etc). Many hundreds of miles of wires are required to carry the energy to the factories and homes. Since wires have resistance, the resistive loss is energy which is completely lost without doing any useful work (except warming up the planet). Since the line voltage is fixed (say 120V into the home), the power draw is represented by a varying I_{load}. The equivalent circuit (Fig. 3.8a) shows that the "load" can be represented by R_{load} and the power loss is proportional to $I^2_{load} R_w$.

Say a house is dissipating $P = 1.2\text{kW}$ of power. If the house operated from a 120V DC source, then that's a current draw of [1]

$$I_{load} = P/V = 1\text{kW}/120\text{V} = 10\text{A}$$

Assume a copper wire of 5mm radius and 100km long

$$R_w = \rho \frac{L}{A} = \rho \frac{L}{\pi r^2} = 1.72 \times 10^{-8} \times \frac{100 \times 10^3}{\pi (5 \cdot 10^{-3})^2} = 22\Omega$$

The power lost to heat is

$$P_{loss} = I^2_{load} R_w = 2.2\text{kW}$$

That's more than the power delivered! The efficiency of this system is very low.

[1] Note we're ignoring the fact that most power delivered is AC rather than DC, but as we'll learn later, that's a simple correction factor

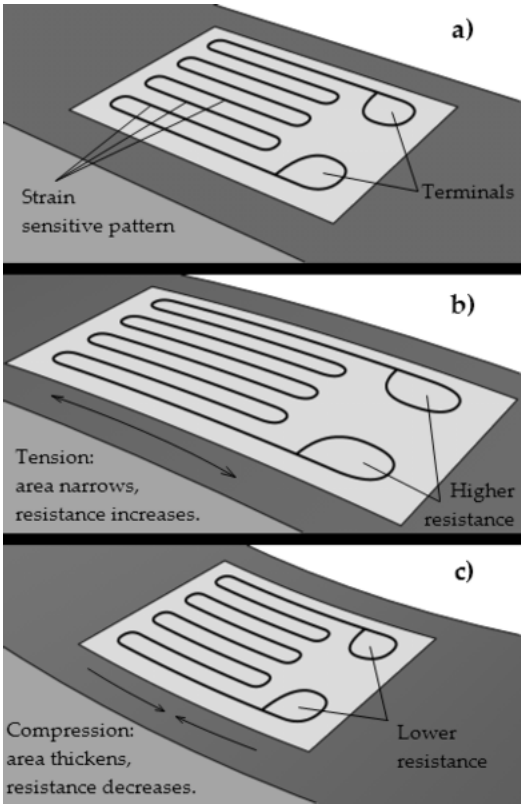
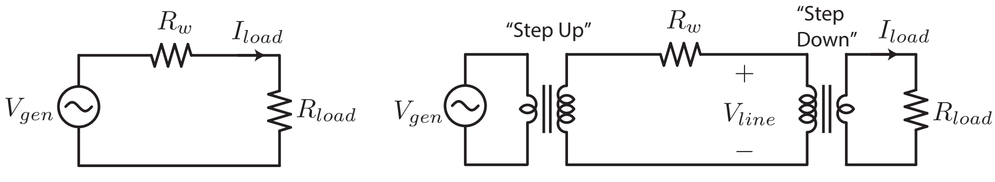


Figure 3.7: A strain gauge measures deformations of a structure by using a strain sensitive resistive pattern (Source: Wikipedia).



(a) (b)

Figure 3.8: (a) A simple model for an electrical generator, the long wires from the generator ot the load represetned by R_w, and the load, R_{load}. (b) In a practical scenario, the voltage is stepped up and down using transformers to reduce the current.

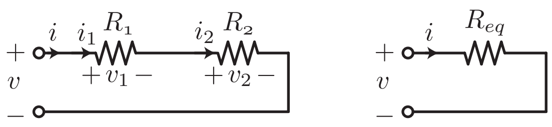


Figure 3.9: Resistors in series can be represented by an equivalent resistance $R_{eq} = R_1 + R_2$.

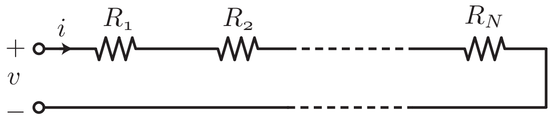


Figure 3.10: Any number of series resistors can be combined into an equivalent resistor $R_{eq} = \sum R_k$.

High-Voltage Lines

The key is to transform the voltage to a much higher value to minimize the current, as shown in Fig. 3.8b. One key reason why we use AC power is so that we can use transformers to easily boost the voltage on the lines, and then to drop the voltage to more reasonable values as we get near residential areas (for safety). In the above example, the voltage is increased to 100kV, and so the equivalent current needed to deliver 1.2kW is decreased substantially

$$I_{load,line} = P/V_{line} = 0.012\text{A}$$

And so that the energy lost to heat is minimized

$$P_{loss} = I^2_{load,line} R_w = 3\text{mW}$$

3.4 Series Resistors

Consider two resistors in series as shown in Fig. 3.9. The voltage across the two resistors is the sum of the voltage across each individual resistor (KVL)

$$v = v_1 + v_2 = i_1 R_1 + i_2 R_2 \tag{3.7}$$

From KCL, we know that the currents through the resistors is equal, and so we can write $i = i_1 = i_2$

$$v = i(R_1 + R_2) = iR_{eq} \tag{3.8}$$

where R_{eq} is an equivalent resistance which describes the behavior of the series connection of the resistors. As you would expect, the net resistance increases when resistors are placed in series.

The concept of series resistors generalizes since we have from KVL and KCL (Fig. 3.10)

$$v = v_1 + v_2 + \cdots + v_N = i(R_1 + R_2 + \cdots R_N) \tag{3.9}$$

$$R_{eq} = \sum R_k \tag{3.10}$$

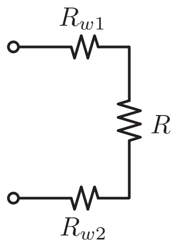


Figure 3.11: Any real circuit uses wires that have non-zero resistance R_{w1} and R_{w2}. These resistors can be absorbed into the load or they should be made to be a small fraction of the total resistance, $R_{w1,2} \ll R$.

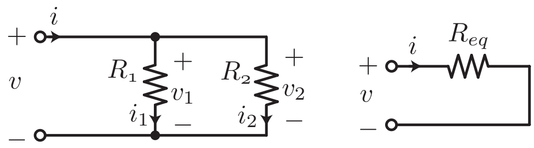


Figure 3.12: Two resistors in parallel can be combined into an equivalent resistance given by $R_{eq} = 1/(R_1^{-1} + R_2^{-1})$, usually written as $R_1||R_2$.

3.4.1 Resistance of Wires

In any real circuit, the wires are made from conductors that have resistance as shown in Fig. 3.11. To be exact, we should include series resistor as shown above to model the wire resistances R_w. For instance, the current into a resistor is given by

$$I = V/R_{eq} = V/(R + R_{w1} + R_{w2}) \tag{3.11}$$

This is tedious because in practice and unnecessary if we can make the wire resistance much smaller than the component resistances: $R \gg R_w$.

3.5 Parallel Resistors

Shunt resistors are the “dual" of series resistors if we consider the conductance of the circuit. From KVL we know that the voltage across the shunt (parallel) components is equal

$$v_1 = v_2 = v \tag{3.12}$$

From KCL we also know that the total current into the network is the sum of the currents of the individual components

$$i = i_1 + i_2 \tag{3.13}$$

$$i = G_1 v_1 + G_2 v_2 = (G_1 + G_2)v = G_{eq} v \tag{3.14}$$

Conductance of shunt resistors add.

Alternative Expressions for Parallel Resistors

Another useful expression is directly in terms of resistance

$$\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2} \tag{3.15}$$

$$R_{eq} = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2}} = \frac{R_1 R_2}{R_1 + R_2} \tag{3.16}$$

It's very common notation to write $R_{eq} = R_1||R_2$, in other words we define an operator '||' that takes two inputs outputs

$$x||y = \frac{xy}{x+y}$$

Note that if $x \gg y$, then

$$x||y = \frac{xy}{x+y} = \frac{y}{1+\frac{y}{x}} \approx y$$

For N resistors in parallel, the result generalizes

$$G_{eq} = G_1 + G_2 + \cdots + G_N \tag{3.17}$$

$$R_{eq} = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \cdots + \frac{1}{R_N}} \tag{3.18}$$

Example 5: "Christmas" Lights

Let's say we want to light up a string of lights and use the minimum number of wires. We can connect the lights in series or in shunt as shown in Fig. 3.13.

When placed in series, if a single light fails and goes into an open circuit state (typical failure mode), then all the lights turn off. On the other hand, if one of the lights short

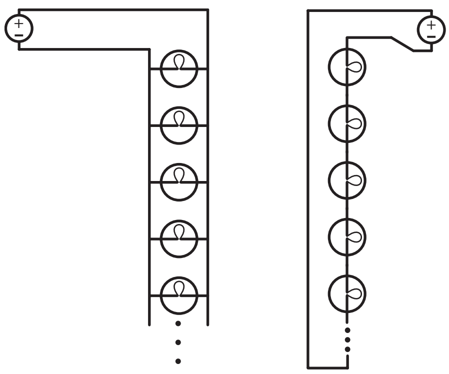


Figure 3.13: Christmas lights constructed with resistors in parallel (shunt) or in series.

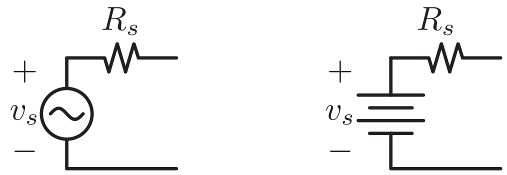


Figure 3.14: Real voltage sources and batteries have a source resistance, which can be represented as a series connection of an ideal source and a resistance.

circuits, then all the other lights would continue to work (albeit with a slightly higher voltage).

If the lights are placed in shunt, then if one light fails (open circuit), all the other lights continue to work unabated. If a short circuit occurs (unlikely), though, then all the lights are shorted and turn off.

3.6 Internal Resistance of Battery

Physical voltage sources such as batteries have internal resistance, which is represented as a series resistor in the schematic (Fig. 3.14). The output voltage therefore drops if you draw more current from the battery.

The maximum current available from the battery is also set by this resistance, because even a perfect short circuit across the battery can only draw $i = v_s/R_s$. When two batteries are connected in parallel, the current draw from the larger voltage battery to the smaller battery is determined by the sum of their internal resistances.

Example 6: Microprocessor Power Supply

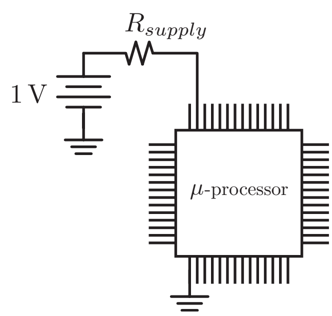


Figure 3.15: A microprocessor is connected to a 1V battery. The supply resistance R_{supply} is a combination of the battery internal voltage and the resistance of the wires and traces that connect the source to the processor.

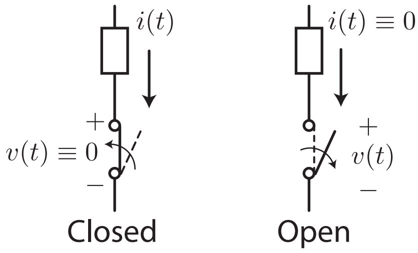


Figure 3.16: An ideal switch allows current to flow when closed but breaks the circuit in open state. Note that when closed the voltage across the switch is zero, whereas when the switch is open the current through the switch is zero.

Consider a high performance microprocessor that operates on a 1V power supply (this is typical for nanoscale CMOS technology), as shown in Fig. 3.15). During peak operation, it dissipates 50W. Calculate the supply resistance in order to incur only a 10% efficiency loss.

Solution: Note that the current draw is huge (due to the low supply voltage): $I = P/V = 50\text{W}/1\text{V} = 50\,\text{A}$. As a load, the microprocessor can be modeled as a resistor of value

$$R_\mu = \frac{1V}{50A} = .02\Omega$$

The efficiency of the system can be written as

$$\eta = \frac{P_\mu}{P_\mu + P_{loss}} = \frac{I^2 R_\mu}{I^2 R_\mu + I^2 R_{supply}}$$

$$= \frac{R_\mu}{R_\mu + R_{supply}} = 90\%$$

$$R_{supply}.9 = 0.1 R_\mu \rightarrow R_{supply} = R_\mu/9 = 0.0022\Omega$$

The supply must have an extremely low source resistance. In practice the current is delivered from multiple supplies each “regulated" (so that the fluctuations from the external supply do not cause problems) with a circuit that uses *feedback* to realize very low source resistance.

3.7 An Ideal Switch

We are all familiar with mechanical switches, for instance turning on and off the lights in the house. These switches basically have two states, open-circuit and short-circuit. When an ideal switch is open (off), the flow of current is interrupted and $I \equiv 0$ (Fig. 3.16). When an ideal switch is closed (on), then current flows readily through the switch but the voltage across the switch is zero, $V \equiv 0$.

R In a fluid flow analogy, the switch is a valve with only two stages, “on" and “off".

An ideal switch cannot absorb power because in either state, the voltage or current is zero so $p(t) \equiv 0$. All real switches have finite on-resistance, which causes a small power loss. In off-state, there is only the small conductivity due to air, which is more or less a true open-circuit unless you are concerned with extremely small currents. If the voltage across the switches in off-state gets too large though, dielectric breakdown will occur, which will result in arcing and a large current flow.

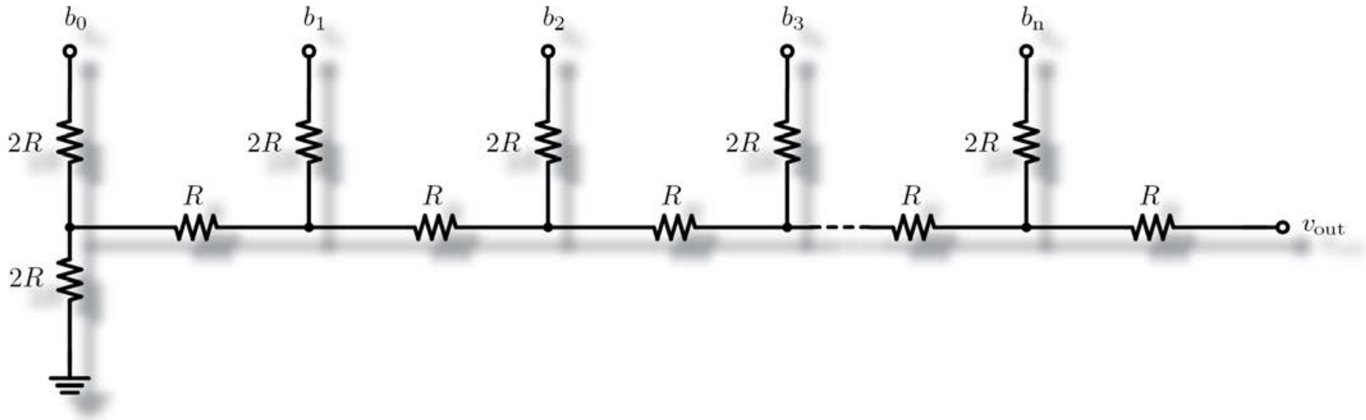
3.8 Is Ohm's Law Strange?

It's also remarkable that this linear relation holds true over a wide range of voltages. In general, one might expect a power series relation to model the current voltage relationship

$$V = f(I) = R_1 I + R_2 I^2 + \cdots \tag{3.19}$$

But for all practical purposes, R_1 is the only term that matters. The voltage drop across a resistor is V. Note that the V represents how much energy is lost by a unit of charge as it moves the the resistor element. The current I we have learned is proportional to the velocity of charge carriers (such as electrons). We would therefore expect a quadratic relation, not linear, between the current and voltage (Kinetic energy).

The answer to this riddle lies in the fact that carriers do not move unimpeded through a conductor (in vacuum they would in fact have a quadratic dependence), but rather the motion is mostly random motion. On average the charge carriers move only a short distance before colliding with atoms (impurities) in the crystal. After the collision all "memory" of the previous path of motion is lost and the energy of the carrier is converted into heat (vibrations in the crystal). The gain in momentum is proportional to the voltage.



4. Voltage and Current Dividers

4.1 Voltage Dividers

A voltage divider is an extremely useful circuit since it allows us to derive any fraction of the source voltage at the output (Fig. 4.1). Notice that the current in the circuit is given by

$$i = \frac{v_s}{R_1 + R_2} \tag{4.1}$$

Suppose the output voltage v_2 is connected across the terminals of R_2. Let's calculate the output voltage v_2 in terms of the source voltage v_s

$$v_2 = iR_2 = \frac{R_2}{R_1 + R_2} v_s = \alpha v_s \tag{4.2}$$

Note that the factor $\alpha \leq 1$. By changing either R_1, R_2, or both, we can vary the attenuation of the circuit. Between R_1 and R_2, the largest resistor wins (gets the majority of the voltage drop).

4.1.1 Potentiometers

A potentiometer, as shown in Fig. 4.2, is typically constructed so that it presents a fixed resistance R across two of its terminals. The third terminal is connected to a point between the two other

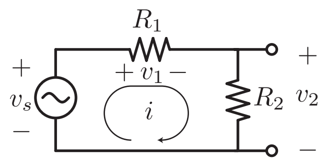


Figure 4.1: A voltage divider can attenuate a voltage to any fractional value by choosting the resistance ratio appropriately.

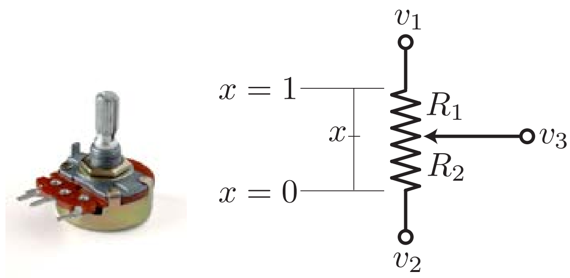


Figure 4.2: A potentiometer is a device with three terminals and an adjustment knob ("volume knob" or slider, for example) that varies the resistance measured between points 3 and either terminal.

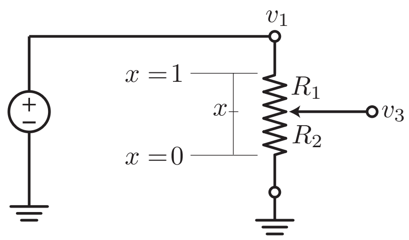


Figure 4.3: A voltage divider constructed with a potentiometer.

terminals in such a way that the resistance R_2 varies linearly (or perhaps logarithmically depending on the application).

For the potentiometer circuit shown in Fig. 4.3, since the total resistance is fixed, we can write

$$v_3 = \frac{R_2}{R_1 + R_2} v_s = \frac{x(R_1 + R_2)}{R_1 + R_2} = x v_s. \tag{4.3}$$

where x is the position of terminal 3 ($0 \leq x \leq 1$).

4.2 Current Dividers

Suppose two conductors are placed in parallel, such as shown in Fig. 4.4. The current i_s then splits into the two conductors. The ratio of the current into say G_1 can be calculated as follows

$$i_1 = G_1 v_s \tag{4.4}$$

$$v_s = (G_1 + G_2)^{-1} i_s \tag{4.5}$$

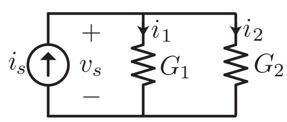


Figure 4.4: A current divider can split the current of a source into two branches. The current division is controlled by adjusting the ratio of G_1 and G_2.

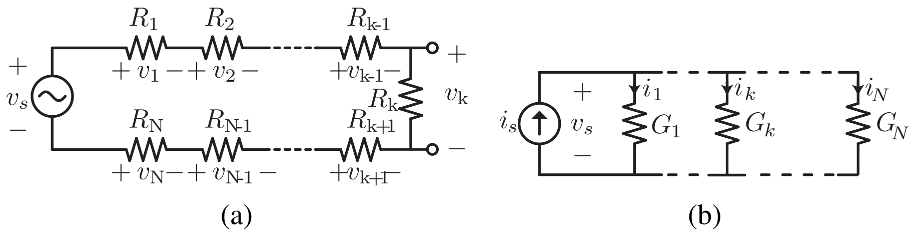


Figure 4.5: A generalized (a) voltage and (b) current divider.

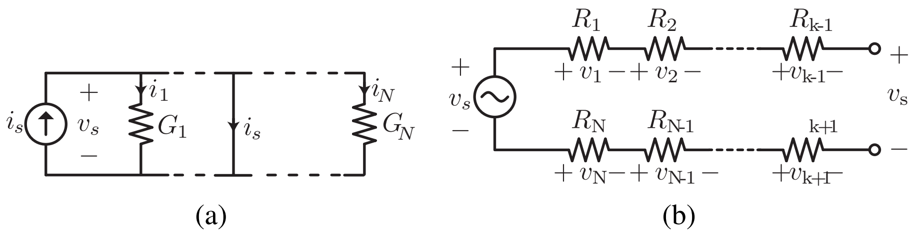


Figure 4.6: (a) In a current divider, a short circuit takes all the current. (b) In a voltage divider, an open circuit takes all of the voltage.

Substituting the above relation, we have that the current into i_1 is given by

$$i_1 = \frac{G_1}{G_1 + G_2} i_s = \frac{R_2}{R_1 + R_2} i_s = \beta i_s \tag{4.6}$$

As before, $\beta \leq 1$ and the largest conductance (smallest resistance) “wins" (get the majority of the current).

4.2.1 Generalization of Dividers

The equations for voltage divider and current divider are easy to generalize to N series or parallel resistors (Fig. 4.5)

$$v_k = \frac{R_k}{\sum R_j} v_s \tag{4.7}$$

$$i_k = \frac{G_k}{\sum G_j} i_s \tag{4.8}$$

R As shown in Fig. 4.6, in a current divider, a short circuit always wins, whereas in a voltage divider, an open circuit always wins.

4.3 (Bad) Application: Light Dimmer

Suppose we wish to build a light dimmer using a variable resistor. By adding a resistor in series, we can control the voltage drop across the lamp, which is modeled as an equivalent resistor. For $R = 0\Omega$, the bulb radiates at full intensity. For $R = \infty\Omega$, the light shuts off.

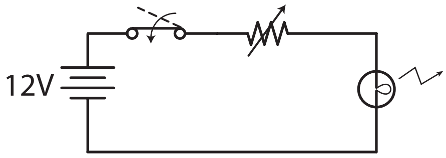


Figure 4.7: A light dimmer uses a series rheosat to dim a light. Note that this a very inefficient solution.

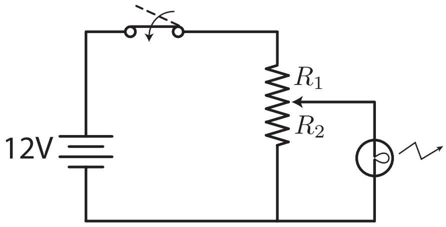


Figure 4.8: A light dimmer using a potentiometer to dim a light. Note that this is also very inefficient.

For example, say a 10W light bulb works off a 12V battery. It has an equivalent DC resistance of

$$\frac{V^2}{R_{eq}} = 10\,\mathrm{W} \rightarrow R_{eq} = \frac{V^2}{10\,\mathrm{W}} = 14.4\Omega \tag{4.9}$$

The lamp draws $I = 10\,\mathrm{W}/12\,\mathrm{V} = 0.83\,\mathrm{A}$ at full intensity. To go to 10% light level, we should reduce the current by a factor of 100. In other words, the equivalent series combination of the rheostat and lamp should present

$$R_T = \frac{12V}{I_{LL}} = \frac{12V}{\frac{10}{1200}} = 1440\Omega \tag{4.10}$$

Which means the rheostat should have a resistance of $R_{max} = 1440\Omega - 14.4\Omega$.

R Why is this a bad idea? What is the efficiency of such a system when the light is at low intensity?

4.3.1 Dimmer with Potentiometer?

You may have wondered why we didn't use a potentiometer to control the light level. The reason is that the equation we derived for the potentiometer neglected the *loading* effect of anything connected to the third terminal. You can now see that the correct voltage divider equation is given by

$$V_{bulb} = \frac{R_{bulb}||R_2}{R_1 + R_{bulb}||R_2} \tag{4.11}$$

This is not as straightforward as we thought. If $R_2 \ll R_{bulb}$, then the desired equation would follow but then a lot of unnecessary power would be wasted in R_1 and R_2.

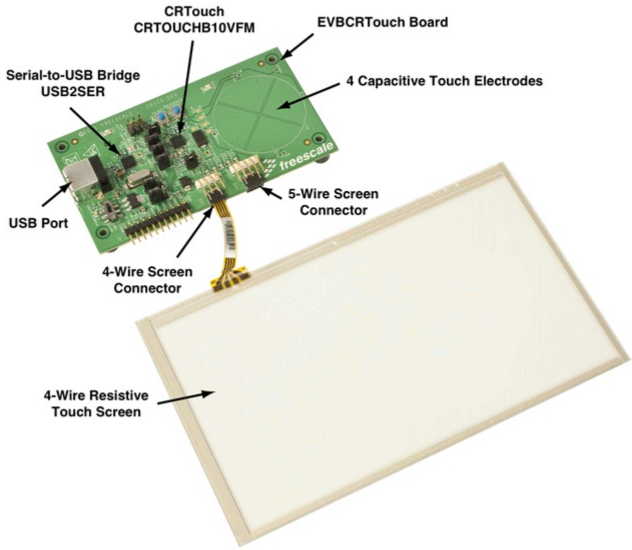


Figure 4.9: A “4-wire" resistive touch screen and accompanying electronics to read the screen. Note that there are four wires connected to the touchscreen. (Freescale)

4.4 Resistive Touch Screen

A commercial touch resistive screen touch screen is a device that can detect the position of a single “touch" anywhere on the screen (Fig. 4.9). One way to build this device is shown in Fig. 4.9, where thin layers of conductors, or bars, are arranged in orthogonal directions, setting up a grid, embedded in a thin sheet of flexible and transparent material. Four electrical wires are connected to the ends of the wire.

Each bar is just a thin film of material that is modeled as a rectangular cross-section conductor. Electrically it’s just a resistor, but we can model the conductive bars as a resistor string (Fig. 4.11a), where each resistor models one segment falling between adjacent y bars. Of course physically it’s just one resistor, but as we will see momentarily, it’s convenient to think of it as a resistor string. The ends of the bars are tied together using a highly conductive layer, where we apply a bias voltage.

Since each bar is a resistive string, the voltage grows linearly from the bottom of the bar to the top of the bar. At the bottom the voltage is zero, due to the ground reference connection, and at the top it’s just the supply voltage V_{bar}. For any point along the bar, we can compute the voltage by using the voltage divider formula (Fig. 4.11b)

$$V_y = V_{bar} \frac{R_{down}}{R_{down} + R_{up}} \tag{4.12}$$

where R_{down} is the total resistance from the point of interest to ground, and R_{up} is the total resistance from the point of interest to the supply connection. Due to the regular arrangement of the structure, for the i’th element we have

$$V_{y,i} = V_{bar} \frac{iR_{unit}}{iR_{unit} + (N-i)R_{unit}} \tag{4.13}$$

since there are N elements in total. Simplifying

$$V_{y,i} = V_{bar} \frac{i}{i + (N-i)} = V_{bar} \frac{i}{N} \tag{4.14}$$

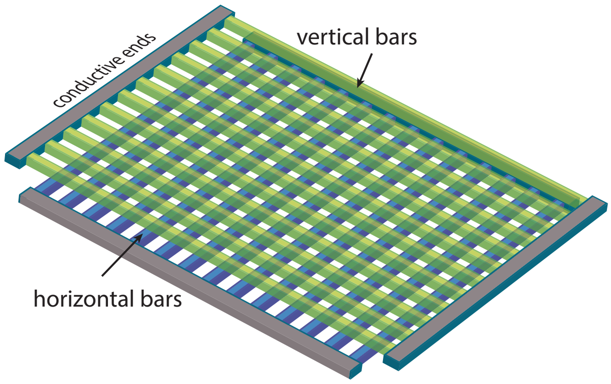
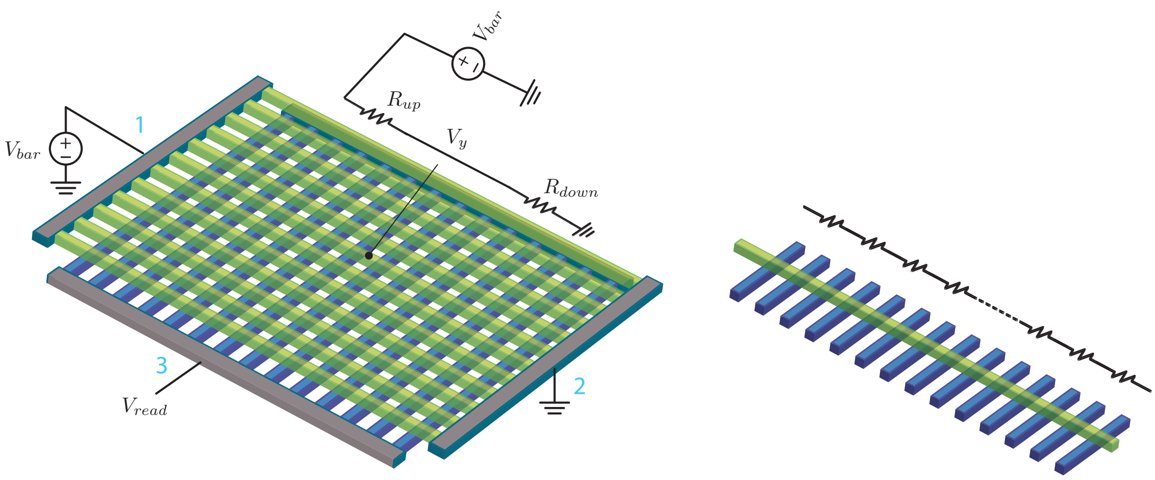


Figure 4.10: A touchscreen built from two flexible transparent layers embedded with bars of conductive material that are normally not in contact with each other. Pressing on the screen presses these layers together.



(a) (b)

Figure 4.11: (a) Measurement of position is accomplished by applying a voltage to one set of bars and reading from the second. (b) The resistive bars in the touchscreen can be modeled as a string of resistors, with each segment consisting of the resistance between two of the orthogonal bars.

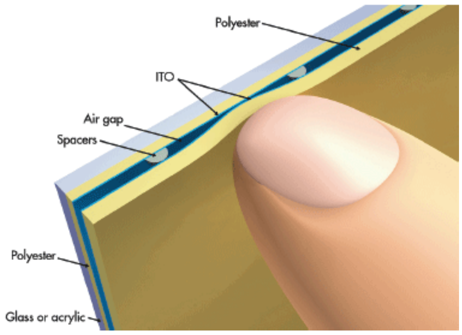


Figure 4.12: The physical pressure causes the two layers in the touch screen to come into contact. (electronicdesign.com)

As shown in Fig. 4.12, if we touch the screen at a particular location, the horizontal bar at that location will be pressed down and connect to the vertical bar, thereby forming a connection. By reading the voltage on the horizontal bars, we would know the y location of the touch. By using Eq. 4.14, the voltage we read can be normalized to a location

$$i = N\frac{V_{read,horz}}{V_{bar}} \tag{4.15}$$

In other words, we can locate the y coordinate of the touch with a resolution determined by the number of horizontal bars N. To detriment the horizontal location of the touch, though, requires more work, since we cannot in any way determine the x location as all horizontal bars are shorted at the ends and read the same voltage.

But if we simply rotate out setup, in other words we connect the horizontal bars to a voltage at one end and ground the other end, and then read the voltage with the vertical bars, we have

$$j = N\frac{V_{read,vert}}{V_{bar}} \tag{4.16}$$

which gives us the (x,y) coordinates

$$(x,y) = (i\delta x, j\delta y) \tag{4.17}$$

where δx and δy are the spacings between the horizontal and vertical segments. A complete setup for the measurement is shown in Fig. 4.13, where switches are used to alternatively bias one set of bars while the other set is used for reading. Two reads are required to narrow down the precise x and y coordinates.

Even though we used bars in the resistive touch screen, it's also clear that if a uniform sheet of a thin conductor were used instead, the functionality of the touchscreen would remain the same, as shown in Fig. 4.14. One simple and inexpensive way to build the touch screen is to use two conductive sheets separated by a spacer layer. Pressure on the top surface then causes the top and bottom layers to touch at a given location, and the setup we've described here can be used to read the x and y coordinates.

It's worth noting that this scheme does not work for multi-touch, or two more more presses on the screen. Multiple touches on the screen create new paths for current to flow and disrupts the uniform grid voltages. We will learn how to create a multi-touch screen when we learn about capacitance.

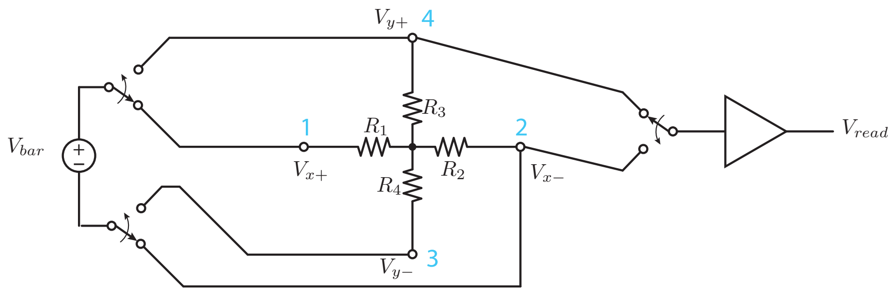


Figure 4.13: Schematic of a 4-wire resistive touchscreen (see Fig. 4.14). Two sets of switches are used to alternatively connected a voltage source across one set of conductors while the other set is used to read the voltage.

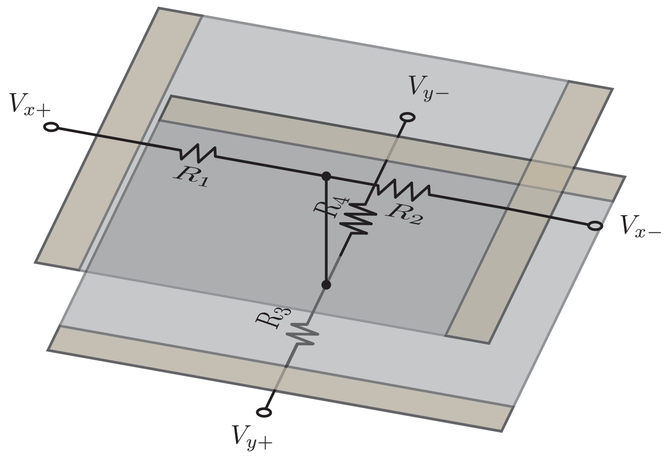


Figure 4.14: A resistive touch screen made from solid sheets of conductors and four terminals for drive and sense. Here a touch connection is shown with a vertical wire and solder dots. In reality, the sheets make physical contact through deformation of the surface.

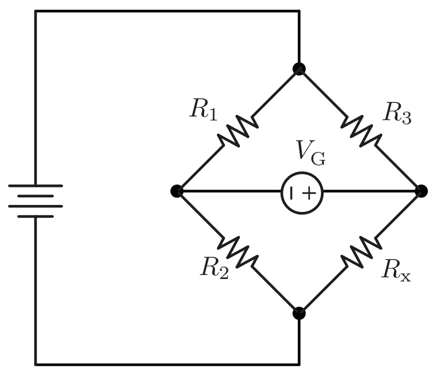


Figure 4.15: A Wheatstone Bridge is a circuit consisting of 3 known resistors, one adjustable, and one unknown resistor labeled R_x. The value of the adjustable resistor R_2 is varied until the circuit is balanced, or when the current through the galvanometer V_G is zero.

4.5 The Wheatstone Bridge

The Wheatstone Bridge, shown in Fig. 4.15, originally invented by Samuel Hunter Christie in 1833 and then popularized by Sir Charles Wheatstone in 1843, is used to measure an unknown resistance. It is highly accurate and only requires an adjustable resistor (or set of well known calibrated resistors) and a method of measuring zero current, such as a galvanometer.

The to the accuracy comes about because we only need to measure if the current is zero, which we can do this very precisely with a galvanometer. Note that if we could measure current accurately to begin with, we could just measure the ratio of voltage and current to determine the resistance. But in many applications, it's the change in resistance that we wish to detect, which can be extremely small, making the Wheatstone bridge the method of choice. Such applications include strain gauges, thermocouples, and other transducers.

The operation of the Wheatstone Bridge is as follows. One leg of the bridge contains an unknown resistance R_x which we would like to find. The other leg contains an adjustable resistor R_2 (of known value). The goal is to adjust the resistor R_2 until the circuit is "balanced", in other words until no current flows through the galvanometer.

Under the balanced condition, there is no current I_g, so the current in R_1 and R_2, say I_1, is the same as the current through R_3 and R_x, call it I_3. By KVL, under the balanced condition $V_g = 0$, we have

$$I_3 R_3 = I_1 R_1 \tag{4.18}$$

and

$$I_3 R_x = I_1 R_2 \tag{4.19}$$

Taking the ratio of these two currents, we have

$$\frac{R_3}{R_x} = \frac{R_1}{R_2} \tag{4.20}$$

Which means that the unknown resistance can be computed from the known resistors

$$R_x = \frac{R_3}{R_1} R_2 \tag{4.21}$$

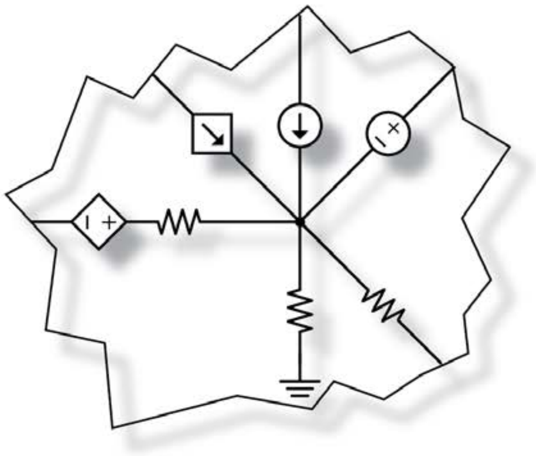
In practice, we vary R_2 until we achieve balance. In some commercial units, the scale factor R_3/R_1 can be changed as well.

Example 7: Suppose that $R_1 = 100\Omega$, R_2 can be adjusted from 1Ω steps from 0Ω to 100Ω, and R_3 can be selected to be 100Ω, $1\,\mathrm{k}\Omega$, $10\,\mathrm{k}\Omega$, or $100\,\mathrm{k}\Omega$. (a) If the bridge is balanced with $R_2 = 36\Omega$ and $R_3 = 10\,\mathrm{k}\Omega$, find R_x. (b) For $R_3 = 10\,\mathrm{k}\Omega$, what is the accuracy of the measurement? (c) What's the largest R_x that can be measured? This is a trivial application of the equations

$$R_x = \frac{R_3}{R_1} R_2 = \frac{10,000}{100} 36 = 3600\Omega$$

(b) The accuracy of the measurement is set by the scale factor. As R_2 changes by 1Ω, the value of R_x changes by $\frac{R_3}{R_1}$, which is equal to 100Ω. (c) The largest R_x is given by

$$R_{x,max} = \frac{R_{3,max}}{R_1} R_{2,max} = \frac{100,000}{100} 100 = 100\,\mathrm{k}\Omega$$



5. Analyzing a Complex Network

Simple circuit can be solved using repeated applications of series-parallel formulas and current/voltage dividers, see for example Fig. 5.1. In the next chapter we'll learn a few more tricks to add to our toolbox. For now we'd like to develop a systematic methodology to solve for the currents and voltages (and hence power) in any arbitrary circuit, such as the one shown in Fig. 5.2.

We'll primarily study nodal analysis, which is a repeated application of KCL to each node of the circuit. By doing this, we can be used to solve for all of the voltages (and subsequently currents) in the circuit. Two other approaches are called Mesh and Loop Analysis. These involve repeated application of KVL can be used to find the currents (and subsequently voltages) in a circuit. In this book we will not use mesh/loop analysis.

5.0.1 Identifying the Reference Node

Recall that voltage is defined as a quantity that measures the potential difference between two nodes in a circuit. We can arbitrarily pick one node of the circuit and define all *node voltages* in reference to this node (Fig. 5.3). Call this node *ground*, or node '0'. In other words, define V_k as the node voltage at node k which is the energy gained per unit charge as it moves from node *gnd* to node k, or in more cumbersome notation, $V_{k,gnd}$. If we subtract the two node voltages, we get

$$V_{j,gnd} - V_{k,gnd} = V_j - V_{gnd} - (V_k - V_{gnd}) = V_j - V_k = V_{j,k} \tag{5.1}$$

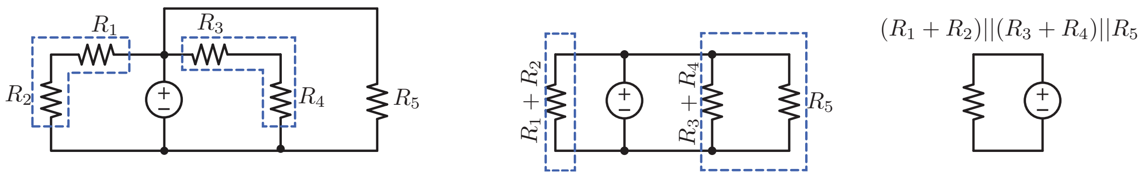


Figure 5.1: This seemingly complicated circuit is easily converted into a simple circuit by combining resistors in series and in shunt.

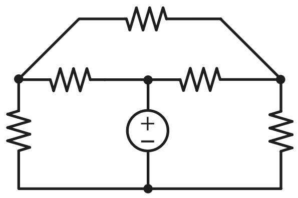


Figure 5.2: A circuit that cannot be easily simplified using series and parallel combination of elements, as none of the resistors are in series or in shunt.

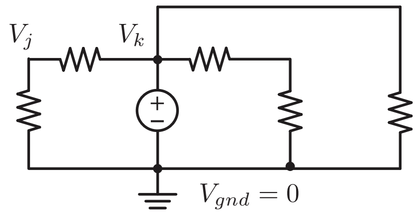


Figure 5.3: Any point in the circuit can be chosen as the reference node, usually denoted with a ground symbol. By definition, the voltage at this node is zero. This reduces the number of unknowns to $n-1$, where n is the number of nodes.

In other words $V_{j,k} = V_{j,gnd} - V_{k,gnd}$, which makes sense since they are defined with respect to the same reference. Note that the reference potential is by definition at zero potential, $V_{gnd} = 0$.

5.1 Nodal Analysis

5.1.1 Node Equations for Resistive Branches

Since the potential at the ground node is known, only $n-1$ unknown node voltages remain. Writing KCL at node j, shown in Fig. 5.4 results in an equation in the following form

$$I_1 + I_2 - I_3 + I_x = 0 \tag{5.2}$$

In a resistive branch, the current can be written in terms of the node voltages. If current I_x flows from node k to node j, it's given by

$$I_x = \frac{V_k - V_j}{R_x} = G_x(V_k - V_j) \tag{5.3}$$

If an independent current source is connected to the node, then the current is known.

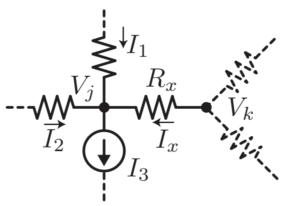


Figure 5.4: An illustrative node where KCL is applied. Note that the direction of current flow is important and sometimes arbitrarily chosen. If our choice is wrong, the resulting current will be negative to make a self-consistent solution.

Example 8:

We are now in a position to solve many circuit problems. For example, in the circuit shown in Fig. 5.5, we proceed as follows. First identify and number the nodes of the circuit.

Next we choose the reference ground potential at the negative terminal of the voltage source. With this choice, we eliminate two unknown node voltages from the circuit: the reference ground, and the positive terminal of the voltage source.

$$v_4 = 0\text{V}$$

$$v_3 = v_s$$

Now there are only two nodes left in the circuit, v_1 and v_2. For each node we write KCL equations. We setup two equations with two unknowns. At node 1 we have

$$\frac{v_1}{R_1} + \frac{v_1 - v_s}{R_2} + \frac{v_1 - v_2}{R_5} + (-i_s) = 0$$

Note that i_s flows into node 1, and since we summed currents out of this node, it is a negative term. At the node 2 we can similarly write

$$\frac{v_2}{R_4} + \frac{v_2 - v_s}{R_3} + \frac{v_2 - v_1}{R_5} + i_s = 0$$

At node 2, since i_s is flowing out of the node, its contribution is now positive.

We have two linear equations, and two unknowns, which we can readily solve.

$$\left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_5}\right) v_1 - \frac{1}{R_5} v_2 = \frac{1}{R_2} v_s + i_s$$

$$-\frac{1}{R_5} v_1 + \left(\frac{1}{R_3} + \frac{1}{R_4} + \frac{1}{R_5}\right) v_2 = \frac{1}{R_3} v_s - i_s$$

This equation can be put into the following matrix form

$$\begin{pmatrix} \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_5} & -\frac{1}{R_5} \\ -\frac{1}{R_5} & \frac{1}{R_3} + \frac{1}{R_4} + \frac{1}{R_5} \end{pmatrix} v = \begin{pmatrix} \frac{1}{R_2} v_s + i_s \\ \frac{1}{R_3} v_s - i_s \end{pmatrix}$$

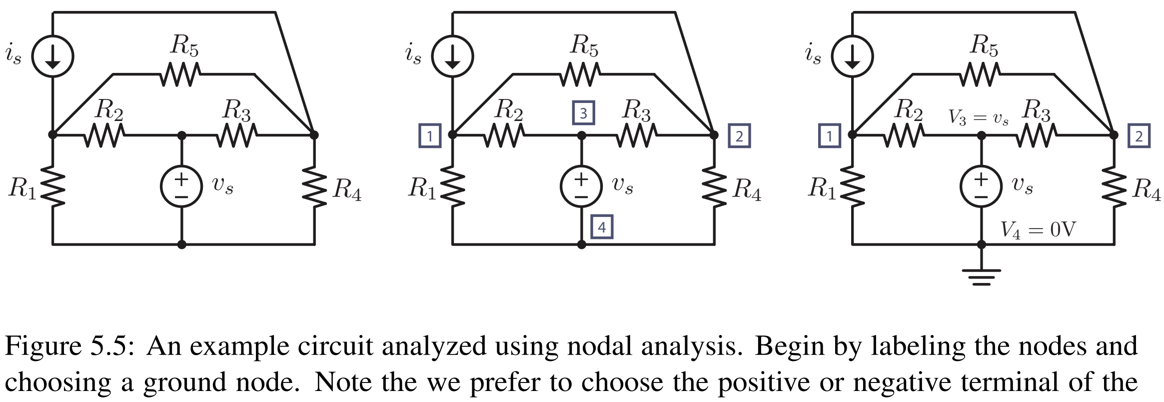


Figure 5.5: An example circuit analyzed using nodal analysis. Begin by labeling the nodes and choosing a ground node. Note the we prefer to choose the positive or negative terminal of the independent voltage source.

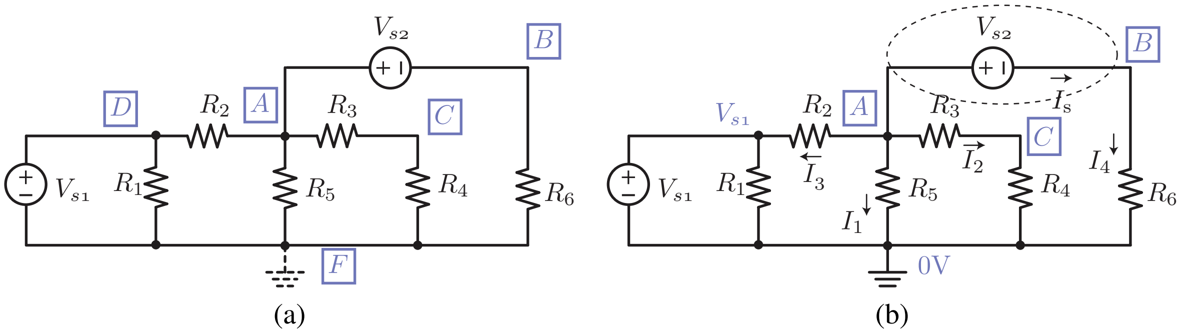


Figure 5.6: (a) A circuit with two independent voltage sources. Node F is arbitrarily chosen as the ground reference node. (b) With the choice of reference node, we now have three nodes to consider (A, B, C), but nodes A and B are actually pinned to a constant difference by the second voltage source V_{s2}. For this reason, we can treat the combined nodes (A,B) as a *super* node.

In the previous example, we saw that the KCL node equations result in a simple linear equation, which we can put in standard form. Put the unknowns on the left hand side and the known currents (due to the independent sources) on the right hand side. Then we have a matrix equation

$$Av = b \tag{5.4}$$

We can solve for the unknown node voltages by matrix inversion

$$v = A^{-1}b \tag{5.5}$$

5.1.2 Dealing with “Floating" Voltage Sources

In the previous example, there was only one independent voltage source, so we used a trick and labeled the negative terminal as ground. That eliminated two nodes immediately from our list of unknowns. Suppose there are two or more voltage sources, as shown in Fig. 5.6a.

To analyze this circuit, we again pick an arbitrary node as ground, and here you can choose either negative terminal and label the nodes as shown, resulting in three nodes (A,B,C). In this circuit, the actual number of unknowns is less than 3 since the voltage between A and B is pinned by the source to a constant difference of V_{s2}. In other words, if we know the voltage at A, we automatically know the voltage at B, since $V_{AB} = V_{s2}$.

Now if we try to apply KCL at one of the nodes, A or B, we encounter a problem. The current through the voltage source can take on any value, which means that other circuit elements determine the current through it.

5.1.3 KCL at a “Super Node"

Since the current of the voltage source V_s is unknown, let’s see if we can eliminate it. Write the KCL equations for A and B. With reference to Fig. 5.6b, at node A

$$I_3 + I_1 + I_2 + I_s = 0 \tag{5.6}$$

and similarly at node B

$$-I_s + I_4 = 0 \tag{5.7}$$

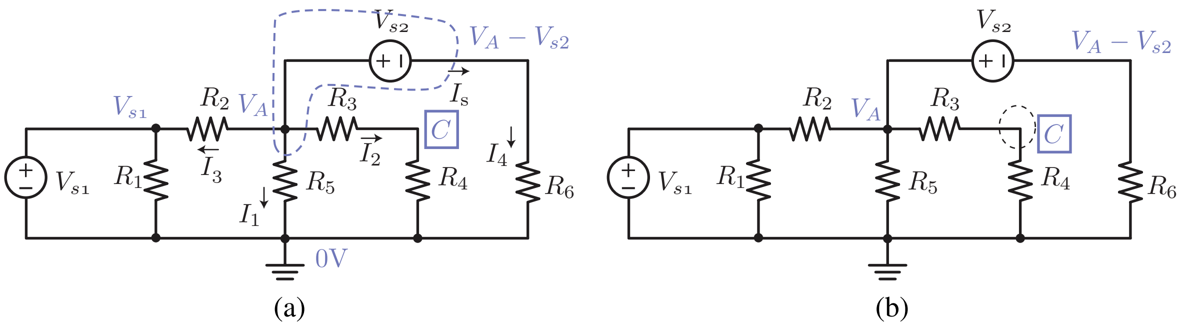


Figure 5.7: (a) Only two unknown voltage nodes remain in circuit when we identify nodes AB as a super node (shown with dotted lines). (b) Note that node C is a trivial node and it can be easily eliminated from the list of unknowns using a voltage divider relation.

Note that this unknown current I_s appears twice. If we simply add these two equations, I_s cancels out

$$I_1 + I_2 + I_3 + I_4 = 0 \tag{5.8}$$

This is an independent equation we can use. It's actually KCL for a "super node", which is what we call nodes A and B together. We have just shown that current continuity applies to a node and also to a super node. This is no surprise, because from charge conservation we know that if we draw a circle around any number of nodes in a circuit, KCL must apply.

Using the steps that we have learned, we can now see that the circuit is actually very easy to analyze since there are only two remaining unknown nodes in the circuit, super node AB and C, as shown in Fig. 5.7a.

5.1.4 Elimination of Trivial Nodes

Any node with less that three elements is in some sense trivial. That's because we can find the node voltage for such a node from the branch current. Thus, it's smart to avoid writing equations for these nodes and to deal with them later. In the above example the number of unknowns has been reduced from 2 to 1 by using this technique (see Fig. 5.7b). KCL at super node AB is then written by inspection

$$\frac{V_A - V_{s1}}{R_2} + \frac{V_A}{R_5} + \frac{V_A}{R_3 + R_4} + \frac{V_A - V_{s2}}{R_6} = 0 \tag{5.9}$$

The two key steps were to eliminate node C by just writing that the current leaving node A is $V_A/(R_3 + R_4)$ rather than $(V_A - V_C)/R_3$ and to note that the voltage at node B is just $V_A - V_{s2}$.

5.2 Nodal Analysis with Dependent Sources

Dependent sources require a bit more work but are generally dealt with in the same way. In the circuit shown in Fig. 5.8, we initially treat the current through the dependent source as an unknown in writing KCL at node k

$$\frac{V_k - V_j}{R_2} + \frac{V_k}{R_5} + \frac{V_k}{R_3 + R_4} - I_x = 0 \tag{5.10}$$

There are now more unknowns than equations. For each additional unknown (dependent current), we must write an additional equation which relates the dependent current to the node voltages. For

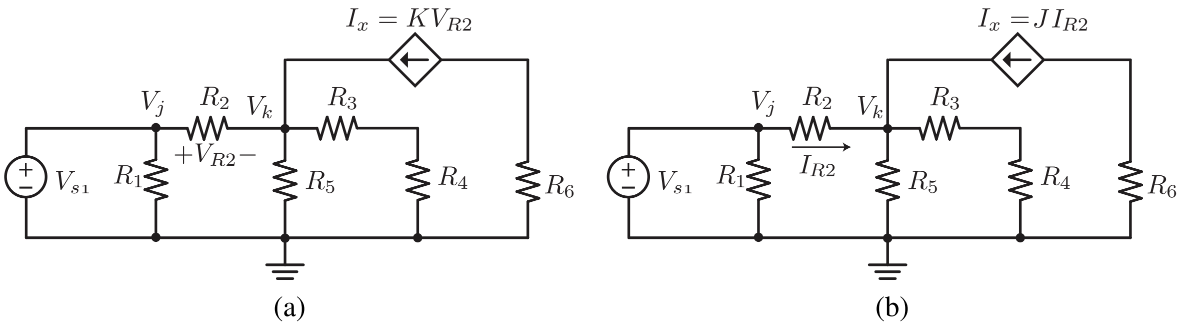


Figure 5.8: Circuits involving (a) voltage controlled current sources (VCCS) and (b) current controlled current sources (CCCS) are treated as ordinary current sources with the exception that an extra equation is required to write the unknown dependent current in terms of the node variables.

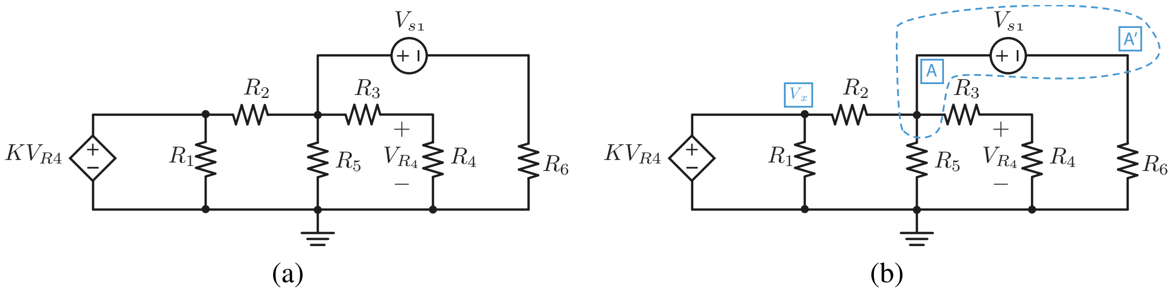


Figure 5.9: (a) A network with a dependent voltage source. (b) KCL can be performed at teh super node AA' while the dependent source is initially treated as a unknown source voltage. Next the voltage of the source is resolved in terms of the node voltages.

instance, for a voltage controlled current source (VCCS) shown in Fig. 5.8a, this is trivial since

$$I_x = KV_{R2} = K(V_j - V_k) \tag{5.11}$$

For a current-controlled current source (CCCS) shown in Fig. 5.8b, we need to calculate the current in terms of the node voltages

$$I_x = JI_{R2} = J(V_j - V_k)/R_2 \tag{5.12}$$

Once you get practice with this extra step, you will learn to write these equations directly into your KCL equations by inspection. In the beginning, though, it's good to write the extra equations explicitly.

Nodes with dependent voltage sources are handled in the same way, as shown in Fig. 5.9. They can also form super nodes and a KCL equation can be written for the super node, just like independent sources. The only difference is that the potential difference between the floating dependent voltage source is not known until all the equations are solved. Writing KCL at the supernode AA'

$$0 = (V_A - V_x)\frac{1}{R_2} + \frac{V_A}{R_5} + \frac{V_A}{R_3 + R_4} + \frac{V_A - V_{s1}}{R_6} \tag{5.13}$$

where V_x is the voltage of the dependent source. We can write this voltage in terms of V_A by noting that the voltage across R_4 is found from a voltage divider between R_4 and R_3

$$V_x = KV_{R_4} = K\frac{R_4}{R_3 + R_4}V_A \tag{5.14}$$

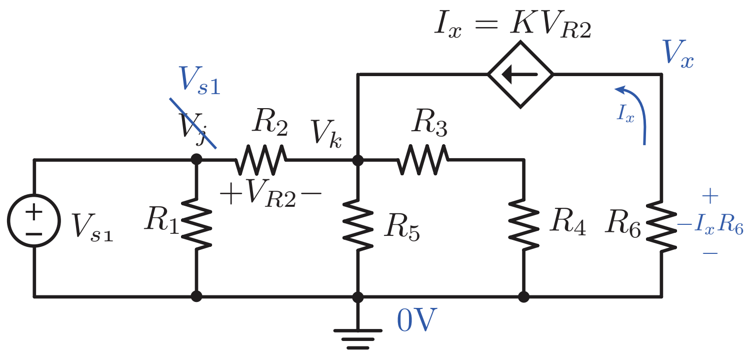


Figure 5.10: Analysis of a circuit with a dependent source. Note that node V_x is not needed since the current through branch R_6 is determined by the dependent current source.

Example 9:

For the circuit shown in Fig. 5.10, we start as always by defining the reference ground node, and then by writing node equations for all the remaining nodes. Initially we define 5 nodes, and quickly eliminate a few nodes. Since the node connected between R_3 and R_4 is trivial, we don't even count it. Also node V_j is pinned by a voltage source, so its voltage is known. Finally node V_x is effectively pinned by the dependent source, since the voltage across R_6 is simply given by the product of I_xR_6. So the only node standing is V_k. Writing KCL for this node we have

$$\frac{V_k - V_{s1}}{R_2} + \frac{V_k}{R_5} + \frac{V_k}{R_3 + R_4} + -I_x = 0 \tag{5.15}$$

The only wrinkle is that the dependent current source I_x, which introduces an extra unknown. To eliminate this extra unknown, we write additional equations expressing the unknown current in terms of the remaining node voltages.

$$I_x = KV_{R_2} = K(V_{s1} - V_k) \tag{5.16}$$

5.3 Nodal Analysis Summary

Nodal analysis is a general technique that can be used to solve any linear circuit. Even though we have used examples involving only resistors and sources, it can be generalized to include all linear elements, including inductors, capacitors, and transformers. The procedure is summarized as follows

1. Identify nodes of a circuit and label each one. Our goal is to find the nodal voltages.
2. If not explicitly defined, eliminate one node in the circuit by defining a reference node (0V). A good choice is the negative terminal of a voltage source since the positive terminal then is a known quantity.
3. Eliminate trivial nodes in the circuit. These nodes are either at a known potential (voltage sources) or nodes with only two elements connected to it.

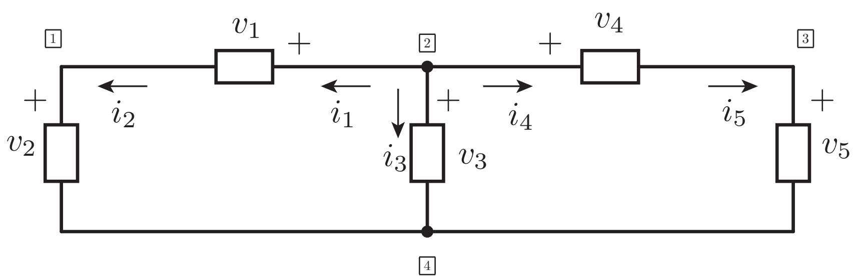


Figure 5.11: An example circuit containing two-terminal elements with branch voltages (v_k) and currents (i_k) defined as shown. The nodes of the network are labeled with boxed numbers.

4. Group nodes that are at a fixed potential difference due voltage sources between them. These are the so called super nodes. Circle these nodes.
5. Apply KCL to the remaining nodes and supernodes.
6. If there are any dependent sources in your circuit, resolve the dependent sources into an equation involving only the unknown node voltages of the circuit.

5.4 Systematic Nodal Analysis

We can convert our nodal analysis procedure into a systematic form that is easily converted into an algorithm that a computer can perform on large circuits. In fact, while by hand we may analyze a circuit with 3 or 4 elements, a computer can easily analyze circuits with millions of nodes.

Recall that the goal of nodal analysis is to find all the node voltages in a circuit, e_k, where k runs over all nodes in the circuit $(1 \cdots n)$. In practice we only need to find $n-1$ node voltages since one node is the reference or "ground" node.

For each branch in our circuit, let us assign a voltage v_k and branch current i_k (see Fig. 5.11). We can assign the direction of the branch voltage arbitrarily (or the positive terminal), but to keep consistent, we assign the current direction to flow into the positive point of the branch. In practice, the voltages and currents will be negative if our choice was "wrong".

5.4.1 Branch Incidence Matrix

From KCL, we immediately know that the branch currents satisfy the following equation at each node k

$$\sum_{j \in N_k} i_j = 0 \tag{5.17}$$

The above notation is unnecessarily confusing since there are two indices, one for each node k, and another for all branches j that intersect with this node. In fact, we can write the above as a matrix relation

$$A_b \vec{i} = 0 \tag{5.18}$$

where $\vec{i}$ is a column vector of branch currents and the matrix A_b is known as the branch incidence matrix. Notice that in general A_b is not square, since the rows of A_b correspond to the nodes of the circuit whereas the columns correspond to the branches, and in general there are as many branches, often more, than nodes. The matrix A_b is fairly sparse, because at any given node, only a small number of branches are incident. In fact, we can state that the matrix element a_{ij} is given by

$$a_{ij} = \begin{cases} +1 & j \text{ is positive incident} \\ -1 & j \text{ is negative incident} \\ 0 & \text{otherwise} \end{cases} \tag{5.19}$$

By "incident" we mean that a branch is connected at a particular node. If the positive terminal is connected, we assign $+1$, whereas if the negative terminal is connected, we assign -1.

Example 10: Consider the simple circuit shown in Fig. 5.11. The branch incidence matrix is constructed by inspection. We number the nodes and then for each node we can observe which branches are incident, placing a $+1$ or -1 in the correct column (corresponding to the branch element)

$$A = \underbrace{\begin{pmatrix} -1 & +1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & -1 & +1 \\ 0 & -1 & -1 & 0 & -1 \end{pmatrix}}_{\text{branch elements}}$$

For illustration, consider the second row of this matrix. Examining the circuit, we see that branch elements 1, 3, and 4 are incident on this node, and so for each element there's a corresponding 1 in the column position. In this particular example all elements have their positive terminal connected to the node, so all elements are positive. By contrast, in the last row of the matrix, another three branches are connected to the fourth node, but all with negative terminals, giving rise to -1s in the last row.

Multiplying this matrix out row by row, we recover KCL equations

$$\begin{aligned} -I_1 + I_2 &= 0 \\ I_1 + I_3 + I_4 &= 0 \\ -I_4 + I_5 &= 0 \\ -(I_2 + I_3 + I_5) &= 0 \end{aligned}$$

It's interesting to perform Gaussin elimination on this matrix to bring it into a triangular form. Adding the first and second row and replacing the second row gives

$$\begin{pmatrix} -1 & +1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & -1 & +1 \\ 0 & -1 & -1 & 0 & -1 \end{pmatrix} \to \begin{pmatrix} -1 & +1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & -1 & +1 \\ 0 & -1 & -1 & 0 & -1 \end{pmatrix}$$

Now adding the second row to the fourth

$$\to \begin{pmatrix} -1 & +1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & -1 & +1 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

And finally adding the third row to the fourth, we end up with a row of zeros

$$\rightarrow \begin{pmatrix} -1 & +1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & -1 & +1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

This demonstrates that the matrix has a null space, as it must since it satisfies KCL. It's also interesting to find the vectors that span the null space

$$\vec{i} = \begin{pmatrix} +1 \\ +1 \\ -1 \\ 0 \\ 0 \end{pmatrix} i_2 + \begin{pmatrix} 0 \\ 0 \\ -1 \\ +1 \\ +1 \end{pmatrix} i_4$$

This null space corresponds to current flowing in the meshes defined by $1 \rightarrow 2 \rightarrow 3$ and $3 \rightarrow 4 \rightarrow 5$. Note that in both cases the loop traverses element 3 in reverse from the defined polarity.

From the above example, we see that the branch incidence matrix has the property that the sum of each column is zero. This is because any given branch current can only be incident on two nodes, once with magnitude $+1$ and a second time -1, since it can only leave one node and return to another node. This immediately tells us that the matrix A is not full rank since a non-trivial linear combination of the rows yields the zero vector. Since row and column rank equal, we can say the column rank is also not full. From this we can conclude that the circuit admits a solution in its null space.

5.4.2 Branch Voltage

It is worth noting that KVL allows us to assign each node of the circuit a unique voltage e_k. There are multiple ways to reach the node k through different branch elements, but KVL guarantees that all paths sum to the same voltage potential, otherwise we would violate the conservation of energy. You could, for example, start at the ground node, make your way to node k through one path, and then return back to ground through another path. If the paths yielded different results for the voltage, one could extract energy from the loop by just traveling in the loop.

For each branch in the circuit, we can therefore find the voltage as the difference between two nodal voltages

$$v_j = e_m - e_n$$

where branch j is connected between nodes m and n. This is somewhat reminiscent of the branch incidence matrix, but in reverse. For each branch, we can write such an equation and if we gather all such equations into a matrix, we have

$$\vec{v} = B\vec{e} \tag{5.20}$$

where the matrix B has the following properties

$$b_{ij} = \begin{cases} +1 & i \text{ is positive incident} \\ -1 & i \text{ is negative incident} \\ 0 & \text{otherwise} \end{cases} \tag{5.21}$$

In other words, $b_{ij} = a_{ji}$, and so the matrix B is actually related to A_b by its transpose

$$B = A_b^T \tag{5.22}$$

This allows us to write two equations relating the branch currents

$$A_b \vec{i} = 0 \tag{5.23}$$

and branch voltages

$$v = A_b^T \vec{e} \tag{5.24}$$

The question is how do we connect the v and the i ? First we need to delete one row of the matrix A_b corresponding to the reference node. Now the matrix A is full rank[1].

5.4.3 Nodal Voltages

We know that for each branch element, there is an I-V relation between the current and voltage. For example, for a simple resistor, the branch current is given by

$$i_k = G_{kk} v_k$$

In fact, for any linear element, we can relate the branch current to the branch node voltages in a manner that admits the following linear relation

$$\vec{i} = G\vec{v} + \vec{i_s} \tag{5.25}$$

where $\vec{i_s}$ are the constant current sources and the matrix G is the branch conductance matrix. If we admit only resistors, G is a diagonal matrix with each diagonal entry corresponding to the branch conductance. On the other hand, if dependent sources are allowed, the matrix has non-diagonal entries. Here we have converted all voltage sources into current sources (Norton's Theorem) to simplify the equations. With this additional fact, we can write

$$A\vec{i} = 0 = AG\vec{v} + A\vec{i_s} \tag{5.26}$$

And now we substitute for the branch voltages

$$A\vec{i} = 0 = AGA^T\vec{e} + A\vec{i_s} \tag{5.27}$$

Re-arranging the above equation, we have

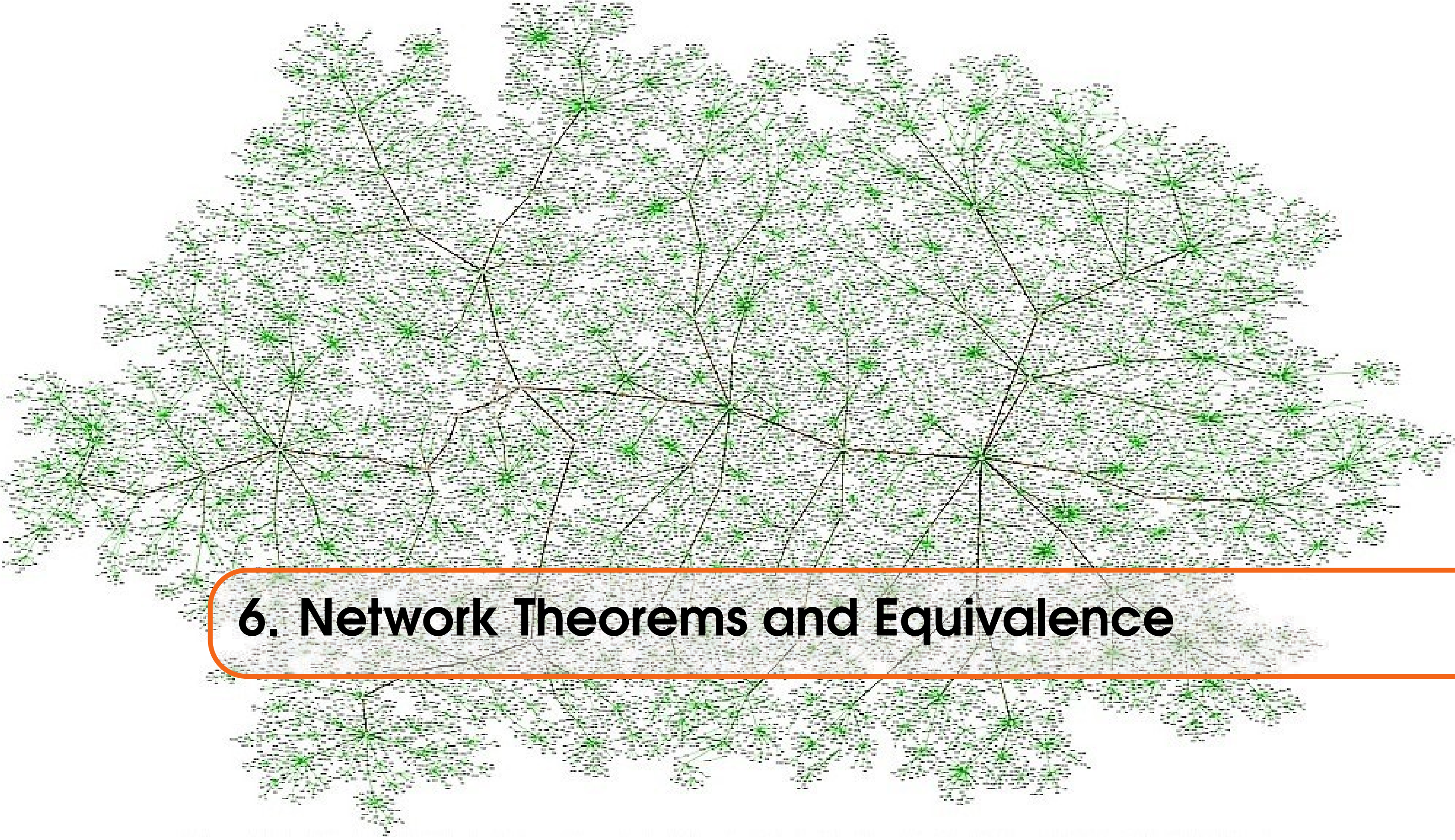
$$(AGA^T)\vec{e} = -A\vec{i_s} \tag{5.28}$$

If the matrix AGA^T is invertible, then we can solve for the node voltages

$$\vec{e} = (AGA^T)^{-1}(-A)\vec{i_s} \tag{5.29}$$

Although we did not prove it, the matrix A with the reference node removed is full rank, and if the matrix G is diagonal (no dependent sources), the solution always exists and is unique.

[1] You can prove this by invoking KCL.



6. Network Theorems and Equivalence

Even though nodal analysis can be applied to any circuit to yield the node voltages, and therefore the branch currents, it's often unnecessary to resort to full blown nodal analysis when the circuit can be analyzed by simpler means, such as a voltage divider or by applying simplifications of series / parallel formulas. We'll introduce a few new tools that we can apply to many circuits, which are applicable to linear circuits in particular, such as the concept of superposition, which allows us to analyze the circuit behavior one source at a time. Even if these techniques are not "faster" to apply, they often result in more insight into the behavior of the circuit, which helps us to understand how to *design* circuits, rather than just analyze them.

The other big idea in this chapter is the concept of equivalence, or the ability to reduce a complex circuit with dozens of elements by only two elements. We have already met this idea in simpler form, such as the fact that a single resistor can model the combination of two more resistors in series. The circuit solution is the same regardless of how we treat these resistors, as an equivalent unit or as a series of individual resistors. As we shall see, this concept generalizes to any arbitrary linear circuit.

6.1 Non-Linear Components and Dependent Sources

Before we begin our journey, let's be clear on what we mean by linear circuits, which is a circuit consisting of only linear elements. A non-linear resistor has a non-linear I-V relation,

$$V = r_1 I + r_2 I^2 + r_3 I^3 \tag{6.1}$$

or

$$V = \cos(I \cdot R_x) \tag{6.2}$$

Or more generally, any non-linear function

$$I = f(V) \tag{6.3}$$

or

$$V = g(V) \tag{6.4}$$

A non-linear dependent source is a non-linear function of one or more independent currents/voltages in the circuit. Some examples of non-linear dependent sources are below. For example, a resistor that depends on the current squared:

$$v_k = K i_j^2 \tag{6.5}$$

Or a dependent source depends on two source, and one in a non-linear fashion (squarer)

$$i_k = K i_j^2 + M i_j \tag{6.6}$$

A dependent source that depends on the product of two voltages, or a multiplier

$$v_k = K v_j \cdot v_m \tag{6.7}$$

6.1.1 Linear Resistor or Conductor

By definition, a linear element means that if two inputs are applied together, then the output of the sum is the sum of the individual outputs. Suppose for example

$$v = Ri \tag{6.8}$$

Then if we take the inputs one at a time, we have

$$v_1 = Ri_1 \tag{6.9}$$

$$v_2 = Ri_2 \tag{6.10}$$

Now the sum of the inputs is applied to the element, and we have

$$v = R(i_1 + i_2) = Ri_1 + Ri_2 = v_1 + v_2 \tag{6.11}$$

Note that a linear resistor/conductor must cross the origin, otherwise it would be a source in series with resistor

$$v = Ri + v_s \tag{6.12}$$

and furthermore it does not satisfy the linearity constraint

$$v = R(i_1 + i_2) + v_s \neq v_1 + v_2 = R(i_1 + i_2) + 2v_s \tag{6.13}$$

6.2 Superposition

If a circuit is linear, then by the principal of superposition, we can analyze the circuit one source at a time. The total response is the sum of the outputs due to the individual sources. This is clear if we re-write the matrix equation as follows

$$Ax = b = b_1 + b_2 + \cdots \tag{6.14}$$

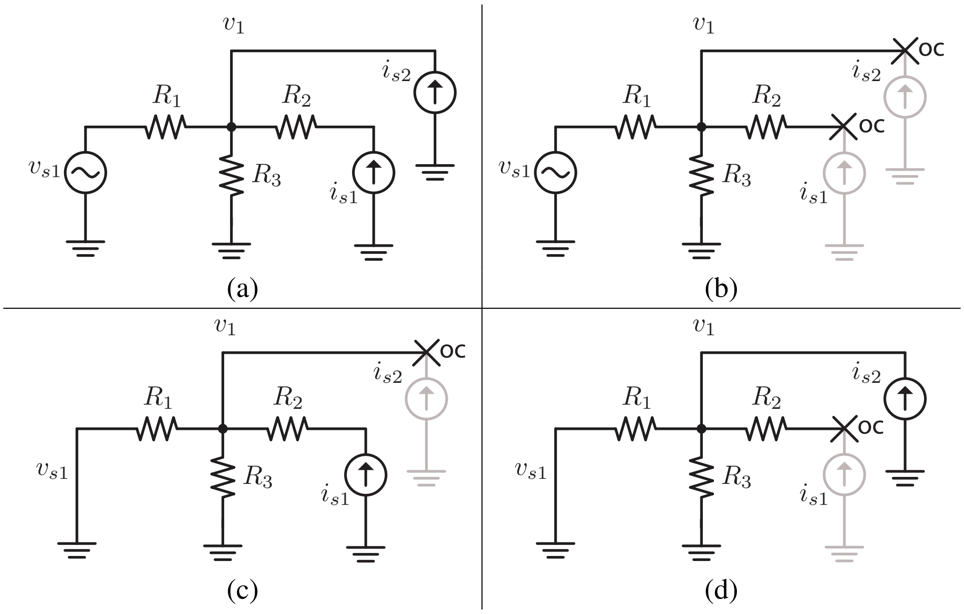


Figure 6.1: (a) A circuit involving multiple sources. Analysis of the circuit using superposition when all sources are zeroed except (b) v_{s1}, (c) i_{s1}, and (d) i_{s2}.

Note that we have partitioned the source terms so that each b_k only contains a single source. Clearly, the solution is given by

$$x = A^{-1}b_1 + A^{-1}b_2 + \cdots = x_1 + x_2 + \cdots \tag{6.15}$$

where x_k is the solution with source k turned on and all other sources set to zero. That means that other voltage sources are short-circuited (zero voltage) and other current sources are open-circuited (zero current). Often analyzing a circuit one source at a time gives us more insight into the behavior of a complex circuit.

Example 11: Analyzing a Circuit with Superposition

In the example shown in Fig. 6.1, there are three independent sources. When we analyze the circuit source by source, the circuit is often simple enough that we can solve the equations directly by inspection. First turn off i_{s1} and i_{s2}. Zero current means that we replace these sources with open circuits. The node voltage v_1 is therefore by inspection

$$v_1^{v_{s1}} = \frac{R_3}{R_1 + R_3} v_{s1} \tag{6.16}$$

Next turn off v_{s1} (short-circuit) and i_{s2} (open-circuit). The current i_{s1} will therefore divide between R_3 and R_1 and establish a voltage at node v_1 (equivalently, it see's a parallel combination of R_1 and R_3

$$v_1^{i_{s1}} = \frac{R_3 R_1}{R_1 + R_3} i_{s1} \tag{6.17}$$

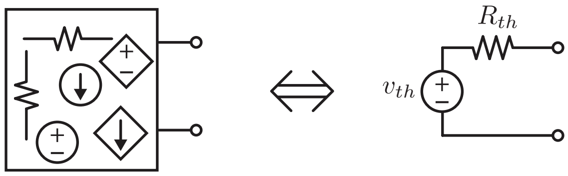


Figure 6.2: An arbitrary circuit (left) built from linear components can be represented by an equivalent circuit involving only a voltage source and a series resistor (right).

Finally, we turn off all sources except i_{s2}. Now only R_1 and R_3 remain (R_2 is dangling)

$$v_1^{i_{s2}} = \frac{R_3 R_1}{R_1 + R_3} i_{s1} \tag{6.18}$$

By superposition, the node voltage v_1 is the sum of the three node voltages due to each source

$$v_1 = v_1^{i_{s1}} + v_1^{i_{s2}} + v_1^{v_{s1}} = \frac{R_3}{R_1 + R_3}(v_{s1} + R_1(i_{s1} + i_{s2})) \tag{6.19}$$

We can verify the solution by performing KCL directly at center node

$$(v_1 - v_{s1})G_1 + v_1 G_3 - i_{s1} - i_{s2} = 0 \tag{6.20}$$

or

$$v_1(G_1 + G_3) = v_{s1}G_1 + i_{s1} + i_{s2} \tag{6.21}$$

The answer here is just as fast, and perhaps even easier, but we don't develop any intuition about the operation of the circuit.

6.3 Thevenin Equivalent Circuit

A powerful theorem in circuit analysis is the Thevenin equivalent theorem, which let us replace a very complex circuit with a simple equivalent circuit model. In the circuit shown in Fig. 6.2 there can be countless resistors, voltage sources (independent and dependent), current sources (independent and dependent), and yet the *terminal* behavior of the circuit is captured by just two elements.

How can this be? This follows from linearity, as all the resistors are linear (follow Ohm's Law) and all dependent sources are also linear, we would expect that the terminal behavior should be relatively simple and given by a “line", or two points. The equivalent circuit representation is often called a “black box", since the details of the circuitry are hidden.

6.3.1 Thevenin Derivation

Since a circuit is linear, then no matter how complicated it is, its response to a stimulus at some terminal pair must be linear. It can therefore be represented by a linear equivalent resistor and a fixed constant voltage source due to the presence of independent sources in the circuit.

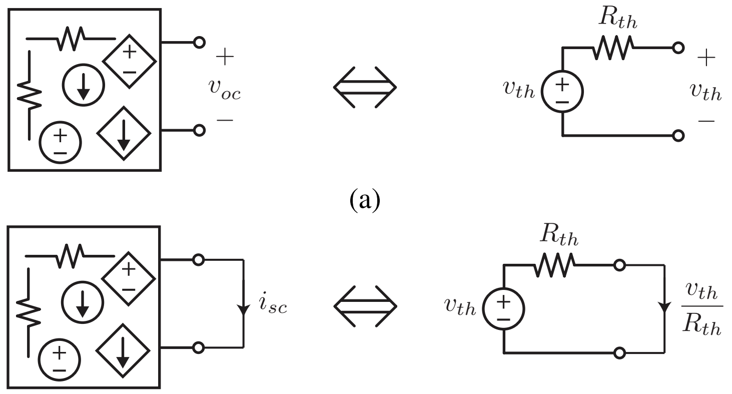


Figure 6.3: For equivalence to hold, (a) the open-circuit behavior of the Thevenin equivalent circuit must be the same as the original circuit. (b) Likewise, the short-circuit current of the original circuit must equate the value from the Thevin equilatent model.

To find the equivalent source value, called the Thevenin voltage source v_{th}, simply observe that the open-circuit voltage of both the “black box" and the original circuit must equal (Fig. 6.3a), which means

$$v_{th} = v_{oc} \tag{6.22}$$

In other words, open-circuit the original circuit, find its equivalent output voltage at the terminals of interest, and that’s v_{th}

6.3.2 Thevenin Source Resistance

To find equivalent Thevenin source resistance R_{th}, notice that in order for the terminal behavior of the two circuits to match, the current flow into a load resistor has to be the same for any load value. In particular, take the load as a short circuit, as shown in Fig. 6.3b. The output current of the Thevenin equivalent under a short circuit is given by v_{th}/R_{th}. Equating this to the short-circuit current of the original circuitry, we have

$$i_{sc} = \frac{v_{th}}{R_{th}} \tag{6.23}$$

$$R_{th} = \frac{v_{th}}{i_{sc}} = \frac{v_{oc}}{i_{sc}} \tag{6.24}$$

Example 12: Thevenin Equivalent Example

In the circuit shown in Fig. 6.4, we will calculate the Thevenin equivalent circuit. We begin by finding the open-circuit voltage. In this case, it’s a simple application of the voltage divider.

$$v_3 = v_2 \frac{R_4}{R_3 + R_4} \tag{6.25}$$

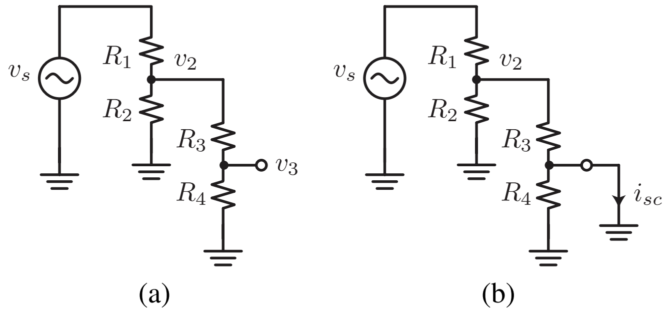


Figure 6.4: (a) Application of Thevenin's theorem to an example circuit at node 3 (referenced to ground). Open-circuit voltage at port three and (c) short-circuit current.

$$v_2 = \frac{R_2||(R_3+R_4)}{R_1+R_2||(R_3+R_4)} v_s \tag{6.26}$$

$$v_{oc} = v_3 = \frac{R_4}{R_3+R_4} \frac{R_2||(R_3+R_4)}{R_1+R_2||(R_3+R_4)} v_s \tag{6.27}$$

$$= \frac{R_4 R_2 (R_3+R_4)}{(R_3+R_4)(R_1(R_2+R_3+R_4)+R_2(R_3+R_4))} v_s \tag{6.28}$$

Next we find the short-circuit current in the original circuit, as shown in Fig. 6.4b. The resistance loading the source under this condition is given by

$$i_{sc} = i_s \frac{R_2}{R_2+R_3} \tag{6.29}$$

$$i_s = \frac{v_s}{R_1+(R_2||R_3)} = \frac{v_s(R_2+R_3)}{(R_2+R_3)R_1+R_2R_3} \tag{6.30}$$

$$i_{sc} = \frac{R_2 v_s}{R_1R_2+R_1R_3+R_2R_3} \tag{6.31}$$

$$R_{th} = \frac{v_{oc}}{i_{sc}} \tag{6.32}$$

Figure 6.5: (a) A circuit with no dependent sources (left) and independent sources zero'ed out. (b) An arbitrary circuit with all independent sources zero'ed out and probed by an external current or voltage source (right).

6.3.3 Calculating R_{th}

If there are no dependent sources in the circuit under consideration, as shown in Fig. 6.5a, then you can calculate the Thevenin equivalent circuit *directly* by setting the independent source values to zero. As before, zeroing out sources means shorting voltage sources (zero voltage) and open-circuiting current sources (zero current). Now just "inspect" to find the equivalent Thevenin resistance.

For the general linear circuit, another approach to find R_{th} is to probe the circuit with an independent voltage/ current source while zeroing out all internal sources. The current / voltage is monitored and the ratio of the test voltage to test current is the equivalent R_{th} as shown in Fig. 6.5b

$$R_{th} = \frac{v_x}{i_x} \tag{6.33}$$

Example 13: By The Direct Method

Let's redo the same example but now we zero out the voltage source and redraw the circuit, as shown in Fig. 6.6. Now we can readily find R_{th} by simply observing that the resistors of the original circuit are in series/parallel:

$$R_{th} = R_4||(R_3 + R_1||R_2) \tag{6.34}$$

$$= \frac{R_4R_3 + R_4R_1||R_2}{R_4 + R_3 + R_1||R_2} \tag{6.35}$$

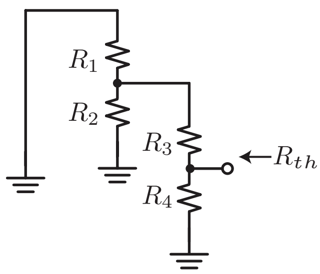


Figure 6.6: The circuit of Fig 6.4 redrawn by zeroing out the sources and by directly observing the equivalent resistance.

$$= \frac{R_4R_3(R_1+R_2)+R_4R_1R_2}{(R_4+R_3)(R_1+R_2)+R_1R_2} \quad (6.36)$$

Example 14:

Consider the circuit shown in Fig. 6.7a. We begin by finding the equivalent open circuit voltage by writing KCL at the intermediate node

$$\frac{v_1-v_{s1}}{R_1}+\frac{v_1}{R_2}+\frac{v_1-V_{OC}}{R_3}=0 \quad (6.37)$$

The last term can be simplified because the current flowing through R_3 under open-circuit conditions is just given by the current of the generator

$$\frac{v_1-v_{s1}}{R_1}+\frac{v_1}{R_2}-Gv_1=0 \quad (6.38)$$

Solving the above equation for v_1, the only unknown, we have

$$v_1(G_1+G_2-G)=v_{s1}G_1 \quad (6.39)$$

$$v_1=\frac{G_1}{G_1+G_2-G}v_{s1} \quad (6.40)$$

The open-circuit voltage, which is the Thevenin equivalent generator, is given by

$$V_{th}=V_{oc}=v_1+Gv_1R_3=\frac{G_1(1+GR_3)}{G_1+G_2-G}v_{s1} \quad (6.41)$$

Now we move to the second part of the solution, or the Thevenin equivalent resistance. To illustrate the probing technique, we zero out all sources and connect a voltage probe,

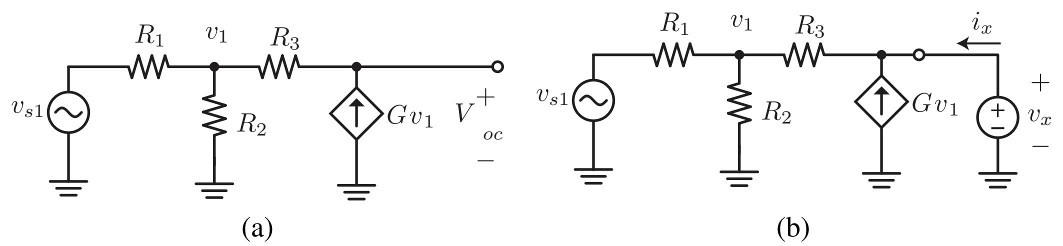


Figure 6.7: (a) An example circuit involving dependent and independent sources will be reduced to a Thevenin equivalent circuit using KCL to find the open-circuit voltage. (b) We zero out all independent sources and probe the terminals to determine the equivalent resistance.

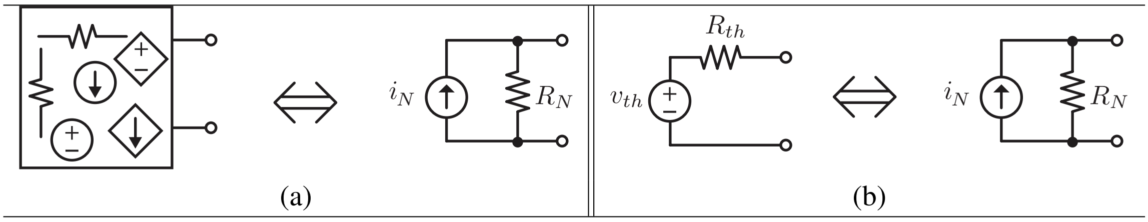


Figure 6.8: (a) The Norton equivalent representation of an arbitrary linear circuit. (b) Applying a Norton / Thevenin transform allows us to transform a voltage source to a current source.

as shown in Fig. 6.7b. Note that dependent sources remain, only independent sources are zero'ed out. In this case, the voltage v_{s1} is replaced with a short-circuit.

Since the voltage at the terminal is pinned by the external probe at v_x, we can find v_1 by inspection by noting that it's just a voltage divider

$$v_1 = v_x \frac{R_1||R_2}{R_1||R_2 + R_3} \tag{6.42}$$

To find the input current i_x, write KCL at the input terminal

$$i_x + Gv_1 + (v_1 - v_x)G_3 = 0 \tag{6.43}$$

or

$$i_x = -(G+G_3)v_1 + G_3 v_x = -(G+G_3)\frac{R_1||R_2}{R_1||R_2 + R_3} v_x + G_3 v_x \tag{6.44}$$

This allows us to find the Thevenin equivalent conductance as

$$G_{th} = \frac{1}{R_{th}} = \frac{i_x}{v_x} = -(G+G_3)\frac{R_1||R_2}{R_1||R_2 + R_3} + G_3 \tag{6.45}$$

6.4 Norton Equivalent Circuit

We can also replace a complex circuit with a Norton equivalent circuit, which contains a current source and a shunt source resistance, as shown in Fig. 6.8a. To find the equivalent source value, we find the short circuit current for both the model and the original circuit and note that

$$i_n = i_{sc} \tag{6.46}$$

Similarly, to find the Norton resistance, we note that if we open-circuit the model, the output voltage is given by

$$v_{oc} = i_n R_n \tag{6.47}$$

This is the same exact equation as before.

R Note that this result is a consequence of circuit *duality*, or the fact that voltages / currents can be interchanged as long as resistors are converted to conductors, and series / parallel connections are interchanged.

6.4.1 Source Transformations

A trivial but often useful application of Norton or Thevenin's Theorem shows us that we can transform from one representation to the other (Fig. 6.8b). For instance, starting from the Thevenin, let's find the Norton. Short-circuit the Thevenin to find

$$i_n = i_{sc} = \frac{V_{th}}{R_{th}} \tag{6.48}$$

This shows that a source can be treated equally as a voltage or current. In practice, when the source resistance is "low", we say the source acts like a good voltage source, meaning that its output voltage does not vary much as we vary the load current. This of course depends on the range of the load, in other words the source is a good voltage as long as $R_L \gg R_{th}$.

We can make the dual argument and say that the source acts like a good current source when its Norton equivalent resistance is large compared to the load, $R_N \gg R_L$. In this scenario, as we vary the load resistance, the load current remains more or less constant, since most of the Norton source current I_N flows into the much smaller load.

Example 15: Norton Equivalent Example

For the circuit shown in Fig. 6.9a, let us find the Norton equivalent circuit. We first start out by finding the short-circuit current, which is in fact the Norton current, as shown in Fig. 6.9b

$$v_1 G_3 + v_1 G_2 + (v_1 - v_s) G_1 - i_{s1} = 0 \tag{6.49}$$

Collecting terms

$$v_1 (G_1 + G_2 + G_3) = G_1 v_s + i_{s1} \tag{6.50}$$

Simplifying

$$v_1 = \frac{G_1}{G_1 + G_2 + G_3} v_s + \frac{1}{G_1 + G_2 + G_3} i_{s1} \tag{6.51}$$

From here the Norton current is easy

$$i_N = i_{sc} = G_3 v_1 = \frac{G_1 G_3}{G_1 + G_2 + G_3} v_s + \frac{G_3}{G_1 + G_2 + G_3} i_{s1} \tag{6.52}$$

To find the Norton equivalent resistance, we can exploit the fact that all sources are independent. If we turn off all the sources, the circuit just is very simple to analyze, as shown in Fig. 6.9c

$$R_N = R_3 + R_1 || R_2 \tag{6.53}$$

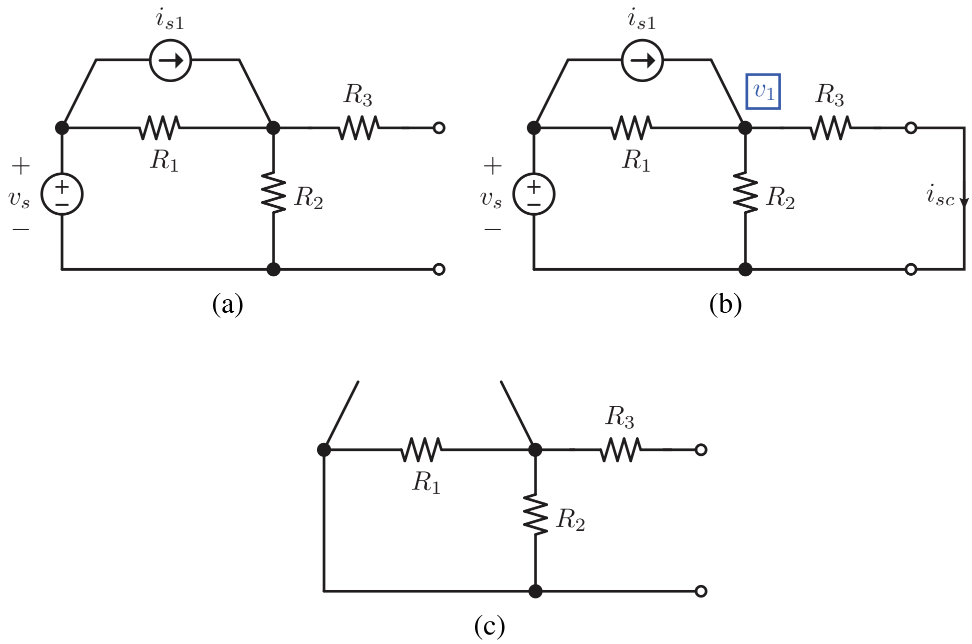


Figure 6.9: (a) Example circuit for calculation of the Norton equivalent circuit. (b) The short-circuit current is derived by applying KCL. (c) Turning off all sources, the circuit simplifies to an easily digestible form.

6.5 Maximum Power Transfer

An important question arises in many electrical circuits when we wish to interface one component to another while maximizing the power transfer to the second component. A good example is a battery with internal resistance and a motor. What's the best "load" resistance to choose in order to maximize the power transfer? Interestingly, if we maximize the current or voltage transfer, the power transfer is exactly zero.

No matter how complicated the black box source, we can represent the source as a Thevenin equivalent circuit, v_{th} and R_{th}, as shown in Fig. 6.10, connected to a load, represented by R_L. The general procedure is to find the power through the load and then to find the optimal load value.

$$P_L = I_L^2 R_L = \left(\frac{v_{th}}{R_L + R_{th}}\right)^2 R_L \tag{6.54}$$

We can take the derivative of the load power with respect to the load resistance and set it equal to zero (occurs at only a maximum or minimum).

$$\frac{dP_L}{dR_L} = \left(\frac{v_{th}}{R_L + R_{th}}\right)^2 - 2R_L v_{th}^2 \left(\frac{1}{R_L + R_{th}}\right)^3 = 0 \tag{6.55}$$

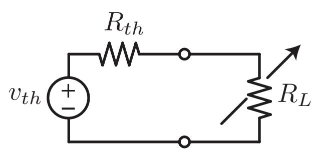


Figure 6.10: The power transfer between two circuits, something with a source and a "load", can be analyzed using a Thevenin equivalent circut.

Or

$$(R_L + R_{th}) = 2R_L \rightarrow R_L = R_{th} \tag{6.56}$$

A second derivative test confirms that it's a peak. This means that maximum power transfer happens when we match the load resistance to the source resistance.

Example 16:

A common example occurs when designing amplifiers in RF applications. For the receiver, shown in Fig. 6.11, we wish to extract the maximum possible power from the antenna (since the received signal can be very weak). We can represent the antenna by its Thevenin equivalent voltage source and resistance R_{th}. The load resistance, which is the input resistance of the amplifier in this case, should present a value that is *matched* to the antenna impedance. Note that we have used the Thevenin equivalent theorem twice, once to simply the source, and next to simplify the load. Since the amplifier does not have any independent sources inside of it, we can represent it as a simple resistor.[1]

---

[1]Later you will learn that an amplifier needs a supply voltage, which means that this statement is not strictly true. But it will work out to be the same since in the AC equivalent circuit of the amplifier, the power supply source will be effectively zero, since it does not vary.

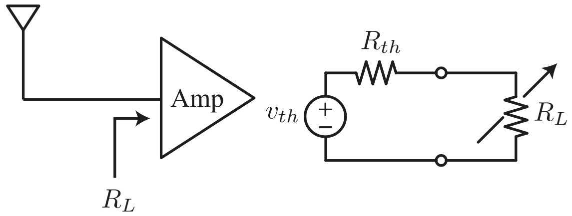


Figure 6.11: An amplifier is connected to an antenna. In this scenario, resistance matching (or more generally impedance matching), is used to maximize the power transfer from the antenna to the amplifier, represented as an equivalent load resistance.

7. Capacitance and Capacitors

7.1 Thought Experiment

Imagine a current source connected to an ideal short circuit as shown in Fig. 7.1a. The waveform for the current is a constant current I_0 for a time T, so a total charge of $Q = I_0 \cdot T$ circulates around but no net work is done since $v(t) = 0$ (short circuit).

Now imagine that we cut the wire and leave a gap between the conductors, as shown Fig. 7.1b. Now let's repeat the experiment by applying the same current waveform. What happens? The current flow is interrupted but the same amount of charge Q leaves the positive terminals of the current source. Where does it go? And how does the voltage across the gap change?

7.1.1 Capacitor Charge

In addition to the positive charge Q leaving the positive terminal, the same amount of charge enters the negative terminal. Equivalently, a charge of $-Q$ leaves the negative terminal and goes into the conductors.

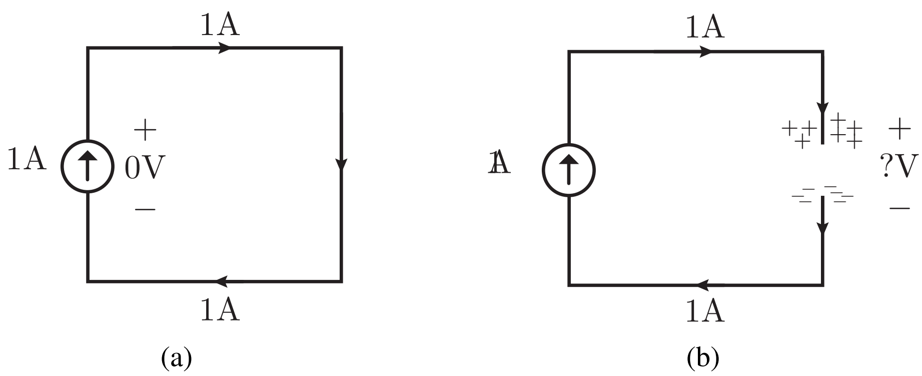


Figure 7.1: (a) A current source is driving a constant current through a wire. (b) The wire is cut and charges begin to accumulate on the wire.

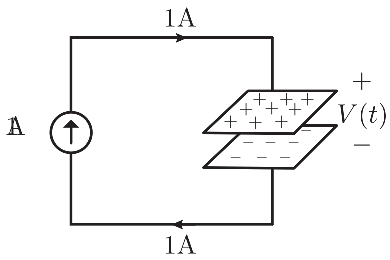


Figure 7.2: Two plates are used to "store" charge accumulating due to the current from the current source. Due to the large surface area of the plates, charges spread out and as a result experience less repulsive forces. Likewise, charges of opposite sign are in near proximity on the nearby bottom plate.

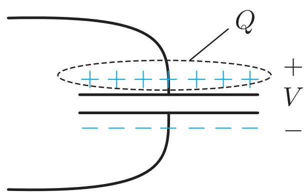


Figure 7.3: The total charge on the capacitor plates.

We see that the charge cannot go anywhere but into the conductors, and therefore the charge is stored there. Since like charges repel, we have to do work to force the charges to accumulate on the conductors. In fact, the smaller the conductor, the more work that we have to do, as the density of charge is higher. In Fig. 7.2, two plates have been inserted into the gap. Due to the large surface area of the plates, the charges can spread out and thus the current source does less work to move a given amount of charge onto the plates compared to the circuit with just wires.

7.1.2 Capacitor Voltage

By definition, the potential V across the capacitor represents the amount of work required to move a unit of charge onto the capacitor plates. This is the work done by the current source.

For linear media, we observe that as we push more charge onto the capacitor with a fixed current, its voltage increases linearly because it's more and more difficult to do it (like charges repel), or

$$V \propto Q \tag{7.1}$$

As shown in Fig. 7.3, the charge is defined as the total charge Q on the top plate, which is equal and opposite in sign to the charge on the bottom plate

$$Q_{top} = -Q_{bot} \tag{7.2}$$

$$Q = |Q_{top}| = |Q_{bot}| \tag{7.3}$$

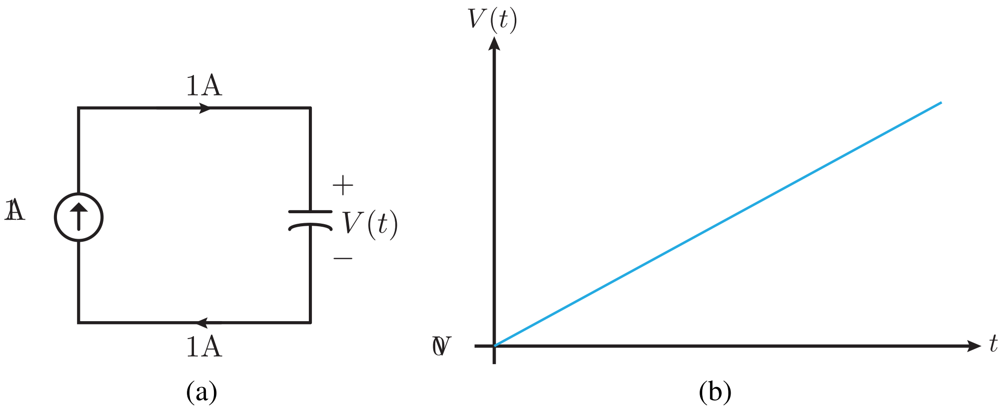


Figure 7.4: (a) A DC current will charge a capacitor with a fixed amount of charge per unit time, so (b) the voltage across the terminals increase linearly.

7.2 Definition of Capacitance

The proportionality constant between the charge and the voltage is defined as the capacitance of the two terminal element

$$q = Cv \tag{7.4}$$

The units of capacitance are given by charge over voltage, or Farads (in honor of Michael Faraday)

$$[C] = \frac{[q]}{[v]} = \frac{\mathrm{C}}{\mathrm{V}} = \mathrm{F} \tag{7.5}$$

We expect that a physically larger conductor should have a larger capacitance $C_1 > C_2$, because it has more surface area for the charges to reside. The average distance between the like charges determines how much energy you have to provide to push additional charges onto the capacitor.

The symbol for a capacitor is shown in Fig. 7.4. Sometimes a $+$ label indicates that a capacitor should only be charged in a given direction. Most capacitors, though, are symmetric and positive or negative charge can be applied to either terminals. A variable capacitor is known as a varactor.

Example 17: Let's return to our thought experiment at the start of the chapter. We can represent the scenario as shown in Fig. 7.4a, where we now recognize that C represents the capacitance of the wires. Every real circuit has some capacitance, even a very small amount, as charges accumulate on the wires and field lines begin to emerge from the positive wires to the negative terminal wires. Since the current is constant, the charge increases linearity, and therefore so should the voltage, as shown in Fig. 7.4b.

7.2.1 Capacitor Analogy

Imagine a tank of water where we pump water into the tank from the bottom (Fig. 7.5a). As we initially pump water, there is no water in the tank and it takes virtually no work. But as the tank

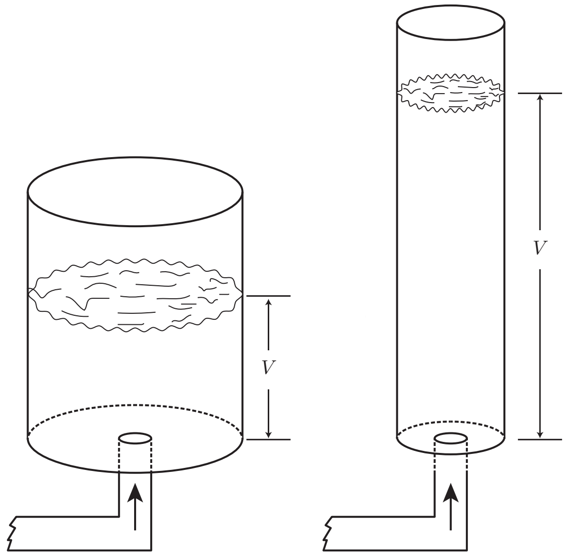


Figure 7.5: (a) The water analogy for a capacitor is a tank pumped from the bottom. (b) The water level (voltage) in a tank with a smaller cross-sectional area will rise more quickly for the same amount of water pumped in, requiring more work to be done by the pump (current source).

fills up, it takes more and more work since the liquid obtains gravitational potential energy, and the act of pumping water in raises the water level (voltage). For the same water pumped, a smaller tank requires more work (it has less capacity) because the liquid column gets higher and higher (Fig. 7.5b). Notice that we can always recover the work by emptying the tank and recovering the stored potential energy.

7.2.2 Capacitor Energy

The incremental of amount of work done to move a charge dq onto the plates of the capacitor is given by

$$dE = vdq \tag{7.6}$$

where v is the potential energy of the capacitor in a given state. Since $q = Cv$, we have $dq = Cdv$, or

$$dE = Cvdv \tag{7.7}$$

If we now integrate from zero potential (no charge) to some final voltage

$$E = \int_0^{V_0} Cvdv = C\frac{v^2}{2}\bigg|_0^{V_0} = \frac{1}{2}CV_0^2 \tag{7.8}$$

This is the energy stored in the capacitor. Just like the water tank, it's stored as potential energy that we can later recover.

7.2.3 Field Lines

Since we get charge separation in a capacitor, we expect that field lines emanate from the positive charges to the negative charges (see Fig. 7.6). The energy of the capacitor is in fact stored in these field lines.

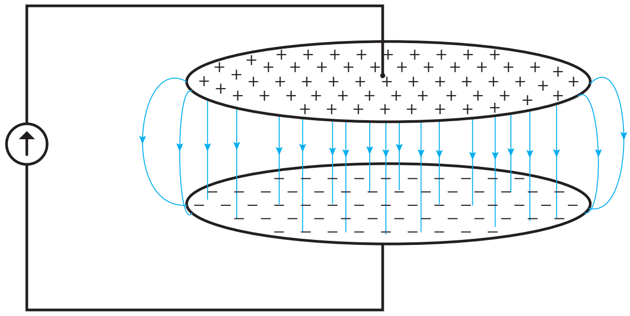


Figure 7.6: The field lines in a capacitor originate from positive charges and terminate on negative charges.

That means that there are two competing charge related force mechanisms in a capacitor. Like charges are forced to reside on the same plate, which requires energy. On the other hand, unlike charges are placed in close proximity, which have attractive forces. So we can see that the charges should bunch up as close as possible to the charges of opposite sign.

The smaller the gap spacing, the more capacity we have in a capacitor, because now the "unhappy" feelings of being cramped up to similar charge is somewhat alleviated by the "happy" feeling due to close proximity of unlike charges. We can therefore see that when we increase the plate width, the voltage drops (Fig. 7.7a), or equivalently the capacitance increases. Similarly, when we bring the plates closer together, the capacitance also increases (Fig. 7.7b).

7.3 Parallel Plate Capacitor

From basic physics, it's easy to show that the capacitance of a parallel plate structure is given by

$$C = \frac{\varepsilon A}{d} \tag{7.9}$$

where A is the plate area, ε is the permittivity of the dielectric (also called the dielectric constant), and d is the gap spacing. The permittivity of free space is $\varepsilon_0 = 8.854 \times 10^{-12}\,\mathrm{F/m}$. Most materials have a higher permittivity which is captured by the unitless relative permittivity $\varepsilon_r = \varepsilon/\varepsilon_0$, with typical materials $\varepsilon \sim 1 - 10$. For instance, air is mostly empty space and so $\varepsilon_r \approx 1$.

7.3.1 Dielectric Constant

Some materials, such as water, have polar molecules that align when an electric field is applied. Thus the dielectric constant is very large. As shown in Fig. 7.8, the plates of a capacitor are separated by a non-conducting medium composed of atoms and molecules that form dipole moments in the presence of the field generated by the charges on the plates. This distortion means that there's a net charge density generated by the dipoles at the top and bottom plate, as shown in Fig. 7.9. It's important to note that inside the medium, the charges cancel out (uniform material assumption), whereas at the discontinuity points, particularly at the boundary of the material, the charge appears and *cancels* the charges on the plates, thus lowering the voltage between the plates. For this reason, we see that the dipole moments tend to increase the capacitance.

7.3.2 Practical Capacitors

Real capacitors are made of large sheets of conductors (to maximize surface area) and thin dielectric layers (to minimize the gap), as shown in Fig. 7.10a. A multi-layer sandwich structure can then

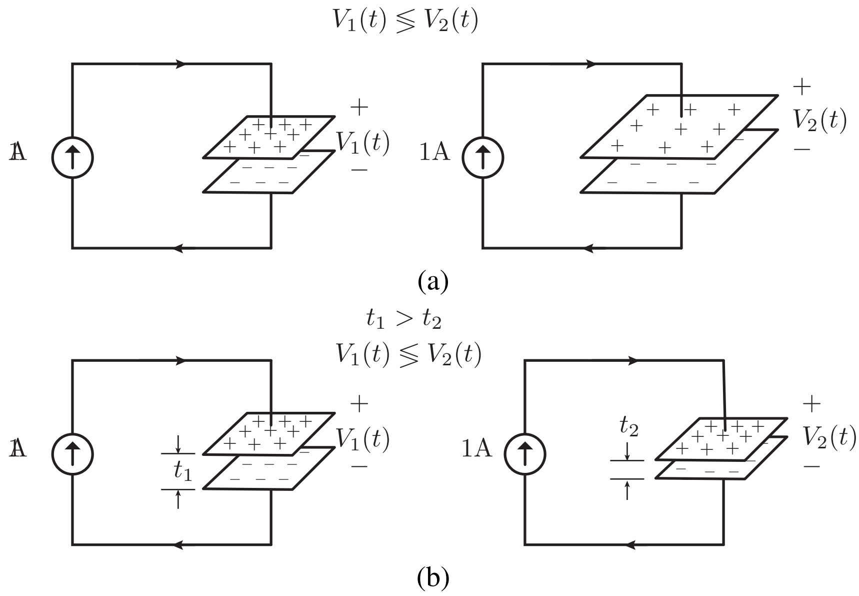


Figure 7.7: When an equal current flows into two capacitors, the voltage will be larger (a) for a smaller width plate and (b) for a larger separation between the plates.

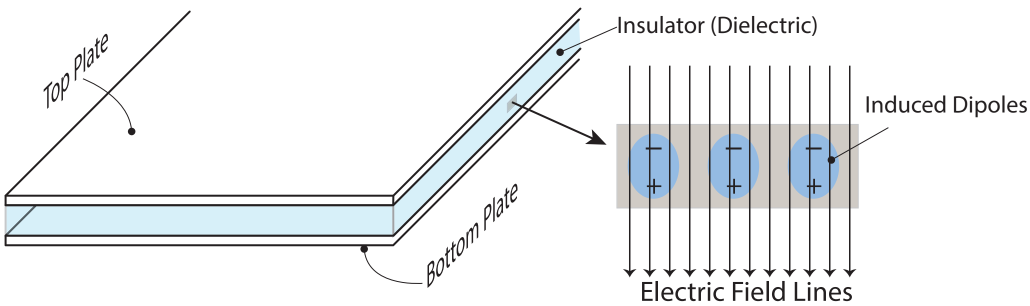


Figure 7.8: In a capacitors, the insulating medium between the plates is known as a dielectric material. Even though it's charge neutral, there's a small charge separation due to the action of the fields setup by the charges on the positive and negative plates.

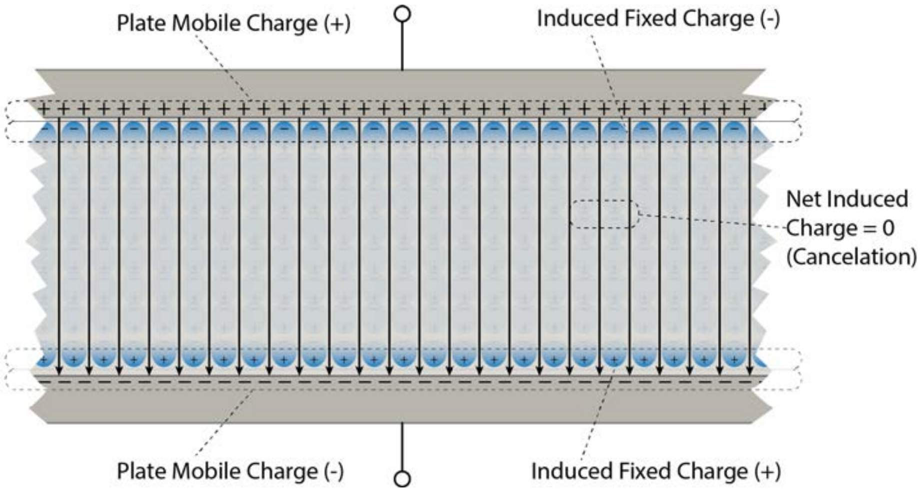


Figure 7.9: A pictorial representation of the induced dipole moments inside of the dielectric. Due to the uniform distribution of matter, the internal dipole moments cancel out but at the boundary of the material a net charge is subtracted from the plate charge.

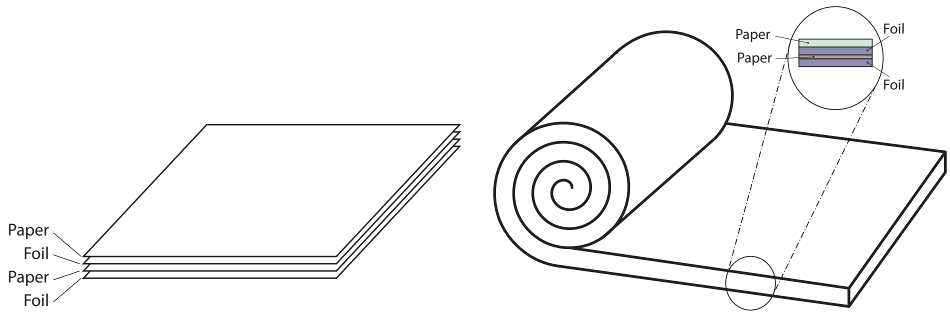


Figure 7.10: Practical capacitors are often cylindrical in shape and consist of plates of flexible and thin conductors wound into a cylinder.

be wrapped together to form a large capacitor, as shown in 7.10b. In integrated circuits or layered PCB modules, thin dielectrics and/or multiple metal fingers in close proximity form a high density capacitor. Super-capacitors use nanoparticles to boost the capacitance even higher.

7.4 Capacitor I-V Relations

7.4.1 Capacitor Current

A steady DC current into a capacitor means that the voltage ramps linearly. In theory it would ramp to infinity but in practice dielectric breakdown (arcing) would short out the capacitor.

Now consider a time-varying current, shown in Fig. 7.11. Since a steady positive current increases the voltage linearity, a steady negative current does the opposite. Thus if we pass a time varying current that has positive and negative polarity, the voltage across the capacitor would increase and decrease. For a fixed current, we can infer the voltage. What about a general relation?

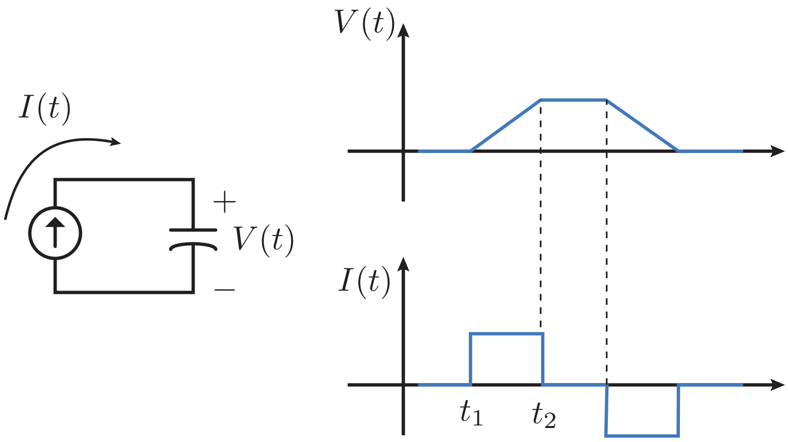


Figure 7.11: A time-varying current flows into a capacitor and the resulting voltage varies in magnitude as shown.

7.4.2 Capacitor Voltage

Since charge flow in time results in current, the relationship between the current and voltage in a capacitor is easily derived

$$q = Cv \tag{7.10}$$

$$i = \frac{dq}{dt} = C\frac{dv}{dt} \tag{7.11}$$

Recall that for a resistor we had $i = Gv$. The capacitor has a "conductance" which is proportional to it's capacitance but it also depends on how quickly the voltage is varied. If a fixed voltage is applied to a capacitor, no current flows.[1] When a slowly varying voltage is applied, the current flow is slow as well since only enough charge needs to flow to change the potential. When the potential is varied very quickly, the current flow is larger because it has to keep up with the voltage changes. Only changes in the voltage require new charge to flow.

We can re-write the capacitor "I-V" relationship as

$$\frac{dv}{dt} = \frac{i}{C} \tag{7.12}$$

If we integrate the above relation, we have

$$\int_{t_0}^{t} \frac{dv}{d\tau} d\tau = \int_{t_0}^{t} \frac{i}{C} d\tau \tag{7.13}$$

$$v(t) = v(t_0) + \frac{1}{C}\int_{t_0}^{t} i(\tau) d\tau \tag{7.14}$$

The integral of current over time is nothing but the *net charge* flow into the capacitor

$$Q = \int_{t_0}^{t} i(\tau) d\tau \tag{7.15}$$

[1] If a capacitor is initially uncharged, or charged to a different voltage, then in theory an enormous (infinite) current would instantaneously flow to charge the capacitor to the right voltage

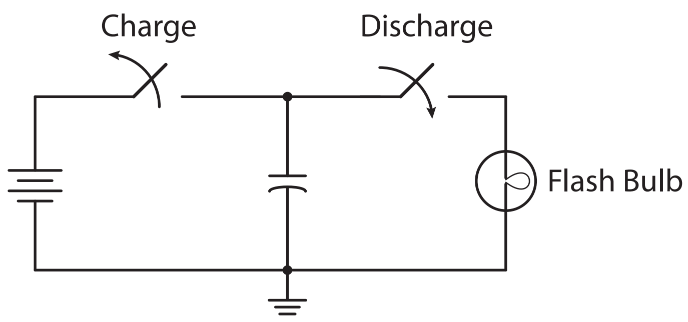


Figure 7.12: The circuit schematic of a flash circuit. The first switch is engaged to charge the flash capacitor. The second switch is used to quickly discharge the capacitor into the flash bulb.

Let's break this down term-by-term

$$\underbrace{v(t)}_{\text{voltage now}} = \underbrace{v(t_0)}_{\text{voltage then}} + \underbrace{Q/C}_{\text{net charge flow over capacitance}} \tag{7.16}$$

Example 18: Consider the circuit shown in Fig. 7.12. The first switch (closest to the supply voltage) is initially closed, charging the capacitor to the same DC voltage. Note that this charge time is very fast if the battery has small internal resistance. Next the first switch is opened and the second switch is closed. The energy stored in the capacitor is therefore released as charges are now allowed to flow from the top plate to the bottom plate through the flash bulb. If the flash bulb has very low resistance, a large current will quickly discharge and cause a flash of light to appear.

7.4.3 Sinusoidal Drive

If we drive the capacitor with an AC sinusoidal voltage, we can calculate the current

$$v(t) = V_0 \cos \omega_0 t \tag{7.17}$$

$$i(t) = C\frac{dv}{dt} = -CV_0\omega_0 \sin \omega_0 t \tag{7.18}$$

Current is out of phase with the voltage. It is exactly 90° out of phase, which is very important. The power flow onto the capacitor is given by

$$p(t) = v(t)i(t) = -CV_0^2\omega_0 \cos \omega_0 t \sin \omega_0 t \tag{7.19}$$

$$= -\frac{CV_0^2}{2}\omega_0 \sin 2\omega_0 t \tag{7.20}$$

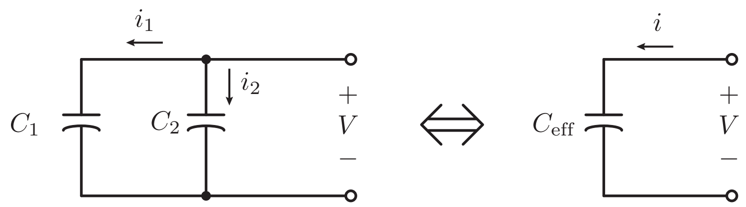


Figure 7.13: Parallel capacitors act like an equivalent larger capacitor $C_1 + C_2$.

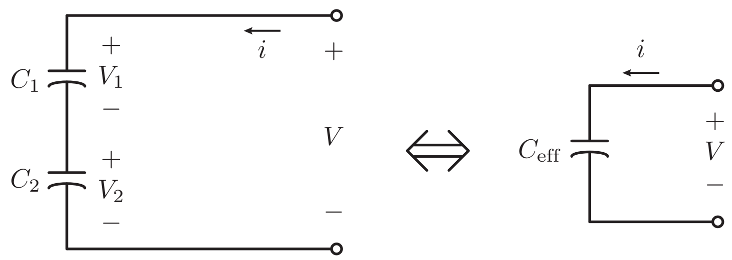


Figure 7.14: Series capacitors act like an equivalent smaller capacitor $C_1||C_2$.

The instantaneous power flow switches signs twice per cycle, as in each cycle energy is first delivered onto the capacitor, but then the energy is returned back to the source. In other words, there is no net energy flow into the capacitor

$$\int_0^T p(\tau)d\tau = -\frac{CV_0^2}{2}\omega_0 \int_0^T \sin 2\omega_0 \tau d\tau \equiv 0 \tag{7.21}$$

7.5 Circuits with Capacitors

7.5.1 Shunt Connection

When capacitors are connected in parallel, they have the same voltage (Fig. 7.13). The total current drawn from the supply is given by

$$i = i_1 + i_2 = C_1 \frac{dv}{dt} + C_2 \frac{dv}{dt} \tag{7.22}$$

$$i = (C_1 + C_2)\frac{dv}{dt} = C_{eff}\frac{dv}{dt} \tag{7.23}$$

In other words, two parallel capacitors act like an effectively larger capacitor $C_{eff} = C_1 + C_2$. This is easily generalized to N capacitors in parallel. This makes sense intuitively since the plates of C_1 and C_2 are both accessible to the charges, and so the capacitance should increase by exactly the increase in the plate area.

7.5.2 Series Connection

When capacitors are placed in series, the same current flows through both capacitors whereas the voltage drop is shared across the two (Fig. 7.14)

$$i = C_1 \frac{dv_1}{dt} = C_2 \frac{dv_2}{dt} \tag{7.24}$$

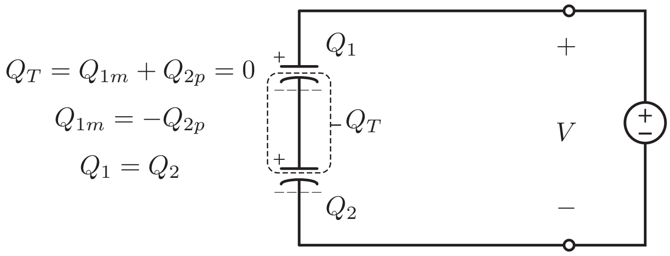


Figure 7.15: In series connected capacitors, the charge stored in each capacitor must be the same. Here Q_{1m} and Q_{2p} represent the charge stored on the negative terminal of C_1 and positive terminal of C_2 respectively.

$$v(t) = v_1(t) + v_2(t) \tag{7.25}$$

$$\frac{dv}{dt} = \frac{dv_1}{dt} + \frac{dv_1}{dt} \tag{7.26}$$

$$= \frac{1}{C_1} i + \frac{1}{C_2} i \tag{7.27}$$

Re-arranging the above equation

$$i = \frac{1}{\frac{1}{C_1} + \frac{1}{C_2}} \frac{dv}{dt} \tag{7.28}$$

If we treat "||" as an operator, then for capacitors in series we have

$$C_{eff} = C_1 || C_2 \tag{7.29}$$

Unlike the case of parallel capacitors, the series case takes more thinking to understand. First putting capacitors in series lowers the overall capacitance. Also, mathematically from the || operator we know that the smaller capacitor dominates. Why ? That's because the smaller capacitor has a higher voltage drop across it. Since the total voltage is the sum of the two voltage drops, it's always worse than just a single capacitor. But do the capacitor both charge up with the same polarity ? How is the charge distributed on each capacitor ?

To answer these questions, the other key observation is that both capacitors have equal charge storage (see Fig. 7.15). This follows from conservation of charge at the "floating" node. If a charge of $+Q_1$ appears on the top plate of C_1, then a charge of $-Q_1$ appears on the bottom plate. By the same token, a charge of $+Q_2$ and $-Q_2$ appears on the plates of the second capacitors.

But the top plate and bottom plate of the second and first capacitors are shorted together. Also, if the system is initially uncharged ($v = 0$), then this node is at ground potential and there is no net charge at this node. This means that

$$-Q_1 + Q_2 \equiv 0 \tag{7.30}$$

Or in other words, $Q_1 = Q_2$. So if bot capacitors hold the same amount of charge, the one with the lower capacitance will have a higher voltage drop.

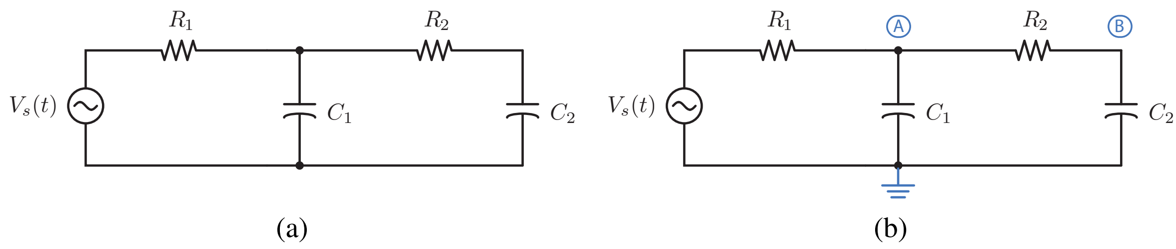


Figure 7.16: (a) A circuit involving capacitors and resistance can be analyzed with (b) nodal analysis.

7.6 KCL with Capacitors

We can continue to write KCL and KVL equations with capacitors. Unfortunately, instead of linear equations we are now dealing with linear differential equations. Later on we'll see that these differential equations can be converted to simple algebraic equations in the sinusoidal steady-state analysis or by using a Laplace Transform.

Example 19: Consider the circuit shown in Fig. 7.16a. First we choose a reference node and label the unknown node voltages A and B as shown in Fig. 7.16b. Since KCL must apply at each node of the circuit, we can use the I-V relationship for a capacitor to write the following KCL equations. At node A

$$\frac{V_A - V_s}{R_1} + \frac{V_A - V_B}{R_2} + i_{C1} = 0$$

but

$$i_{C1} = C_1 \frac{dV_A}{dt}$$

so we have

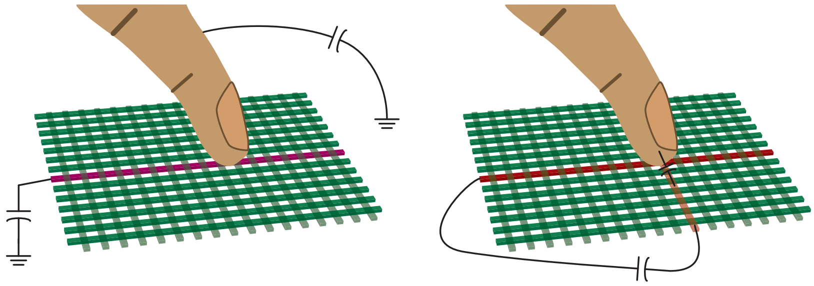
$$\frac{V_A - V_s}{R_1} + \frac{V_A - V_B}{R_2} + C_1 \frac{dV_A}{dt} = 0$$

Similarly, at node B

$$\frac{V_B - V_A}{R_2} + C_1 \frac{dV_B}{dt} = 0$$

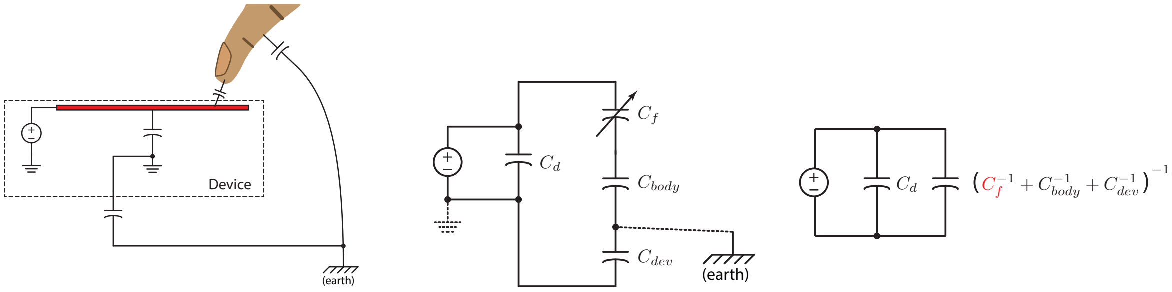
7.7 Application: Capacitive Touch Screen

As we have seen in Chapter 3, we can build touch screens with resistive materials, but such screens suffer from issues such as requiring relatively hard touch for activation and difficulty in realizing multi-touch. An alternative technology called a capacitive touch sensor exploits the capacitive properties of the finger/body in order to realize a more versitile and sensititive touch screen. Two



(a) (b)

Figure 7.17: Capacitive touch screen employing (a) self- and (b) mutual capacitance.



(a) (b)

Figure 7.18: (a) Model of the capacitance touch measurement system. (b) Equivalent circuit schematic.

approaches are shown in Fig. 7.17, a self-capacitive touch screen and a mutual capacitance touch screen.

In a self-capacitance touch screen, an array of small capacitors are used to detect the presence of a finger. As we shall see, the capacitance changes when someone touches the screen by placing their finger in the vicinity of one of the plates. Note that electrical contact is not necessary, only close proximity between the finger and the electrode. In the case of a mutual capacitance touch screen, the capacitance between adjacent filaments, or between vertical and horizontal filaments, is measured, and due to the fringing fields that are intercepted by the finger, the increase in capacitance is detected.

7.7.1 Self Capacitance Touch Sensors

A model of this scenario is shown in Fig. 7.18, where the capacitance to ground is increased due to the presence of the body-to-ground capacitance. There are two confusing points about this figure that we should resolve. First off, what is the capacitance of the human body? Second, why are we considering a capacitance to “ground"? What if the user of the device is not electrically in the vicinity of ground?

To address the first question, the human body consists of layers of fat and muscle that have electrical capacitance and resistance. The human body itself has the potential to store electric charge like any conductor, and so when a part of the body is brought into close proximity with a charged conductor, electrical charge in the body will redistribute and couple to the charged electrode, as shown in Fig. 7.18. Note that this behavior can be modeled as a series of capacitors which couples from the device to the body, and from the body to earth ground. We are assuming for simplicity that the conductivity of the body turns the body into an equipotential surface. In practice the resistance of the body is critical when understanding current flow through the body and how it interacts with muscles (hence electrical safety). But in this scenario, current is not flowing through the body but simply moving around the body in response to an electrical field.

Second, the concept of ground is confusing here. If the device is battery powered, then it is definitely not connected to earth ground. But it will have its own internal ground, which we'll call "ground". The way that the body then couples to the device is in two ways, both by capacitive coupling to a common earth ground, and by direct coupling between the body and the device. To illustrate this, in Fig. 7.18 we use two symbols for ground, a device ground and an earth ground.

If we examine this model carefully, then, we see that there is a parallel capacitance formed by the body which interacts with the device. If the finger is far away from the touch sensor, then the capacitance added by the body is very small, and effective the self capacitance of the plate electrode is given by

$$C_{eff} \approx C_d$$

On the other hand, when the finger comes near the electrode, an additional capacitance is effectively in parallel with the circuit and the capacitance incrases to

$$C_{eff} = \left(C_f^{-1} + C_{body}^{-1} + C_{dev}^{-1}\right)^{-1}$$

Note that the above equation is valid in both scenarios if we let C_f vary from a small value to a large value. In this equation, the smallest capacitance dominates (C_f) and therefore can have a dramatic impact on the overall capacitance of the parallel path.

What happens if earth ground disappears? Does the capacitance sensor still work? Yes, since the filaments are charged with respect to a known reference inside the device, and since the amount of charge flowing into the capacitance will vary due to the presence of the finger and body. To fully appreciate this requires some knowledge of electrostatics, but let us observe that charges inside the body can move around and if a filament is charged to a voltage and the fingers are in close proximity, the local distribution of charge inside the body can change (even though the body remains neutral), and this redistribution of charge is the key to understanding the touch sensor.

7.7.2 Mutual Capacitance Touch Screen

Conceptually, the mutual capacitance multi-touch system is much easier to understand since there is no ground reference needed, or even consideration of the presence or absence of earth ground. In a system using mutual capacitance (Fig. 7.19), we can measure the capacitance between two adjacent thin filaments or between an x and y direction filament (Fig. 7.19)[2]. The key concept is that some of the electric field lines fringe out and reach the surface of the device where the finger interacts with the fields. Due to the dielectric response of the finger, the molecules of the finger become polarized and partially cancel the fields of the capacitor, thus causing the effective capacitance to increase.

7.7.3 Capacitive Touch Sensor Detection

In both sensing approaches, a circuit will continuously scan the array of capacitors, measuring each one by one, and comparing to a threshold value. The threshold value can be determined dynamically by comparing the average capacitance of an array element, as most elements will not be loaded by the finger. Switches are needed to select x and y filaments. Various circuits can be used to measure the capacitance. For example, a current source can charge the filaments and the resulting voltage is measured. Since the entire array is scanned rapidly, multi-touch can be easily detected with this scheme.

[2] The filaments are thin so that they do not signficantly alter the light passing through the touch panel.

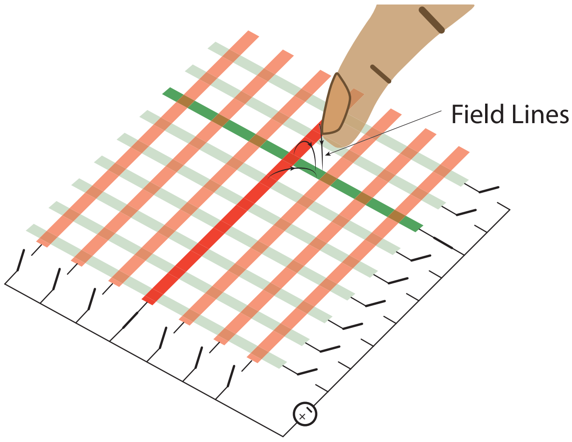


Figure 7.19: Capacitive touch sensors employing mutual capacitance between filaments can use adjacent filament capacitance in both the x and y directions. As shown, the system uses the mutual capacitance between vertical and horizontal filaments. We can consider the finger as a high dielectric constant material that interacts with the fringing fields.

7.8 Capacitors Everywhere!

A typical circuit may employ a lot of capacitors. Many of them are simply connected to the power supply. Why is that? Capacitors can act as small batteries with very small internal resistance, so they can deliver a lot of power.

The current available from a capacitor is related to the *Effective Series Resistance* (ESR). This is partly why we use a bank of capacitors, from small values to large values, in parallel on important power supply nodes. The small capacitors can deliver current much faster, but only a limited amount. They take care of high frequency transients. The big capacitors have more charge but they respond more slowly. They respond to slow transients. Finally, the battery, which is "far away" (due to inductance), is by far the slowest and keeps these capacitors charged.

These concepts may not make much sense until we study inductance and transient circuit response. But for now, you'll at least know why people do seemingly silly things such as putting a huge capacitor in parallel with a tiny one.

7.8.1 Can Capacitors Replace Batteries?

Since capacitors store charge, can we use them as energy sources to replace batteries? The advantage is that they can be charged up much faster than a battery (consider an electric car), so there is definitely good reason to ask this question.

A small AA battery has a cell voltage of 1.5V and a capacity of about 3000mA-hr. The total energy is therefore:

$$E = Q \cdot V = 3000\,\mathrm{mA} \cdot \mathrm{hr} \cdot \frac{3600\,\mathrm{s}}{1\,\mathrm{hr}} \cdot 1.5V = 16.2\,\mathrm{kJ} \tag{7.31}$$

Compare that to a small 1-oz steak, which has about 65 kCal or 273 kJ ! Now using about the same volume, you can buy a "super capacitor" with a capacitance of 100 F. The energy you can pack into this depends on the voltage rating, which is 2.5 V in this case

$$E = \frac{1}{2}CV^2 = \frac{1}{2}1002.5^2 = 313\,\mathrm{J} \tag{7.32}$$

This energy is considerably less than a battery and it costs about $25! Table 7.1 is an interesting comparison of the energy density of various elements and materials.

Technology	Volume (mL)	Weight (g)	Energy (kJ)	Cost ($)
Meat	29	28	273	.25
AA Battery	30	23	16	.25
Super Cap	61	22	.3	25
Gasoline	30	23	1050	.03

Table 7.1: A comparison of the energy density and cost of various energy sources.

7.9 Non-Linear Capacitors

For some elements, the charge/voltage relationship is non-linear. A good example is a semiconductor diode. We can then write

$$q(v) = f(v) \tag{7.33}$$

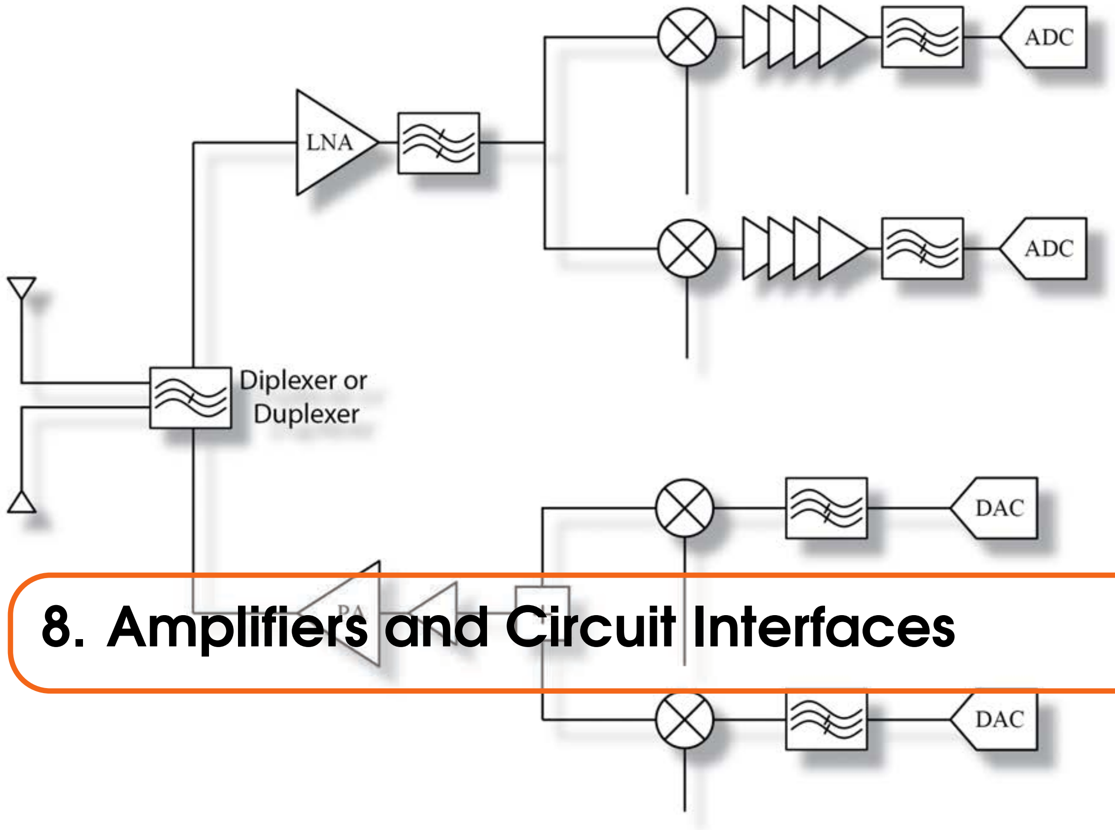
where f is an arbitrary function. The current through the device is given by

$$i = \frac{dq}{dt} = \frac{dq}{dv}\frac{dv}{dt} = C(v)\frac{dv}{dt} \tag{7.34}$$

which allows us to define a voltage-dependent incremental capacitance

$$C(v) = \frac{dq}{dv} \tag{7.35}$$

These non-linear capacitors find wide application in circuits that require a variable capacitance.



8. Amplifiers and Circuit Interfaces

8.1 Amplifiers

An amplifier is a multi-terminal element (Fig. 8.1). It usually has an input terminal (v_{in}), an output terminal (v_{out}), both referenced to a common reference (ground), which is supplied through another terminal, labeled *GND*. There is also a positive supply terminal, labeled V_{pos}, V_{supply}, or V_{CC}, or V_{DD}. In some cases, the amplifier has a negative supply terminal, labeled as V_{neg}, V_{SS} or V_{EE}. The supplies are also known as the rails of an amplifier. In practice, often the only signals shown are the input and output.

Inside, the amplifier is constructed with various devices such as resistors, capacitors, and transistors. Sometimes inductors and transformers also appear inside, especially in high frequency amplifiers. In this chapter we neglect such details and just picture the amplifier as a “black box" that amplifies signals. This perspective is analogous to the Thevenin representation of a one port where all the internal details can be lumped into an equivalent source and resistance. We will find a similar

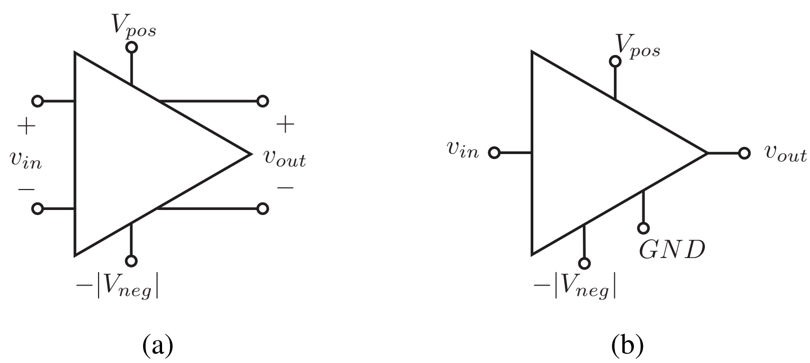


Figure 8.1: (a) Schematic symbol for an amplifier. Many times only the input (v_{in}) and output (v_{out}) are shown, since the other ports do not contain active signals but simply provide voltage supplies. (b) The amplifier input and output are implicitly referenced to ground.

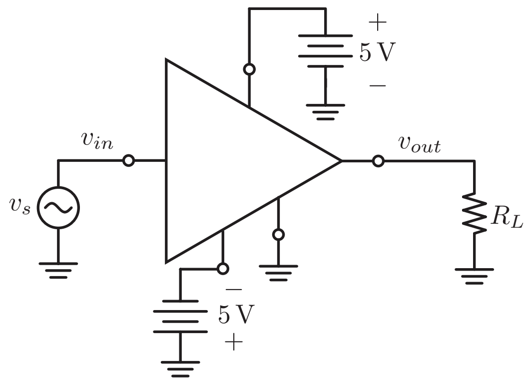
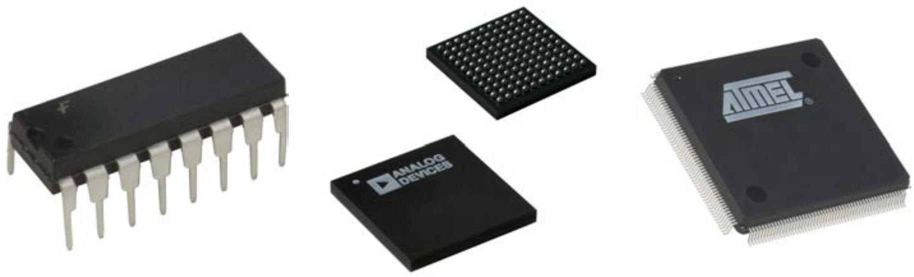


Figure 8.2: An amplifier schematic including the input source, an output load, and batteries connected to the supplies.



(a) (b) (c)

Figure 8.3: Different packaging options. (a) A dual in-line package (DIP) is used when the number of pins is limited. (b) If a package has many pins, or if higher frequency operation is desired, a ball grid array (BGA) can be used. (c) A plastic quad flat pack has shorter leads and connections on all four sides.

representation for an amplifier (sometimes called a 2-port since there are two sets of terminals, i.e. the input and output).

8.1.1 Unilateral Gain

The symbol for the amplifier is a triangle (Fig. 8.2) because the amplifier nominally amplifies only in one direction. We call this behavior “unilateral” gain, because a signal at the input is amplified and appears at the output. Signals applied to the output have no effect on the input of the amplifier.

The supplies are connected to batteries or supply voltages (Fig. 8.2). If a negative supply is not needed, the negative terminal of the battery, usually ground, is connected there. The current drawn from the supplies is usually a DC current.

The input signal is applied to the input terminal. It is usually connected to a voltage or current source and it can be a DC or an AC signal. The output is often connected to a load, which is usually modeled by a resistor. The load resistance can model a transducer (speakers, or a screen), an analog-to-digital converter, or another amplifier (cascade amplifiers). The output signal is usually a faithful reproduction of the input.

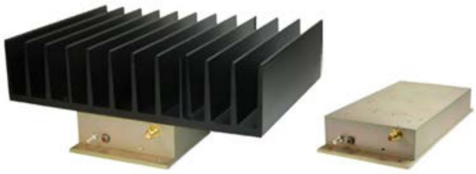


Figure 8.4: (a) A high power amplifier module with a visibly large heat sink. (b) An amplifier module housed in a metal box for precision applications.

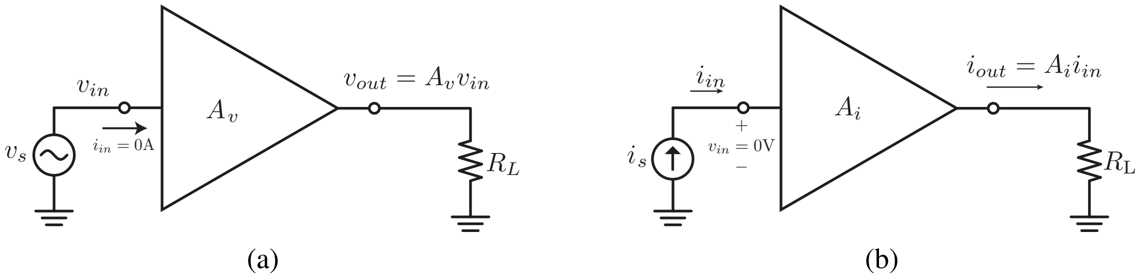


Figure 8.5: (a) An ideal voltage amplifier. (b) An ideal current amplifier.

8.1.2 Amplifier Packaging

Depending on the frequency of operation, different packaging options are available for the amplifier. Often a DIP (dual in-line package) is used to package integrated circuit amplifiers. Inside the package an integrated circuit die is wire-bonded to the package leads. For larger number of pins, pin grid array (PGA) or plastic quad flat packages (PQFP) are used. Ball grid array (BGA) can be surface mounted using solder balls (see Fig. 8.3).

8.1.3 Amplifier Modules

High frequency amplifiers and precision amplifiers are often packaged using a metal housing to limit the amount of interference in the package (Fig. 8.4). Often co-axial cables are used to deliver the input and output signals. Coaxial cables use a ground shield to prevent noise pickup. High power amplifiers have heat sinks to keep the temperature cool.

8.2 Ideal Amplifiers

8.2.1 Ideal Voltage Amplifiers

In an ideal voltage amplifier (Fig. 8.5a), the input signal is a voltage, and the amplifier faithfully “copies" the input to the output while increasing the magnitude of the voltage (regardless of the load)

$$v_{out} = A_v v_{in} \tag{8.1}$$

where A_v is the voltage gain of the amplifier. The voltage gain is usually specified using a dB scale

$$A_v[\text{dB}] = 20 \cdot \log(A_v) \tag{8.2}$$

so a gain of 100 is 40dB. The input current flowing into the amplifier is ideally zero

$$i_{in} \approx 0 \tag{8.3}$$

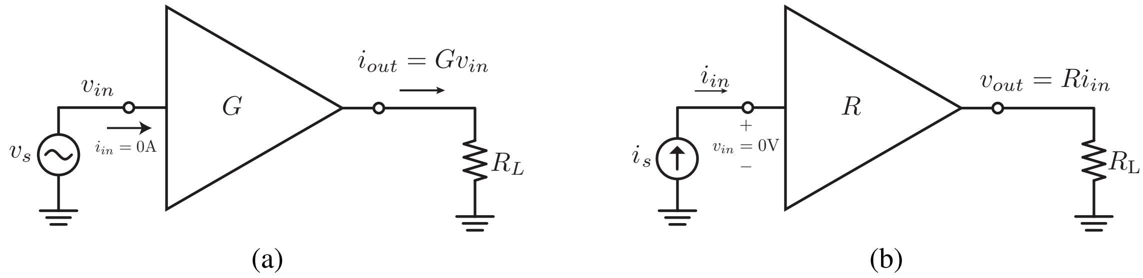


Figure 8.6: (a) An ideal transconductance amplifier. (b) An ideal transresistance amplifier.

which means that the *input resistance of the amplifier* is infinite. Note that the power flow into the input is therefore zero, which means the amplifier has infinite power gain.

8.2.2 Ideal Current Amplifiers

An ideal current amplifier (Fig. 8.5b), the input signal is a current, and the amplifier faithfully “copies" the input to the output while increasing the magnitude of the current (regardless of the load)

$$i_{out} = A_i i_{in} \tag{8.4}$$

where A_i is the current gain of the amplifier. The current gain is usually specified using a dB scale

$$A_i[\text{dB}] = 20 \cdot \log(A_i) \tag{8.5}$$

so a gain of 1000 is 60dB. The input voltage into the amplifier is ideally zero

$$v_{in} \approx 0 \tag{8.6}$$

which means that the *input resistance of the amplifier* is zero. Note that the power flow into the input is therefore zero, which means the amplifier has infinite power gain.

8.2.3 Ideal Transconductance Amplifiers

An ideal transconductance amplifier (Fig. 8.6a), the input signal is a voltage, while the output is a current. The amplifier faithfully “copies" the input to the output (regardless of the load)

$$i_{out} = Gv_{in} \tag{8.7}$$

where G is the transconductance gain of the amplifier. The transconductance gain is usually specified using a dBS scale

$$G[\text{dBS}] = 20 \cdot \log(G/1S) \tag{8.8}$$

so a gain of 10^4S is 80dBS. The input current into the amplifier is ideally zero

$$i_{in} \approx 0 \tag{8.9}$$

which means that the *input resistance of the amplifier* is infinite. Note that the power flow into the input is therefore zero, which means the amplifier has infinite power gain.

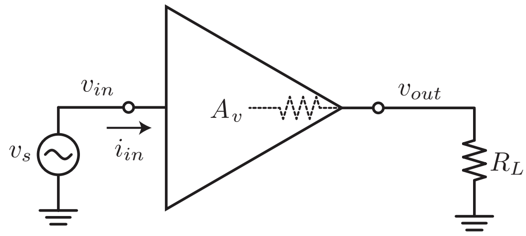


Figure 8.7: A real voltage amplifier will draw current from the source i_{in}, and the output voltage will vary when the load is varied, meaning that it too has an effective "source" resistance.

8.2.4 Ideal Transresistance Amplifiers

An ideal transresistance amplifier (Fig. 8.6b), the input signal is a current, while the output is a voltage. The amplifier faithfully "copies" the input to the output (regardless of the load)

$$v_{out} = Ri_{in} \tag{8.10}$$

where R is the trans-resistance gain of the amplifier. The trans-resistance gain is usually specified using a dBΩ scale

$$R[\text{dB}\Omega] = 20 \cdot \log(R/1\Omega) \tag{8.11}$$

so a gain of $10^6\Omega$ is 120dBΩ. The voltage current into the amplifier is ideally zero

$$v_{in} \approx 0 \tag{8.12}$$

which means that the *input resistance of the amplifier* is zero. Note that the power flow into the input is therefore zero, which means the amplifier has infinite power gain.

8.3 Real Amplifiers

8.3.1 Real Voltage Amplifier

A real voltage amplifier, shown in Fig. 8.7, has three important imperfections. First, the input current is non-zero, which means it has finite power gain. This input current also leads to a *loading effect*, because the input voltage applied to the amplifier is now smaller than the source voltage due to non-zero source resistance. Recall that even a battery has internal source resistance, meaning that if you draw higher current, the voltage of the source drops.

Likewise, the output of the amplifier will vary if the loading at the output varies. For a fixed load resistance, this means that the gain will be lower than expected. The output voltage is also limited in range. Typically it cannot exceed the positive/negative supply rails. We'll return to this point in Section 8.4.

8.3.2 Equivalent Circuit Model

The complete equivalent circuit model for a voltage amplifier is shown in Fig. 8.8. The input and output resistance model the fact that the amplifier is subject to gain reduction due to the loading effects due to source/load resistance. The dependent source models the linear gain from the input to the output.

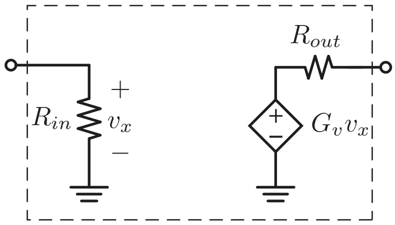


Figure 8.8: The linear equivalent circuit for a non-ideal amplifier that includes input resistance R_{in}, output resistance R_{out}, and gain G_v.

Input Resistance

The input current can be modeled by an input resistance. A test source is connected to the input terminals and the current flow at the input terminal is modeled by a Thevenin equivalent resistance

$$R_{in} = \frac{v_{in}}{i_{in}} \tag{8.13}$$

When a voltage supply with source resistance R_s is connected at the input, the actual input voltage is reduced by the voltage divider formula

$$v_{in} = \frac{R_{in}}{R_{in} + R_s} v_s \tag{8.14}$$

which means the performance is the most ideal with $R_{in} \to \infty\Omega$ or in practice, $R_{in} \gg R_s$.

Output Resistance

Since the output voltage varies depending on the load connected at the output, the output terminal has an equivalent Thevenin resistance at the output terminals. To find this resistance, we can take the ratio of the open circuit voltage to the short circuit current, or connect a voltage (current) source and monitor the output current (voltage)

$$R_{out} = \frac{v_{out}}{i_{out}} \tag{8.15}$$

When a load is connected to the output, the actual load voltage is reduced by the “loading” effect

$$v_L = \frac{R_L}{R_L + R_{out}} v_{th} \tag{8.16}$$

An ideal voltage amplifier has $R_L \gg R_{out}$.

Effective Gain

Because of the loading effects, the actual gain of an amplifier is going to be lower than A_v, the advertised gain. A_v applies only under an open-circuit load (for a voltage or trans-resistor amplifier) when driven by an ideal source (say a voltage source with no source resistance). To calculate the actual gain, we can simply take into account the loading effects at both the input and output

$$A_{eff} = \frac{R_{in}}{R_{in} + R_s} A_v \frac{R_L}{R_{out} + R_L} \tag{8.17}$$

The first factor is a voltage divider that takes a fraction of the input voltage and applies it to the amplifier. The last term also accounts for the loading at the output. Since both factors are ≤ 1, the amplifier effective gain is $\leq A_v$.

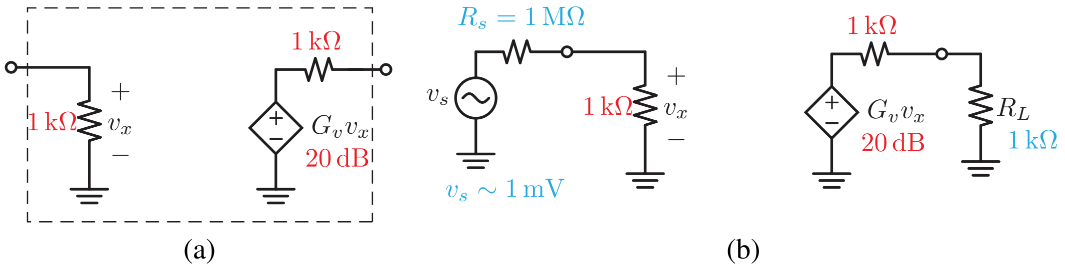


Figure 8.9: (a) An example amplifier with input and output resistance. (b) A source/load connected to the example amplifier results in loading.

Example 20: Amplifying Weak Signals

Suppose that we wish to amplify a weak signal, on the order of 1mV, to a range appropriate for digitization. Suppose the signal originates from a sensor (such as an ECG electrode) with higher source resistance, say $1\,\mathrm{M\Omega}$. The digitizer is an analog-to-digital converter (ADC), which takes an analog input voltage and produces a digital representation of the voltage, which can be stored using digital memory.

The ADC usually can only amplify signals larger than a certain range. In this examples, the range of the ADC is given by $1\,\mathrm{mV} < V_{in} < 1\,\mathrm{V}$. Also, the ADC has an input resistance of $1\,\mathrm{k\Omega}$. We select the amplifier shown in Fig. 8.9a to perform this job

$$A_v = 20\,\mathrm{dB}, R_{in} = 1\,\mathrm{k\Omega}, R_{out} = 1\,\mathrm{k\Omega} \tag{8.18}$$

What is the gain of the amplifier ? 20 dB is the ideal gain when the source and amplifier are not loaded. We will now calculate the effective gain as shown in Fig. 8.9b.

In this example, R_s is much larger than the input resistance R_{in} of the amplifier, which results in significant loading

$$A_{eff} = \frac{10^3}{10^6 + 10^3} \cdot 10 \cdot \frac{10^3}{10^3 + 10^3} \tag{8.19}$$

$$\approx \frac{10^3}{10^6} \cdot 10 \cdot \frac{1}{2} = 5 \times 10^{-3} \tag{8.20}$$

The gain is now less than 1, which is exactly opposite of what we're trying to do. To solve this problem, we need a buffer (next lecture) or an amplifier with a higher input resistance.

8.3.3 Cascade of Amplifiers

When several ideal amplifiers are placed in cascade (Fig. 8.10), the gain increases, as you might expect

$$A_v = A_{v,1} \cdot A_{v,2} \cdots \tag{8.21}$$

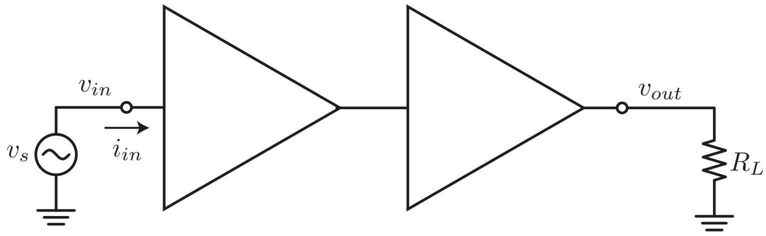


Figure 8.10: Cascading amplifiers to obtain higher gain.

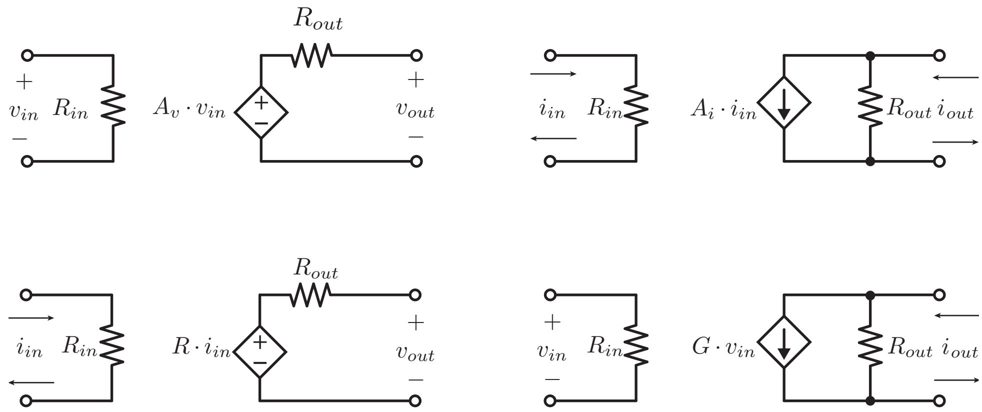


Figure 8.11: Equivalent circuits for the four possible amplifiers including the voltage amplifiers ($V \to V$), current amplifiers ($I \to I$), transresistance ($I \to V$) amplifier, and transconductance amplifiers ($V \to I$).

When real amplifiers are cascaded, we have to take loading effects into account. For instance, for the cascade of the two amplifiers from the previous example, we can derive a new amplifier equivalent circuit model. The input and output resistances are given by the first and last amplifiers (this is only true for unilateral amplifiers)

$$R_{in} = R_{out} = 1\,\mathrm{k}\Omega \tag{8.22}$$

whereas the gain is may suffer due to internal loading effects. In this example, for instance, the input and output resistances are equal, which results in a gain drop factor of $1/2$

$$A_v = \frac{1}{2} A_{v,1} A_{v,2} = 50 \tag{8.23}$$

8.3.4 Loading with Other Amplifier Topologies

Although we did not explicitly discuss loading effects in other amplifiers, such as current, transresistance, and transconductance amplifiers, the concepts of input and output impedance are general and pass naturally to these amplifiers. The equivalent circuit for all four amplifier types is shown in Fig. 8.11. When the output signal is a current, it's natural to use a Norton representation of the output of the amplifier.

Note that when the source is a current, we prefer to use a current or trans-resistance amplifier with low input resistance to minimize loading. The same is true when the output is a current, as the ideal load is a short circuit. Then opposite is true for voltage inputs and output.

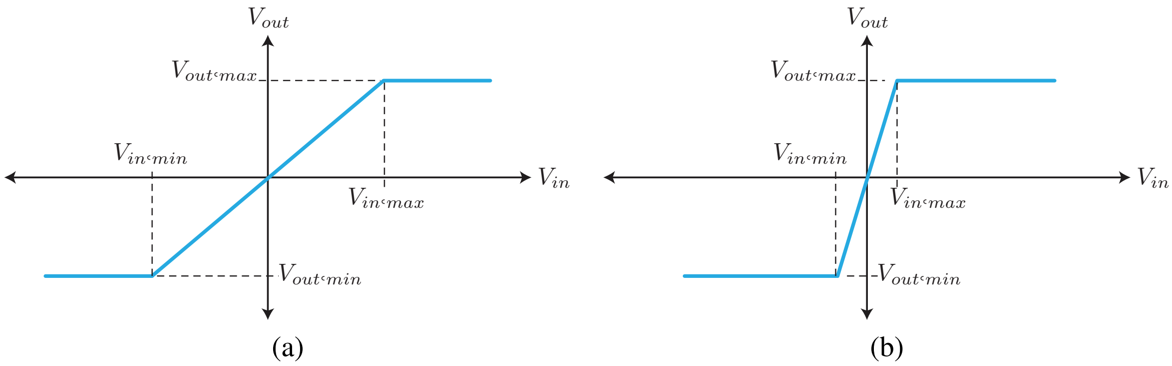


Figure 8.12: (a) The input/output curve for every amplifier eventually clips the output signal due to the limited voltage range supplied by the power supply connections. To avoid distortion, the signal is usually limited to lie within the linear range of the amplifier. (b) A higher gain amplifier has a smaller input range.

8.4 Amplifiers Non-Idealities

Because of the limited output range of the amplifier, the input-output curves saturate, usually at a voltage lower or equal to the supply rails. This means that the input voltage cannot exceed a certain range before the amplifier behavior becomes non-linear.

Input/Output Curve

The maximum input signal that we can apply to the amplifier is thus related to the supply voltage and the gain of the amplifier (see Fig. 8.12b)

$$v_{in,max} \leq \frac{V_{sup}}{G_v} \tag{8.24}$$

Similarly, the smallest input has to be larger than

$$v_{in,min} \geq \frac{-V_{sup}}{G_v} \tag{8.25}$$

For a high gain amplifier, this is a very limited input range. Unless the supply voltage can be increased (limited by technology), the output dynamic range is fixed.

Clipping

When the input signal goes beyond the linear range, the output can exhibit clipping, which is very non-linear. In a video application, any intensity variations beyond the clipping level will appear at a constant maximum intensity, which represents information loss and image distortion. In audio applications, a strong noticeable distortion due to clipping will degrade the sound quality.

8.4.1 Distortion

When the input range exceeds the linear input range, the amplifier behavior becomes much more non-linear. All real amplifiers are non-linear and the input / output relationship can be modeled by a Taylor series (for small signals)

$$v_{out} = a_1 v_{in} + a_2 v_{in}^2 + a_3 v_{in}^3 + \cdots \tag{8.26}$$

This non-linearity produces *distortion* in the signal, which can change the signal enough to make the system operation poor or unacceptable. A non-linear audio amplifier produces unwanted distortion (harmonics) which alters the quality of the sound.

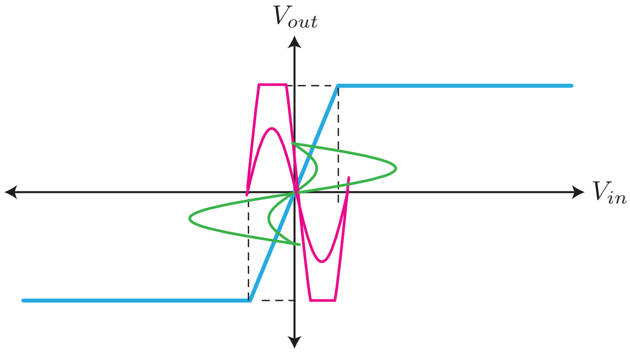


Figure 8.13: When the input voltage is below the maximum clipping limits, a faithful (amplified) copy of the waveform is generated. When the input is larger, the waveform is clipped.

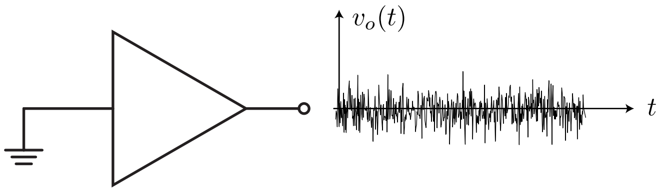


Figure 8.14: Real amplifiers generate noise. When the signal is sufficiently large, we may not even notice this noise, but when no signal is present, the noise is clearly visible on a sensitive detector.

Gain Reduction

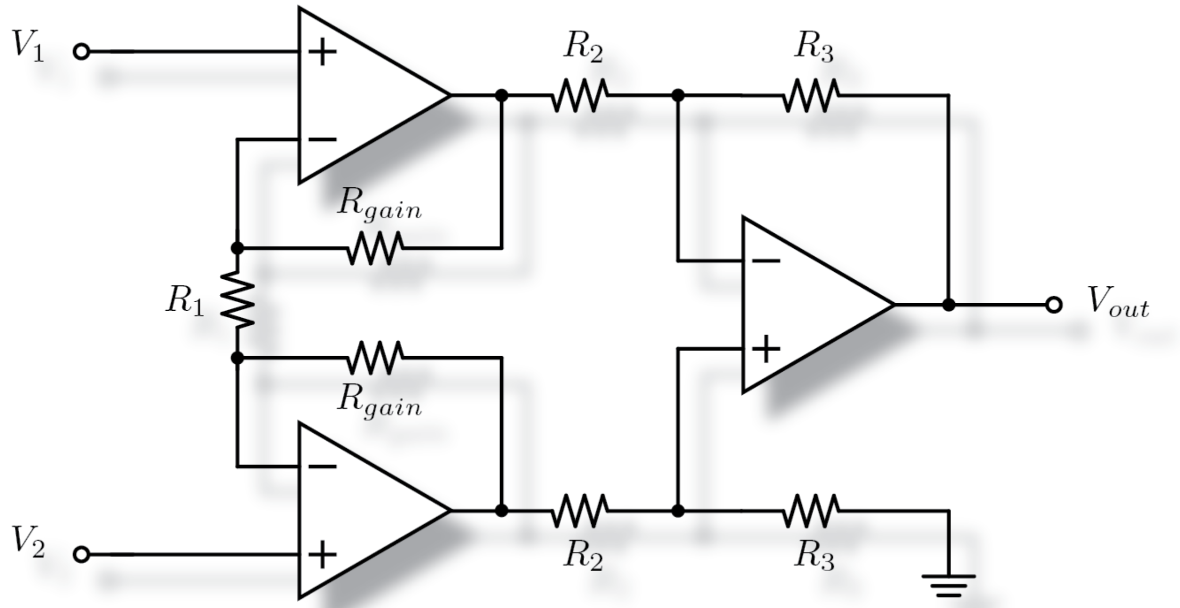
Note too that the effective gain of the system reduces as the input exceeds the input range of the amplifier. That's because the output saturates at V_{sup} whereas the input keeps increasing

$$G_{eff} = \frac{V_{sup}}{v_{in}} \leq G_v \tag{8.27}$$

8.4.2 Noise

Even though the subject of noise is beyond the scope of this course, it's important to realize that all real amplifiers produce noise at their output. In Fig. 8.14, the input of the amplifier is shorted, but the output is non-zero due to noise. We must exercise care when amplifying weak signals. If we choose the wrong amplifier, the output noise can be larger than the output signal!

R The sound of static on an analog radio or the “snot" on an analog television screen are audio and visual manifestations of noise.



9. Operational Amplifiers

The Operational Amplifier, or op-amp, was invented in 1941by Bell Labs engineer Karl D. Swartzel Jr. using vacuum tubes. It found wide application in WW-II. An op-amp is an amplifier with some unique properties, but otherwise an ordinary voltage amplifier amplifier. Many of the unique and useful properties emerge from the fact that it has a very large voltage gain, usually millions of times larger than the voltage presented at the inputs. Today op-amps are ubiquitous low cost components used in countless applications for analog and mixed-signal signal processing (gain, filtering, signal conditioning).

9.1 Operational Amplifiers

The basic op-amp schematic symbol is shown in Fig. 9.1. In schematics it has at least 3 pins, but in practice it has many more. The three important pins are the inputs, V^+ and V^-, called the non-inverting and inverting inputs, and the output V_{out}. These are the signal pins that are often shown in schematics, and the other pins are omitted to keep things simple.

The reason that we often omit the other pins is that they are DC signals such as the power

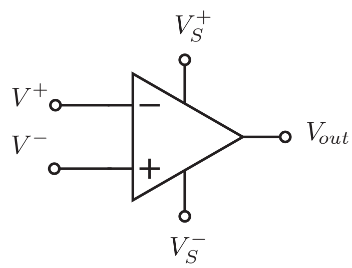


Figure 9.1: The schematic symbol for an op-amp. Three essential signal pins are the inputs V^+, V^-, and V_{out}. The supply pins, usually connected to a positive and negative supply $V_S^\pm$ are often omitted from the schematic. All voltages referenced to a common ground.

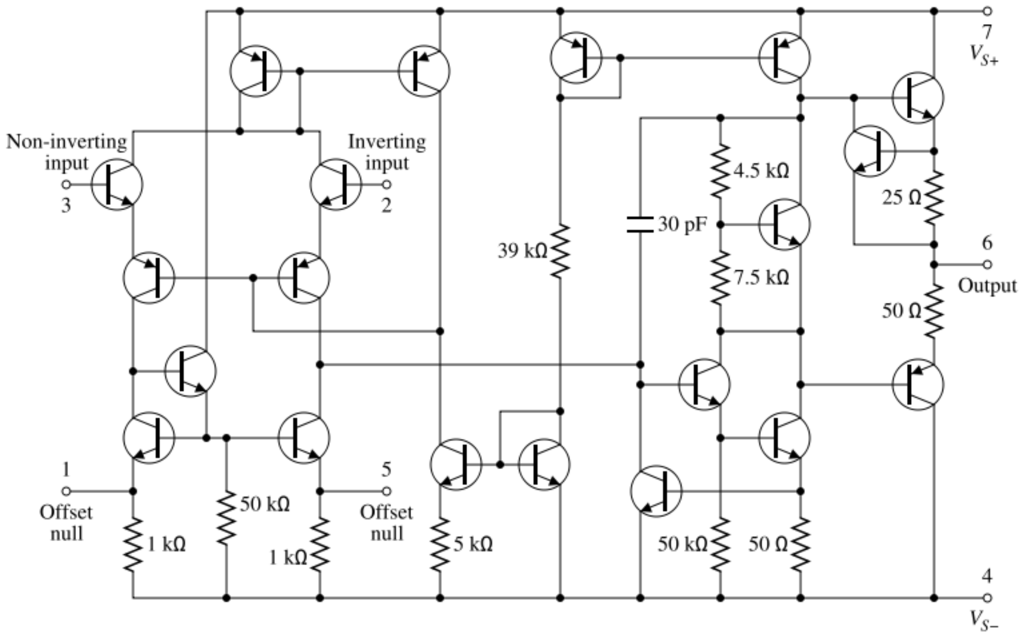


Figure 9.2: The classic 741 op-amp internal schematic implemented with bipolar junction transistors (BJTs). The inputs (pins 2 and 3) and output (pin 6), and supply pins (4, 7) are labeled. Pins 1 and 5 are used to zero out the offset voltage, a non-ideality we will ignore in this chapter.

supply pins[1], and they are needed to make the op-amp function, but otherwise they are not active in the operation of the circuit. On the other hand, the signal pins are usually AC voltages containing signals of interest.

An op-amp is a high gain amplifier with a *differential input*

$$v_o = A \cdot (v^+ - v^-) \tag{9.1}$$

We say *differential* because in practice the input voltage must be very small, otherwise the output would clip due to the high gain.

9.1.1 Classic 741 Schematic

The first monolithic Integrated Circuit (IC) op-amp was designed by Bob Widlar at Fairchild Semiconductor, shown in Fig 9.2. The 741 op-amp is perhaps the best known op-amp in the world. Many other op-amps use the same pin configuration as the 741. Internally, the op-amp is constructed from dozens of transistors (bipolar junction transistors, or BJTs). In this class we will ignore the internal details and instead work with an equivalent circuit model of the op-amp.

9.2 Equivalent Circuit Model

We model the complex op-amp by using the simple equivalent circuit shown in Fig. 9.3. The most salient features are the high gain A (typically a million or more), very high input resistance R_{in}, and low output resistance R_{out}. Often the output resistance R_{out} is neglected.

As previously noted, because of the large gain, only a few microvolts of input signal is required to saturate the op-amp output. Thus the amplifier is very impractical if used without *feedback*.

[1] Some op-amps work with a single supply, in which case the negative rail is ground.

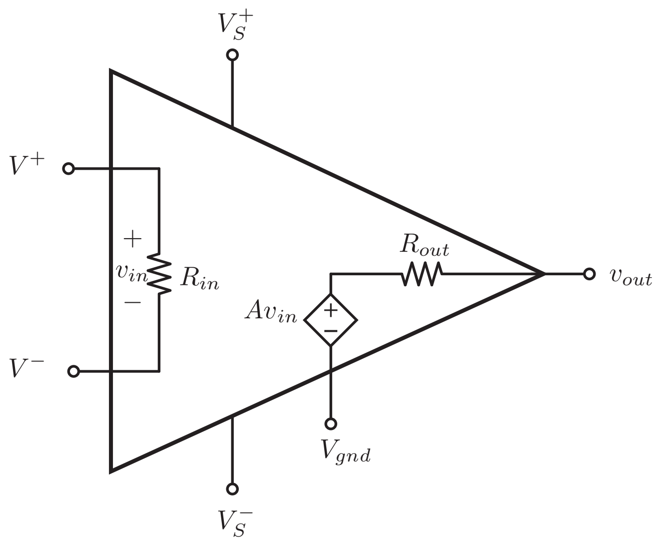


Figure 9.3: Like any unilateral amplifier, the op-amp can be modeled by its input resistance R_{in}, output resistance R_{out}, and voltage gain A. One unique aspect of an op-amp is the presence of two inputs of dual polarity. A good op-amp has high gain $A \sim 10^4 - 10^6$ and very high input resistance R_{in}.

In fact, the gain of the op-amp is a very poorly controlled parameter, often varying wildly with temperature or from part-to-part. How do you design with such an imperfect component? *Feedback.*

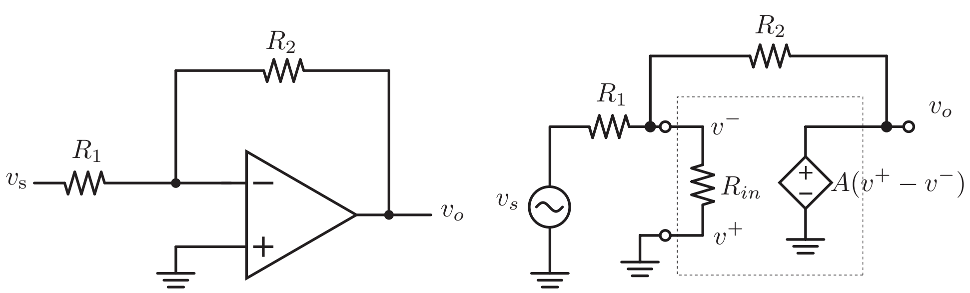
9.2.1 Op-Amp Inverting Amplifier

The example circuit shown in Fig. 9.4, is a typical op-amp configuration where the output signal is fed-back to the negative input terminals. This is called negative feedback. This seems strange at first because we are subtracting the output from the input, but as we shall see, this is a self-regulation mechanism that results in a very precise amplifier.

Write KCL at the input node of the amplifier

$$(v^- - v_o)G_2 + v^- G_{in} + (v^- - v_s)G_1 = 0 \tag{9.2}$$

But the output voltage in this case is simply given by $v_o = -Av^-$, where A is very large, which



(a) (b)

Figure 9.4: (a) An op-amp configured as an inverting amplifier (gain is negative). (b) To analyze the circuit, the op-amp is replaced with a simple model. The output resistance of the op-amp is neglected for simplicity.

means that $v^- = -v_o/A$ is a very small voltage

$$(\frac{-v_o}{A} - v_o)G_2 + \frac{-v_o}{A}G_{in} + (\frac{-v_o}{A} - v_s)G_1 = 0 \tag{9.3}$$

which allows us to write the complete expression for gain

$$\frac{v_o}{v_s} = \frac{-AG_1}{G_2(A+1) + G_{in} + G_1} \tag{9.4}$$

Assuming that the op-amp has a very large gain, the above equation simplifies

$$\frac{v_o}{v_s} \approx \frac{-AG_1}{G_2(A+1)} \approx \frac{-G_1}{G_2} = \frac{-R_2}{R_1} \tag{9.5}$$

Notice that the final expression for the gain does not involve A at all. This is good actually because the gain A is large, but widely varying from part to part, and as a function of temperature. The gain is precisely set by resistors R_1 and R_2, and more specifically their ratio, which is something we can control very well.

So what did we gain by using an op-amp with such a high value of gain when the final gain does not even depend on A? We trade gain for precision. In the process we have an extremely versatile device that can be reconfigured to have any gain range by simply selecting the feedback components (R_1 and R_2), so it's a user programmable gain device. As we'll see shortly, there are many other building blocks such as summers, difference circuits, and integrators and differentiators that can be made with an op-amp.

9.2.2 Differential rather than Difference Amplifier

We call an op-amp a differential amplifier rather than a difference amplifier. Why? In the inverting amplifier configuration, we can calculate the effective input voltage by

$$(v^+ - v^-) = \frac{v_o}{A} = \frac{\frac{-R_2}{R_1}v_s}{A} \approx 0 \tag{9.6}$$

We see the input voltage must be small for the op-amp to operate correctly (hence a "differential" voltage). We will build a true *difference* amplifier with op-amps later.

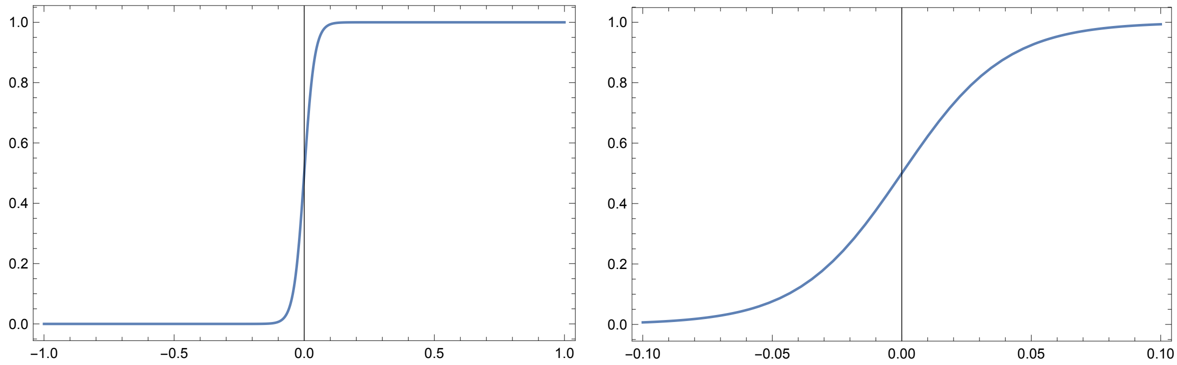
The reason the inputs must be small can be understood by noting that that in most amplifiers the output cannot be larger than the supply voltages, so that limits the input range of the amplifier

$$|v^+ - v^-| < \frac{V^\pm}{A} \tag{9.7}$$

9.2.3 Dynamic Range of Amplifier

Even though we modeled the op-amp as a device with very high gain, in reality the transfer function is only linear in a very small regime, and the output clips if the inputs are increased, as shown in Fig. 9.5. In fact, without zooming in, the op-amp appears to be a perfect "slicer", meaning that it splits the input range into two distinct regions that result in an output of either V^+ or V^-.

R *(Subtle)* For a feedback amplifier, the input linear range is expanded since regardless of the input voltage source magnitude, the differential input of the op-amp is always small $(v^+ - v^-) = v_o/A \approx 0$.



(a) (b)

Figure 9.5: (a) The transfer characteristics for an op-amp over a wide range of input voltages. The output clips to the positive and negative rails if the input is larger than approximately the rail voltage divided by the gain. (b) If we zoom in to a very narrow range of inputs, we see the op-amp transfer characteristic has a linear region with a very high gain (slope of the curve).

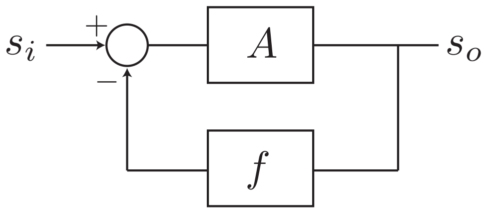


Figure 9.6: Model for a general feedback system consisting of a gain element A, a feedback element f, which subtracts the output signal s_o from the input s_i.

9.3 Feedback System

In our derivation of gain, we had a nice result that the voltage gain of the overall circuit is just set by the ratio of two resistors, which can be made very precise and can track temperature. The internal gain of the amplifier A does not appear in the final expression, which means if it varies due to temperature or from part to part, it plays a negligible role in setting the gain.

So we sacrificed gain to arrive at a solution that is much more robust. This is the concept of negative feedback and it is used widely in electronic systems (biological, chemical, and mechanical systems use it too). The idea is to sample a fraction of the output and compare it to the input. By forcing equality between the input and the fraction of the output, the gain is determined by the fraction rather than by the raw gain of the amplifier. Note that positive feedback is not used here, since it has a saturating (rather than regulating) characteristic. There are applications for positive feedback as well, which we will discuss in Section 9.8.

We can model a general feedback system as shown in Fig. 9.6. Here the gain of the system without feedback is given by A,

$$s_o = As_\varepsilon \tag{9.8}$$

and the input and output are labeled s_i and s_o, and s_ε is the error signal, or the difference between the input and the feedback signal. The feedback takes a fraction f and subtracts it from the input

$$s_o = A(s_i - fs'_o) \tag{9.9}$$

Notice the recursive nature of this equation as the current output depends on the input but also on the output itself, s'_o. In reality the output in the error signal is actually the previous value of the output s'_o, not the current output s_o, but due to the fast propagation delays of signals in the loop, we can assume that it's actually the current output and write

$$s_o = A(s_i - fs_o) \tag{9.10}$$

Solving this equation

$$s_o(1 + Af) = as_i \tag{9.11}$$

$$s_o = \frac{A}{1 + Af} s_i \tag{9.12}$$

The denominator is the *loop gain*, or the gain in the loop from the input, to the output, and then back to the input. In most feedback systems using op-amps, this gain is very large

$$Af \gg 1 \tag{9.13}$$

which allows us to write

$$s_o \approx \frac{A}{Af} s_i = \frac{1}{f} s_i \tag{9.14}$$

Example 21: A common example of a feedback system is an environmental temperature control system, such as a thermostat in your home of office, which is used to heat or cool a room. An "open loop" temperature control system has several settings from low to high which cause the heater to operate at various levels. In order to use such a system without feedback, you have to have knowledge about the heat generation rate of the heater, the heat loss due to air flow (convection), conduction, and radiation, and other complicated factors. Since these factors change from day to day, and sometimes even hour to hour, it's virtually impossible to get the temperature at the desired point.

A system with feedback solves this problem by placing a temperature sensor in the room and by monitoring the actual temperature of the room. The heater can be turned on or off, or the level can be varied continuously, in order to reach the desired temperature. There's no need to know anything about the room, the source of heat loss, etc., only it is needed to measure the temperature and to make corrections.

9.4 Ideal Op-Amp: Golden Rules

The ideal op-amp model introduced thus far is useful for circuit simulation, but too complicated for hand analysis. In fact, since the gain of the op-amp is so large, we can make several simplifying assumptions:

1. Both inputs are at the same voltage.
2. No current flows through either input.

As a consequence of the second rule, the input impedance of the amplifier is nearly infinite.[2] It is important to note that the first rule applies only when a circuit has negative feedback, whereas the second rule is true for a well designed op-amp irrespective of the feedback configuration.

Example 22: In the analogy of feedback in a factory producing widgets, note that if your factory were very productive, much more than needed to produce x widgets a day, then you would reduce the rate of input materials fed into the factory (through feedback) to a very low level, practically approaching zero in the case of an extremely productive factory. This is analogous to the Golden Rules.

9.5 Op-Amp Circuits

9.5.1 Inverting Amplifier

Let's redo the calculations for the so-called "inverting amplifier" using the Golden Rules. By the first golden rule, the inverting input of the op-amp must be at ground potential ($v^+ = v^-$), which is

[2] Even if the amplifier has a relatively modest input impedance, when feedback is applied, or in "closed loop" configuration, the input impedance is driven to very high values.

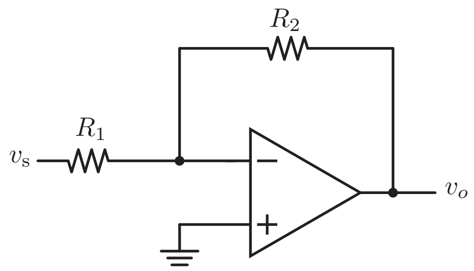


Figure 9.7: An op-amp based inverting amplifier.

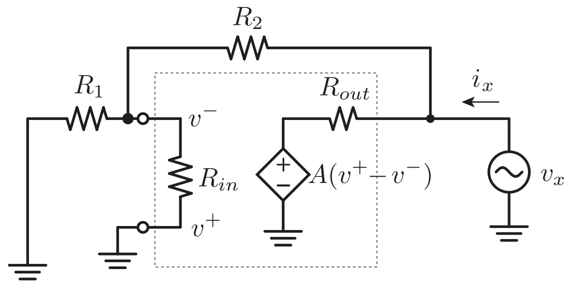


Figure 9.8: Derivation of the output resistance of an op-amp buffer is performed by calculating the current drawn by a test source when the input is set to zero.

often called a "virtual ground". That's because this voltage moves very little as an input signal is applied.

Write KCL at the input of the op-amp (which is at the virtual ground potential):

$$\frac{v_s}{R_1} = -\frac{v_o}{R_2} \tag{9.15}$$

Note that the term for the input current of the op-amp is missing, due to the golden rule. Then we have

$$\frac{v_o}{v_s} = \frac{-R_2}{R_1} \tag{9.16}$$

This agrees with our previous result, but it was much faster to derive it. Note this result is only true in the limit that the op-amp has infinite gain, but in practice it's very close to the correct answer. As an exercise, compare the error in the inverting amplifier gain if $A = 10^6$.

Inverting Amp Input/Output R

Since the input of the amplifier is at a virtual ground, the voltage source v_s only "sees" the resistance R_1, which is approximately the input resistance of the circuit.

At the output, the action of the feedback lowers the output resistance, and so the output looks like a nearly ideal voltage source, which means that the op-amp has transformed the voltage source into a nearly perfect voltage source. In other words, the op-amp buffers the source and presents it as a nearly perfect source (with very small source resistance).

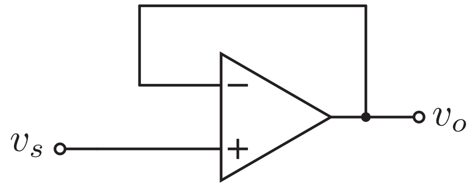


Figure 9.9: A volage buffer, or unity-gain follower.

Example 23: The last point is subtle, and not obvious. To see how the feedback transforms the op-amp into a nearly perfect voltage source, consult the model of the inverting amplifier, re-drawn in Fig. 9.8, where we connect a voltage source v_x at the output in order to derive the Thevenin equivalent output resistance of the circuit.

Note that the voltage at the inverting terminal is easily found using a voltage divider

$$v^- = \frac{R_1||R_{out}}{R_1||R_{out}+R_2} v_x \tag{9.17}$$

To find i_x, we sum the two components, one from the dependent source, the other from the feedback resistors

$$i_x = \frac{v_x - (-Av^-)}{R_{out}} + \frac{v_x}{R_1||R_{out}+R_2} \tag{9.18}$$

Let's call $R_1||R_{out} = R_1'$. Substituting for v^-

$$i_x = v_x \frac{1+\frac{AR_1'}{R_1'+R_2}}{R_{out}} + v_x \frac{1}{R_1'+R_2} \tag{9.19}$$

Focusing only on the second term containing A, we see that

$$i_x = v_x \frac{AR_1'}{R_{out}(R_1'+R_2)} + \cdots \tag{9.20}$$

In a well designed op-amp, the gain A is very large, which makes this term very large. In other words, the op-amp equivalent circuit has high (infinite if A is infinite) conductance, like an ideal voltage source.

9.5.2 Voltage Follower

The voltage follower has an input-output relation that at first seems trivial:

$$v^s = v^+ = v^- = v_o \tag{9.21}$$

Which means the output voltage is just the same as the input voltage, or it *follows* the input. But notice that the input impedance seen by the source is nearly infinite, since no current flows into the op-amp. This means that the op-amp does not load the source in this configuration. Likewise, the output impedance is very low, which means that an imperfect source can be buffered and made to appear as an ideal voltage source.

This circuit is an example of a *buffer*, or a circuit that allows us to read a voltage without disturbing the voltage. Note that if we connect an amplifier with a non-infinite input impedance to a node, current will be drawn from that node, which disturbs the node voltage, and so we would be reading the wrong voltage. The buffer makes a copy of the voltage without loading.

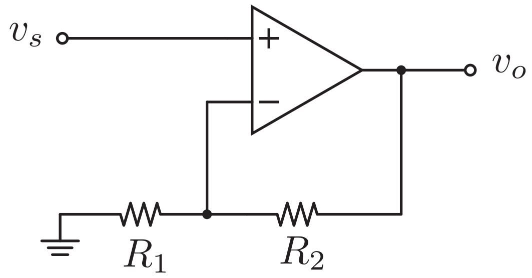


Figure 9.10: A non-inverting amplifier.

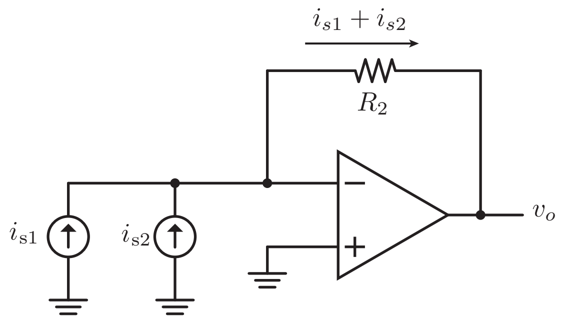


Figure 9.11: Current summing amplifier. An input current is converted to an output voltage, which gives rise to the name trans-resistance amplifier.

9.5.3 Non-Inverting Amplifier

For the non-inverting amplifier shown in Fig. 9.10, we still apply feedback to the negative terminal. The input is supplied to the positive terminal. Applying the first golden rule as before, we have

$$v^- = v^+ = v_s \tag{9.22}$$

Now applying the second golden rule, since the input current of the op-amp is zero, there is a perfect voltage divider from the output of the op-amp to the negative terminal

$$v^+ = v_o \frac{R_1}{R_1 + R_2} \tag{9.23}$$

Which means that

$$\frac{v_o}{v_s} = 1 + \frac{R_2}{R_1} \tag{9.24}$$

9.5.4 Current Summing Amplifier

An important observation is that in the inverting amplifier, the current injected into the negative terminal of the op-amp is routed to the output and converted into a voltage through R_2. If multiple currents are injected, as shown in Fig. 9.11, then the *sum* of the currents is converted to a voltage.

This circuit finds wide application in optical receivers in fiber optic communication systems. The light signal on the fiber is sensed by a photodiode, which produces a current proportional to the light intensity. This modulation is converted to a voltage with this trans-resistance amplifier[3]

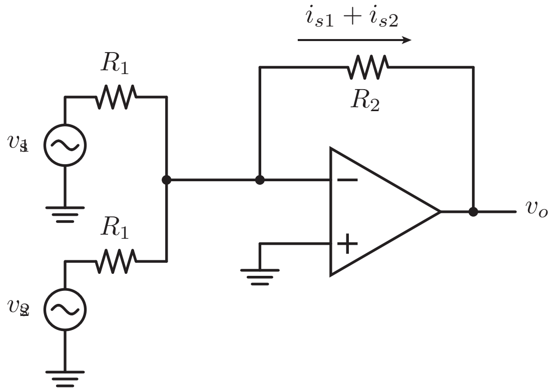


Figure 9.12: A voltage summing circuit converts voltages into currents and sums using the trans-resistance op-amp configuration. Note that the virtual ground at the input keeps the sources isolated.

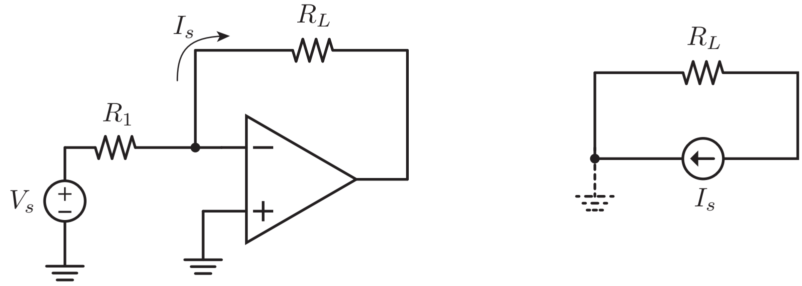


Figure 9.13: A current source can be built from a voltage source by exploiting the virtual ground of the op-amp to create a fixed current.

9.5.5 Voltage Summing Amplifier

In Fig. 9.12, each source is converted into a current and then summed in a similar fashion as the currents in a trans-resistance amplifier. Note that the virtual ground means that no current is "lost" when the currents are put in parallel (due to the output resistance).

9.5.6 A Current Source

Recall that a current source delivers a steady current regardless of the load. While a good battery or voltage source approximates an ideal voltage source, it's hard to find examples of an ideal current source. But using an op-amp, we can build an ideal source as follows. In Fig. 9.13, note that a voltage is converted to a current as before, and the current is fixed since the virtual ground at the op-amp inverting terminal is fixed at ground potential. This current cannot flow into the op-amp, so it flows through the load resistor R_L. It does not matter how we change R_L, the current is fixed by the ratio of V_s and R_1, and as long as the voltage drop across R_L does not cause the op-amp to saturate, the current delivered is constant.

9.6 Application: Instrumentation Amplifiers

In many systems, the desired signal is weak but it's accompanied by a much larger undesired signal. A good example is the ECG measurement on the human body (Fig. 9.14). The human body

[3] More commonly known as trans-impedance amplifier, or TIA.

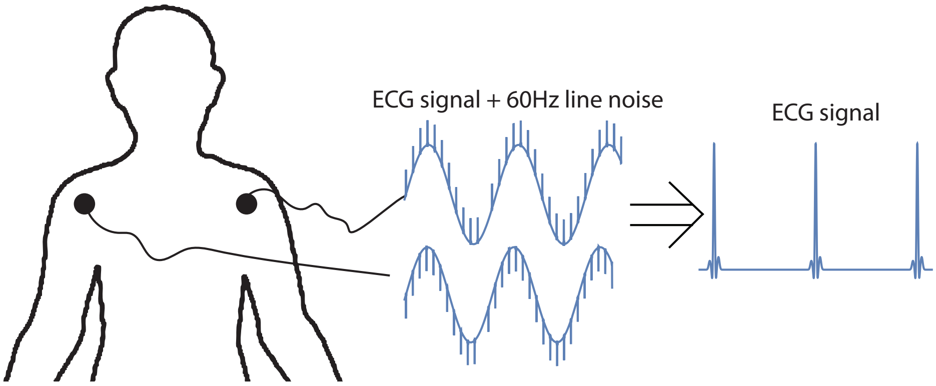


Figure 9.14: ECG signals are measured by detecting the small voltage difference between different points on the body. These signals originate from the heart muscle activity and can be used to diagnose many heart ailments. Due to noise pickup from AC lines, and many other sources, the small ECG signal is riding on top of a large interference signal.

picks up a lot of 60 Hz noise (due to capacitive pickup) and so a very weak ECG signal (mV) is accompanied by a large signal ($\sim 100mV$) that we wish to reject.

Fortunately the noise pickup, as shown in Fig. 9.15, is in *common* with both leads of the ECG because the body is essentially an equipotential surface for the noise pickup. If we take the difference between two points, it disappears.

9.6.1 Common-Mode and Differential-Mode Gain

We thus need an amplifier that can detect a small *differential mode* voltage in the presence of a potentially very strong *common mode* voltage. For example, if each lead picks up a voltage

$$v_1 = V_c + v_{ecg1} \tag{9.25}$$

$$v_2 = V_c + v_{ecg2} \tag{9.26}$$

where V_c is a common voltage coupled from power lines, then the difference voltage

$$v_1 - v_2 = v_{ecg1} - v_{ecg2} \tag{9.27}$$

is the desired signal, which measures the potential difference between two points on the body, and it is free from the unwanted interference V_c. What we desire is a difference amplifier.

9.6.2 Don't Do This

Even though the op-amp output is the difference between the inputs amplified, the circuit shown in Fig. 9.16 does not work for several reasons. The op-amp in open loop configuration has very unreliable gain. Also, the gain is probably too high for this application and the amplifier may rail out. In closed loop configurations, the op-amp input voltage is nearly zero ($v^+ - v^- \approx 0$). What we need is a way to amplify potentially large voltage differences.

9.6.3 Difference Amplifier

A difference amplifier is shown in Fig. 9.17. Let's analyze this circuit quickly using superposition. For port 1, the amplifier is simply an inverting stage

$$v_o^1 = v_1 \frac{-R_2}{R_1} \tag{9.28}$$

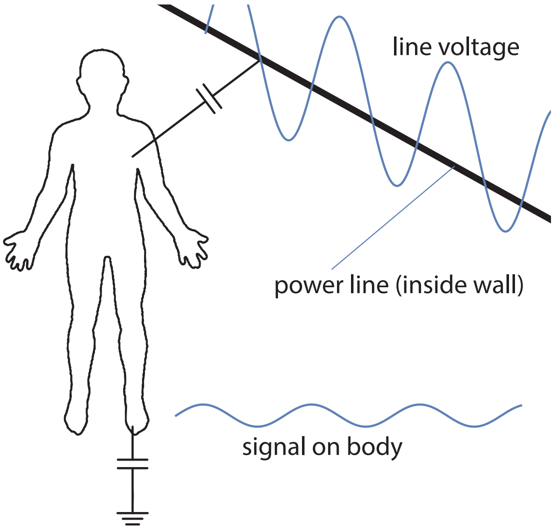


Figure 9.15: Interference from power lines couples to the body capacitively. The body acts like an equipotential surface and the interference appears uniformly throughout the body, producing a common mode noise signal on top of the ECG signal.

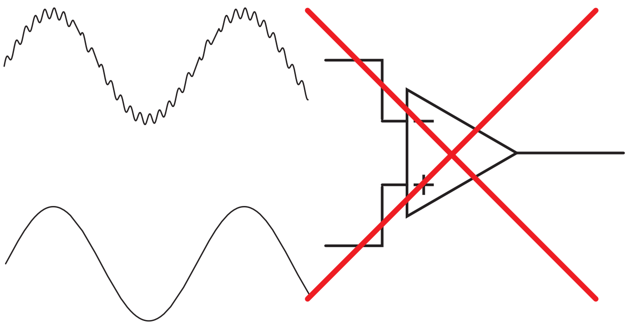


Figure 9.16: An op-amp cannot be used as a difference amplifier without applying feedback. Due to the large gain, the output is likely to rail out. Even if the input signal is small enough, the circuit transfer function will vary with temperature and from part to part.

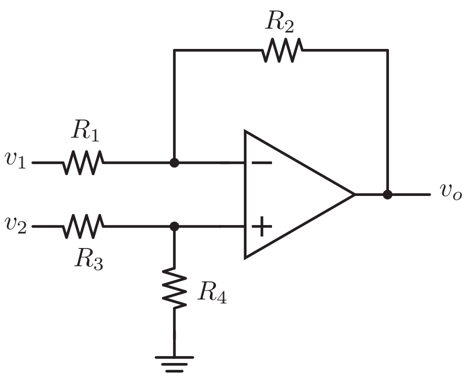


Figure 9.17: A difference amplifier has a transfer function $v_o = v_1 - v_2$. The circuit functions as an inverting amplifier for v_1 and a non-inverting amplifier for v_2. Note that the input v_2 is first scaled by the divider formed by R_3 and R_4 in order to equalize its gain with the inverting path.

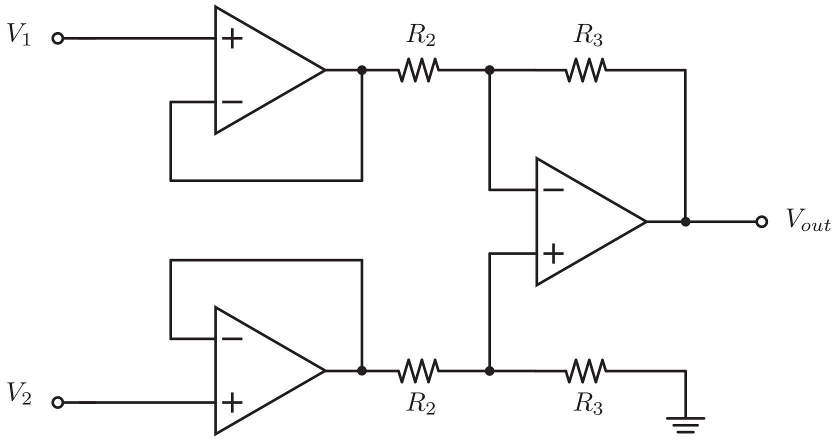


Figure 9.18: An instrumentation amplifier (IA) uses two unity gain buffers in order to have minimal loading on the signals v_1 and v_2. A difference amplifier is at the core of the IA.

For port 2, it's a non-inverting stage, except we only tap off a fraction of v_2

$$v_o^2 = v_2 \frac{R_4}{R_3 + R_4}(1 + \frac{R_2}{R_1}) \tag{9.29}$$

Take the sum and set $R_1 = R_3$ and $R_2 = R_4$, we have

$$v_o = \frac{R_2}{R_1}(v_2 - v_1) \tag{9.30}$$

9.6.4 Instrumentation Amplifier

In a difference amplifier, the input resistance is not infinite since the virtual grounds make each input look like a resistance R_1 to ground. If we add two unity-gain buffers, as shown in Fig. 9.18, we can effectively buffer the input signal. Next a difference amplifier is used to provide a gain of R_3/R_2.

9.6.5 Improved Instrumentation Amplifier (Subtle)

Fig. 9.19 is an improved instrumentation amplifier. First we get some "free" gain from the input buffers. The overall gain is therefore

$$v_o = \left(1 + \frac{2R_1}{R_{gain}}\right) \frac{R_3}{R_2}(v_2 - v_1) \tag{9.31}$$

The input buffers are actually configured as a non-inverting amplifier. Why does the resistor R_{gain} appear between the amplifiers and not as two separate resistors to ground?

This is a subtle point, and to understand it requires you to think of the input as a superposition of a balanced input, the signal we wish to amplify, and a common-mode signal, a signal that we want to detect. In other words, let's re-write the inputs as follows

$$V_1 = \frac{V_d}{2} + V_c \tag{9.32}$$

$$V_2 = -\frac{V_d}{2} + V_c \tag{9.33}$$

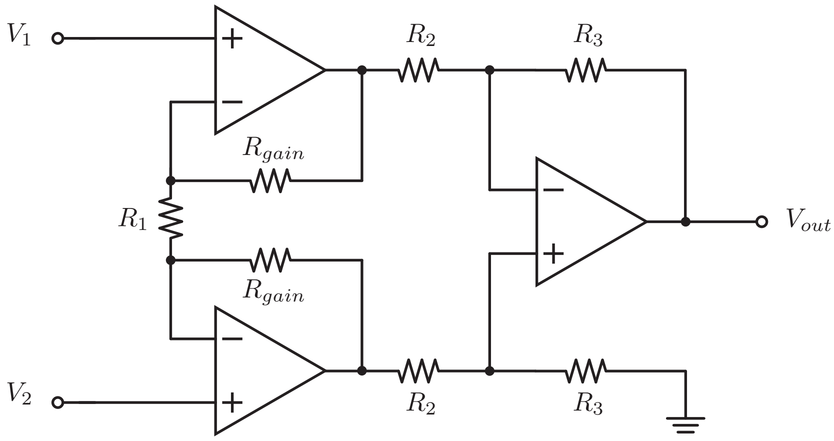


Figure 9.19: An improved IA uses non-inverting amplifiers to both buffer the input and to provide gain, but only to the balanced signal. Common-mode signals see the input stage as a simple follower. In this manner, the rejection of common mode signals is enhanced by this gain.

So the difference signal is given by

$$V_1 - V_2 = \frac{V_d}{2} + V_c - (-\frac{V_d}{2} + V_c) = V_d \tag{9.34}$$

The common-mode signal is given by

$$V_1 + V_2 = \frac{V_d}{2} + V_c + (-\frac{V_d}{2} + V_c) = 2V_c \tag{9.35}$$

So for any signal waveform, we can decompose it into these two parts. The purpose of the instrumentation amplifier is to amplify the differential signal V_d. Suppose we drive a resistance with a balanced signal, as shown in Fig. 9.20a.

$$V_{mid} = \left(\frac{R_1/2}{R_1/2 + R_1/2}\right)\frac{V_d}{2} + \left(\frac{R_1/2}{R_1/2 + R_1/2}\right)\frac{-V_d}{2} = 0 \tag{9.36}$$

Since the voltage at the center of the resistor is zero, which means that we can split the resistor in half. On the other hand, for a common-mode excitation (Fig. 9.20b), no current flows through the resistor, and so it's as if the resistor is not even there.

With this knowledge, we see that for the balanced input, the resistor is split in two and grounded in the middle (another virtual ground). This means that differential signals are gained through the input amplifiers. But for common-mode inputs, no current flows and the resistor is not there, so there's no common-mode gain, the signal is just buffered. This provides additional rejection of the common-mode signal, which is important in applications such as ECG, since the common mode signal can be 10^5 times larger in magnitude and need to be rejected. Even though in theory the difference amplifier will completely null out this common-mode signal, in practice the common-mode rejection is limited by the matching in the resistors between the two op-amps.

9.7 Digital Signal Processing

Signal processing, such as amplification, filtering, rejection of common-mode signals, etc. can be done both in the analog domain, or by capturing a continuous wave signal and representing as discrete and quantized samples, and processing the signal in the digital domain using either dedicated signal processing circuits or a general purpose microprocessor.

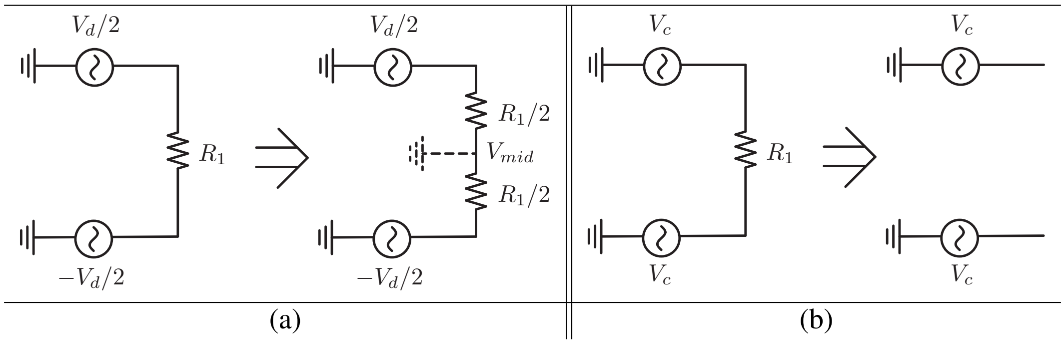


Figure 9.20: (a) For a balanced signal driving a resistor, the mid-point of the resistor is always at zero potential and acts like a virtual ground, splitting the resistor into two equal sized halves. (b) For a common-mode signal drive, no current flows through the resistor and it can be removed from the circuit.

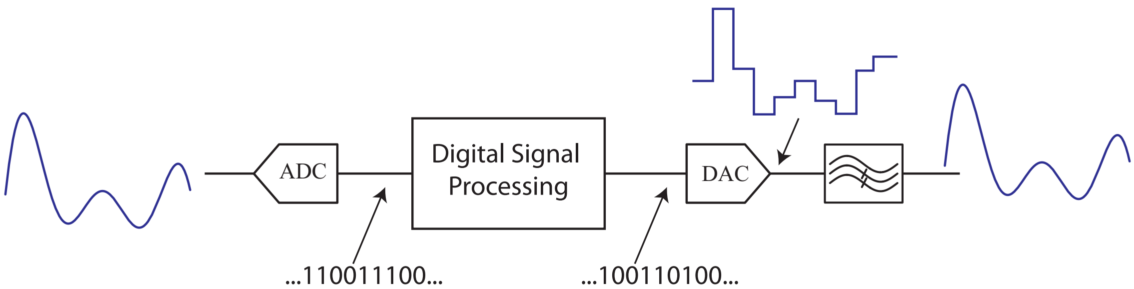


Figure 9.21: A general “mixed-signal" processing system captures signals in the analog domain, converts the signal to the digital domain where custom digital or a general purpose computer processes the data, and the data may be converted once again into analog form. The analog-to-digital converter (ADC) digitizes the signals by converting the signals to a finite precision discrete binary sequence of numbers. A pre-filter is needed to remove high rate transitions in the signal which would distort the digital sequence. The digital-to-analog (DAC) does the opposite. Filters are again needed to smooth the sharp transitions.

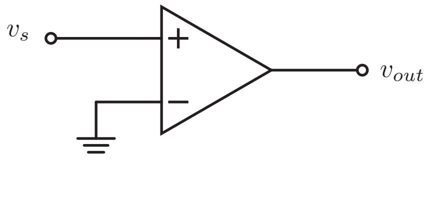


Figure 9.22: An op-amp can act as a zero-crossing detector (left), or a comparator, detecting when the signal crosses a given threshold (right) V_{ref}. The circuit takes advantage of the high gain of the amplifier to produce binary outputs. A positive rail output signal signifies that the input is larger than the threshold, and a negative rail output signal signifies the opposite.

Fig. 9.21 shows an Analog-to-Digital Converter (ADC), which converts a continuously varying analog signal into digital form (1's and 0's), which can be read into a computer, stored, processed (filtered or otherwise manipulated). Note: You'll need a pre-filter to avoid aliasing, something that will be covered in a system level course.

A Digital-to-Analog Converter (DAC) does the reverse process. A digital input is converted into a continuous waveform. Most DAC's hold their value over the entire clock period ("zero-order hold"). By filtering this data we can reconstruct the original waveform.

9.7.1 Comparator / Zero Crossing Detector

In theory an op-amp shown in Fig. 9.22 in "open-loop" configuration can be turned into a comparator, a circuit that compares the inputs and produces either a "high" signal ($v^+ > v^-$) or a "low" signal ($v^+ < v^-$). This is useful in converting small analog signals into a digital signal, or to detect when a signal crosses a given threshold.

The circuit works because of the very high gain of the op-amp, which causes the output to rail to one or the other supply as the input crosses zero (or a given bias). In practice, a specialized comparator should be used since it can operate much faster than an op-amp.

9.7.2 Op-Amp Based DAC

A very simple DAC is shown in Fig. 9.23. Each voltage source represents a binary digit. A 1 corresponds to the maximum voltage (rail) and a 0 corresponds to the ground potential. Each voltage is converted into a current. We weight the voltages by the binary digit position (2^j). The weighted sum is the desired analog voltage.

9.7.3 Comparator Analog-to-Digital Converter

The simplest ADC is shown in Fig. 9.24. A resistor string is essentially a voltage divider, where the tap points provide a uniform reference voltages for comparison. All the comparators with reference voltage below the input go high. The remaining comparators go low. This output can be translated into standard binary representation.

9.8 Positive Feedback

Most of the circuits we have seen so far use negative feedback. The output is scaled and fed into the *negative* terminal of the op-amp. As we have seen, this leads to a very stable operating point. We will now study a circuit that uses positive feedback to realize a useful function.

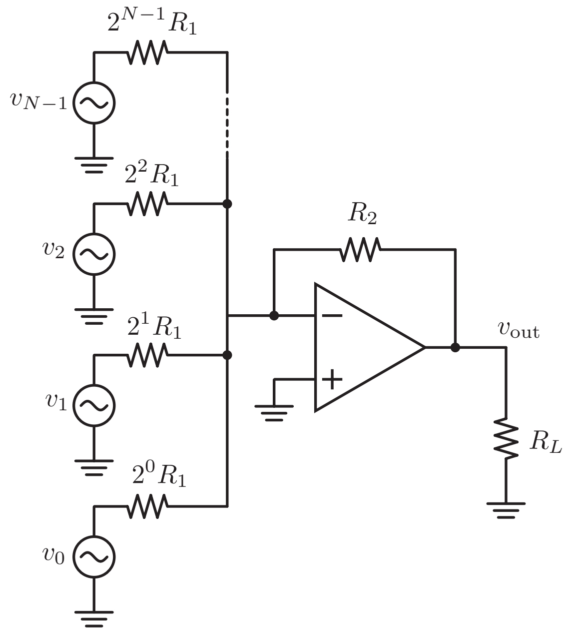


Figure 9.23: A current summer can be used to convert a binary number represented by the sequence $(v_0, v_1, \cdots, v_{N-1})$ into a analog voltage v_{out}. Each binary voltage is converted into a binary current through the weighted resistors 2^iR_1, and then converted to a voltage through the trans-resistance amplifier.

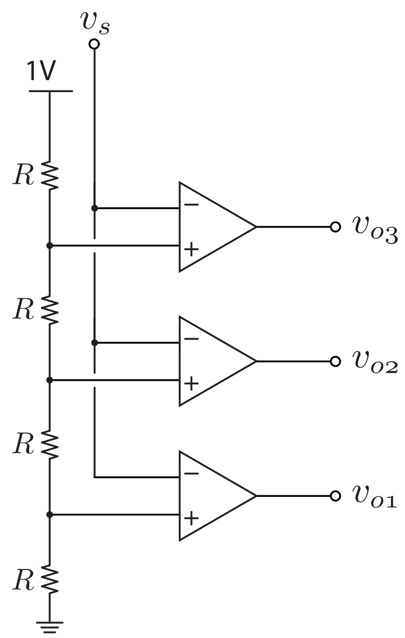
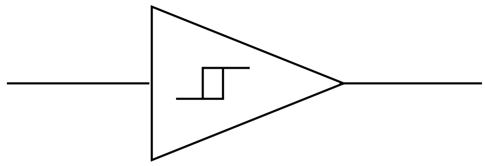
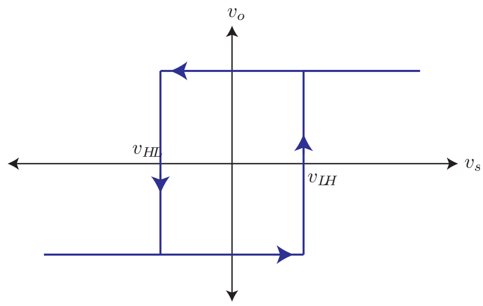


Figure 9.24: An analog-to-digital converter (ADC) uses a string of resistors to build a uniform sequence of reference voltages corresponding to the discrete levels of the binary signal. The signal v_s is compared against each code, and the output code $(v_{o1}, v_{o2}, \cdots)$ is a digital (two-level) binary sequence that can be converted to binary coded form.





(a) (b)

Figure 9.25: (a) Symbol for a Schmitt trigger, or a circuit with transfer function shown in (b). Note that the transfer function has hysteresis, in other words the output depends not only on the input, but also the past value of the output.

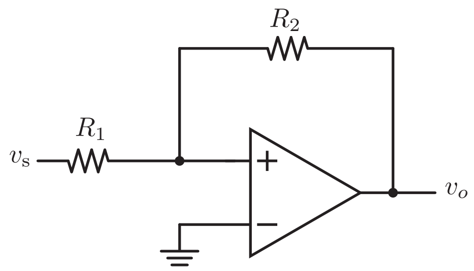


Figure 9.26: An op-amp based Schmitt trigger. In order to cause the output to toggle, the voltage at V^+ must cross zero. If the previous state of the Schmitt trigger is "high" (positive rail), then output tends to keep the the signal V^+ positive, even without the input signal v_s. We must pull v_s sufficiently negative to overcome the action of the output acting on the input in order to cause a high-to-low transition.

9.8.1 Schmitt Trigger

The Schmitt trigger is a circuit that employs positive feedback, shown in Fig. 9.25. It is similar to a comparator, but due to the feedback, it has *hysteresis*. In other words, the transfer characteristic of the amplifier is not static, but depends on the history or past output value. The symbol for a Schmitt trigger conveys the hysteresis curve.

A Schmitt Trigger can be constructed using an op-amp as shown in Fig. 9.26. The action of the feedback is to set the "comparison" voltage based on the current output voltage. The op-amp output will change if v^+ cross zero. At any given time, the voltage v^+ can be found by Nodal analysis (no current flows into the input terminals of an ideal op-amp

$$\frac{v^+ - v_s}{R_1} + \frac{v^+ - v_o}{R_2} = 0 \tag{9.37}$$

Solve for v^+ to obtain

$$v^+ \left(\frac{1}{R_1} + \frac{1}{R_2} \right) = \frac{v_s}{R_1} + \frac{v_o}{R_2} \tag{9.38}$$

$$v^+ (R_2 + R_1) = v_s R_2 + v_o R_1 \tag{9.39}$$

$$v^+ = \frac{v_s R_2 + v_o R_1}{(R_2 + R_1)} \tag{9.40}$$

Schmitt Trigger Thresholds

We find that v^+ is a function of both the input v_s and the output v_o. This second dependence gives rise to the hysteresis.

One can say that the input v^+ is an "arm wrestle" between the input and the current output. For instance, if the output is already high, then the input v_s has to come sufficiently low to cause v^+ to cross zero. We can find the "high-to-low" threshold v_{HL} by setting $v_o = V_{sup}$ and solve for $v^+ = 0$

$$v^+ = \frac{v_{HL} R_2 + V_{sup} R_1}{(R_2 + R_1)} = 0 \tag{9.41}$$

$$V_{HL} = -V_{sup} \frac{R_1}{R_2} \tag{9.42}$$

Similarly, the "low-to-high" threshold v_{LH} can be found by setting $v_o = -V_{sup}$ and solving for $v^+ = 0$

$$v^+ = \frac{v_{LH} R_2 - V_{sup} R_1}{(R_2 + R_1)} = 0 \tag{9.43}$$

$$V_{LH} = V_{sup} \frac{R_1}{R_2} \tag{9.44}$$

Schmitt Trigger Application

First notice that a Schmitt Trigger is a *bi-stable* device, with two stable operating points. It's a memory cell ! In practice we can build memory cells with far fewer transistors than an op-amp, but this is a useful way to think about the circuit.

The most common application is to increase the noise immunity of a circuit. In a comparator, there is only a single threshold voltage. If the input is noisy, then the output will bounce if the noise causes the input to cross the threshold. In a Schmitt Trigger, though, once the output transitions (say high), the threshold to bring the output back down to zero is more negative, which means it's less likely that a noise signal will cause a false transition.